Genetic association data are broadly consistent with stabilizing selection shaping human common diseases and traits

June 19, 2024

E. Koch[1,2*], N. Connally[1,2*], N. Baya[2], M. P. Reeve[3], M. Daly[1,2,3,4], B. Neale[1,2,4], E. S. Lander[1,2], A. Bloemendal[2], S. Sunyaev[1,2]

[1] Harvard Medical School, Boston, MA
[2] Broad Institute, Cambridge, MA
[3] Institute for Molecular Medicine Finland HiLIFE, University of Helsinki, Helsinki, Finland
[4] Massachusetts General Hospital, Cambridge, MA
* These authors contributed equally to this work

Abstract

Results from genome-wide association studies (GWAS) enable inferences about the balance of evolutionary forces maintaining genetic variation underlying common diseases and other genetically complex traits. Natural selection is a major force shaping variation, and understanding it is necessary to explain the genetic architecture and prevalence of heritable diseases. Here, we analyze data for 27 traits, including anthropometric traits, metabolic traits, and binary diseases—both early-onset and post-reproductive. We develop an inference framework to test existing population genetics models based on the joint distribution of allelic effect sizes and frequencies of trait-associated variants. A majority of traits have GWAS results that are inconsistent with neutral evolution or long-term directional selection (selection against a trait or against disease risk). Instead, we find that most traits show consistency with stabilizing selection, which acts to preserve an intermediate trait value or disease risk. Our observations also suggest that selection may reflect pleiotropy, with each variant influenced by associations with multiple selected traits.

Introduction

Many human traits, including diseases, are known to be highly heritable, and affect the lifespan or reproduction of individuals. These qualities are the basis of natural selection, which shapes the genetic architecture of complex traits (those with many associated genetic loci). With complex traits, a long-standing question in evolutionary biology is how substantial heritable variation is maintained in the presence of selection (1; 2). In human biomedical genetics, the same question appears in the seeming contradiction between complex diseases' observed reduction of fitness, and their high prevalence in the population.

It has been proposed that some conditions, such as breast cancer, have persisted due to their post-reproductive onset (3). The argument is that a trait with no effect on the chance or ability to reproduce would—despite being harmful to individuals—have no effect on fitness. Such a trait would be evolving neutrally (i.e., without selection), and could occur at high frequency. The premises of this explanation are debatable (4; 5; 6). However, they cannot explain traits that occur earlier in life, such as schizophrenia, which is estimated to reduce reproduction by up to 75% (7; 8), but continues to have a population prevalence of 1% (9).

For schizophrenia—and other complex human diseases—it is natural to assume a model in which selection acts to remove risk-increasing alleles, but is balanced by the incessant influx of new mutations (mutation-selection-drift balance). This assumption of selection against risk alleles (directional selection) seems to be supported by the observation that variants causing monogenic familial versions of diabetes, breast cancer, and other diseases are rare in the population (10; 11). For quantitative traits, large effect coding mutations, on average, lead to a decrease in height and hand grip—presumably counterbalanced by directional selection (12; 13; 14). Recent studies of GWAS datasets show that lower frequency alleles have higher effect sizes (15; 16; 17; 18; 19; 20; 21; 22). This observation may be evidence of selection, but it has not been shown that selection on common variants acts directionally.

In order to explain the maintenance of trait variation and disease in a population, we must consider multiple models of selection. Directional selection reduces maladaptive traits in a population, such as high risk of a binary disease, or harmfully high levels of a quantitative trait (4; 5; 23; 2). Another model, stabilizing selection works to maintain a trait at its current level in the population, while also reducing variance (24; 25; 26; 27; 23; 1; 2). If stabilizing selection is acting on many traits and diseases, it could help explain their counter-intuitive prevalence. For example, schizophrenia would be relatively prevalent because stabilizing selection has prevented schizophrenia risk from becoming too high or too low. The assumption of stabilizing selection appears to be natural for some quantitative traits, such as blood pressure, which can be dangerous at either extreme. However, this model has not been deeply explored for diseases, as it is not intuitive that individuals with unusually low risk for disease would have a fitness disadvantage.

The discovery of associated variants from genome-wide association studies (GWAS) offers a window into the evolutionary dynamics and genetic architecture of many traits (1; 2). We take a set of traits with sufficient numbers of associated variants (at least 100), and examine the relationship between the effect size and frequency of the trait-increasing (or risk-increasing) allele. Visualizing this relationship in a scatter plot reveals that most traits share a similar distribution, with a U-shaped curve that resembles a "smile" (Figure 1). In measuring the relationship between signed effect sizes and frequency, these plots are similar to "trumpet" plots (12; 28), but are organized to facilitate observations about evolution and selection.

The "smiles" curves suggest some initial inferences about selection on these traits. The lower edge of the curve represents the limits of statistical power, and is not reflective of selection. The upper edge of the curve is the result of an absence of large-effect variants at intermediate frequencies.

This is evidence of selection—under neutrality, large-effect variants could drift randomly to any frequency, which selection prevents them from doing. If selection were directional, acting either to decrease or increase the trait, we would expect to see large-effect variants at one side of the distribution (where the effects of selection are balanced against the influx of new mutations), but not at both sides (Figure 2b). The symmetry of the smile is consistent with the symmetry of stabilizing selection, which acts against both trait-increasing and trait-decreasing variants. Additionally, the location of the upper edge and spread of observed variants may be influenced by pleiotropy—the propensity of variants to experience selective pressure from multiple associated traits. Models of stabilizing selection in the case of high pleiotropy have been investigated theoretically (29; 30; 31; 22) and tested against the empirical data on some human quantitative traits (31; 22), although not against disease traits.

Methods and results

Testing each form of selection required correcting for confounding variables that have the potential to alter or even invent signals of selection. In our analysis, we were primarily concerned with four confounders. 1) Linkage disequilibrium—the correlation of GWAS statistics for variants near one another—limits our ability to identify which specific variants influence the trait. 2) The evolutionary demographic history of a population can alter the distribution of variants, even in the absence of selection (32). 3) Ascertainment bias refers to the fact that GWAS have limits on power, and identify only a subset of trait-associated variants. 4) And the associations that are identified are likely to have measured effect sizes larger than their true effects—a phenomenon referred to as winner's curse (33). These confounders were simulated in order to test their effects. When necessary, their effects are accounted for in GWAS data before models of selection are tested (Supplements A, B.3, D).

We analyzed the consistency of observed GWAS data with models of neutrality and four different forms of selection: directional selection (34; 2), stabilizing selection (26; 34; 1; 2), a combination of directional and stabilizing selection (35), and pleiotropic stabilizing selection (30; 31; 22). Each of these models can be characterized by the change in frequency of a variant under selection (26; 34; 1; 36) (Table 1). The dynamics of frequency changes can be used to predict the distribution of trait-associated variants for the model (Supplement B.2) under the assumption of mutation-selection-drift equilibrium. This will be true when the form of selection has been relatively constant over evolutionary time. By varying the selection parameters shown in Table 1, we can find the best possible fit for a given model for the observed data. The strength of this fit is compared against the neutral model using a likelihood ratio test. As a conservative strategy, we use only variants with genome-wide significance (5×10^{-8}) and a minimum variance contribution (Supplements B.1 & B.3.1).

We initially fit each model to summary statistics from 15 GWAS, 10 quantitative traits and 5 binary disease traits. All traits showed greater consistency with stabilizing selection than with neutral or directional models (Figure 3). The pleiotropic model of stabilizing selection (31) was consistently a better fit than the single-trait stabilizing selection model, though the difference was sometimes small.

In order to validate this finding, we selected an additional three quantitative traits and nine disease traits. Three of the additional diseases were underpowered to distinguish selection from neutrality. Across the other disease and quantitative traits, the stabilizing selection models remained the most consistent, with the pleiotropic model still tending to provide a better fit than the single-trait model.

There is only one trait (hypothyroidism) for which directional selection is more consistent with

the data than the neutral and stabilizing models. Interestingly, for five traits, directional selection is more consistent than the neutral model, though stabilizing selection provides a much better fit. Additionally, there is one trait (malignant neoplasms) that is most consistent with simultaneous directional and stabilizing selection. However, it is possible that the directional component reflects a greater power to detect risk variants than protective ones. Regardless of whether this is the case, our replication confirmed models of stabilizing selection are more consistent with GWAS data than the directional selection model.

We used maximum likelihood estimates of selection parameters (Table 1) to investigate the implied strength of selection on GWAS variants resulting from our model fits. We present these as selection coefficients scaled by an approximate ancestral effective population size of 10,000 (a common scaling for selection measured in population genetics studies). While there are some differences between traits, we generally infer that ascertained variants with minor allele frequency $> 10\%$ have scaled selection coefficients in the range 1-10, which reflects relatively weak selection (Figure 4). Variants with lower minor allele frequencies can experience much stronger selection. GWAS discoveries therefore span a range of selection from approximately neutral to strong selection.

Discussion

Our results show that across traits, selection is common and is consistent with stabilizing models. Across 27 traits, 24 show evidence of selection. There is a single trait (hypothyroidism) for which directional selection is distinguishable from neutrality and has a higher likelihood than stabilizing selection. Every other selected trait is consistent with stabilizing selection.

The intuitively appealing model of directional selection against disease is not supported by the observed genetic architectures of traits. This is despite a power imbalance in the ability to detect risk vs. protective variants, with risk variants being easier to observe for any binary trait with a low prevalence.

Generally, traits are consistent with two forms of stabilizing selection: single-trait or pleiotropic. Earlier work has argued that quantitative traits are under pleiotropic, not single-trait stabilizing selection. We note that the variance contribution test previously applied to human quantitative traits has a limited ability to discriminate between forms of selection (31) (Supplement F). Still, we found that pleiotropic stabilizing selection models had higher likelihoods than single-trait models for all quantitative traits and six out of the nine diseases under stabilizing selection. However, the distinction between the single trait and pleiotropic models in our analysis may be overstated. We simulated GWAS data under the two models and found much more limited power to distinguish between them (Supplement E.1, Supplementary Figures S9–S11). It may be that other sources of heterogeneity in allelic effect sizes and frequencies (ascertainment, stratification, etc.) are better absorbed into the pleiotropic likelihood. For instance, it is difficult to model uncertainty in effect sizes due to linkage disequilibrium, which is evident from variation among independent GWAS studies. Alternatively, our inference may correctly point to the high degree of pleiotropy, but the true model allows for a greater variation in selection given a particular effect size than the chosen parametrization.

There are further models of selection we did not test. In particular, a broad range of scenarios may be considered apparent stabilizing selection, a phenomenon in which an unselected trait can appear selected because its associated variants experience selection due to their associations with other traits (37). If the resulting scaling between effect sizes and selection is sufficiently similar to that predicted under stabilizing selection, then they would not be identifiable using only allele frequencies and effect sizes. Gaining more insight into the precise form of selection will be valuable,

but regardless of the specifics of the model, our work shows that protective variants for traits are often selected against.

The finding of selection against some protective variants has several implications. First, it helps explain disease prevalence and variation in traits. Diseases, including post-reproductive traits, are not broadly neutral. For selection to decrease the genetic component of disease risk would require it to increase the prevalence of low-frequency protective genetic variants—which our results show does not occur.

Secondly, this selection may provide a warning about limitations on genetics-guided drug discovery. Efforts to identify potential drug targets often examine existing genetic variation—"experiments of nature"—for protective variants (38). Selection against these variants may indicate that lowering the risk of a disease is inherently associated with a fitness cost or tradeoff. Alternatively, even if lowering disease risk in isolation is itself associated with minor costs, it may be that all possible genetic routes (and corresponding drug targets) to doing so lead to harmful changes in other, seemingly unrelated, traits. The single trait model of stabilizing selection appears to suggest the first of these scenarios, while the pleiotropic model may suggest the latter.

As the sizes of GWAS grow, more traits will have a sufficient number of associated variants for the type of analysis presented here. Indeed, trumpet plots are becoming a standard part of the presentation of well-powered GWAS results. When possible, quick, simple selection analyses like those presented here may provide useful insight into the mode and strength of selection on identified variants. Combining better-powered studies with more advanced models may also allow us to better distinguish between models of selection that produce similar genetic architectures, expanding on the scope of this study. These models may leverage sub-significant variants, combine multiple traits, or leverage linkage disequilibrium between causal variants. Stabilizing selection models make explicit predictions that can be tested. For instance, more pleiotropic variants should, all else equal, experience stronger selection, and haplotypes carrying opposite effect variants should be found at higher frequencies than those with concordant effect variants (39; 40).

We have shown that, across a broad array of traits, selection is common, but does not appear to have acted directionally over extended periods of evolutionary time. Instead, the genetic architecture of these traits is consistent with forms of stabilizing selection, indicating that selection has kept trait values and disease risk at intermediate levels. This does not rule out adaptive shifts, but implies that they were accomplished without major changes to genetic architectures and did not result in large-scale fixation and establishment of a strong mutational bias. Our findings help explain the surprisingly high prevalence of some diseases and other health-associated traits. It also suggests limitations that may be present in genetically informed drug discovery, and raises important questions about the relationship between traits and fitness.

Figures

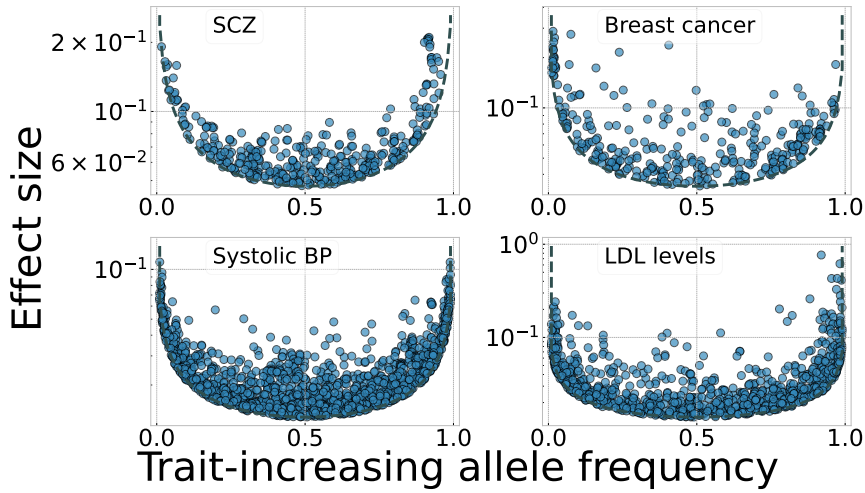


Figure 1: A U-shaped distribution for the effect sizes and allele frequencies after polarizing by direction of effect.
For four example traits we plot the GWAS-estimated effect sizes of trait or disease risk-increasing alleles against their frequencies. For disease traits, the effect size is calculated as the odds-ratio minus one. We include variants that are both genome-wide significant and have an estimated variance contribution exceeding a threshold variance associated with 5×10^{-8}.

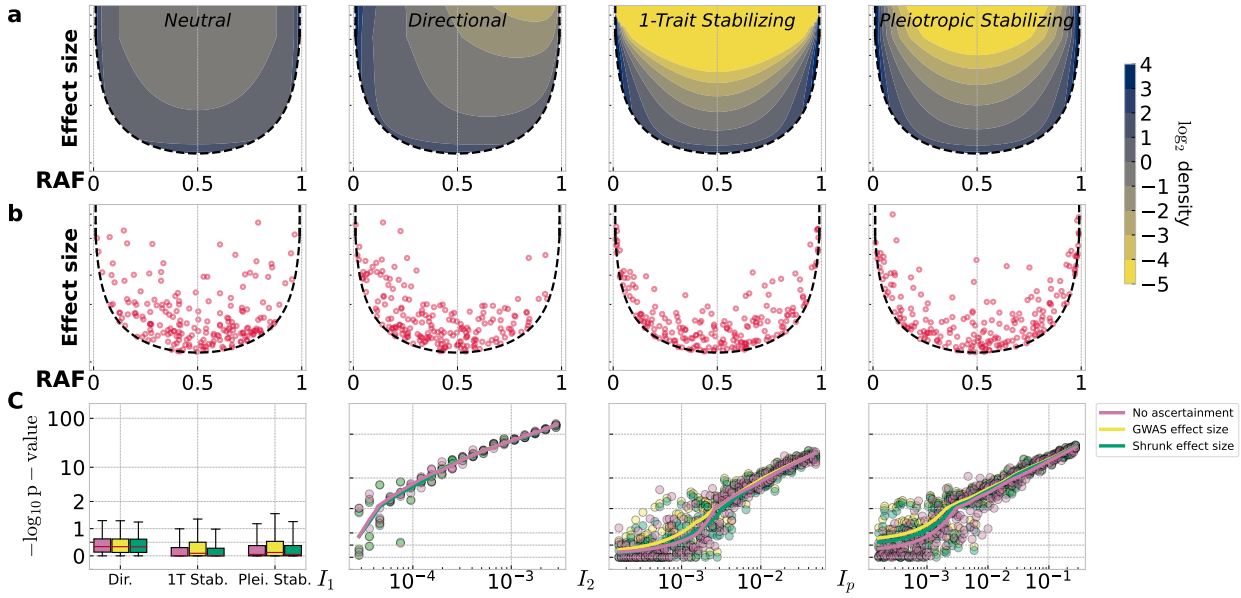


Figure 2: Selection leaves an identifiable pattern in the distribution GWAS hits. a) The probability densities for the joint distributions of variant frequency and effect size show the expected shapes of smiles under different models. Neutrality produces a more filled-in smile. Directional selection produces a more asymmetric smile. Single- trait stabilizing selection produces a symmetric smile that is not filled-in. Pleiotropic stabilizing selection produces a thicker smile than stabilizing selection on one trait. b) Simulations show examples of the distributions of variant effect sizes and risk-increasing (or trait-increasing) allele frequencies (RAF) expected under different models. Ascertainment is modeled using a variance-contribution threshold. An absence of selection results in a large number of large-effect common variants. Directional selection is modeled assuming mutational equilibrium. New mutations are highly biased by their effect direction on the trait, and most are deleterious. Stabilizing selection is considered at single-trait (1D) and pleiotropic (hD) extremes. Both result in a U-shape, with symmetry in RAF space and a lack of high-frequency, large-effect variants. c) Simple models using variant frequencies conditional on effect size are able to detect selection. Evidence for selection is quantified using p-values relative to the neutral model where selection intensities are zero, values of 1 and above can be considered strong evidence. Neutral simulations rarely produce false selection signals. Selection is reliably detected in all models once the intensity exceeds a threshold. I_1, I_2, and I_p represent the intensity of selection in the directional, single-trait stabilizing, and pleiotropic stabilizing models respectively. Simulation results presented here used the distribution of effect sizes estimated for common IBD variants and a model of European population history from Tennessen et al. (2012). Each simulation was fixed at 200 ascertained loci.

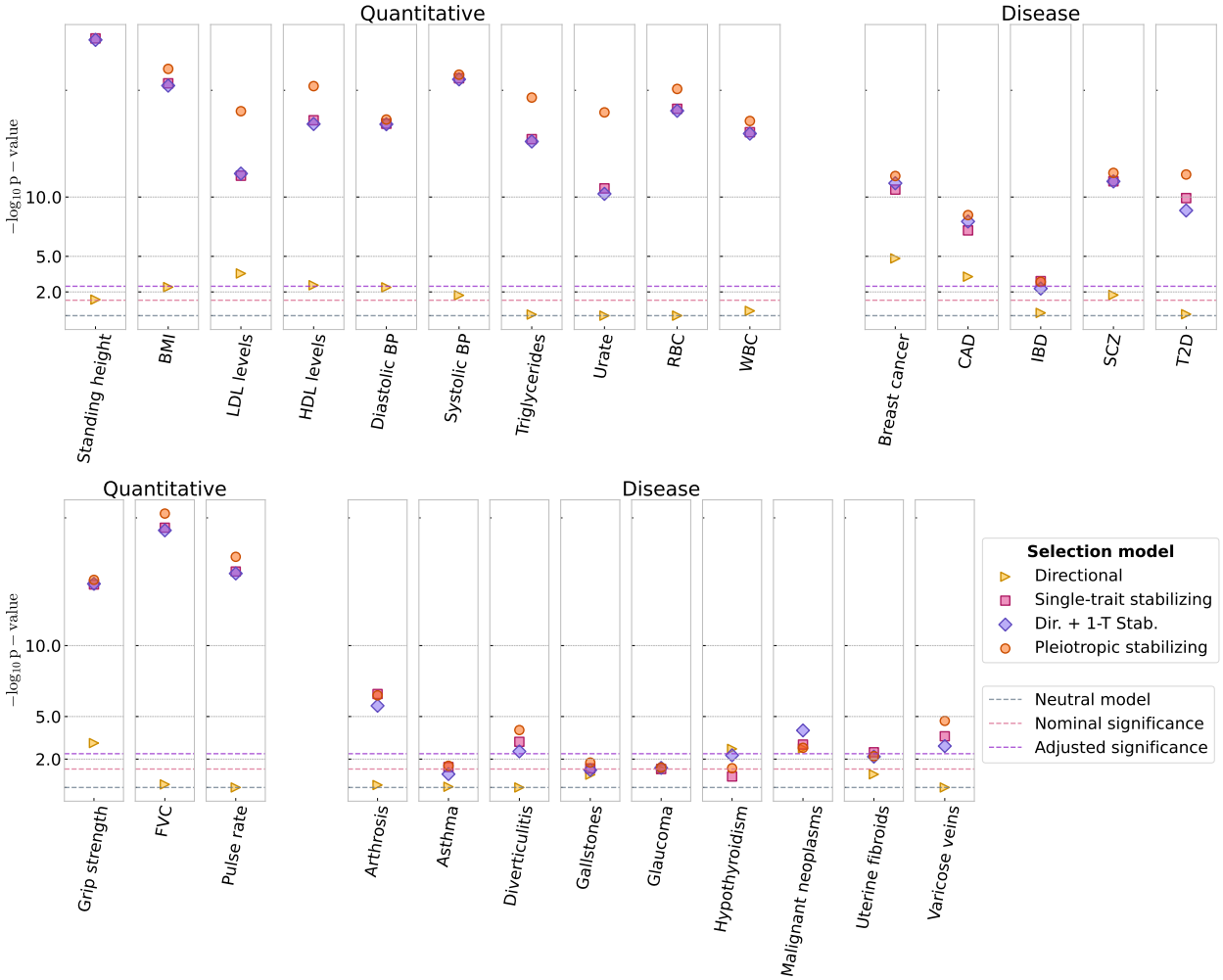


Figure 3: Model comparison favors stabilizing selection across categories of traits. We fit models of selection to each trait, estimating the intensity of selection under each model, and summarized results using p-values for a likelihood ratio test against a neutral model. To account for ascertainment bias in effect sizes we shrunk effect sizes by fitting the prior distribution of effect sizes using ash and computing posterior means from estimated GWAS effect sizes and standard errors. We use a variance-contribution threshold corresponding approximately to the p-value threshold 5×10^{-8} and also filter explicitly on p-value. In addition to the directional and stabilizing models shown in Figure 1, we also fit a model that combined directional and single-trait stabilizing selection. a) The original 15 traits tested. b) The 12 traits tested as a replication set.

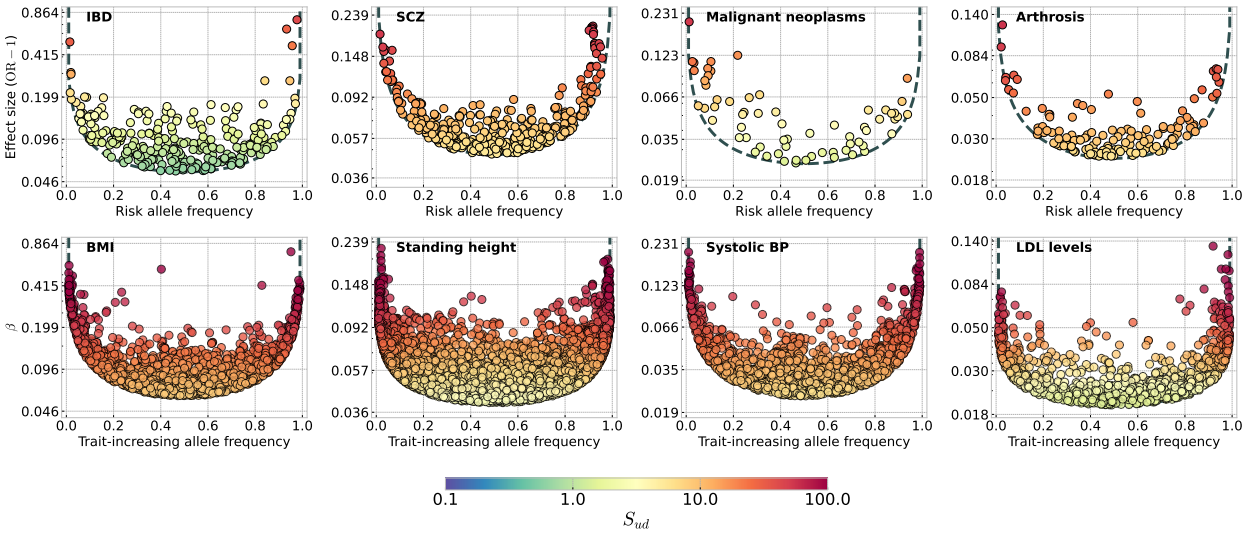


Figure 4: Estimated stabilizing selection on disease and quantitative traits.
We used maximum-likelihood estimates of selection intensities to predict the strength of selection on observed GWAS variants for disease traits. Estimates represent median selection coefficients under the model of high-dimensional stabilizing selection. We show GWAS effect sizes for disease traits as the odd-ratio minus one. Selection estimates were calculated using shrunk effect sizes. The scaled strength of selection is calculated as $S_{ud} = 2N_e\ s_{ud}$, with Ne=10,000 used to give a sense for the scaled selection coefficients in an ancestral human population.

Tables

Table 1: A summary of different selection models, their effects on smiles plots, and their parameters.
Each selection model assumes a different relationship between traits and fitness, which leads to different effects on trait-associated variants, and qualitatively different smiles plots. The parameters S, S_{ud}, S_P are the coefficients for directional selection, underdominant selection (which occurs when a trait is under stabilizing selection), and pleiotropic selection. They are used to model the expected change in allele frequency, $E[\Delta x] = sx(1-x)+s_{ud}x(1-x)(\frac{1}{2}-x)$. Directional examples were written as though selection is acting against a maladaptive trait, but would be reversed if selection is acting to promote a beneficial trait.

Type of selection	**Trait value with highest fitness**	**Effect on trait-associated variants**	**Type of smile observed**	**Notes**	$E[\Delta x] = Sx(1-x) + S_{ud}x(1-x)(\frac{1}{2}-x)$	
					S	S_{ud}
Neutrality (no selection)	The trait does not affect fitness, so any value is equivalent.	Variants are ignored by selection, so alleles can drift to any frequency	Symmetric, but "filled in" — some large-effect variants exist at intermediate frequencies	Because neutrality is the absence of selection, it can be used as a null model for testing the statistical evidence for selection.	0	0
Directional selection	The optimal trait value is lower than any that occurs, so the lowest values in the population are the most advantageous.	Trait-increasing alleles diminish in frequency	Asymmetric, with more variants observed at low frequencies	Traits under directional selection can remain constant over time if the effects of selection are balanced by the effects of new mutations.	$I_1\beta$	0
Stabilizing selection	The average trait value of the population	Whichever allele is less common (the minor allele) diminishes in frequency	Symmetric and not filled in	Like neutrality, stabilizing selection does not exert pressure on the average trait value. But a neutral trait can change randomly; stabilizing selection reduces deviation in either direction.	0	$I_2\beta^2$
Stabilizing + directional selection	A value near the population average, but slightly lower. Though the optimal trait value is not the average, it is within the range of population values	The minor allele diminishes in frequency, but a minor trait-increasing allele diminishes more quickly than a minor trait-decreasing allele	More symmetric than under directional selection alone, but less symmetric than under stabilizing selection alone	This model is a combination of the previous two	$I_1\beta$	$I_2\beta^2$
Pleiotropic stabilizing selection	The average trait value of the population. However, each trait-associated variant also affects other traits, which adds other selection pressures	The minor allele diminishes in frequency, but because an allele may affect multiple traits, the rate at which it diminishes is noisier	Symmetric, but thicker than stabilizing selection without pleiotropy	Though variants affect sets of traits, each affects a different set, so the strength of selection is represented by a distribution	0	$\sim \text{Levy}(I_P\beta^2)$

Trait	Source
Height BMI LDL cholesterol HDL cholesterol Diastolic blood pressure Systolic blood pressure Triglyceride levels Urate levels Red blood cell count White blood cell count Grip strength Forced vital capacity Pulse rate	Pan-UK Biobank (41)
Arthrosis Asthma Diverticulitis Gallstones Glaucoma Hypothyroidism Malignant neoplasms Uterine fibroids Varicose veins	Pan-UK Biobank + Finngen meta-analysis (42)
Breast cancer	(43)
Coronary artery disease	(44)
Inflammatory bowel disease	(45)
Schizophrenia	(46)
Type II diabetes	(47)

Table 2: Sources of GWAS data.
For GWAS from the Pan-UK Biobank, only associations, effect sizes, and allele frequencies corresponding to CEU ancestry were used. When using results from the meta-analysis of Pan-UK Biobank and Finngen results, we used the P-values and effect size estimates of the meta-analysis, but plotted them against the allele frequencies for UK Biobank participants with CEU ancestry.

References

[1] Sella, G. & Barton, N. H. Thinking about the Evolution of Complex Traits in the Era of Genome-Wide Association Studies. *Annual Review of Genomics and Human Genetics* **20**, 461–493 (2019).

[2] Koch, E. M. & Sunyaev, S. R. Maintenance of complex trait variation: Classic theory and modern data. *Frontiers in Genetics* **12** (2021). URL `https://www.frontiersin.org/articles/10.3389/fgene.2021.763363`.

[3] Partridge, L. & Gems, D. Mechanisms of aging: public or private? *Nature Reviews Genetics* **3**, 165–175 (2002).

[4] Charlesworth, B. Patterns of age-specific means and genetic variances of mortality rates predicted by the mutation-accumulation theory of ageing. *Journal of Theoretical Biology* **210**, 47–65 (2001).

[5] Wright, A., Charlesworth, B., Rudan, I., Carothers, A. & Campbell, H. A polygenic basis for late-onset disease. *Trends in Genetics* **19**, 97–106 (2003).

[6] Pavard, S. & Coste, C. F. Evolutionary demographic models reveal the strength of purifying selection on susceptibility alleles to late-onset diseases. *Nature Ecology & Evolution* **5**, 392–400 (2021).

[7] Laursen, T. M. & Munk-Olsen, T. Reproductive patterns in psychotic patients. *Schizophrenia Research* **121**, 234–240 (2010). URL `https://www.sciencedirect.com/science/article/pii/S0920996410013393`.

[8] Power, R. A. *et al.* Fecundity of Patients With Schizophrenia, Autism, Bipolar Disorder, Depression, Anorexia Nervosa, or Substance Abuse vs Their Unaffected Siblings. *JAMA Psychiatry* **70**, 22–30 (2013). URL `https://doi.org/10.1001/jamapsychiatry.2013.268`.

[9] Charlson, F. J. *et al.* Global Epidemiology and Burden of Schizophrenia: Findings From the Global Burden of Disease Study 2016. *Schizophrenia Bulletin* **44**, 1195–1203 (2018). URL `https://doi.org/10.1093/schbul/sby058`. `https://academic.oup.com/schizophreniabulletin/article-pdf/44/6/1195/26088417/sby058.pdf`.

[10] Salguero, M. V. *et al.* Monogenic forms of diabetes. *Diabetes in America [Internet]* (2023).

[11] Linder, J. E. *et al.* Prospective, multi-site study of healthcare utilization after actionable monogenic findings from clinical sequencing. *The American Journal of Human Genetics* **110**, 1950–1958 (2023).

[12] Marouli, E. *et al.* Rare and low-frequency coding variants alter human adult height. *Nature* **542**, 186–190 (2017).

[13] Ganna, A. *et al.* Quantifying the impact of rare and ultra-rare coding variation across the phenotypic spectrum. *The American Journal of Human Genetics* **102**, 1204–1211 (2018).

[14] Huang, Y. *et al.* Rare genetic variants impact muscle strength. *Nature Communications* **14**, 3449 (2023).

[15] Yang, J. *et al.* Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. *Nature genetics* **47**, 1114–20 (2015). URL http://www.ncbi.nlm.nih.gov/pubmed/26323059http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4589513. Publisher: Nature Publishing Group.

[16] Mancuso, N. *et al.* The contribution of rare variation to prostate cancer heritability. *Nature genetics* **48**, 30–5 (2016). URL http://www.ncbi.nlm.nih.gov/pubmed/26569126http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC7534691. Publisher: Nature Publishing Group.

[17] Gazal, S. *et al.* Functional architecture of low-frequency variants highlights strength of negative selection across coding and non-coding annotations. *Nature genetics* **50**, 1600–1607 (2018). URL http://www.nature.com/articles/s41588-018-0231-8http://www.ncbi.nlm.nih.gov/pubmed/30297966http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC6236676. ISBN: 6082621603 Publisher: Springer US.

[18] Zeng, J. *et al.* Signatures of negative selection in the genetic architecture of human complex traits. *Nature Genetics* **50**, 746–753 (2018). URL http://dx.doi.org/10.1038/s41588-018-0101-4. Publisher: Springer US.

[19] Schoech, A. P. *et al.* Quantification of frequency-dependent genetic architectures in 25 UK Biobank traits reveals action of negative selection. *Nature Communications* **10**, 790 (2019). URL https://www.biorxiv.org/content/early/2017/09/13/188086?rss=1http://www.nature.com/articles/s41467-019-08424-6. Publisher: Springer US.

[20] Speed, D., Holmes, J. & Balding, D. J. Evaluating and improving heritability models using summary statistics. *Nature genetics* **52**, 458–462 (2020). URL http://dx.doi.org/10.1038/s41588-020-0600-yhttp://www.ncbi.nlm.nih.gov/pubmed/32203469. Publisher: Springer US.

[21] Zeng, J. *et al.* Widespread signatures of natural selection across human complex traits and functional genomic categories. *Nature communications* **12**, 1164 (2021). URL http://dx.doi.org/10.1038/s41467-021-21446-3http://www.ncbi.nlm.nih.gov/pubmed/33608517http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC7896067. Publisher: Springer US.

[22] Simons, Y. B., Mostafavi, H., Smith, C. J., Pritchard, J. K. & Sella, G. Simple scaling laws control the genetic architectures of human complex traits. *bioRxiv* 2022–10 (2022).

[23] Walsh, B. & Lynch, M. *Evolution and selection of quantitative traits* (Oxford University Press, 2018).

[24] Wright, S. The analysis of variance and the correlations between relatives with respect to deviations from an optimum. *Journal of Genetics* **30**, 243–256 (1935). URL http://link.springer.com/10.1007/BF02982239.

[25] Robertson, A. The effect of selection against extreme deviants based on deviation or on homozygosis. *Journal of Genetics* **54**, 236–248 (1956). URL http://link.springer.com/10.1007/BF02982779.

[26] Bulmer, M. G. The genetic variability of polygenic characters under optimizing selection, mutation and drift. *Genetical Research* **19**, 17 (1972). URL http://www.journals.cambridge.org/abstract_S0016672300014221. ISBN: 0016-6723 Publisher: University of Chicago.

[27] Lande, R. Natural Selection and Random Genetic Drift in Phenotypic Evolution. *Evolution* **30**, 314 (1976). URL http://www.jstor.org/stable/2407703%5Cnhttp://www.jstor.org/stable/2407703?seq=1&cid=pdf-reference#references_tab_contents%5Cnhttp://www.jstor.org/page/http://www.jstor.org/stable/2407703?origin=crossref. ISBN: 0014-3820 _eprint: arXiv:1011.1669v3.

[28] Aragam, K. G. *et al.* Discovery and systematic characterization of risk variants and genes for coronary artery disease in over a million participants. *Nature genetics* **54**, 1803–1815 (2022).

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

[30] Zhang, X.-S. & Hill, W. G. Multivariate stabilizing selection and pleiotropy in the maintenance of quantitative genetic variation. *Evolution; international journal of organic evolution* **57**, 1761–75 (2003). URL http://doi.wiley.com/10.1111/j.0014-3820.2003.tb00584.xhttp://www.ncbi.nlm.nih.gov/pubmed/14503618. ISBN: 0014-3820.

[31] Simons, Y. B., Bullaughey, K., Hudson, R. R. & Sella, G. A population genetic interpretation of GWAS findings for human quantitative traits. *PLOS Biology* **16**, e2002985 (2018). URL http://dx.plos.org/10.1371/journal.pbio.2002985. ISBN: 1111111111.

[32] Tennessen, J. A. *et al.* Evolution and Functional Impact of Rare Coding Variation from Deep Sequencing of Human Exomes. *Science* **337**, 64–69 (2012). URL https://www.science.org/doi/10.1126/science.1219240. Publisher: American Association for the Advancement of Science.

[33] Stephens, M. False discovery rates: a new deal. *Biostatistics* **18**, 275–294 (2017). URL https://academic-oup-com.ezp-prod1.hul.harvard.edu/biostatistics/article/18/2/275/2557030. Publisher: Oxford Academic.

[34] Bulmer, M. Maintenance of genetic variability by mutation–selection balance: a child's guide through the jungle. *Genome* **31**, 761–767 (1989).

[35] Martin, G. & Lenormand, T. A general multivariate extension of Fisher's geometrical model and the distribution of mutation fitness effects across species. *Evolution* **60**, 893–907 (2006). URL http://onlinelibrary.wiley.com/doi/10.1111/j.0014-3820.2006.tb01169.x/abstract.

[36] Koch, E. M. The Effects of Demography and Genetics on the Neutral Distribution of Quantitative Traits. *Genetics* **211**, genetics.301839.2018 (2019).

[37] Kondrashov, A. S. & Turelli, M. Deleterious mutations, apparent stabilizing selection and the maintenance of quantitative variation. *Genetics* **132**, 603–618 (1992). ISBN: 0016-6731 (Print).

[38] Plenge, R. M., Scolnick, E. M. & Altshuler, D. Validating therapeutic targets through human genetics. *Nature reviews Drug discovery* **12**, 581–594 (2013).

[39] Bulmer, M. The effect of selection on genetic variability. *The American Naturalist* **105**, 201–211 (1971).

[40] Zhang, M. J. *et al.* Pervasive correlations between causal disease effects of proximal snps vary with functional annotations and implicate stabilizing selection. *Research Square* (2023).

[41] Karczewski, K. J. *et al.* Pan-uk biobank gwas improves discovery, analysis of genetic architecture, and resolution into ancestry-enriched effects. *medRxiv* 2024–03 (2024).

[42] Kurki, M. I. *et al.* Finngen provides genetic insights from a well-phenotyped isolated population. *Nature* **613**, 508–518 (2023).

[43] Michailidou, K. *et al.* Association analysis identifies 65 new breast cancer risk loci. *Nature* **551**, 92–94 (2017). URL `https://www.nature.com/articles/nature24284`. Number: 7678 Publisher: Nature Publishing Group.

[44] Nelson, C. P. *et al.* Association analyses based on false discovery rate implicate new loci for coronary artery disease. *Nature Genetics* **49**, 1385–1391 (2017). URL `https://www.nature.com/articles/ng.3913;`. Number: 9 Publisher: Nature Publishing Group.

[45] Liu, J. Z. *et al.* Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nature Genetics* **47**, 979–986 (2015). URL `https://www.nature.com/articles/ng.3359`. Number: 9 Publisher: Nature Publishing Group.

[46] Trubetskoy, V. *et al.* Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature* **604**, 502–508 (2022). URL `https://www.nature.com/articles/s41586-022-04434-5`. Number: 7906 Publisher: Nature Publishing Group.

[47] Mahajan, A. *et al.* Multi-ancestry genetic study of type 2 diabetes highlights the power of diverse populations for discovery and translation. *Nature genetics* **54**, 560–572 (2022).

Supplemental Methods and Figures

A GWAS Summary Statistics and Pipeline

A.1 GWAS processing

GWAS summary statistics were taken from the studies referenced in 2. All variants other than biallelic SNPs were removed, and the 28,500,000 - 35,000,000 bp region of chromosome 6 was removed in order to avoid complications from the extended linkage disequilibrium and unusual selection properties of the human leukocyte antigen (HLA) region.

A.2 Effect size shrinkage using ASH

Ascertainment of GWAS variants can cause traits evolving neutrally to display signals of selection. As described in the main text, the concept of winner's curse reflects that some variants have a perceived effect in a GWAS sample that is higher than their causal effect. These variants will be inherently more likely to be discovered in GWAS, and because most variants in a population are relatively rare, most variants identified through winner's curse will also be rare, potentially causing or exaggerating a U-shaped curve.

To reduce the effects of winner's curse, we used the adaptive shrinkage (ASH) method implemented in the *ashr* package (33). This Bayesian method first constructs a model of GWAS variant effect sizes and standard errors—to balance computational tractability with sufficient information across a range of effect sizes, we construct the model using every fifth variant tested in each GWAS. The Pan-UK Biobank and Finngen meta-analysis reported only significant associations, so our ASH model was based on the Pan-UK Biobank data alone. Because these standard errors are generally larger than those of the meta-analysis, this strategy is conservative and will lead to a greater reduction in the effect size estimates of genome-wide significant associations. Our model was a mixture of normal distributions with an additional pointmass at 0. A separate model was calculated for each GWAS, and then used to re-estimate effect sizes. *Ashr* produces a posterior distribution of effect sizes for each variant. In our analysis we used the median of the distribution, but the results did not change when we tested randomly sampling from each distribution instead.

A.3 Selecting variants for analysis

Each plot and model test uses linkage-independent variants. These were selected with a stepwise procedure. The most significant variant is selected, and all other variants within 100 kb are removed from the data. The most significant variant remaining is selected, and neighboring variants are again removed. This process is repeated until no two significant variants fall within 100 kb of one another. The stepwise method prioritizes variants by their original significance; prioritizing based on ASH-correct significance would create a bias toward variants with higher minor allele frequencies. Traits based on Pan-UK Biobank and Finngen meta-analysis use a window size of 500 kb, as these are the publicly reported summary statistics.

A.4 Risk allele frequencies are robust to linkage disequilibrium

Most trait-associated variants in GWAS are not themselves causative but arise due to tagging of a causative variant by LD; furthermore, the fine mapping task of identifying causative variants is not straightforward. Our selected LD-independent variants thus serve as proxies for causative variants driving associations at the respective loci. A key concern is then how well the proxy allele

frequencies and effect sizes used in inference approximate those of the respective causative variants. In particular, while it is intuitive and well known that variants in tight LD have similar minor allele frequencies and similar absolute effect sizes, a question remains as to sign, namely whether the *risk* allele frequency might "flip", for example from around 0.1 to around 0.9. The answer is that it does not. We provide both an analytic argument and an empirical demonstration.

The analytic argument rests on the observation that, with respect to a given trait, *a causative risk allele and a tagging risk allele are in positive LD.* In detail, let X_a, X_b be causative and tagging variants respectively, coded $\{0, 1, 2\}$ as usual with respect to reference/alternate alleles A/a, B/b defined so that the alternate allele is increasing for trait Y in the sense that the GWAS regression coefficient $\beta_a = \mathrm{Cov}(Y, X_a)/\mathrm{Var}(X_a) > 0$ and similarly $\beta_b > 0$ (for simplicity here we omit additional covariates). The signed LD is the Pearson correlation coefficient

$$r_{ab} = \mathrm{Corr}(X_a, X_b) = \frac{p_{ab} - p_a p_b}{\sqrt{p_a q_a p_b q_b}},$$

where $p_a = 1 - q_a$, etc. are allele frequencies and p_{ab} is the frequency of the ab haplotype. Writing $Y = \beta_a X_a + \varepsilon$ with $\mathrm{Cov}(X_a, \varepsilon) = \mathrm{Cov}(X_b, \varepsilon) = 0$ (for simplicity here we assume a single causal variant per locus), one finds $\beta_b = r_{ab}\sqrt{\frac{p_a q_a}{p_b q_b}}\beta_a$. It follows that $r_{ab} > 0$. Incidentally, we note that the effect size relationship simplifies nicely on the variance explained scale: with $V_a = 2p_a q_a \beta_a^2$ and $V_b = 2p_b q_b \beta_b^2$, we have $V_b = r_{ab}^2 V_a$.

The classical bounds on LD in terms of marginal allele frequencies may be summarized

$$-\sqrt{\frac{p_a p_b}{q_a q_b}} \vee -\sqrt{\frac{q_a q_b}{p_a p_b}} \;\leq\; r_{ab} \;\leq\; \sqrt{\frac{p_a q_b}{q_a p_b}} \wedge \sqrt{\frac{q_a p_b}{p_a q_b}};$$

these represent four inequalities, two of which are nontrivial in a given instance, and which saturate respectively when one of the four haplotype frequencies $p_{ab}, p_{AB}, p_{aB}, p_{Ab}$ drops to zero. In particular, $r_{ab} = 1$ only when $p_a = p_b$, while $r_{ab} = -1$ only when $p_a = q_b$. The former extends quantitatively to the case $r_{ab} > 0$, where we obtain the bounds

$$r_{ab}^2 \frac{p_a}{q_a} \;\leq\; \frac{p_b}{q_b} \;\leq\; \frac{1}{r_{ab}^2}\frac{p_a}{q_a}.$$

Inverting the monotonic function $p \mapsto p/(1-p)$ via $t \mapsto t/(1+t)$ yields bounds on p_b in terms of p_a and r_{ab}. For example, with $p_a = 0.2$ and $r_{ab}^2 = 0.9$ we find $0.18 \leq p_b \leq 0.22$; with $p_a = 0.2$ and $r_{ab}^2 = 0.5$ we find $0.11 \leq p_b \leq 0.34$.

For the empirical demonstration, we looked at actual GWAS results for human height, selecting a number of significantly associated loci at random. For each locus we selected the sentinel (most significantly associated) variant and then considered all variants in LD $r^2 > 0.6$ with the sentinel variant. For these variants we plot variance explained against risk allele frequency. As one can see in the sample loci in Figure S1, the plots for each locus cluster tightly along both axes; in particular we find no "shadow cluster" at the complementary allele frequency. Note that this demonstration recapitulates none of the simplifying assumptions from the analytic argument (covariates, causal architecture).

B Model and Likelihood

B.1 Observables and notation

The pipeline described in section A yields a set of approximately independent loci. Each locus has the accompanying summary statistics of allele frequency, estimated effect size, standard error,

and p-value. These are given for the variant at that locus with either the smallest p-value (sentinel variant). The allele frequency available for each locus is that of the trait-increasing or risk-increasing allele (RAF) (section A.4), which we denote using y. We treat the derived allele frequency, denoted by x, as a latent variable. The effect size estimated by the GWAS is denoted by $\hat{\beta}$, the posterior median after shrinkage by $\widetilde{\beta}$, and the true latent effect size by β. Standard errors and p-values are denoted by SE and p.

The variance contribution of a locus is defined as

$$v = 2\beta^2 y(1-y), \tag{1}$$

where we may also substitute x for y. Similarly, let $\hat{v}$ be the GWAS-estimated variance contribution, and $\widetilde{v}$ be the posterior median variance contribution after shrinkage. Ascertainment in GWAS based on a p-value threshold can be approximated as a variance-contribution threshold. If v^* is the threshold for a particular study and p-value, variants face an allele frequency threshold conditional on their effect size

$$x^* = \max\left(x_{\min}, \left(1 - \sqrt{1 - v^*/\hat{\beta}^2}\right)/2\right). \tag{2}$$

Even if the GWAS is powered to discover some rare variants, we apply the minor allele frequency threshold $x_{\min}$.

B.2 The allele frequency distribution under different models of selection

We first describe the structure of the model in the absence of any uncertainty associated with measurement through GWAS. This provides intuition and is also used in simulations to demonstrate the theoretical limitations of our approach. Our inferences are based on the RAF distribution, conditional on effect size

$$f(y|\beta) = \frac{\mathbb{P}(+ \mid \beta)\tau(y \mid \beta) + \mathbb{P}(- \mid \beta)\tau(1-y \mid -\beta)}{\int_{x^*}^{1-x^*} \mathbb{P}(+ \mid \beta)\tau(x \mid \beta) + \mathbb{P}(- \mid \beta)\tau(1-x \mid -\beta)\, dx}. \tag{3}$$

This distribution accounts for the fact that the ancestral/derived status of the trait-increasing allele is unknown. $\mathbb{P}(+ \mid \beta)$ is the probability that a new mutation with effect size magnitude $|\beta|$ is trait-increasing, while $\mathbb{P}(- \mid \beta) = 1 - \mathbb{P}(+ \mid \beta)$ is the probability this effect is trait-decreasing. $\tau(x \mid \beta)$ is the site-frequency spectrum (SFS) of derived allele frequencies with (signed) effect size β. In an equilibrium population this is proportional to the sojourn time a new mutation spends at frequency x, while in more general scenarios it gives us the expected number of variants at this frequency. How the SFS depends on β varies between different models of phenotypic selection. The denominator normalizes by the expected number of variants in the discoverable range defined by v^* to yield a probability density on the RAF. Normalization also removes any dependency on mutation rate under an infinite-sites model.

Describing a model of selection amounts to specifying $\mathbb{P}(+ \mid \beta)$ and $\tau(x \mid \beta)$. We do this for five models which we hope capture a broad range of possible phenotypic selection scenarios. The SFS, $\tau(x \mid \beta)$, is determined by the first two moments of the change in allele frequency. Directional selection at the phenotypic level generates additive selection at the genic level, while stabilizing selection at the phenotypic level generates underdominant selection at the genic level. We include one selection coefficient for each effect

$$\mathbb{E}[\Delta x] = sx(1-x) + s_{ud}x(1-x)(1/2 - x). \tag{4}$$

The variance is given by the standard formula for the Wright-Fisher model

$$\mathrm{Var}[\Delta x] = \frac{x(1-x)}{2N_e}. \tag{5}$$

B.2.1 Neutral

For a neutral model with a total absence of selection on the focal trait, we specify both that there is no relationship between effect sizes and selection coefficients and that all variants discovered by GWAS have selection coefficients of zero. We therefore have $\tau(x \mid \beta) = \tau(x)$ for any observed β, and $\tau(x) \propto 1/x$ at equilibrium. There is a question of how to set the mutational balance between $+$ and $-$ alleles under neutrality. We opt for $\mathbb{P}(+ \mid \beta) = \mathbb{P}(- \mid \beta) = 0.5$ because strong biases here would lead to rapid changes in the mean trait value over longer evolutionary periods due solely to mutational bias.

B.2.2 Directional selection

For a model of directional selection on the focal trait we consider the scenario where the optimum trait value exists outside of the range of phenotypes present in the population. Each mutation experiences selection proportional to its effect size:

$$s = I_1 \beta. \tag{6}$$

Under a directional model of selection on disease liability the optimal genotype would contain no risk-increasing alleles. Mutational bias is generated because most sites are more likely to be fixed for the risk-decreasing allele. The population mean sits off the optimum phenotype due to a polygenic mutational bias towards disease.

In an equilibrium population the probability a site with effect size β is fixed for the risk-decreasing allele is equivalent to the probability that a new mutation with that effect size is risk-increasing:

$$\mathbb{P}(+ \mid \beta) = \frac{1 - e^{4N_e s}}{1 - e^{8N_e s}}. \tag{7}$$

We combine this with the standard expression for the SFS under additive selection (section C) or one obtained numerically account for nonequilibrium demography (section D).

B.2.3 Single-trait stabilizing selection

Stabilizing selection applies to the situation where the population mean for focal trait is very close to the optimal value. Stabilizing selection will be approximately one-dimensional when only the focal trait and those strongly correlated with it are under selection. Each mutation receives a selection coefficient proportional to its squared effect size

$$s_{ud} = I_2 \beta^2, \tag{8}$$

and the form of this selection is underdominant rather than additive. We constrain the sign of I_2 so that selection is never overdominant. Because selection is symmetric with respect to the sign of β we have $\mathbb{P}(+ \mid \beta) = \mathbb{P}(- \mid \beta) = 0.5$. This also agrees with the intuition that if the population mean is close to the optimum there should not be a strong mutational bias

B.2.4 Combined directional and single-trait stabilizing selection

We also consider a model where both directional and stabilizing selection at the trait level make a contribution to the selection on individual alleles. Such a model may apply if the population mean is off the optimum value but not far enough that stabilizing selection can be ignored. It is less well-justified than the other models considered but could be useful in cases where selection

is relatively strong while mutational asymmetry is mild. We simply allow alleles to have both an additive and underdominant selection coefficient. The stabilizing component of selection remains symmetric in β, so $\mathbb{P}(+ \mid \beta)$ is determined only by the additive selection coefficient.

B.2.5 Pleiotropic stabilizing selection

The single-trait model of stabilizing selection arises when the variants influencing the focal trait, and those traits highly correlated with it, meaningfully affect no other traits under stabilizing selection. Simons *et al.* (31) developed a model where variants affect many other traits under stabilizing selection that are mutationally uncorrelated with the focal trait. We use the terms single-trait and pleiotropic to refer to the one-dimensional and high-dimensional limits of this model. In the limit as the number of uncorrelated traits becomes large, the effect size on the focal trait conditional on the selection coefficient is

$$\beta | s_{ud} \sim N(0, c s_{ud}), \tag{9}$$

where c depends on the number of traits and the strength of selection on each. Because our likelihood is formulated conditional on the effect size, to use the high-dimensional model of pleiotropy we must make assumptions about the distribution of fitness effects of new mutations (DFE). Rather than attempting to also fit this distribution using the data, we put a flat prior on $\log(s_{ud})$

$$f(s_{ud}) \propto 1/s_{ud}. \tag{10}$$

This yields a Lévy distribution on s_{ud} conditional on β

$$s_{ud} | \beta \sim \text{Lévy}(I_p \beta^2). \tag{11}$$

The parameter I_p is analogous to the selection intensities I_1 and I_2 defined above for directional selection and single-trait stabilizing selection. We can connect this parameter back to the average fitness effects of mutations with particular trait effects through

$$\text{Median}\left(s_{ud} | \beta^2\right) = \frac{\beta^2 \mathcal{I}_p}{2\text{erfc}^{-1}(1/2)^2} \approx 0.455 \beta^2 I_p. \tag{12}$$

To obtain the SFS we integrate over the distribution of selection coefficients conditional on the effect size

$$\tau(x \mid \beta) = \int_0^\infty \tau(x \mid s_{ud}) f(s_{ud} \mid \beta^2; I_p)\, ds_{ud}. \tag{13}$$

By the same logic given for one-dimensional stabilizing selection, we also assume equal proportions of trait-increasing and trait-decreasing mutations.

B.3 GWAS and ascertainment bias

The truncated RAF distribution in equation (3) assumes that if a variant with true effect size β passes the minor allele frequency threshold x^* defined by v^* it will be observed in the GWAS sample. In reality, variants that pass either a p-value or variance contribution threshold will be subject to winner's curse and estimates of their effect sizes ($\hat{\beta}$) will be biased upwards. We developed strategies to partially account for the effect of winner's curse while avoiding fitting the distribution of effect sizes and its dependence on allele frequency. We also performed simulations to investigate some of the impacts of ignoring this ascertainment bias.

B.3.1 Choosing the variance contribution threshold

In order to truncate the frequency distribution and ultimately calculate the likelihood contribution of each locus, we need to compute the variance contribution threshold v^*. The relationship between variance contributions and p-values is approximately

$$v = F^{-1}_{\chi^2_1}(1 - p\text{-value})/n_{\text{eff}}, \tag{14}$$

where $F^{-1}_{\chi^2_1}$ is the inverse CDF of a chi-squared distribution with one degree of freedom and n_{eff} is the effective sample size of the GWAS. In order to find a v^* corresponding to a particular p-value threshold, we first calculated the median n_{eff} for the GWAS by using the observed set of $\{v_i, p\text{-value}_i\}$ value to calculate a local $n_{\text{eff},i}$ for each genome-wide significant variant in the processed summary statistics: $n_{\text{eff},i} = F^{-1}_{\chi^2_1}(1 - p\text{-value}_i)/v_i$ Using the median n_{eff}, we compute v^* for various p-value thresholds using equation (14).

The relationship between p-values and effective sample sizes for each trait is shown in Figure S3. A single effective sample size is not truly sufficient because the GWAS regressions used to estimate effect sizes and compute p-values include confounding variables as well as genetic structure (e.g. principal components). Variants that are correlated with these covariates will require a large sample for their effects to be detected and therefore have a lower effective sample size. We use this procedure to generate a v^* that roughly corresponds to a p-value threshold. In addition to the variance contribution threshold, we also removed variants that do no satisfy the corresponding p-value threshold. Applying the p-value threshold removes variants with large standard errors relative to their estimated effect sizes, while making sure that they fall within the frequency range defined by v^*.

B.3.2 Winner's curse strategies

The first strategy simply ignores winner's curse and treats $\hat{\beta}$ as if it were the true effect size. This strategy uses overestimates of effect sizes, and more so for low frequency variants with larger standard errors where winner's cures will have a stronger effect (Figure S4). We simply substitute $\hat{\beta}$ for β in equation (3)

$$f(y|\hat{\beta}) = \frac{\mathbb{P}(+ \mid \hat{\beta})\tau(y \mid \hat{\beta}) + \mathbb{P}(- \mid \hat{\beta})\tau(1 - y \mid -\hat{\beta})}{\int_{x^*}^{1-x^*} \mathbb{P}(+ \mid \hat{\beta})\tau(x \mid \hat{\beta}) + \mathbb{P}(- \mid \hat{\beta})\tau(1 - x \mid -\hat{\beta})\, dx}. \tag{15}$$

The second strategy treats the shrunken effect size (*ash* posterior median) $\widetilde{\beta}$ as the true effect size, while still using $\hat{\beta}$ to calculate the frequency cutoff x^*.

$$f(y \mid \hat{\beta}, \text{SE}) = \frac{\mathbb{P}(+ \mid \widetilde{\beta})\tau(y \mid \widetilde{\beta}) + \mathbb{P}(- \mid \widetilde{\beta})\tau(1 - y \mid -\widetilde{\beta})}{\int_{x^*(\widetilde{\beta})}^{1-x^*(\hat{\beta})} \mathbb{P}(+ \mid \widetilde{\beta})\tau(x \mid \widetilde{\beta}) + \mathbb{P}(- \mid \widetilde{\beta})\tau(1 - x \mid -\widetilde{\beta})\, dx}. \tag{16}$$

The effects of shrinkage on ascertained variants for all traits is shown in Figure S5 for schizophrenia. The degree of shrinkage depends on its standard error, which is influenced by frequency and local effective sample size.

If ascertained variants are under selection, then we would expect that higher frequency variants have lower effect sizes on average. Prior distributions were fit to pruned variants with minor allele frequencies greater than 0.01. The posterior distribution would therefore be overestimated for high frequency variants and underestimated for low frequency variants. This should have a flattening

effect on the effect size - frequency distribution and lead to conservative inferences of selection. Ultimately, because we do not make inferences about either the underlying distribution or effect sizes or the DFE, our approach must make a guess about the true effect size at each locus.

The above strategies describe different ways to account for uncertainty in β. However, even if we knew the true value of β there would still be a distortion to the frequency distribution if the variance contribution threshold is based on $\hat{v} = 2\hat{\beta}^2 x(1-x)$. We can see this by also conditioning on this event

$$f(y \mid \beta, \hat{v} > v^*) = \frac{f(y, \hat{v} > v^* \mid \beta)}{\mathbb{P}(\hat{v} > v^* \mid \beta)}. \tag{17}$$

The RAF density should in theory be adjusted by the probability that the estimated effect size exceeds a threshold dependent on y

$$f(y, \hat{v} > v^* \mid \beta) = f(y \mid \beta)\, \mathbb{P}\left(|\hat{\beta}| > \sqrt{\frac{v^*}{2y(1-y)}} \mid y, \beta\right). \tag{18}$$

GWAS effect sizes should be approximately normally distributed conditional on the true effect size, frequency, and effective sample size.

$$\hat{\beta} \sim N\left(\beta, 1/\sqrt{2n_{\text{eff}} x(1-x)}\right) \tag{19}$$

In practice, we do not make use of this adjustment and rely on simulations to evaluate the extent of the bias that results from ignoring it.

B.4 Likelihoods and model comparison

For each trait, the data consist of a set of RAFs and effect sizes $\{y_i, \hat{\beta}_i\}$ after filtering based on a variance contribution, and possibly also p-value, threshold. Depending on the model, the parameters Θ are either empty (neutral), I_1 (directional), I_2 (single-trait / one-dimensional stabilizing), (I_1, I_2) (directional plus stabilizing), or I_p (pleiotropic / high-dimensional stabilizing). Assuming independence among loci, conditional on model parameters, the likelihood is

$$f(\{y_i, \hat{\beta}_i\} \mid \Theta) = \prod_{i=1}^{m} f(y_i \mid \hat{\beta}, \text{SE}_i), \tag{20}$$

where m is the number of loci passing the threshold.

Since the models only contain either one or two parameters, we calculate likelihoods over a grid of parameter values chosen to span the range from where all ascertained variants are neutral, to that where all ascertained variants are under strong selection. For I_1 and I_2, the maximum value was chosen so that $|2N_e I_1^{\max} \hat{\beta}_{\min}| = 1\,000$ and $|2N_e I_2^{\max} \hat{\beta}^2_{\min}| = 1\,000$, while the minimum values were chosen so that $|2N_e I_1^{\min} \hat{\beta}_{\max}| = 0.01$ and $|2N_e I_2^{\min} \hat{\beta}^2_{\max}| = 0.01$. Both positive and negative values of I_1 were included. For high-dimensional stabilizing selection, a larger range of I_p was considered to account for the variance in β values given a particular selection coefficient: $|2N_e I_p^{\max} \hat{\beta}^2_{\min}/F^{-1}(0.995)^2| = 1\,000$ and $|2N_e I_2^{\min} \hat{\beta}^2_{\max}/F^{-1}(0.495)^2| = 0.01$, where F^{-1} is the inverse CDF of the standard normal distribution. We used a grid size of 1 000 for I_1 and I_2 for models of directional and one-dimensional stabilizing selection, a grid of size 100 for both I_1 and I_2 in the model combining directional and stabilizing selection, and a grid size of 2 000 for I_p in the model of high-dimensional stabilizing selection.

All models considered here converge to neutrality as the selection intensity parameters become small. We therefore use likelihood ratio tests to compute p-values for each model relative to the

neutral model. To compare models of selection we directly compare likelihoods by computing Akaike information criteria.

$$-\Delta\text{AIC}_{\text{model 1}-\text{model 2}} = 2\log f(\{y_i, \hat{\beta}_i\} \mid \hat{\Theta}_1) - 2\log f(\{y_i, \hat{\beta}_i\} \mid \hat{\Theta}_2) - 2k_1 + 2k_2, \tag{21}$$

where $\hat{\Theta}_1$ and $\hat{\Theta}_2$ are the maximum likelihood parameter values and k_1 and k_2 are the numbers of free parameters. ΔAIC values presented in figures are on the scale of twice the log-likelihood differences.

C The equilibrium SFS under additive and underdominant selection

Although not used in any of the main results, we present some equilibrium expressions for $\tau(x|\beta)$. The primary reason for this is to give a solution when both additive and underdominant selection are acting on a single mutation.

The equilibrium site frequency spectrum under both directional and stabilizing selection can be derived by computing the sojourn time that a mutation spends at each frequency before loss or fixation (Ewens, 2004). The infinitesimal mean and variance are

$$\mathbb{E}[\delta x] \approx a(x)\delta t$$

and

$$\text{Var}[\delta x] \approx b(x)\delta t.$$

Following equation (4), the expected change in allele frequency per generation is

$$a(x) = (s + s_{ud}(1/2 - x))x(1 - x),$$

while the variance in allele frequency change due to binomial Wright-Fisher sampling is (equation (5))

$$b(x) = \frac{x(1-x)}{2N}.$$

Measuring time in units of $2N_e$ generations we have

$$a(x) = (2N_e s + 2N_e s_{ud}(1/2 - x))x(1 - x),$$

and

$$b(x) = x(1 - x).$$

To simplify algebra, let $\sigma_1 = 2N_e s$ and $\sigma_2 = 2N_e s_{ud}$:

$$a(x) = (\sigma_1 + \sigma_2(1/2 - x))x(1 - x). \tag{22}$$

From equation (4.23) in Ewens (2004), the sojourn time at frequency x of a mutation with initial frequency p is

$$\tau(x, p) = 2P_1(p)[b(x)\psi(x)]^{-1} \int_x^1 \psi(y)\, dy, \tag{23}$$

where $P_1(p)$ is the probability of fixation from starting frequency p.

$$P_1(p) = \frac{\int_0^p \psi(y)\, dy}{\int_0^1 \psi(y)\, dy}, \tag{24}$$

where

$$\psi(x) = \exp\left(-2\int_0^x \frac{a(y)}{b(y)}\,dy\right). \tag{25}$$

For our model

$$\psi(x) = \exp\left(-2\int_0^x \frac{(\sigma_1 + \sigma_2(1/2 - y))y(1-y)}{y(1-y}\,dy\right) \tag{26}$$

$$= \exp\left(-2\sigma_1 x - \sigma_2 x(1-x)\right). \tag{27}$$

The indefinite integral is

$$\int \psi(x)\,dx = -\exp\left(\frac{-(2\sigma_1+\sigma_2)^2}{4\sigma_2}\right)\frac{\sqrt{\pi}}{2\sqrt{|\sigma_2|}}\mathrm{erf}\left(\frac{-2\sigma_1+\sigma_2(2x-1)}{2\sqrt{|\sigma_2|}}\right). \tag{28}$$

and the fixation probability

$$P_1(p) = \frac{\mathrm{erf}\left(\frac{-2\sigma_1-\sigma_2}{2\sqrt{|\sigma_2|}}\right) - \mathrm{erf}\left(\frac{-2\sigma_1+\sigma_2(2p-1)}{2\sqrt{|\sigma_2|}}\right)}{\mathrm{erf}\left(\frac{-2\sigma_1-\sigma_2}{2\sqrt{|\sigma_2|}}\right) - \mathrm{erf}\left(\frac{-2\sigma_1+\sigma_2}{2\sqrt{|\sigma_2|}}\right)}. \tag{29}$$

The other part of the expression for sojourn time equation is

$$\int_x^1 \psi(y)\,dy =$$
$$\exp\left(\frac{-(2\sigma_1+\sigma_2)^2}{4\sigma_2}\right)\frac{\sqrt{\pi}}{2\sqrt{|\sigma_2|}}\left[\mathrm{erf}\left(\frac{-2\sigma_1+\sigma_2(2x-1)}{2\sqrt{|\sigma_2|}}\right) - \mathrm{erf}\left(\frac{-2\sigma_1+\sigma_2}{2\sqrt{|\sigma_2|}}\right)\right]. \tag{30}$$

To convert the sojourn time to a frequency spectrum that gives the intensity of derived mutations in any frequency range, note that mutations enter the population at a rate $2NL\mu$ per generation at frequency $\frac{1}{2N}$. This corresponds to a rate of $2N \times 2NL\mu$ mutations per $2N$ generations. The intensity of mutations at a frequency x is the rate at which mutations enter the population times the sojourn time at x. Take the limit where $2NL\mu \to \theta$ as $N \to \infty$. $P_1(\frac{1}{2N})$ is the portion of $\tau(x, \frac{1}{2N})$ that depends on N.

$$\lim_{N\to\infty} 2NP_1(1/2N) = \frac{\frac{2\sqrt{|\sigma_2|}}{\sqrt{\pi}}\exp\left(\frac{(2\sigma_1+\sigma_2)^2}{4\sigma_2}\right)}{\mathrm{erf}\left(\frac{-2\sigma_1-\sigma_2}{2\sqrt{|\sigma_2|}}\right) - \mathrm{erf}\left(\frac{-2\sigma_1+\sigma_2}{2\sqrt{|\sigma_2|}}\right)}. \tag{31}$$

Plugging all this into equation (23) and multiplying by θ we get

$$\tau(x \mid \sigma_1, \sigma_2) = 2\theta\frac{\exp(2\sigma_1 x + \sigma_2 x(1-x))}{x(1-x)}\frac{\mathrm{erf}\left(\frac{-2\sigma_1+\sigma_2(2x-1)}{2\sqrt{|\sigma_2|}}\right) - \mathrm{erf}\left(\frac{-2\sigma_1+\sigma_2}{2\sqrt{|\sigma_2|}}\right)}{\mathrm{erf}\left(\frac{-2\sigma_1-\sigma_2}{2\sqrt{|\sigma_2|}}\right) - \mathrm{erf}\left(\frac{-2\sigma_1+\sigma_2}{2\sqrt{|\sigma_2|}}\right)} \tag{32}$$

In the case where only directional selection is acting ($\sigma_2 = 0$) we have the usual expression for the equilibrium SFS under additive selection.

$$\tau(x \mid \sigma_1) = 2\theta\frac{\exp(2\sigma_1 x)}{x(1-x)}\frac{1-\exp(2\sigma_1(1-x))}{1-\exp(2\sigma_1)} \tag{33}$$

Alternatively, when only stabilizing selection is acting (σ_1=0) we have

$$\tau(x \mid \sigma_2) = 2\theta \frac{\exp(\sigma_2 x(1-x))}{x(1-x)} \left(1 + \frac{\mathrm{erf}(\sqrt{|\sigma_2|}(\frac{1}{2} - x))}{\mathrm{erf}(\sqrt{|\sigma_2|}/2)}\right). \tag{34}$$

Equation (32) interpolates between the fully additive (equation (33)) and fully underdominant (equation (34) extremes (Figure S6). Because there are now two selection coefficients, equation (32) is really just an expression for the SFS with arbitrary dominance, though the form is helpful for thinking about the different effects of directional versus stabilizing selection.

C.1 Computation

There are aspects of (32) that make it a difficult in computation. Consider the scenario in which $\sigma_1 < 0$ and $\sigma_2 << \sigma_1$ (directional selection dominates and induces negative selection on the derived variant). The error function rapidly approaches one for values greater than two ($\mathrm{erf}(3) \approx 0.99998$) and we have to take the quotient of two very small numbers. For $x > 1$ we can compute the error function using the series

$$\mathrm{erf}(x) = 1 - \frac{e^{-x^2}}{\sqrt{\pi}} \left(x^{-1} - \frac{1}{2}x^{-3} + \frac{3}{4}x^{-5} + \dots\right). \tag{35}$$

For large x

$$\mathrm{erf}(x) \approx 1 - \frac{e^{-x^2}}{\sqrt{\pi}x}. \tag{36}$$

We can approximate the error function portion of equation (32) as

$$\frac{\frac{e^{-[\sigma_1/\eta - \frac{1}{2}\eta]^2}}{-2\sigma_1 - \eta^2} - \frac{e^{-[\sigma_1/\eta - \eta(x-\frac{1}{2})]^2}}{-2\sigma_1 - \eta^2(2x-1)}}{\frac{e^{-[\sigma_1/\eta - \frac{1}{2}\eta]^2}}{-2\sigma_1 - \eta^2} - \frac{e^{-[\sigma_1/\eta + \frac{1}{2}\eta]^2}}{-2\sigma_1 + \eta^2}}, \tag{37}$$

where $\eta = \sqrt{|\sigma_2|}$. Some algebra gives

$$\frac{\frac{1}{-2\sigma_1 - \eta^2} - \frac{e^{2\sigma_1(1-x) + \eta^2 x(1-x)}}{-2\sigma_1 - \eta^2(2x-1)}}{\frac{1}{-2\sigma_1 - \eta^2} - \frac{e^{2\sigma_1}}{-2\sigma_1 + \eta^2}}. \tag{38}$$

Importantly, as $\eta \to 0$ and only directional selection operates this reduces to

$$\frac{1 - e^{2\sigma_1(1-x)}}{1 - e^{2\sigma_1}}, \tag{39}$$

which when plugged into equation (32) gives the usual site frequency spectrum under directional selection.

$$\tau(x) = 2\theta \frac{e^{2\sigma_1 x}}{x(1-x)} \frac{1 - e^{2\sigma_1(1-x)}}{1 - e^{2\sigma_1}}.$$

This verifies that equation (32) is correct and suggests a way to stably compute the error function term.

Rewrite the error function term as

$$\frac{\mathrm{erf}(-\sigma_1/\eta - \eta(x - \frac{1}{2})) - \mathrm{erf}(-\sigma_1/\eta - \frac{1}{2}\eta)}{\mathrm{erf}(-\sigma_1/\eta + \frac{1}{2}\eta) - \mathrm{erf}(-\sigma_1/\eta - \frac{1}{2}\eta)}. \tag{40}$$

Computation will be unstable if $-\sigma_1/\eta+\frac{1}{2}\eta$ and $-\sigma_1/\eta+\frac{1}{2}\eta$ have the same sign and if$|-\sigma_1/\eta+\frac{1}{2}\eta|$ and $|-\sigma_1/\eta-\frac{1}{2}\eta|$ are large. This is the situation when directional selection dominates. If stabilizing selection is strong both terms in the denominator have the same sign and there is no divide-by-zero issue. If we Extend the series approximation to the error function for $x > 1$, then we can rewrite the error function term as

$$\frac{H(B) - H(A)\exp(2\sigma_1(1-x) + \eta^2 x(1-x))}{H(B) - H(C)\exp(2\sigma_1)}, \tag{41}$$

where

$$\begin{aligned}
H(X) &= x^{-1} - \frac{1}{2}x^{-3} + \frac{3}{4}x^{-5} + \dots \\
A &= -\sigma_1/\eta - \eta(x - \frac{1}{2}) \\
B &= -\sigma_1/\eta - \frac{1}{2}\eta \\
C &= -\sigma_1/\eta + \frac{1}{2}\eta.
\end{aligned}$$

The overall site frequency spectrum is therefore

$$\begin{aligned}
\tau(x) &\approx 2\theta \frac{\exp(2\sigma_1 x - \eta^2 x(1-x))}{x(1-x)} \frac{H(B) - H(A)\exp(2\sigma_1(1-x) + \eta^2 x(1-x))}{H(B) - H(C)\exp(2\sigma_1)} \\
&\approx 2\theta \frac{H(B)\exp(2\sigma_1 x - \eta^2 x(1-x)) - H(A)\exp(2\sigma_1)}{x(1-x)[H(B) - H(C)\exp(2\sigma_1)]}.
\end{aligned}$$

When directional selection on the derived variant is positive this will also be numerically unstable and we are better off with

$$\tau(x) \approx 2\theta \frac{H(B)\exp(-2\sigma_1(1-x) - \eta^2 x(1-x)) - H(A)}{x(1-x)[H(B)\exp(-2\sigma_1) - H(C)]}.$$

D Non-equilibrium demography

The GWAS from which we obtained summary statistics for our analyses were all performed using individuals of predominantly European ancestry. The demographic history shaping the genealogical relationships among these individuals and their genomes is complex and not well-described by an equilibrium population model. Two major events shaping current distributions of allele frequencies in contemporary European genomes are the out-of-Africa (OOA) bottleneck and more recent population growth. To account for these, we computed likelihoods using a demographic model of non-Finnish European history fit by Tennessen *et al.* (32) (Figure S7).

We computed the site frequency spectrum under this model by iterating the discrete-time Wright-Fisher transition matrix. To speed up computations we rescaled time and population size downwards by a factor of 20. We computed the SFS for a grid of (s, s_{ud}) values such that the $2N_e s$ and $2N_e s_{ud}$ values at the oldest ancestral human population size ranged between 0.1 and 1 000. 100 grid points were used for each selection coefficient. To calculate the SFS for a particular value of (s, s_{ud}) during inference we look up values from this grid and use linear interpolation.

E Simulations

E.1 Trait simulations

To evaluate the ability of these statistics to distinguish between different models of selection we performed simulations designed to provide approximately realistic distributions of effect sizes and allele frequencies. Our simulation strategy used the prior distribution of effect sizes, $\mathbb{P}(\beta)$, fit using *ash* for a given trait. We fix the number of ascertained variants at some number m and simulate allele frequencies and effect sizes until it is reached. We first sample a set of effect sizes $\{\beta_i\}$ from $\mathbb{P}(\beta)$ and compute the allele frequency distribution $f(y|\beta_i)$ (equation (3)) for the selection intensity parameters (I_1, I_2, I_p) corresponding to the simulation. Rather than using the MAF cutoff x^* corresponding to a variance contribution threshold v^*, we first apply a global MAF threshold of $x^* = 0.01$. This is equivalent to the assumption that $\mathbb{P}(\beta)$ represents the effect size distribution of all alleles with MAF greater than x^*. Using $f(y|\beta_i)$, we sample trait-increasing allele frequencies from the range $(x^*, 1 - x^*)$, yielding a large set of $\{y_i, \beta_i\}$.

We take variants from the set of allele frequencies and effect sizes to approximate GWAS ascertainment using a variance-contribution threshold. For simulations without ascertainment bias, we simply apply v^* to $2\beta_i^2 x_i(1 - x_i)$, where v^* corresponds to the given trait and a particular p-value threshold. Otherwise, we add GWAS noise to β_i using

$$\hat{\beta}_i \sim N(\beta_i, 1/(2n_{eff}x_i(1 - x_i)),$$

and we assume that the correct standard error $\mathrm{SE}_i = \sqrt{1/(2n_{eff}x_i(1 - x_i)}$ is available as a summary statistic. Variants are then added to the ascertainment set when $2\hat{\beta}_i^2 x_i(1-x_i) > v^*$. We add variants until the m is reached, and draw new β_i from the prior and repeat the process if not enough variants from the current set pass ascertainment.

When calculating likelihoods using simulated data we face the choice of what to treat at the true effect size. In simulations without ascertainment bias this is always the true effect size β_i. In simulations with ascertainment bias we calculate likelihoods using β_i, $\hat{\beta}_i$, and the posterior median calculated using the known $\mathbb{P}(\beta)$ and SE_i for that simulation. Making calculations for each of these allow us to investigate the effects of ascertainment bias on allele frequencies and uncertainty in effect sizes somewhat separately. Simulation results were qualitatively similar for different traits and we present results only for IBD.

E.2 Simulations under Simons et al. (2022) DFE

Simons *et al.* (22) found that estimated DFEs for different quantitative traits have similar shapes. They presented a single shared distribution (SSD) that fit the data from all analyzed traits reasonably well (Figure S16). We performed simulations using this DFE to evaluate our model's ability to discriminate between one-dimensional and high-dimensional stabilizing selection. We sampled $\mathrm{s_{ud}}$ from $\mathrm{f_{SSD}}$ then sampled a derived allele frequency using either a constant effective population size or a demographic model. Under the high-dimensional pleiotropy model (31), the effect size (β) conditional on $\mathrm{s_{ud}}$ is

$$\beta \sim N\left(0, \frac{w^2}{n}\,\mathrm{s_{ud}}\right).$$

We transformed β so that variance contributions would be in units of scaled selection coefficients

$$\beta_s := \beta\sqrt{2N\frac{n}{w^2}}, \tag{42}$$

and

$$\beta_s \sim N\left(0, 2N\,\mathrm{s_{ud}}\right).$$

The variance contribution of an allele with effect size β_s is $v_s = 2\beta_s^2 x(1-x) = 2vNn/w^2$. Again following Simons *et al.* (31), the expected variance contribution in trait units of a pleiotropic locus under strong selection is

$$\gamma = 2\frac{w^2}{nN}.$$

We model ascertainment by keeping alleles with a variance contribution exceeding some threshold v^*. To aid with interpretation, we choose this threshold relative to the expected contributions from strongly selected alleles:

$$v^* = c\gamma.$$

On the selection-transformed scale this threshold is

$$v_s^* = 4c.$$

To simulate GWAS noise and ascertainment bias in the sample, we also need to simulate $\hat{\beta}_s|\beta_s$. GWAS noise can be modeled as being approximately normally distributed so that

$$\hat{\beta} \sim N\left(\beta_s, \frac{V_p/m}{2x(1-x)}\right),$$

where V_p is the phenotypic variance in the population and m is the sample size. For transformed effect sizes this is

$$\hat{\beta}_s \sim N\left(\beta_s, \frac{4}{\gamma}\frac{V_p/m}{2x(1-x)}\right).$$

The approximate variance contribution threshold corresponding to a particular p-value p^* is therefore

$$v_s^* = \frac{4}{\gamma}\frac{V_p}{m}2\left(\mathrm{erf}^{-1}(1-p^*)\right)^2.$$

Since we have already specified that $v_S^* = 4c$, we can define a scaled effective sample size n_{eff}^s such that

$$n_{eff}^s := \frac{\gamma}{4}\frac{m}{V_p} = \frac{\left(\mathrm{erf}^{-1}(1-p^*)\right)^2}{2c}. \tag{43}$$

GWAS noise was then simulated as $\hat{\beta}_s \sim N(\beta_s, (2x(1-x)n_{eff}^s)^{-1})$ and alleles with $\hat{v}_s$ exceeding v_s^* were retained as part of the sample. All of the above holds for the case of single trait (one-dimensional) stabilizing selection, $\beta_s = \pm\sqrt{2N\,\mathrm{s_{ud}}}$, with trait-increasing and trait-decreasing alleles equally likely. For the demographic model of non-Finnish European history we let N be the ancestral population size.

To create a sample of $\{y_i, \beta_i\}$ we used a similar approach to the trait-based simulations but instead first sampled selection coefficients from $\mathrm{f_{SSD}}(\mathrm{s_{ud}})$. This was weighted by the probability of observing an allele with that selection coefficient above the MAF threshold $x^* = 0.01$: $\int_{x^*}^{1-x^*} \tau(x|2N\,\mathrm{s_{ud}})\,d\,\mathrm{s_{ud}}$. We then sampled a frequency from the truncated SFS, β_i from $N\left(0, 2N\,\mathrm{s_{ud}}\right)$ and $\hat{\beta}_i$ from $N(\beta_s, (2x(1-x)n_{eff}^s)^{-1})$. We computed likelihoods without ascertainment, and with ascertainment bias only treating $\hat{\beta}$ as the true effect size. We used a p-value threshold of 5×10^{-8}, v_s^* at $(1,3,5,7)$, and 500 ascertained variants in each simulation. Increasing v_s^* while holding the p-value threshold and number of ascertained variants constant corresponds to decreasing the power of the GWAS while increasing the mutational target size to ensure that a large number of strongly

selected variants are ascertained. In simulations with ascertainment bias this also increased the effect of winner's curse by lowering n_{eff}.

DFE-based simulations were used to evaluate the power to distinguish the low- and high-dimensional models of stabilizing selection. All simulations showed strong evidence ($-\Delta\text{AIC} > 50$) to reject a neutral model of selection on GWAS variants (Figure S10). The evidence for selection and power to detect the pleiotropy of stabilizing selection were higher in simulations using a constant population size ($Ne = 10\,000$), likely due to the effect of the out-of-Africa bottleneck allowing alleles under moderate selection with moderate effect sizes to drift to higher frequencies (Figure S11). However, even under a constant population size a substantial fraction of simulations would fail to show strong evidence for the true pleiotropy model. We also used trait-based simulations to evaluate the power to detect the true pleiotropy model and found that it was limited until the overall evidence for selection became strong (I_2 or I_p large, $-\Delta\text{AIC} > 100$) (Figure S13).

The distribution $p(\mathrm{s_{ud}} \mid \beta)$ used to compute the likelihood in the pleiotropic (high-dimensional) model of stabilizing selection assumed a flat (improper) prior on $\log \mathrm{s_{ud}}$. We checked whether results were robust to this choice by instead using the SSD as a prior and calculating likelihoods for DFE-based simulations. Changing the prior had little effect on the likelihood even though this was the prior used to generate the data (Figure S12).

F The distribution of variance contributions

Simons *et al.* (31) derived the distribution of variance contributions exceeding the threshold v^* under the single-trait and pleiotropic models of stabilizing selection. These results apply to the case when selection is strong and the population is at equilibrium. They argued that, because strongly selected variants should be discovered first by GWAS, the likelihoods of the observed genome-wide significant v values under the two models could provide a test for whether selection is highly pleiotropic or not. We implemented this test and applied it to simulated GWAS data from both trait- and DFE-based simulations as well as to genome-wide significant variants from all traits. We also derived the limiting form of the tail distribution for both models as the variance contribution threshold becomes large relative to the average variance contribution of a strongly site.

Following Simons *et al.* (31), the distribution of variance contributions for strongly selected sites under the single-trait stabilizing selection model is

$$\rho(v) \propto \frac{\exp(-2v/\gamma)}{v}, \tag{44}$$

where γ is the expected variance contribution of a site under strong selection. Because only variants with variance contribution exceeding v^* are ascertained, it make sense to work with the distribution of $z = \frac{v-v^*}{v^*}$ so that

$$f(z) \propto \frac{\exp\left(-2\frac{v^*}{\gamma}(z+1)\right)}{z+1}. \tag{45}$$

Also define $\eta = \frac{v^*}{\gamma}$:

$$f(z) \propto \frac{\exp\left(-2\eta(z+1)\right)}{z+1}. \tag{46}$$

Applying the same approach to the pleiotropic variance distribution

$$\rho(v) \propto \frac{\exp(-2\sqrt{v/\gamma})}{v}, \tag{47}$$

we get

$$f(z) \propto \frac{\exp\left(-2\sqrt{\eta(z+1)}\right)}{z+1}. \tag{48}$$

F.1 Simulation results

The inclusion of variants under weaker selection or at higher frequencies violates the assumptions leading to the distributions in equations (46) and (48). The assumption of a constant population size is also not consistent with allele frequencies in humans. We applied the Simons et al. (2018) test to simulated data from trait- and DFE-based simulations in order to determine its robustness to these violation. To focus only on the effects of changes to the population genetics model, we only used simulations free of ascertainment bias. For each simulation, we calculated the maximum likelihood value of η for both the one- and high-dimensional distributions and compared likelihoods.

At the neutral end of trait-based simulations using the IBD-derived distribution of effect sizes, the high-dimensional pleiotropy model always had a significantly higher likelihood (Figure S15). For strong selection in one-dimensional simulations, the log-likelihood difference between the models decreased towards zero. In high-dimensional simulations the evidence for the high-dimensional model also decreased as selection became strong. In DFE-based simulations a comparable bias towards the high-dimensional model was also observed (Figure S16). Overall, the inclusion of high-frequency variants strongly favored the model of high-dimensional stabilizing selection. These appear in our simulation data due to a combination of including alleles under weak/moderate selection and the out-of-Africa bottleneck allowing some strongly selected variants to drift to high frequencies. That the inclusion of high-frequency variants would favor the high-dimensional model makes intuitive sense, as this model allows the possibility of occasional large effect variants experiencing weaker selection when their effects on pleiotropic traits are small by chance.

F.2 Application to trait data

We observed similar results when applying the Simons *et al.* (31) model to estimated variance contributions from our set of traits using the v^* corresponding to p–value $= 5 \times 10^{-8}$ as well as the p–value cutoff itself. The high-dimensional model had a substantially higher likelihood for all traits, even those like breast cancer and IBD where there we detected no significant deviation from neutrality (Figure S14). For some traits, like schizophrenia and systolic/diastolic blood pressure, the variance contribution distribution fit the data reasonably well, but for most traits we observed a fatter tail of large variance contributions than in the best-fit one- and high dimensional models.

F.3 Limiting distributions

We derived limiting forms for the distributions of variance contributions as the variance contribution threshold becomes large relative to the average variance contribution of a strongly site. That is, as η become large. GWAS will likely only uncover the tail of strongly selected mutations affecting a given trait, so it makes sense to ask how the distributions corresponding to the two models of pleiotropy differ in their tail behavior.

Starting with the one-dimensional model, calculating the normalizing constant for equation (46) yields an incomplete gamma function $\Gamma(0, 2\eta)$. When η is large, this is approximately

$$\Gamma(0, 2\eta) \approx \frac{\exp(-2\eta)}{2\eta}. \tag{49}$$

The density of z can then be approximated as

$$f(z) \approx \frac{2\eta \exp\left(-2\eta(z+1)\right)}{\exp(-2\eta)(z+1)} = \frac{2\eta \exp\left(-2\eta z\right)}{z+1}. \tag{50}$$

If η is large enough, the density will be concentrated at low values of z as larger variance contributions will be increasingly suppressed by selection. We can ignore z in the denominator so that

$$f(z) \approx 2\eta \exp(-2\eta z). \tag{51}$$

This is an exponential distribution with rate 2η.

For the high-dimensional model, the normalizing constant is also an incomplete gamma function, $2\Gamma(0, 2\sqrt{\eta})$. We can apply the same approximation as in equation (49), though in this case it must be that $\sqrt{\eta}$ is large. Doing this for equation (48) gives

$$f(z) \approx \sqrt{\eta} \frac{\exp\left(-2\sqrt{\eta}\sqrt{z+1} + 2\sqrt{\eta}\right)}{z+1} \tag{52}$$

Assuming that the density is concentrated at small z, we can use $\sqrt{z+1} \approx 1 + z/2$, giving

$$f(z) \approx \sqrt{\eta} \exp(-\sqrt{\eta} z). \tag{53}$$

This is again an exponential distribution, although the rate is proportional to $\sqrt{\eta}$ in this case. The two models are not distinguishable when v^* is sufficiently high because both models have exponential tails and γ is unknown. The one-dimensional model will approach the exponential limit faster than the high-dimensional model. This implies an asymmetry in the ability to distinguish the two where it is more likely for traits under one-dimensional stabilizing selection to look identical to a high-dimensional model with a lower γ.

G Code and data availability

Processed GWAS data and code for analyses, simulations, and figures can be accessed at `https://github.com/emkoch/smilenfer`.

H Figures

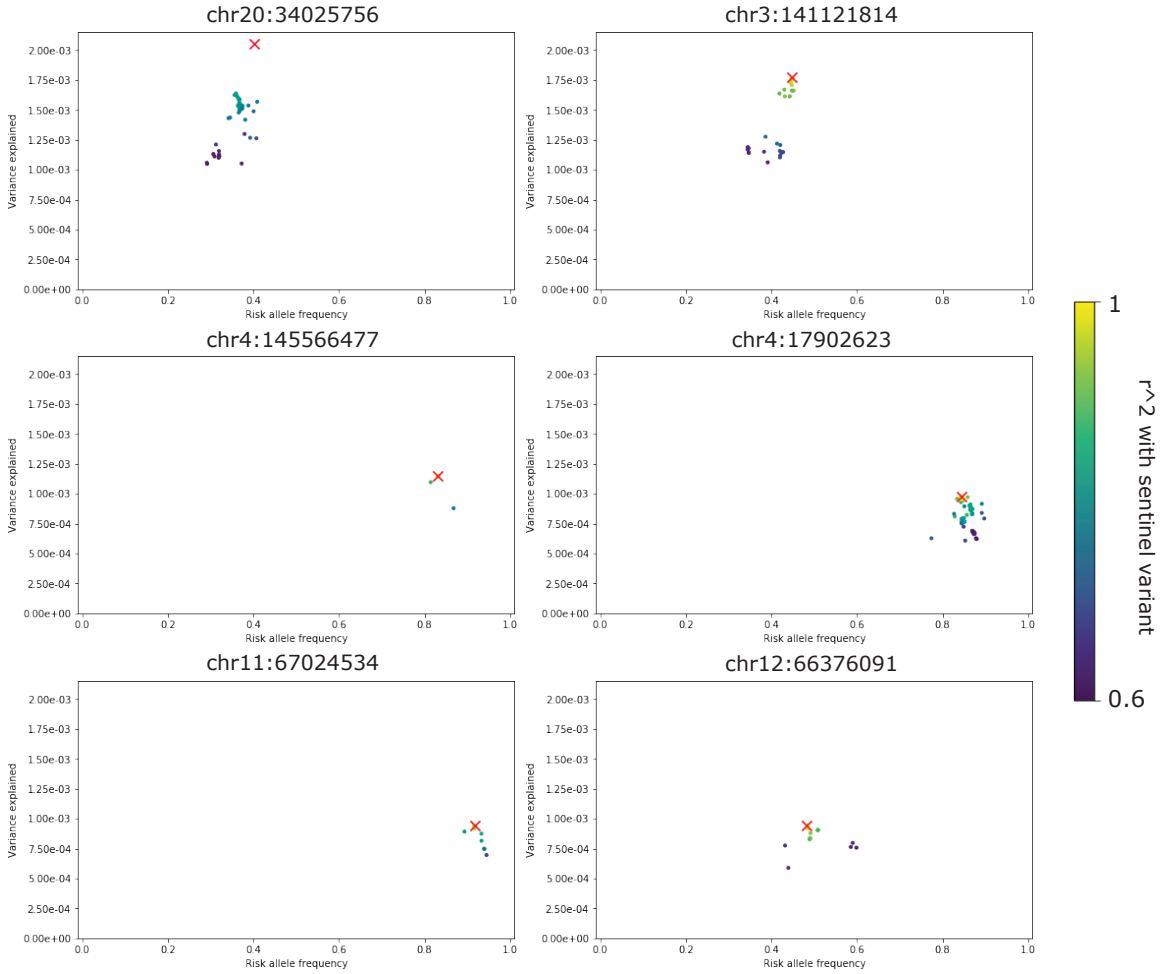


Figure S1: Sentinel variants cluster tightly with linked variants in high LD. A sample of six genome-wide significant loci for standing height. The sentinel variant (lowest p-value) is shown alongside nearby variants in high LD. Only variants with similar risk allele frequencies have a high r^2 with the sentinel variant.

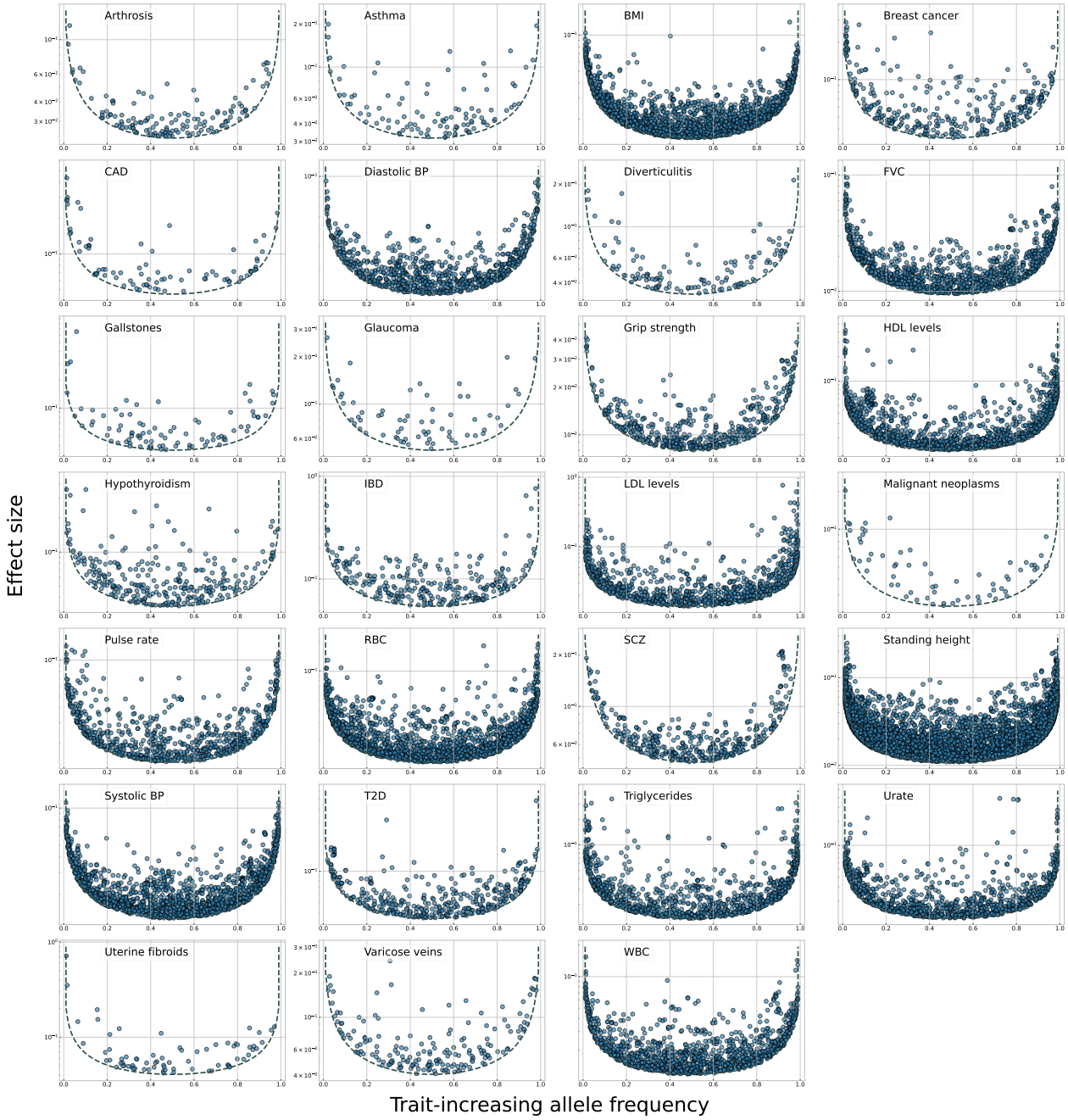


Figure S2: U-shaped distributions of trait or disease risk-increasing allele frequencies and effect sizes for all analyzed traits. For disease traits, the effect size is calculated as the odds-ratio minus one. Raw GWAS estimates are shown here without shrinkage. We include variants that are both genome-wide significant and have an estimated variance contribution exceeding a threshold variance associated with 5×10^{-8}. The dashed line shows the effect size cutoff imposed by the variance contribution threshold.

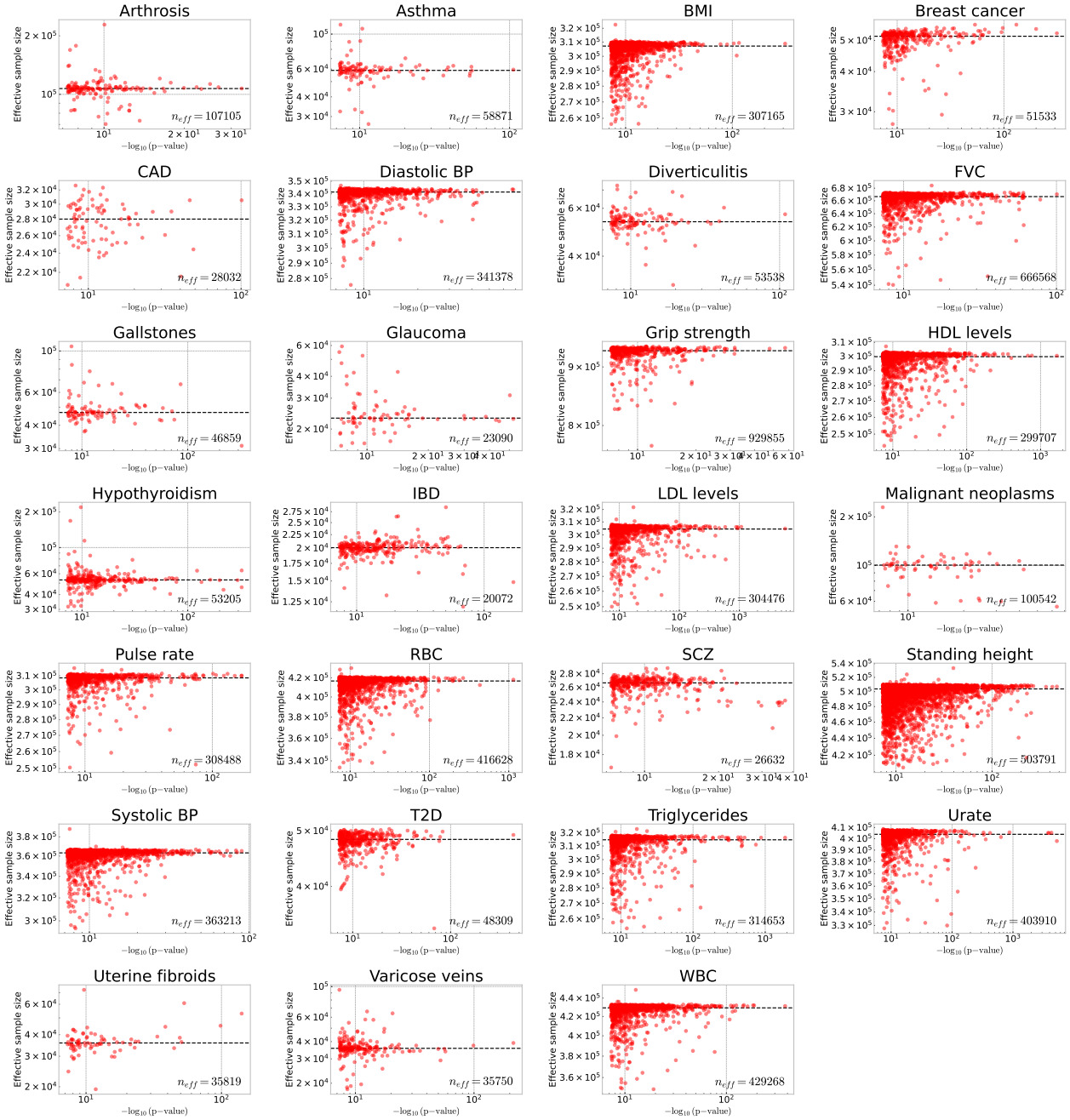


Figure S3: Effective sample size and p-value relationship for all traits. For each trait under consideration we plot the relationship between effective sample sizes calculated using GWAS effect sizes $\hat{v} = 2\hat{\beta}x(1-x)$ and p-values $\hat{n_{\text{eff}}} = F^{-1}_{\chi^2_1}(1 - p\text{-value})/\hat{v}$. Only genome-wide significant variants (p-value$< 5 \times 10^{-8}$) are shown. Dashed lines and values in the bottom right correspond to the median effective sample size among genome-wide significant variants.

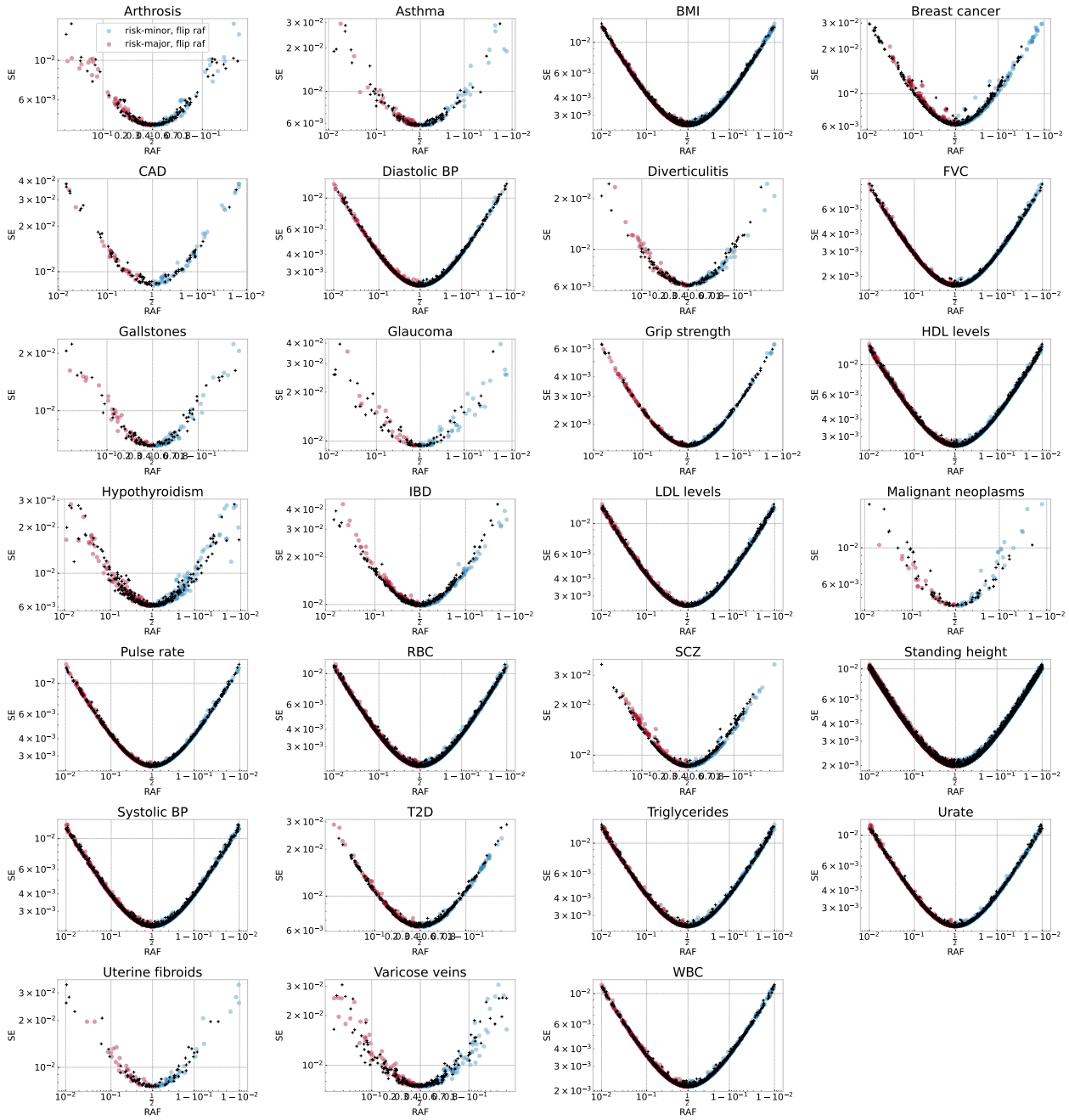


Figure S4: Relationship between standard errors and trait or disease risk-increasing allele frequencies. Black points show standard errors obtained from GWAS summary statistics plotted as a function of the trait or risk increasing allele frequency (RAF). To visualize asymmetries in between variants where the minor or major allele is correlated with increased risk, we also flip the plot along the RAF= 0.5 line (blue and red points). Equation (19) predicts that standard errors should depend only on minor allele frequencies and the effective sample size of the study. In practice, noise results from correlations between the variants and other genotypes and covariates. Asymmetry is also apparent for disease traits where there is greater power to detect low-frequency risk alleles than low-frequency protective alleles. RAFs are plotted on a logit scale.

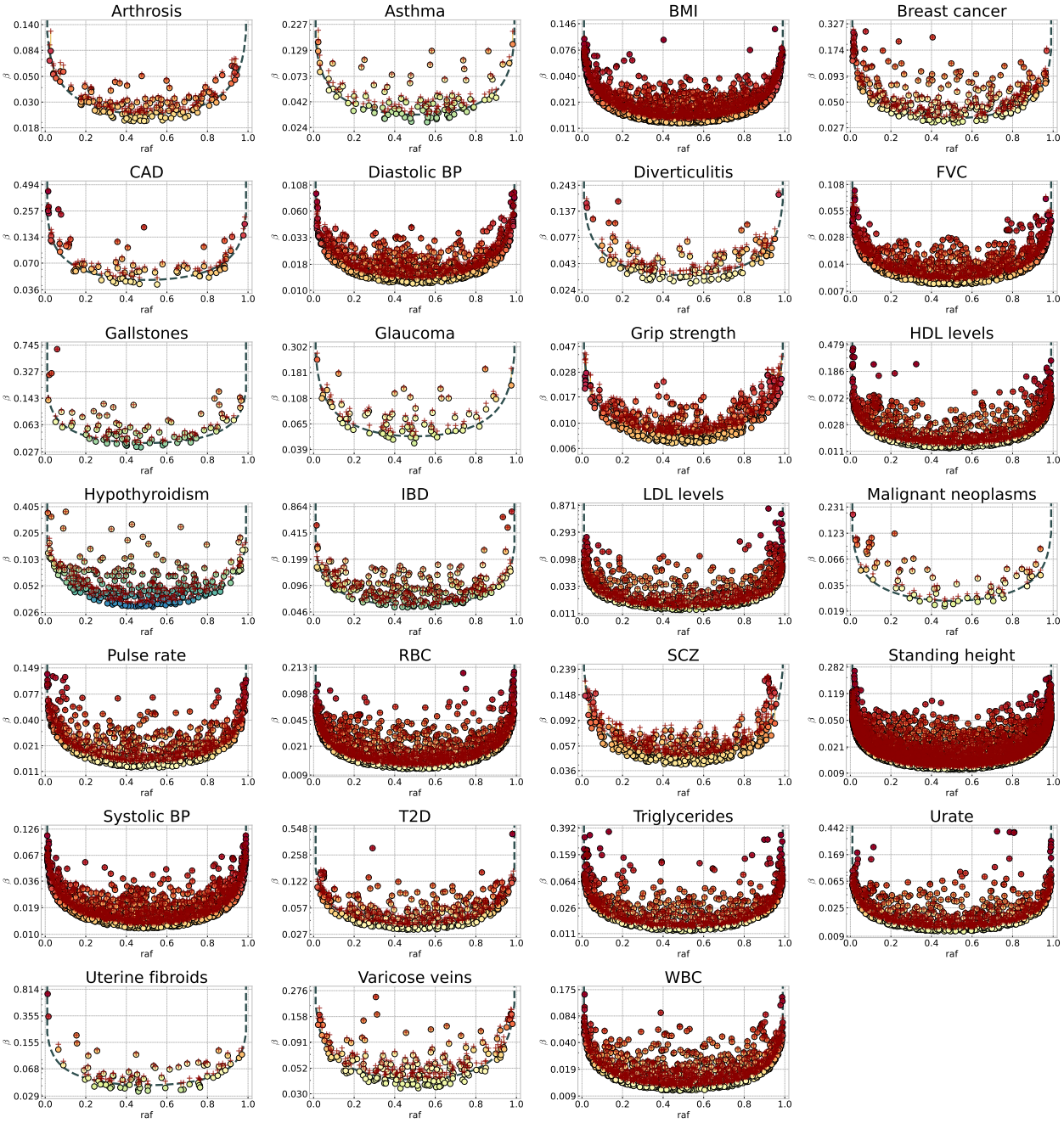


Figure S5: The effect of shrinkage on effect sizes. The fit of the pleiotropic stabilizing selection model to schizophrenia applying both the variance contribution and raw p-value threshold for 5×10^{-8}. The dashed line shows the frequency cutoff at each effect size. Red crosses indicate the raw effect sizes while circles indicate posterior median effect sizes after shrinkage with *ash*. Frequency thresholds are computed based on raw effect sizes, while the strength of selection is fit using shrunken estimates.

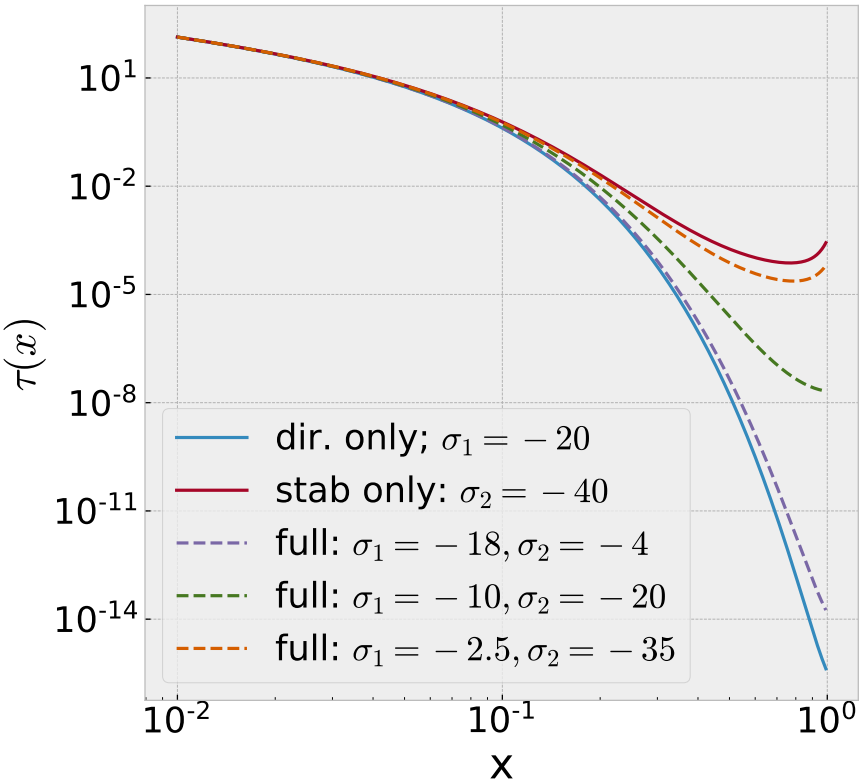


Figure S6: The allele frequency distribution with both additive and underdominant selection interpolates between the two models. Examples of the SFS were calculated using equation (32). At low frequencies the SFS under directional (additive) and stabilizing (underdominant) selection look similar, but more high frequency variants are expected for stabilizing selection. An underdominant scaled selection coefficient twice as large is required to match to additive selection.

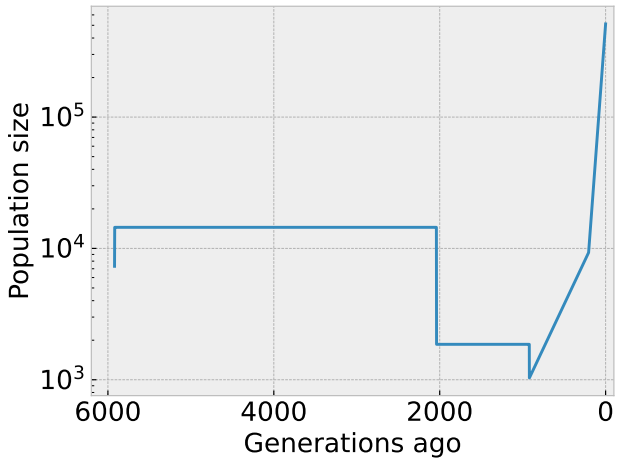


Figure S7: Demographic history model used to generate allele frequency distributions. We implemented the demographic model of non-Finnish European population history fit by (32). The original model also included periods of migration with African and East-Asian populations that we have omitted. Populations sizes were calibrated using a mutation rate of 1.25×10^{-8} per site per generation.

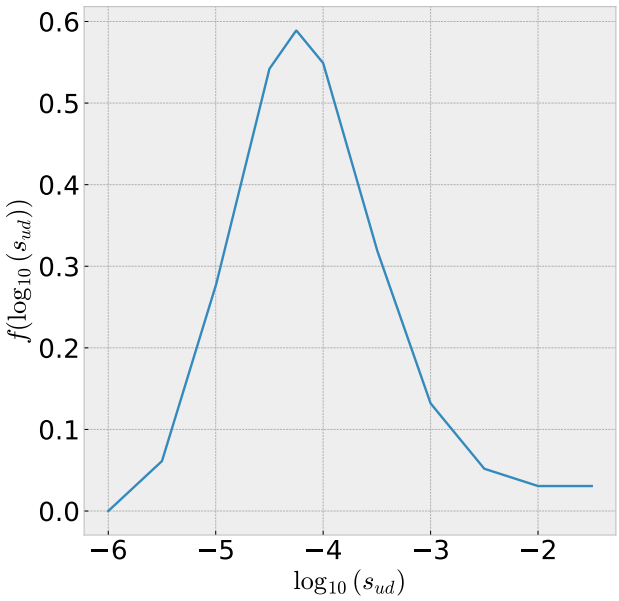


Figure S8: Approximate single shared DFE (SSD). We approximated the DFE inferred by Simons et al. (2022) (22) using a piece-wise linear function on their Figure 3b. The density shown is a probability distribution on $\log_{10}(s_{ud})$, rather than the density of s_{ud} plotted on a log scale.

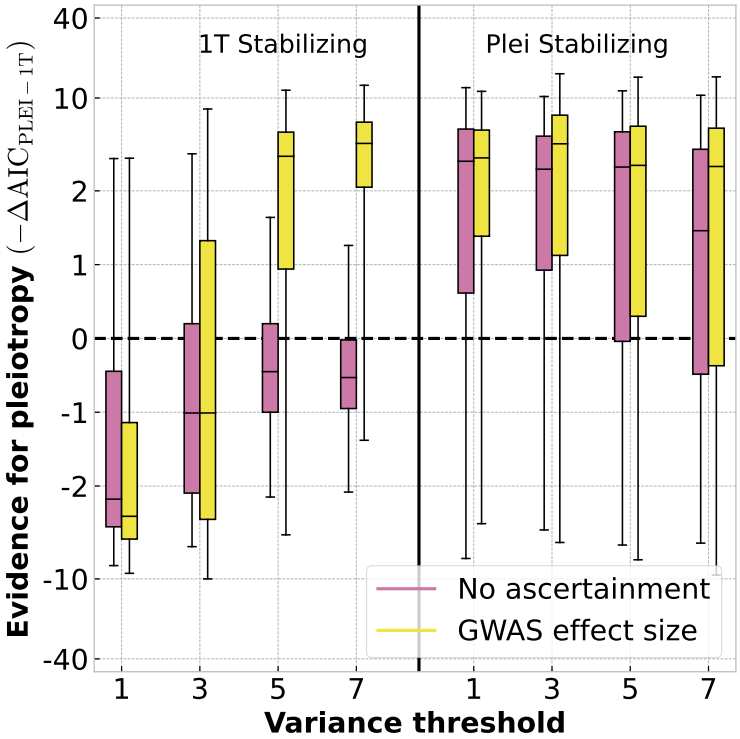
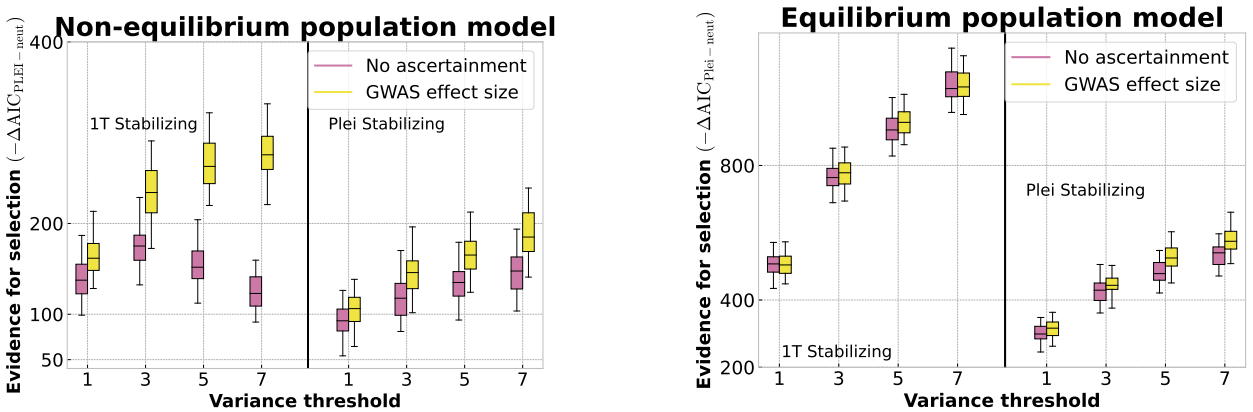


Figure S9: Discerning pleitropic versus single-trait stabilizing selection in simulations. Simulations performed using the DFE estimated by Simons *et al.* (22) under models of single-trait and pleiotropic stabilizing selection. Positive values indicate support for the pleiotropic model and negative values indicate support for the single-trait model. We compare simulations where effect sizes are known without error to a model of GWAS ascertainment at a p-value threshold of 5×10^{-8}. Increasing the variance threshold corresponds to more strongly selected variants being included and greater winner's curse when ascertainment bias is modeled. Each simulation was fixed at 500 ascertained variants. In the absence of ascertainment (pink boxes), the true model does have a higher likelihood in our method on average. However, there is substantial variance such that this is not true of every simulation. The power to discriminate between the two stabilizing selection models decreases as the ascertainment threshold is raised. This is because strongly selected variants behave similarly under both models. In simulations that included GWAS noise, when single-trait selection was the true model, it was possible for the pleiotropic model to have a higher likelihood when the variance threshold was high.



(a) (b)

Figure S10: All SSD simulations have strong evidence for selection. Panel (a) shows the increase in likelihood for the stabilizing selection models relative to neutrality for DFE simulations using the NFE demography, while panel (b) shows results for the same DFE and an effective population size of $N_e = 10\,000$. The body of each boxplot represents the 25 and 75 percentiles of 500 simulations, while whiskers represent the 2.5 and 97.5 percentiles. Without modeling ascertainment bias, the evidence for selection initially increases as the variance-contribution threshold increases because alleles with larger effect sizes, and therefore variance contributions will be under stronger selection. However, the average evidence actually declines for one-dimensional selection and the NFE model as ascertained variance concentrate at the bottom of the U-shaped ascertainment curve. Ascertainment bias has a minor effect in equilibrium simulations, but favors the selection model in NFE simulations.

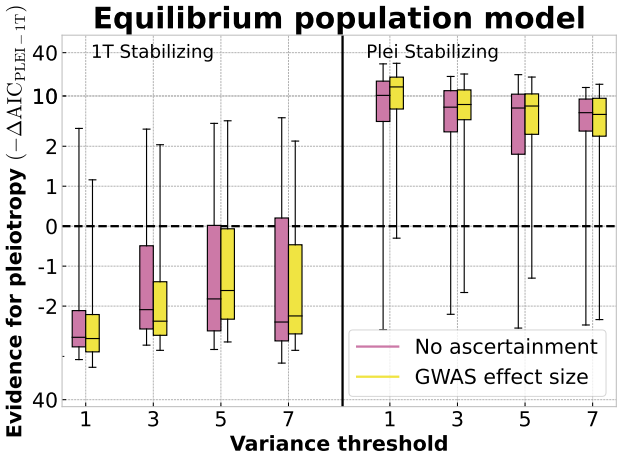
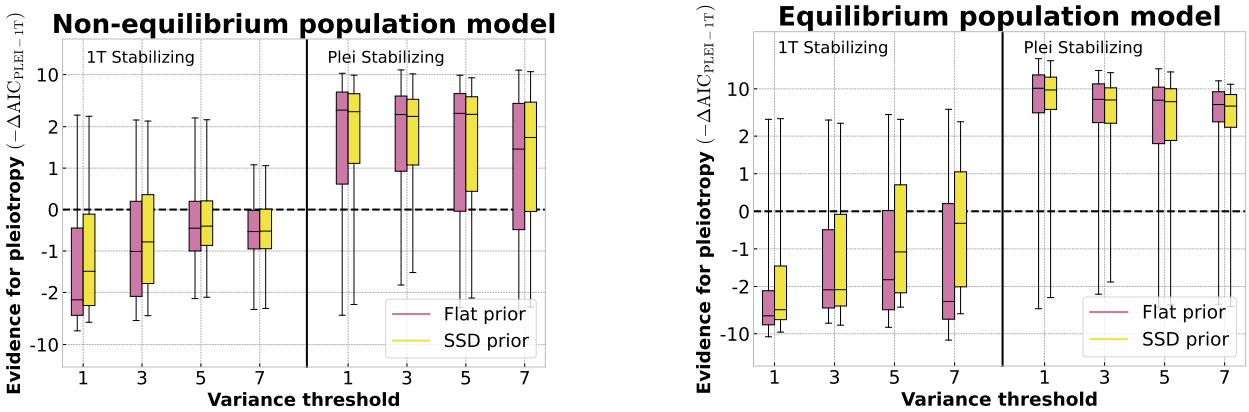


Figure S11: Ability to discriminate single-trait and pleiotropic models of stabilizing selection in an equilibrium population. Log-likelihood differences between the two stabilizing selection models in simulations of a population at equilibrium ($N_e = 10\,000$) using the SSD DFE. The body of each boxplot represents the 25 and 75 percentiles of 500 simulations, while whiskers represent the 2.5 and 97.5 percentiles. There is greater power to discriminate between models than under simulations using the NFE demography, but the probability of false-positives remains non-negligible. In contrast to non-equilibrium simulations, ascertainment bias does not strongly affect results.



(a) **(b)**

Figure S12: Little effect of using the SSD as a prior distribution during likelihood computations. Log-likelihood differences between the two stabilizing selection models in simulations of a population at equilibrium ($N_e = 10\,000$) using the SSD DFE the simulate variance. The body of each boxplot represents the 25 and 75 percentiles of 500 simulations, while whiskers represent the 2.5 and 97.5 percentiles. During inference, we test the effect of using the SSD as a prior when calculating $p(s_{ud} \mid \beta) \propto p(\beta \mid s_{ud})p(s_{ud})$ rather than a flat prior on $\log s_{ud}$.

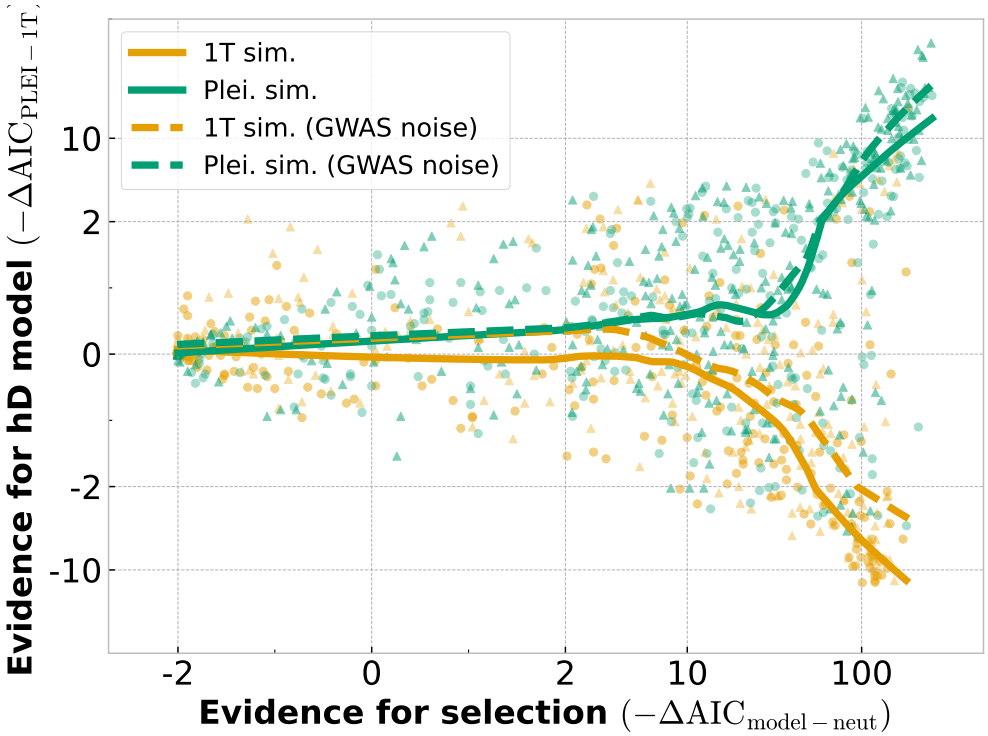


Figure S13: Power to discriminate between single-trait and pleiotropic stabilizing selection in simulation. Figure is based on trait-based simulations using the estimated prior distribution of IBD and a range of I_2 and I_p values. The evidence for selection is determined by both the strength of selection and sampling noise. Circles and bold lines indicate simulations without GWAS noise, such that effect sizes are known without error. Triangles and dashed lines are for simulations where ascertainment bias was simulated and estimated effect sizes were treated as true effects for likelihood calculations. Lines show LOESS curves fit to each simulation category.

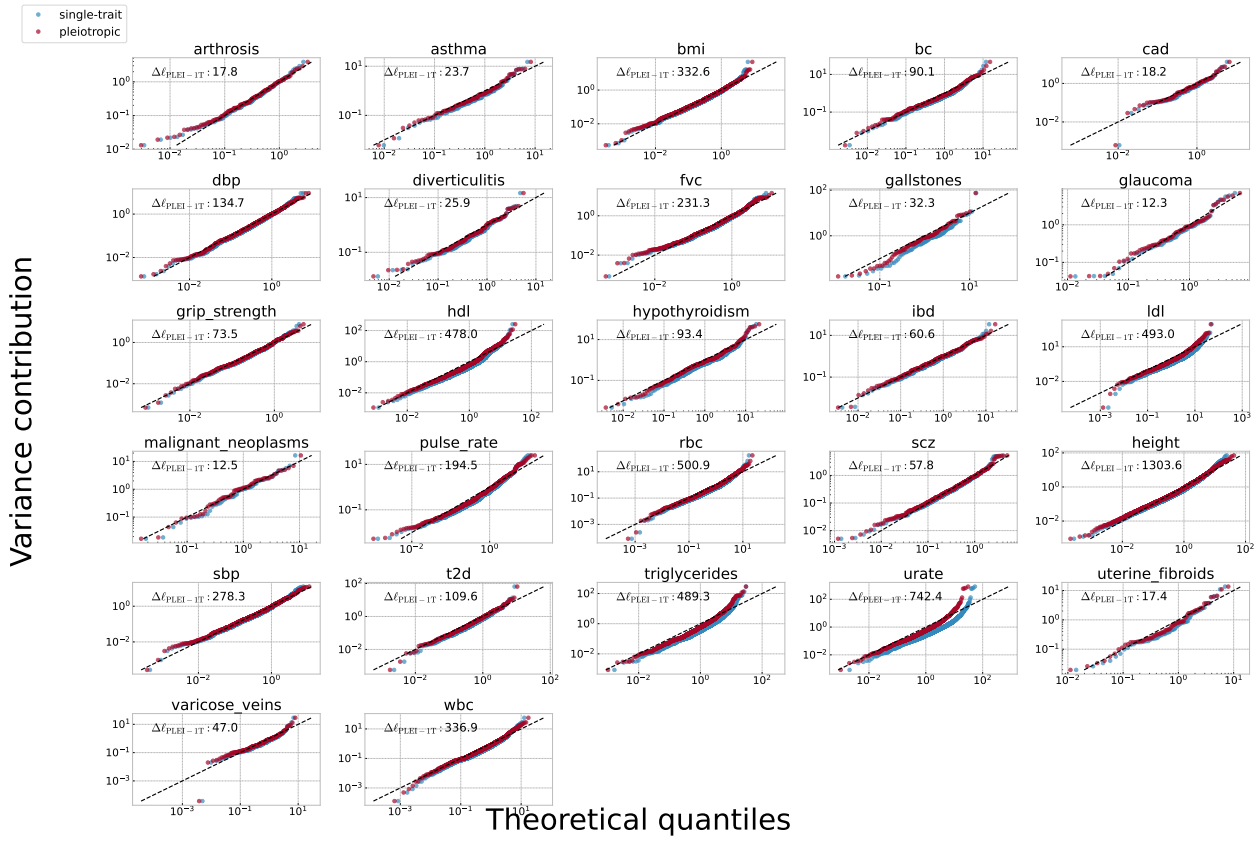


Figure S14: QQ plots for fits of single-trait and pleiotropic variance-contribution distributions to all trait data. The top left corner of each plot contains the excess log-likelihood for the high-dimensional variance contribution distribution. Variance contributions were transformed as $(v - v^*)/v^*$. The variance contribution threshold approximately corresponding to a p-value cutoff 5×10^{-8}.

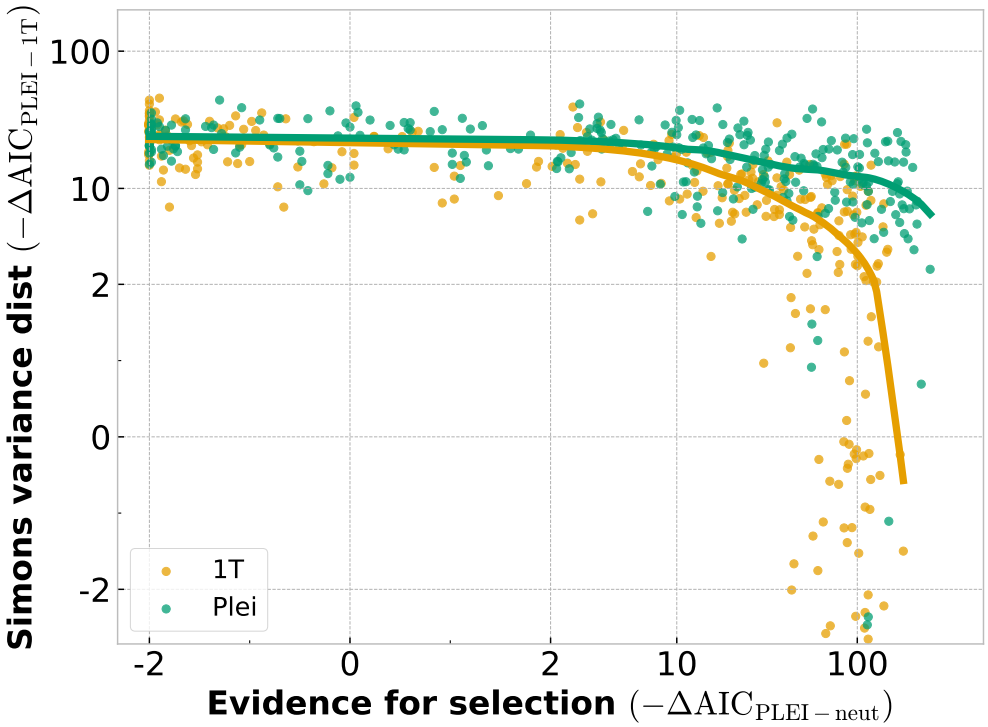


Figure S15: Application of variance-contribution test to trait simulations. The variance contribution test was applied to trait-based simulations of 200 GWAS hits using the estimated IBD effect size prior and a range of I_2 or I_p values ranging from neutral to strong selection. The simulations displayed in this figure did not include ascertainment bias. The evidence for selection in each simulation was summarized using the ΔAIC value for the true simulated mode (single-trait or pleiotropic stabilizing) versus neutrality. Likelihoods for the x-axis were computed using equation (3). AIC was used instead of p-values in order to put both axes on the same scale. The y-axis shows the evidence for pleiotropic stabilizing selection with likelihoods computed for the tail distributions under the single-trait and pleiotropic models derived by Simons *et al.* (31). Lines are LOESS curves showing the average behavior under each model as selection on ascertained variants becomes very strong.

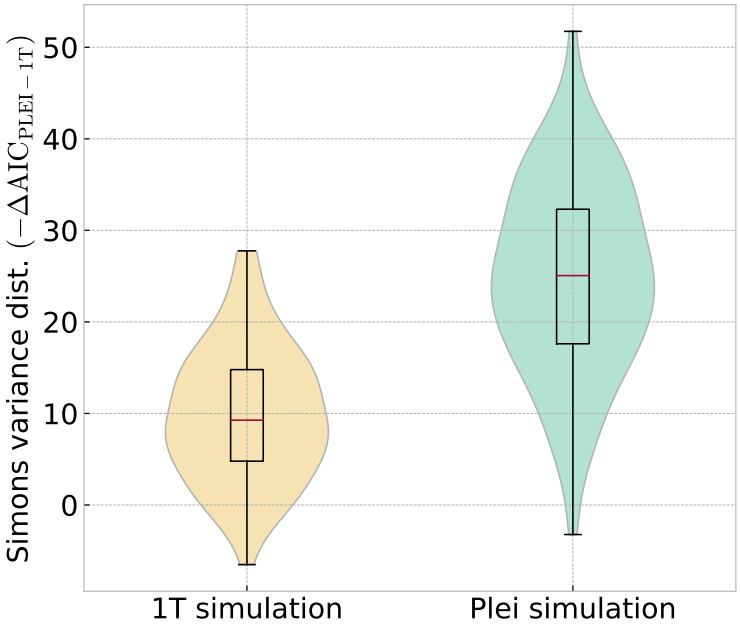


Figure S16: Application of variance-contribution test to DFE simulations. The variance contribution test was also applied to DFE-based simulations of 500 GWAS hits for the single-trait and pleiotropic stabilizing models. Results are displayed for a selection-scaled variance contribution cutoff of one. Boxplots summarize 100 simulations each.