Amplification is the Primary Mode of Gene-by-Sex Interaction in Complex Human Traits

Carrie Zhu1,2, Matthew Ming1,2, Arbel Harpak1,2+

1 Department of Population Health, The University of Texas at Austin

2 Department of Integrative Biology, The University of Texas at Austin

+ To whom correspondence should be addressed: [arbelharpak@utexas.edu](mailto:arbelharpak@utexas.edu)

# Abstract

Sexual dimorphism is observed in many complex traits and diseases and is suspected to be in large part due to widespread gene-by-sex interactions (GxSex). To date, empirical evidence for GxSex in GWAS data—with attempts focused on identifying imperfect genetic correlations between the sexes—has been underwhelming. We hypothesized that GxSex may indeed be pervasive, but largely missed by current approaches, because it acts primarily through sex differences in the magnitude of many genetic effects (“amplification”), regulated by a shared cue such as a sex hormone, rather than differences in the identity of causal variants or the direction of their effect (“correlation”). To test this hypothesis, we inferred the genetic covariance structure between males and females across 27 traits in the UK Biobank, and decoupled GxSex signals into correlation and amplification. Amplification was a more pervasive mode of GxSex. For example, we estimate that 38% of causal variants have a greater effect on urate levels in females than males. In addition, we investigated whether testosterone levels, a continuous measure, may underlie the observed mediation by binary biological sex. For many traits, notably in body mass related traits, testosterone levels are strongly associated with the magnitude of genetic effects in both males and females–but the association is different in magnitude and sign between the sexes. Finally, we developed a novel test of sexually-antagonistic viability selection linking GxSex signals from GWAS and allele frequency divergence between males and females. Using independent allele frequency data from both Finnish and Ashkenazy-Jewish samples, we find subtle evidence for contemporary sexually-antagonistic selection on variants associated with several body mass related traits. In summary, our results suggest that the systematic amplification of genetic effects is a primary mode of GxSex that is germane for understanding the genetics and evolution of sexual dimorphism.

# Introduction

Genetic effects are not exerted in a vacuum, but in environments that mediate them. Several lines of evidence suggest large context dependency of genetic effects on polygenic (complex) human traits (Ge et al., 2017; Mostafavi et al., 2020; Patel et al., 2021; Young et al., 2019). At the same time, unlike studies in organisms and systems where the environment can be tractably manipulated (des Marais et al., 2013; Falconer & Mackay, 1996; Huang, Campbell, et al., 2020; Huang, Carbone, et al., 2020; Hudson et al., 2022; Long et al., 2013; Lynch & Walsh, 1998; MacQueen et al., 2021), there are but a few cases in which GxE models explain data—such as Genome-Wide Association Study (GWAS) data—better than parsimonious models that assume independent contributions of genetic and environmental factors (Duncan et al., 2019; Mars et al., 2020; Mostafavi et al., 2020; Patel et al., 2021; Young et al., 2018). This duality may point to a pervasive mediation of polygenic effects by the environment—alongside the inability of our current toolbox to quantify polygenic GxE (Mills et al., 2020; Nagpal et al., 2022).

Perhaps nowhere is this duality more apparent than in the case of gene-by-sex interaction (GxSex). Sexual dimorphism is observed in many complex traits and diseases and is suspected to be in large part due to widespread GxSex involving autosomal loci (Arnqvist & Rowe, 2005; van Doorn, 2009). Sex differences in polygenic effects on traits are clearly of high importance for healthcare (Harper et al., 2021; Mokkonen & Crespi, 2015; Oliva et al., 2020; Power et al., 2013; Schroderus et al., 2010), evolution (Barson et al., 2015; Cheng & Kirkpatrick, 2016; Connallon et al., 2010; Harrison et al., 2015; Kidwell et al., 1977) and other areas in biology. But with the exception of testosterone levels (Bernabeu et al., 2021; Carole Hooven, 2021; Flynn et al., 2021; Sinnott-Armstrong et al., 2021), the genetic basis of sexual dimorphism is not well-understood (Oliva et al., 2020). To date, empirical evidence for GxSex in GWAS data—with attempts focused on identifying large GxSex effects in individual loci, or instead, with a polygenic prism, imperfect genetic correlations between the sexes—have been underwhelming.

Sex as an “environment” variable offers a methodological advantage as well. The study of context-dependency is often complicated by study participation biases (leading to genetic ancestry structure in the data), reverse causality between the phenotype and environment variable, collider bias, gene-by-environment correlation and other confounders. In this sense, sex provides a useful case study for context-dependency: Reverse causality and other such “usual suspects” confounders involving the phenotype causally affecting sex, are highly unlikely; and sex differences in ancestry composition due to study participation biases are small (Benonisdottir & Kong, 2022) (though detectable (Pirastu et al., 2021)).

In this work, we set to characterize GxSex across complex human traits. We begin by suggesting a mode of GxSex that may have gone largely undetected by previous studies: A systematic difference in the magnitude of genetic effects across a large set of variants, which we refer to as “amplification”. Amplification can happen for a large set of variants regulating a specific pathway if the pathway responds to a shared cue. In classic hypothesis testing approaches that test for a GxE effect separately in each variant, the signal of amplification may be crushed under the multiple hypothesis burden. On the other hand, even state-of-the-art tools designed with complex traits in mind may miss amplification signals: They often treat genetic correlation (between GWAS estimates based on samples from two environments) as a telltale of whether effects are the same in two groups (Brown et al., 2016; B. Bulik-Sullivan et al., 2015; Galinsky et al., 2019; Ni et al., 2018; Shi et al., 2017); but correlations are scaleless and thus amplification signals could be missed entirely.  To quantify the role of genetic amplification in sex differences, we developed a new approach for flexibly characterizing a mixture of male-female genetic covariance relationships, and applied it to 27 traits in the UK Biobank. We conclude that amplification is pervasive across traits, and that it is the primary mode of GxSex. We find evidence that modulators of the size of effects, or amplifiers, are shared among environmental and polygenic effects, and that together, such amplification effects explain phenotypic sexual dimorphism. Finally, we develop a novel test for sexually-antagonistic polygenic selection, which connects GxSex signals—including amplification—to signals of contemporary viability selection, and find evidence of sexually-antagonistic selection on variation associated with body-mass traits.

# Results

## The limited scope of single-locus analysis

We conducted sex-stratified GWASs for 27 continuous traits using a sample of ~150K biological males and ~150K biological females in the UKBB. We chose to analyze traits with SNP heritabilities over 0.075 in the combined sample of both males and females, to allow for high statistical power.

In our analysis of the pairs of sex-stratified GWAS, we examined Miami plots after annotating highly-associated SNPs with the closest gene using *Ensembl Variant Effect Predictor* (McLaren et al., 2016) (REFX **Methods**). Among the 27 traits, we observed substantial differences in the associated genes only for testosterone and waist:hip ratio (whether or not it was adjusted for BMI; REFX **Fig. S1**). In testosterone, we found multiple gene annotations specifically related to one sex, but not the other. As studied in a previous analysis, these genes partake in separate mechanisms in males and females, and the small overlap of top associations may reflect the sexual dimorphism that arises in large part through distinct biological pathways (Flynn et al., 2021; Sinnott-Armstrong et al., 2021). For example, in females, CYP3A7 partakes in the hydroxylation of testosterone, resulting in its inactivation. In males, FKBP4 plays a role in the downstream signaling of testosterone on the hypothalamus (REFX **Fig. 1**).

Diagram

Description automatically generated

**Figure 1: Inferring polygenic covariance structure between males and females. (A)** Our analysis of the polygenic covariance between males and females is based on sex-stratified GWAS. Shown for illustration, is a “Miami plot” for testosterone. Highly associated SNPs were annotated by obtaining the gene with the closest transcription start site less than 5000 bp away.  **(B)** We modelled the sex-stratified GWAS estimates as sampled, with error, from true effects arising from a mixture of hypothesized covariance relationships between female and male genetic effects—see examples in red frames. Each box shows the weight (± SE) we infer for one hypothesis matrix, corresponding to the probability that a variant’s effects adhere to that covariance relationship. The boxes are shown along axes indicating the relative magnitude (amplification) and correlation between males and females, which together fully specify the covariance relationship.  **(C)** The x and y axis are a condensed version of the x and y axes from (B) for testosterone, as well as the condensed version for diastolic blood pressure. The weights are the percentage of non-null weights, i.e. the weight divided by the sum of all weights except for the weight on the all-zero matrix corresponding to no effect in either sex. For example, the square in purple sums over all 12 weights for matrices corresponding to larger effects on testosterone in males that are negatively correlated with female effects, 5.1%, divided by the total weight on matrices with nonzero effects, 32%.

For waist:hip ratio, we saw multiple genes associated with females only, such as ADAMTS9, a gene involved in insulin sensitivity (Lumish et al., 2020). As previous work established (Bernabeu et al., 2021; Flynn et al., 2021; Sinnott-Armstrong et al., 2021), testosterone and waist:hip ratio are the exception, not the rule: Most traits did not display many sex differences in top associations. For instance, arm fat-free mass, a highly heritable dimorphic trait, showed near-perfect concordance in significant loci (REFX **Fig. S1**). Previous studies (Bernabeu et al., 2021) examining top SNPs between males and females have found relatively low numbers of unique loci (<20) in the 84 continuous traits they studied apart from waist:hip ratio. Considering recent evidence on the vast polygenicity of additive effects on complex traits (Boyle et al., 2017; Sella & Barton, 2019; Shi et al., 2016), it stands to reason that a look beyond lead associations, through a polygenic prism, can aid in the characterization of non-additive effects (such as GxSex) as well.

## The limited scope of analyzing GxSex via heritability differences and genetic correlations

We therefore turned to consider the polygenic nature of GxSex; first, employing commonly used approaches. We used LD Score Regression (B. Bulik-Sullivan et al., 2015; B. K. Bulik-Sullivan et al., 2015) to estimate sex-specific SNP heritabilities () and the genetic correlation between the sexes for each trait. In most traits (17/27), males and females had a genetic correlation greater than 0.9. Testosterone had the lowest genetic correlation of 0.01, which suggests very little sharing of signals between males and females.

For the majority of traits (18/27), male and female heritabilities were both greater than the both-sex heritability. For instance, in arm-fat free mass (right), the heritability in the both-sex sample was 0.232 (± 0.009), while the heritabilities for male and female were 0.279 (± 0.012) and 0.255 (± 0.011), respectively. In particular, all body mass-related traits, excluding BMI-adjusted waist:hip ratio, had greater sex-specific heritabilities.

Graphical user interface, application

Description automatically generated

**Table 1. Polygenic Models of GxSex.** We examine different models of the nature of GxSex in complex traits that link to previous studies and motivations. Each model points to different expectations from the analysis of heritability and genetic correlations (REFX **Fig. 2**). The illustration depicts examples of directions and magnitudes of genetic effects as described by the model.

Graphical user interface

Description automatically generated with low confidence

**Figure 2: Heritabilities and Genetic Correlations Cannot Fully Distinguish Models of GxSex.** **(A)** The genetic correlation between the male and females are shown in descending order. **(B)** The x-axis represents the relative heritability, i.e. the heritability divided by the heritability estimated in the both-sex sample. Therefore, the purple values, indicating relative both-sex heritability all fall on the 1 line. Traits marked with red asterisks indicate body-mass related traits with greater sex-specific than both-sex heritabilities.

In addition, we noticed a trend in which as correlation decreased, the difference between the heritabilities of each sex relative to the both-sex heritability tended to increase (Pearson r = -0.88, p = 10-9; REFX **Fig. 2**). However, several traits with genetic correlation above 0.9, such as diastolic blood pressure and arm fat free mass (left), also present relatively large heritability sex differences of 5.2% and 3.4%, respectively. Such discrepancies are incompatible with a model of pervasive uncorrelated genetic effects driving sex-specific genetic contributions to variation in the trait (REFX **Table 1**, second model). We therefore considered two other alternative hypotheses under a simple additive model of variance in a trait within each sex,

|  |  |
| --- | --- |
| , | (1) |

where represent the trait value, additive polygenic effect, and environmental effect (including all non-genetic context aside from sex) in sex , respectively. Under this model, the sex-specific heritability hz is

|  |  |
| --- | --- |
|  | (2) |

and therefore, sex differences in heritability may be due to sex differences in genetic variance, environmental variance, or a combination of the two. If genetic effects are similar, differences in environmental variance alone may explain heritability differences entirely (REFX **Table 1**, first model). But under such a model, the heritability in the combined sample cannot be smaller than both sex-specific heritabilities (REFX **Methods**)

|  |  |
| --- | --- |
|  | (3) |

where represents the environmental effect and sex the combined sample environmental variance,

|  |  |
| --- | --- |
|  | (4) |

must be less than the max environmental variance of either sex. If genetic effects are similar, then

|  |  |
| --- | --- |
|  | (5) |

Therefore, the observation of higher sex-specific heritabilities for most traits suggests that genetic variance must differ between males and females. Assuming, further, that there is little-to-no interaction of sex and participation in the UKB (Benonisdottir & Kong, 2022), UKB allele frequencies in males and females are expected to be near-identical and thus this observation suggests that causal genetic effects differ somewhat between males and females for most traits we analyzed.

A less appreciated possibility for GxSex is that the identity and direction of effects may often be shared (leading to high genetic correlation), but there are pervasive differences in the magnitude of genetic effects—e.g., systematically larger genetic effects on blood pressure in females—which in turn lead to large differences in genetic variance (REFX **Table 1**, third model). To evaluate the evidence for systematic differences in the magnitude of genetic effects (“amplification”) and compare it with evidence for uncorrelated effects as a main mechanism of GxSex, we next set to infer the genome-wide covariance of genetic effects among the sexes.

## Amplification of genetic effects is the primary mode of GxSex

We jointly assessed amplification and correlation of genetic effects between males and females using multivariate adaptive shrinkage (*mash*) (Urbut et al., 2019), a tool that allows the inference of genome wide frequencies of genetic covariance relationships (REFX **Table 1**, fourth model). Namely, we model marginal SNP effects as sampled (with SNP-specific, sex-specific noise) from a mixture of possible covariance relationships (2x2 Variance-Covariance matrices for male and female effects). We pre-composed hypothesis covariance matrices that span a wide array of amplification and correlation relationships, and use *mash* to infer the mixture weights. These weights can be roughly interpreted as the proportion of variants that follow the pattern specified by the covariance matrix (REFX **Fig. 1B, C**). Our hypothesis covariance matrices ranged from -1 to 1 in correlation, and multiple levels of male greater, female greater, and equal magnitude. We also included a matrix representing no effect in both conditions (REFX **Fig. 1B**).

We first examined testosterone as a “positive control”, in the sense that previous research sets our expectation for polygenic M-F covariance: In terms of magnitude, the vast majority of effects should have much greater effect in males. In terms of correlation, we expect a class of genetic effects through largely independent and uncorrelated pathways alongside a class of effects through shared pathways (Sinnott-Armstrong et al., 2021). Independent pathways include the role of hypothalamic-pituitary-gonadal axis in male testosterone regulation and the contrasting role of the adrenal gland in female testosterone production. Shared pathways can be found involving sex hormone-binding globulin (SHBG), which decreases the amount of bioavailable testosterone in both males and females. Indeed, we found that mixture weights for testosterone concentrated on greater magnitudes in males and largely uncorrelated effects. Out of the 32% total weights on matrices with an effect in at least one sex, 98% of the weights were placed on matrices representing larger effects in males, including 20.4% (± 0.7%) having no effect in females (REFX **Fig. 1B**).

While testosterone is unusual in the relative lack of correlation and the extent of male lob-sided amplification, about half (13/27) of the traits analyzed had the majority of weights placed on either female-greater or male-greater magnitudes (REFX **Fig. 3**). For instance, BMI-adjusted waist:hip ratio displayed higher percentages for greater effects in females (64%; REFX **Fig. S3AA**) and arm fat-free mass traits (right) showed higher percentages of greater effects in males (92%; REFX **Fig. S3C**). Both traits had weights concentrated on highly correlated genetic effect matrices.

Chart

Description automatically generated

**Figure 3. Phenotypic variance strongly correlates with amplification.** The x-axis represents the amplification of effect sizes, and is calculated by taking the difference between the sum of weights on matrices with male effects greater in magnitude than female effects (M>F) and sum of weights of M<F matrices.

In addition, none of the 27 traits we studied had more than 10% of the non-zero weights on matrices with negative correlation, save for testosterone, which had 17%. Most weights were rather on perfect or partially positive correlation matrices. For example, diastolic blood pressure and eosinophil percentage had 66% (REFX **Fig 1C**) and 68%, respectively, of the non-zero weights on matrices representing perfect correlation. Overall, the low weights on matrices representing negative correlation do not support opposite directions of effects being a major mode of GxSex (REFX **Fig. S7**).

We also found that, across traits, the difference between male and female amplification weights and the ratio correlates strongly with male-to-female phenotypic variance ratio (Pearson r = 0.873, p = 10-9 after removing testosterone as an outlier; REFX **Fig. 3**). This observation is consistent with our hypothesis of amplification GxSex leading to differences in genetic variance and thereby contributing substantially to sex differences in phenotypic variance. Together, these observations point to amplification, rather than uncorrelated effects, as a pervasive mechanism of polygenic GxSex.

## Low utility of sex-specific models for complex phenotype prediction

The pervasiveness of GxSex that we infer, alongside the plurality of covariance relationships across the genome for each trait, may mean that GxSex is important to consider in phenotypic prediction. To test this possibility, we compared the prediction accuracy of three polygenic scores (PGS) for various traits, together with covariates (REFX **Fig. 4**):

1. Additive PGS, assuming no GxSex
2. Sex-specific, but M-F covariance-naïve PGS: Based on stratified GWAS independently for each sex.
3. Sex-specific, covariance-aware: Using *mash*, we refine marginal SNP effect estimated, by modelling them as sampled from the mixture of covariance metrices with the weights we estimated previously (REFX **Methods**).

A picture containing chart

Description automatically generated

**Figure 4. The predictive utility of G-by-Sex aware polygenic scores.** We compare the incremental R2 of the regression of the phenotypic values on the polygenic score for three models. Incremental R2 is obtained by taking the difference in R2 between a prediction with covariates and the polygenic score and a prediction with only covariates. We choose the highest R2 over the four p-value thresholds (REFX Methods) for comparison. The additive (red) and sex-specific, covariance-naïve (yellow) models are calculated from effect estimates generated in GWAS, with the latter in sex-stratified samples. The sex-specific, covariance aware model (blue) refers to a polygenic score constructed using mash posterior effect estimates, which considers the covariance in genetic effects between the sexes. Sex-specific models contained a median sample of around 123k males and 147k females, whereas the additive model retained the full training set of around 271k individuals. Incremental R2 values and ±3 SE-wide error bars were computed using 5-fold cross-validation.

In almost all cases (25/27), the sex-specific, covariance-naïve PGS model had the lowest predictive power—measured as the squared correlation between the predictor () and trait value. For some traits, like testosterone, height, arm fat-free mass, and hip circumference, the sex-specific covariance-aware PGS outperformed the additive PGS by a slight margin (significant at p < 0.05 level in a two-sample t-test for only height and arm fat-free mass (right)). However, for most traits (22/27), the covariance-aware PGS model did not predict better than the additive model (REFX **Fig 4**); including a significantly better performance of the additive model in 14 of the 22 traits.

Based on these results, there was little improvement in prediction accuracy even with a detailed consideration of polygenic GxSex. We can hypothesize that the persistent advantage of the additive model is due to the reliance on double the sample size, outweighing the utility of refined, covariance-aware sex-specific effect estimates. We chose not to match sample sizes for the additive model to mimic the competition between prediction models in practice. At this stage, we find little utility for sex-specific models in complex trait prediction; yet it seems plausible that future approaches—that better make up for the loss of accuracy in capturing the additive effect because of higher model complexity—could surpass the performance of additive models.

## Testosterone as an underlier

Thus far, we treated the genetic interaction we observe as discretely mediated by biological sex. One mechanism that may underlie our observation is a cue or exposure that modulates the magnitude (and less often, the direction) of genetic effects, and varies in its distribution between the sexes. One natural candidate is testosterone. Testosterone may be a key instigator since the hormone is present in distinctive pathways and levels between the sexes, and a known contributor to the development of male secondary characteristics, and therefore could modulate genetic causes on sex-differentiated traits.

For each sex, we first binned individuals into 10 bins of testosterone levels. Then, for each trait, we quantified the amplification of total genetic effect as the linear regression coefficient, within the bin, of phenotypes to the additive PGS for the trait (**REFX Methods**; see **Fig. S10** for results obtained using sex-specific PGS).

There was a strong correlation between testosterone and the magnitude of genetic effect, for both males and females and for multiple traits–especially ones related to body mass (REFX **Fig. 5**). BMI (REFX **Fig. S9**) and whole body fat mass (REFX **Fig. 5A**) showed significant correlations (Pearson p < 0.05) for both sexes. For all the body mass related traits, there was a negative correlation between the magnitude of genetic effect and testosterone for males, and a positive correlation for females.  For example, Pearson's correlation coefficient for whole body fat mass in males was -0.86 (Pearson p = 0.001) and 0.76 for females (Pearson p = 0.011). For these traits, since the relationship with testosterone is clearly not continuous across the sexes, a model of testosterone as a continuous underlier of the binary sex mediator would be false. These observations may help explain previous reports of positive correlations between obesity and free testosterone in women, and negative correlations in men (Pasquali, 2006). Therefore, in body-mass related traits, testosterone may be modulating genetic effects in a sexually-antagonistic manner.

sGraphical user interface

Description automatically generated with medium confidence

**Figure 5. Amplification of total genetic effect in relation to testosterone levels.** **(A)** The relationship between testosterone level bins and genetic effect—measured as the slope from the regression of phenotypic values to polygenic scores in that bin—is depicted for three different traits. The hollow data point signify a testosterone level bin with overlapping range between males and females, and are based on fewer individuals (XXX compared to YYY in other bins). **Figs. SX** show all other traits analyzed. **(B)** The correlation coefficient for each sex (90% CI) is plotted for all 27 traits. Traits are ordered in descending order of male-female differences in R.

## A model of shared polygenic and environmental amplification

If polygenic amplification is a common mechanism, an immediate question of interest is whether environmental effects are amplified across the sexes via modulators shared with polygenic effects. Consider the example of human skeletal muscle. The impact of resistance exercise varies between males and females. Resistance exercise can be considered as an environmental effect since it upregulates multiple skeletal muscle genes present in both males and females such as IGF-1, which is in turn involved in muscle growth (Liu et al., 2010). However, after resistance exercise at similar intensities, males experienced sustained upregulation of such genes, while in females, levels returned sooner to the resting state (REFX **Fig. S11**). It is plausible that shared modulators, such as sex hormones, drive a difference in the magnitude of effect of core genes such as IGF-1 and thus all its genetic and environmental regulators.

To generalize this intuition into a model: If amplification mechanisms are shared, then the genetic and environmental effect may be modeled as having the same scalar multiplier effect on genetic and environmental effects. Next, we specify the details of a null model of shared amplification, present a prediction that it yields and test this prediction across the 27 traits analyzed.

Consider the within-sex additive model of REFX **Eq. 1** in the section “**The limited scope of analyzing GxSex via heritability differences and genetic correlations**” above. For a phenotype value in sex

|  |  |
| --- | --- |
| , | (6) |

where is the environmental effect and

|  |  |
| --- | --- |
|  | (7) |

is the polygenic effect where is the effect of the causal allele and  is the effect-allele count (Harpak & Przeworski, 2021). We assume here for simplicity that male genetic effects relate to female effects solely through a shared polygenic amplification constant, α,

|  |  |
| --- | --- |
|  | (8) |

Allele frequencies are once again assumed to be close to equal between males and females, since due to random segregation of alleles during meiosis, genotype frequencies at autosomal sites are is independent of sex; and further assuming no substantial sex biases in UKB participation with respect to ancestry (Benonisdottir & Kong, 2022). Consequently, differences in polygenic effect distributions between males and females are solely based on GxSex, and thus:

|  |  |
| --- | --- |
| . | (9) |

The model we would like to test is one where the amplification of environmental effects can also be simplified to the same scalar multiplier,

|  |  |
| --- | --- |
| , and | (10) |

Hence, with equal amplification,

|  |  |
| --- | --- |
|  | (11) |

To test the fit of this theoretical prediction to our data, we used our estimates of sex-specific phenotypic variance and SNP heritabilities to estimate the ratios of genetic and environmental variances (REFX **Methods**). We note that environmental variance is proxied here by all variance not due to additive genetic effects, and caution should be taken with interpretation of this proxy.

22 of the 27 traits were consistent with the null model of equal amplification (within ± 3SE; REFX **Fig. 6B**). This may suggest a sharing of pathways between polygenic and environmental effects for these traits (REFX **Fig. 6A**). Interesting exceptions include BMI-adjusted waist:hip ratio, creatinine, and systolic and diastolic blood pressure—which was the strongest outlier (p=3.06e-12 from one-sample t-test), excluding testosterone.

A picture containing diagram

Description automatically generated

**Figure 6. Testing a model of shared amplification between environmental and polygenic effects**. **(A)** A model of equal amplification of genetic (G) and environmental (E) effect, that produces the dimorphism shown in the phenotype, Y. G and E both act through a core gene that is amplified in a sex-specific manner. **(B)** The blue 1:1 line depicts the theoretical expectation under a simple model of equal amplification of genetic and environmental effects in males compared to females. Error bars show (± 3 SE). Traits in blue are consistent (<3 SE) with the prediction of 1:1 relationship.

## Sexually-antagonistic selection

A hypothesized driver of sexual dimorphism, and of the maintenance of GxSex genetic variance, is sexually-antagonistic selection, where alleles are beneficial in one sex yet deleterious in the other (Cheng & Kirkpatrick, 2016; Connallon et al., 2010; Kasimatis et al., 2021; Kidwell et al., 1977). Sexually-antagonistic selection is difficult to study using traditional population genetics methods because Mendelian inheritance equalizes autosomal allele frequencies between the sexes at conception in each generation—thereby erasing informative signals. One way around this limitation is to examine allele frequency differences between the sexes in the current generation, known as “selection in real time” (Cheng & Kirkpatrick, 2016; Ruzicka et al., 2020, 2021). In this section, we consider a model of sexually-antagonistic selection acting on a polygenic trait and use it to estimate the strength of contemporary viability selection acting on the 27 traits we analyzed.

Our findings on pervasive amplification suggest that variant effects tend to have concurrent signs, and still set the stage for substantial sexual dimorphism. Taken together, these results suggest the somewhat-surprising intuition that alleles affecting a trait under stabilizing selection—even if for similar trait optima in males and females—may frequently experience sexually antagonistic selection, as we illustrate in REFX **Fig. 7A**. We developed a theoretical model of sexually-antagonistic viability selection on a single trait which extends on this intuition. The model ties sex-specific effects on a complex trait and the divergence in allele frequency between males and females (measured as (Weir & Cockerham, 1984; Wright, 1951)) due to viability selection “in real time”, i.e., acting between conception and adulthood. We derive the expected relationship for each site ,

|  |  |
| --- | --- |
|  | (12) |

where is the phenotypic variance due to GxSex at site and is a constant shared across all variants and can therefore be interpreted as the effect of sexually-antagonistic selection–on variant sites underlying variation in a trait—on male-female divergence at these variants (REFX **Methods**). We estimated for all sites in the UKB genotype array across subsamples of various ancestry groups in the gnomAD dataset (Karczewski et al., 2020). To estimate at each site and for each trait, used our previous sex-stratified GWAS. Since there is large heterogeneity in uncertainty of GxSex-genetic variance estimates, we use a variance-weighted linear regression to estimate A (see REFX **Methods** for the derivation of the variance of estimates and for further details on the estimation procedure)

Diagram, table

Description automatically generated

**Figure 7. Testing for sexually-antagonistic selection. (A)** A model of sexually-antagonistic selection yielding the prediction of **Eq. 12**. In this specific example, selection drives trait values towards the optimum in males, and away from the optimum in females. The selection coefficients, and , are linear with the additive effect on the trait and sexually-antagonistic selection is at equilibrium such that . **(B)** Two examples of the weighted least-squares linear regression preformed to estimate the strength of sexually-antagonistic selection on variants associated with a trait ( in panel A and Eq. 12). **(C)** Z-scores (± 90% non-parametric CI) estimated through 1000 resampling iterations of the weighted linear regression of panel B for each trait. The two colored estimates match the corresponding examples in panel B.

In two of three gnomAD subsamples that are the closest in their genetic ancestry to our UKB sample (Privé et al., 2022; results for other ancestry groups are shown in REFX **Figs. SXXX**), we find marginally significant evidence for sexually-antagonistic selection on body mass related traits: The three highest Z-scores (for the null hypothesis **in REFX eq. 12**) in the Ashkenazi-Jewish sample are whole-body fat mass (Z=XXX±YYY), waist circumference (Z=XXX±YYY) and hip circumference (Z=XXX±YYY; REFX **Fig. 7C**); for the Finnish samples the three top traits are whole-body fat mass (Z=XXX±YYY), waist circumference (Z=XXX±YYY) and BMI (Z=XXX±YYY; REFX **Fig. 7C**): For Non-Finish Europeans—the largest subsample—we find no significant evidence for sexually-antagonistic selection (**REFX Fig. 7C**).

# Discussion

Departing from previous studies identifying GxSex (or GxE more broadly) through single loci or heritability analysis, our results show that examining GxSex through a polygenic lens and a mixture of relationships across the genomes, reveals pervasive context-dependency of complex traits. Our results suggest that not only is GxSex pervasive, but also largely acts through the systematic amplification of effects, an underaprreciated mode of GxE. For some traits such as weight or arm fat-free mass previously considered non sex-specific because of high genetic correlations or concordance in significantly-associated SNPs, show substantial GxSex through amplification. Furthermore, the strong relationship between amplification and phenotypic values suggests that differences in magnitude play a primary role in sexual dimorphism. Evidence of widespread amplification may provide clues to possible avenues in which GxSex is acting.

Consequently, we proposed a model in which amplification is regulated by a shared modulator between the sexes. We found a strong, but non-continuous correlation between testosterone and genetic effect on multiple body mass related traits such as BMI and whole body fat mass. These results, with genetic effect increasing in females and decreasing in males with increasing testosterone levels, may give reason to the non-continuous nature between testosterone and body-mass at the phenotypic level. Testosterone or other hormones may play an important role in modulating amplification. Knowledge of such non-continuous mediators is crucial in both research and healthcare since an increase in one mediator may result in opposite effects between the sexes.

With this in consideration, we examined if amplification due to a modulator also applied to environmental effects. We found that most traits were consistent with a model of equal amplification of genetic and environmental effects. Co-amplification may point to a shared biological pathway mediated by the same amplifier. We provide an example with muscle mass, in which the effect of resistance exercise (environmental effect) is amplified greater in males, along with genetic regulators for muscle mass by affecting a core gene, IGF-1. Further research into specific biological pathways may provide more concrete understanding in potential amplifiers and the interaction between polygenic and environmental effects. Specifically, partitioned analysis for amplification may better identify those shared pathways or core genes.

However, the utility of polygenic covariance structures and amplification does not substantially carry over to predictive ability. Out of our three models—additive; sex-specific, covariance naïve; and sex-specific covariance aware—the additive model still had the highest predictive value for most of the traits. Given that the additive model had approximately twice the sample size (both-sex vs one sex), we hypothesized that the advantage in power is still greater than the benefit of utilizing covariance-aware estimates.

Although the utility of covariance-aware estimates for prediction is low at this stage, analyzing GxSex in a polygenic lens aid by breaking it apart into its magnitude and correlation structure aids interpretation and understanding behind the mechanism of the interaction. We see that amplification of effect sizes by sex plays a primary role in sexual dimorphism and potentially acts equally on environmental effects, hinting at shared pathways. Stratified analysis can help localize these pathways or factors of amplification. We can extend this to future studies to examine interaction terms with other environments. Amplification may apply to other environment-specific approaches and provide important insight into biological mechanisms.

# Methods

## UK Biobank

### Sample characteristics

The UK Biobank is an extensive database that contains a wide variety of phenotypic and genotypic information of around half a million participants aged 40-69 at recruitment (Bycroft et al., 2018).

In this study, we considered 337,111 individuals who passed quality control (QC) checks, which included the removal of samples identified by the UK Biobank with sex chromosome aneuploidy or self-reported sex differing from sex determined from genotyping analysis. We excluded related individuals (3rd degree relatives or closer) as identified by the UK Biobank in data field 22020 as individuals used in UK Biobank’s genetic Principal Component Analysis. To reduce potential population structure confounding, we further limited our sample to individuals identified by the UK Biobank as “White British” in data field 22006. These are individuals who both self-identified as white and as British and were additionally very tightly clustered in the genetic principal component space. Individuals who had withdrawn from the UK Biobank were removed by the time of this study. For each phenotype, we also removed individuals who had missing data for the specified phenotype. These procedures left us with between 255,426 to 336,551 individuals included in the analysis for each trait.

### Genotype data

We focused on autosomal bi-allelic SNPs with an INFO score greater than 0.8. Using *plink 2.0 alpha* (Chang et al., 2015; Graffelman & Moreno, 2013; Graffelman & Weir, 2016; Purcell et al., 2007; Purcell & Chang, n.d.; Wigginton et al., 2005), we further retained variants with calling rate > 0.95, Hardy-Weinberg equilibrium test p-value > 10-10, minor allele frequency (MAF) > 0.001, following a QC procedure used by the Neale Lab on version 3 of UK Biobank (Abbot et al., n.d.). These were computed on the aforementioned quality-controlled sample set and resulted in 9,607,691 SNPs.

### Phenotype Data

Our analysis consisted of 27 continuous traits for their relatively high SNP heritability estimates, based on LD Score regression (Abbot et al., n.d.; B. K. Bulik-Sullivan et al., 2015). BMI-adjusted waist:hip ratio (WHR) was calculated by regressing WHR on BMI and obtaining the residuals, as shown in the following commands in R:



### GWAS

We performed all GWAS using *plink 2.0 alpha*, adjusting for birth year, sex, and the first 10 principal components (PC) provided by the UK Biobank as covariates. Covariates were standardized to mean 0, variance 1 (using the flag --covar-variance-standardize). We generated sex-specific GWAS summary statistics for each trait by separating the sample by males and females and following the same model. Any variants with missing values in the summary statistics were removed from further analysis.

## Miami plots

We used the *Ensembl Variant Effect Predictor (VEP)* (McLaren et al., 2016) based on the GRCh38 genome build to annotate SNPs with p < 5x10-8 in Miami plots (REFX **Figs S1**). Using the --nearest flag, we retrieved the gene with the closest protein-coding transcription start site within 5,000 bp up- and down-stream each SNP.

## Heritability and genetic correlation estimation using LD Score Regression

We estimated the SNP heritability of each trait for both-sex, female-specific, and male-specific GWAS, as well as the genetic correlation between sexes using LD Score Regression (B. Bulik-Sullivan et al., 2015; B. K. Bulik-Sullivan et al., 2015). Since our GWAS summary statistics were based on White British individuals of primarily European ancestry, we used the precalculated LD scores computed by Bulik-Sullivan et al. (B. K. Bulik-Sullivan et al., 2015).

## Multivariate adaptive shrinkage (mash)

We utilized multivariate adaptive shrinkage (*mash*) to examine correlation and differences in magnitude of SNP effects between males and females (Urbut et al., 2019). *mash* is a statistical method that improves upon previous methods of estimating and comparing effects across multiple conditions by flexibly allowing for a mixture of covariance patterns between conditions and requiring only summary statistics from each condition (including a point estimate of the effect and corresponding standard error for each SNP and condition). The method adapts to patterns of sparsity, sharing, and correlation among the conditions to compute improved effect estimates.

In this study, we set two conditions, male and female, and provided effect estimates and corresponding standard errors from our male-specific and female-specific GWAS. *mash* learns from the data by estimating mixture proportions of various predefined covariance matrices representing different patterns in effects. Using maximum-likelihood, mash assigns low weights to matrices that capture less patterns in the data, and vice versa.

## Mixture weights for covariance structure between male and female effects

To interpret patterns of SNP effects between males and females, we input 66 hypothesis-based covariance matrices (REFX **Fig. S2**) spanning a range of correlations and relative magnitudes of effects between males and females. We used a random subset of all SNPs for mash to learn the covariance mixture weights. In order for the random subset to contain approximately independent SNPs and capture the weight of SNPs with no effect (REFX **Fig. S2**), we created a subset of SNPs for each trait by taking a random SNP from each of 1701 approximately independent LD blocks (Berisa & Pickrell, 2016). *mash* can also generate data-driven covariance matrices that capture SNP effects in the data, but we did not use this feature since the data-driven matrices had negligible differences from our hypothesis matrices and were less interpretable.

For each trait, we repeat this weight-learning step 100 times, sampling the SNPs from the 1701 LD blocks without replacement to fit the mash model and generate mixture proportions. We then take the average proportion for each covariance matrix as an estimate of its weight, effectively treating each of the 100 samples as iid draws.

### Choice of SNPs used to estimate male-female effect covariance

We examined the effect of using a random subset taken from different p-value thresholds [1, 5e-2, 1e-5, 5e-8] while selecting from LD blocks. By doing so, we can examine differences in the distribution of weights across the p-value thresholds. We performed this test on height, BMI, testosterone, and BMI-adjusted waist:hip ratio. For each trait, weight placed on the no-effect matrix decreased as we reduced the p-value threshold (REFX **Fig. S3**). Patterns of weights for non-null effect matrices varied across the traits (REFX **Fig. S3**). Since *mash* considers the proportion of null effects and patterns may be found in less significant SNPs, we decided on using the whole set of SNPs to sample from when estimating mixture proportions.

## Posterior estimates of sex-specific effect sizes

*mash* can also apply adaptive shrinkage after learning patterns in effect sizes to improve marginal point estimates—estimated using the posterior mode—and measures of significance. These posterior estimates could reduce noise by shrinking effects towards zero and possibly reveal greater or less variation in effect sizes between males and females. Therefore, using the average of the fitted mixture model over 100 repetitions (REFX **Mixture weights for covariance structure between male and female effects**), we computed posterior estimates for each trait to be used in further analysis (REFX **Evaluating the performance of sex-specific polygenic scores**). Mash calculates a local false sign rate (lfsr) for each effect, which is similar to a local false discovery rate and is defined as the cumulative density of the posterior distribution of values with a sign that differs from that of the posterior mean. We mapped lfsr to pseudo p-values by ranking the SNPs according to lfsr values, and then ascribing the same-rank pseudo p-value (REFX **Fig. S5**).

## Environmental variance simulation for *mash*

To ensure that *mash* was not mistaking sex differences in environmental variance to be differences in the magnitude of genetic effects, we performed a simulation study where, in short, samples of males and females are generated under the model in REFX **Eq. 1**, where genetic effects are exactly the same but the environmental variance differs across the sexes; we then perform a GWAS on both samples and input the simulated GWAS results into *mash*, and test whether *mash* identifies no GxSex in the data. We performed this simulation with various levels of heritability [0.05, 0.5], female to male environmental variance ratio [1, 1.5, 5], and numbers of causal SNPs [100, 1k, 10k].

First, we created a sample of 300k individuals with randomly assigned sex. We then created a matrix of genotypes for all individuals consisting of 20K SNPs by sampling from a binomial distribution of allele frequencies from UK Biobank’s imputated data (Sudlow et al., 2015). From the 20K SNPs, we portioned out the predetermined number of causal SNPs and assigned effect sizes by sampling from a Normal distribution with mean 0, standard deviation 1. We estimated the male environmental variance for each causal SNP using the equation,

|  |  |
| --- | --- |
|  | (12) |

where and represents the simulated environmental effect, genetic effect, effect size at site , allele frequency at site , and heritability, respectively. We multiplied by the environmental variance ratio to obtain the environmental variance for females, . Afterwards, for each individual, we sampled environmental effects from a Normal distribution with mean zero, and variance set as or according to the sex of the individual. Phenotypes were then simulated using the following,

|  |  |
| --- | --- |
|  | (13) |

where represents the phenotype and represents the genotype of the causal SNP for the individual. With the phenotype, genotype and environmental effect, we obtained the estimated effect size, , using linear regression for all 20k SNPs as input into *mash*.

For nearly all parameters, out of the weights on matrices with effect, the majority was on the matrix for perfect correlation, equal magnitude (REFX **Fig S6**). As the number of causal SNPs increased, the weight on the no-effect covariance matrix decreased accordingly. Based on these results, we determined that mash was not capturing differences in environmental variance when estimating mixture proportions.

## Evaluating the performance of sex-specific polygenic scores

We calculated polygenic scores (PGS) for three models to test differences in prediction:

1. Additive model based on both-sex GWAS summary statistics
2. Sex-specific, additive model based on sex-specific GWAS summary statistics
3. sex-specific, covariance aware model based on posterior sex-specific effect estimates (REFX Section **Posterior estimates of sex-specific effect sizes**)

The process is illustrated in the following diagram (REFX **Fig. S8**), and, for models (a) and (b), parallels the procedure described in (Choi et al., 2020).

First, for each phenotype, we split the sample of unrelated White British individuals into a test set of 25k individuals randomly sampled from each sex and a training set with the remaining individuals. Second, we re-ran GWAS on the training set following the same procedure above (REFX **GWAS**), generating both-sex, female-specific, and male-specific summary statistics. Third, to obtain posterior estimates, we input the effect sizes and standard errors from the male-specific and female-specific GWAS into *mash*, and took the average of 100 mixture proportions to fit the *mash* model and estimate the posterior mean and lfsr. We then sorted SNPs by lfsr and matched it to a sorted list of p-values taken from the same GWAS summary statistics.

Afterwards, we performed clumping three times for the additive model (both-sex, male-specific, and female-specific GWAS) and two times for the covariance-aware model (male and female posterior estimates) to produce a subsets of SNPs that are relatively independent using *plink 1.9 beta*’s --clump command, removing SNPs with LD threshold or within 250kb. For LD values, we used a sample of 187 unrelated European (population code: GBR and CEU) individuals from 1000 Genomes phase 3 (Auton et al., 2015).

Using the resulting subset of SNPs, we estimate PGS for the individuals in the test set by summing the number of effect alleles an individual has weighted by the allelic effect sizes. We use *plink 2.0 alpha’s* --score command along with the --q-score-range flag to repeat the PGS computation over a range of p-value thresholds [1, 0.01, 1e-5, 1e-8]. Therefore, a total of 20 PGS runs are performed for each trait over the combination of five models and four p-value thresholds (REFX **Fig. S8B**).

Finally, to assess the prediction accuracy, we computed for the following,

|  |  |
| --- | --- |
|  | (14) |

The covariates used for the regression are the same as those used in our GWAS: sex, birth year, and the first 10 PCs of the genotype matrix (UKB data field 22009). We performed regressions both separately by sex and together with both sexes in the test set. For each of the three predictors—additive model, sex-specific additive model, and sex-specific covariance aware model—we selected the p-value threshold with the greatest . We also calculated the incremental , which is the increment in after adding the PGS to a null model with only the covariates. (REFX **Fig. S8**).

### Cross-Fold Validation

We performed five-fold cross validation for our PGS procedure. Each prediction was performed on a different randomly sampled 25k female and 25k male test set, with the remaining used at the training set. For each model, we averaged the greatest from the p-value thresholds over the five folds for comparison across all models.

## Model of Shared Amplification

To test the model of shared amplification between environmental and polygenic effects (REFX **Eq. 11**) we obtained the genetic and environmental variance for males and females based on the following relationships,

|  |  |
| --- | --- |
| , and  , | (15) |

where , and refers to the genetic effect, environmental effect, and phenotype value, respectively. Heritability, , was obtained from previous estimates using LD Score Regression (REFX **Heritability and genetic correlation estimation**).

Representing male genetic or environmental variance as , and the corresponding female variance as , we derived standard errors for the ratio of male to female variance using the Taylor approximation for

|  |  |
| --- | --- |
|  | (16) |

assuming independence between and since they are estimates of the same statistic. The standard error of the genetic and environmental variance was estimated using the law of total expectation for a product of two random variables, and ,

|  |  |
| --- | --- |
|  | (17) |

and using the and as approximations for and respectively,

|  |  |
| --- | --- |
|  | (18) |

In this case, represents the phenotypic variance, and represents for estimation of genetic variance or for estimation ­of environmental variance. Lastly, to obtain the standard error of the phenotypic variance, we used bootstrapping with 100 repeated samples (with replacement) of the phenotypic variance,

|  |  |
| --- | --- |
|  | (19) |

## Sexually Antagonistic Selection

### The relationship between male-female allele frequency divergence at adulthood and GxSex

We developed a model relating sex differences in additive effects on a trait at a biallelic locus (and ) and divergence in allele frequencies. Our model resembles that of Cheng and Kirkpatrick (Cheng & Kirkpatrick, 2016) who developed a similar model relating allele frequency differences and sex bias in gene expression. In short, we model sexually-antagonistic, post-conception viability selection on a focal complex trait. We assume allele frequencies in adult males, , and adult females, , are at equilibrium. Under these conditions, we derive the relationship

|  |  |
| --- | --- |
|  |  |

where (Wright, 1951) is the fixation index with respect to the male and female subpopulations, i.e. the proportion of heterozygosity in the population that is due to allelic divergence between the sexes; and is the contribution to phenotypic variance due to GxSex,

|  |  |
| --- | --- |
|  | (20) |

where is the mean allele frequency at adulthood. is parameter that, importantly, is shared across all variants affecting the trait and can be thought of as the potential for sexually antagonistic selection acting on genetic variation for the trait in question.

Allele frequencies at the autosomal locus are assumed to be equal in males and female zygotes. If we assume equally-sized male and female subpopulations, at adulthood takes the form

|  |  |
| --- | --- |
|  | (21) |

Sexually-antagonistic selection acting on viability will cause divergence in allele frequencies between adult males and females. We write the relative viabilities of the homozygote for the reference allele, the heterozygote and the homozygote for the effect allele as for each sex . The selection coefficient and dominance coefficient can be frequency-dependent, in which case these coefficients take their values at equilibrium. We can write the additive selection coefficient of the effect allele as

|  |  |
| --- | --- |
|  | (22) |

Assuming that zygotes are at Hardy-Weinberg equilibrium, the allele frequency in each sex at adulthood is

|  |  |
| --- | --- |
|  | (23) |

where we neglected terms of order (Gillespie, 2004). Plugging **Eq. 23** into **Eq. 21**, the divergence between males and females post-selection is

|  |  |
| --- | --- |
|  | (24) |

We model the strength of viability selection acting on males and females as linear with the additive effect on a focal trait in each sex,

|  |  |
| --- | --- |
|  | (25) |

and make the simplifying assumption that allele frequencies are at equilibrium under sexually-antagonistic viability selection at the locus, such that selection favoring an allele in one sex is balanced by selection against that allele in the other sex,

|  |  |
| --- | --- |
|  | (26) |

Assuming that does not strictly equal , **Eq. 25, 26** together imply

|  |  |
| --- | --- |
|  | (28) |

Finally, using **Eq. 25***,*

|  |  |
| --- | --- |
|  | (29) |

which together with **Eq. 28** gives

|  |  |
| --- | --- |
|  | (30) |

We denote the heritability due to GxSex at the locus as and the parameter relating this contribution to the differentiation in allele frequencies as

|  |  |
| --- | --- |
|  | (31) |

and plug **Eq. 29** into **Eq. 25**, we get

|  |  |
| --- | --- |
|  | (32) |

### Gathering and Filtering Allele Frequency Data

We downloaded allele frequency data from the gnomAD dataset (Karczewski et al., 2020) a consortium of several ongoing global research projects. Specifically, we used *gnomAD v3.1.2* which consists of 76,156 whole genome sequences mapped to the GRCh38 reference. This data is labeled by gnomAD and available for download with information about the sample demographics. The data is divided into ten ancestry groups: African/African American samples (abbreviated “afr” in the gnomAD files); Amish (“ami”); Latino/Admixed American (“amr”); Ashkenazi Jewish (“asj”); East Asian (“eas”); Finnish (“fin”); Non-Finnish European (“nfe”); Middle Eastern (“mid”); South Asian (“sas”); and samples not assigned to any population, designated Other (“oth”). The data is also separately divided into two chromosomal sexes, labeled XX and XY, based on coverage of X and Y chromosomes for the given individual. Aneuploid individuals (e.g., X or XXY) are not included in the dataset. For the purposes of this study, we refer to XX as Female and XY as Male. Total numbers of individuals sampled can be found on the gnomAD website’s help page (<https://gnomad.broadinstitute.org/help>).

The gnomAD data is available as direct download in variant call format (.vcf) from their online browser (https://gnomad.broadinstitute.org/downloads). We downloaded files for all autosomes and used the command-line program *VCFTools* (Danecek et al., 2011) to parse the information. The original gnomAD .vcf files contain many more INFO fields than were necessary for the scope of our study, so we selected only the following specific columns: CHROM and POS, the chromosomal location; AN, the Allele Number, the total number of alleles sampled for the given locus; AC, the Allele Count, the number of alternative alleles found in the sample; and AF, the Allele Frequency, the frequency of alternative alleles in the sample (i.e., AC/AN). All of these fields (except for CHROM and POS) are given for each ancestry group and sex and are labeled as such in the .vcf fields e.g., AN\_afr\_XX.

We then passed this data through several filtering steps to remove loci which would not be useful to the analysis. During the gnomAD data acquisition process, we only chose sites for which the alternative alleles were single nucleotide polymorphisms (SNPs) as opposed to structural mutations such as insertions or deletions, and only kept those sites which were bi-allelic. After selecting the specific columns of interest but before performing any other filtering, our whole-genome dataset consisted of 2,288,867 sites. We filtered the data by first simply removing any loci with missing data (labeled as a “?” in the .vcf). This removed 3,698 sites, bringing the total number down from 2,288,867 to 2,285,169 sites. In an effort to avoid confounding results that could arise from population substructure, we split the data into the different ancestry groups labeled by gnomAD and worked with the data in each subpopulation separately from this point forward. We removed any loci where there were less than 1,000 alleles in the sample for that site in each specific ancestry group. The number of loci we removed at this step depended on the sample size of each group, as we removed fewer loci from large populations and removed more loci in small populations (REFX **Table S1**). Importantly, from this filtering step we completely removed the Amish and Middle Eastern populations because their low sample sizes meant that there were no loci which had more than 1,000 alleles in the sample. For this reason, we disregarded these populations for all future work.

### Estimating Male-Female FST

We estimated Male-Female for every site remaining in our dataset following **Eq. 21**. We note that this is an upward biased estimator—especially for low levels of , as it only takes non-negative values. To get and , we could directly use the AF\_XX column for each subpopulation. To get , we use (AC\_XX + AC\_XY) / (AN\_XX + AN\_XY), such that is given as the number of alternate alleles in the sample divided by the total number of alleles in the sample.

We removed any loci which resulted in “NA”, indicating that 4pq = 0, or in other words there were no samples with the male or female tag with the alternative allele for that site. This step resulted in a significant reduction in locus number by one or two orders of magnitude (REFX **Table S2**).

### Integrating with GWAS Data and Calculating VGxSex

Having completed these preliminary filtering steps, we integrated the gnomAD data with the GWAS data. We removed loci that were not found in both the gnomAD dataset and the UKB GWAS dataset. The number of loci removed through this step varied greatly across ancestry groups (REFX **Table S3**).

We used the point estimates and the standard errors of the sex-stratified GWAS (REFX **Table SXX**) for 27 physiological or physical traits. We also got an estimate of using **Eq. 20**, where and are the GWAS effect estimates and is the total alternate allele frequency as above.

Our final step of filtering involved filtering loci by GWAS p-value. To investigate the contribution of on a trait, we used four different p-value thresholds at 1e-3, 1e-5, 1e-8 and 1 (i.e., all SNPs; see REFX **Table S4** for the number of sites remaining for each p-value threshold). In the main text, we arbitrarily focus on the 1e-5 threshold reasoning that it strikes a good middle ground between sample size and noise. Results for other p-value thresholds are shown in REFX **Figs. XXX**.

### Estimating of the potential for sexually-antagonistic selection acting on variation in a trait ()

For each trait and population, we estimated using weighted least squares linear regression of to , with weight inversely proportional to our site-specific estimate of noise in the estimate of .

|  |  |
| --- | --- |
|  | (33) |

To simplify the estimation of we treat the allele frequency as perfectly estimated, and as independent of the allele frequency in the GWAS sample—as different data are used in the GWAS (UK Biobank) and in the allele frequency estimation (gnomAD). Under these assumptions,

|  |  |
| --- | --- |
|  | (34) |

and thus the task at hand is estimating . Using the law of total variance,

|  |  |
| --- | --- |
|  | (35) |

We begin with the argument of the first term,

|  |  |
| --- | --- |
|  | (36) |

where we denoted

|  |  |
| --- | --- |
|  | (37) |

for each sex Plugging **Eq. 36** into the first term of **Eq. 35**,

|  |  |
| --- | --- |
|  | (38) |

where the first and second step follow from the fact that is a constant. We can take note of the fact that is Normally distributed around , and in particular that it has no skewness. Therefore,

|  |  |
| --- | --- |
|  | (39) |

where is the skewness of . We can also note that

|  |  |
| --- | --- |
|  | (40) |

where we defined

|  |  |
| --- | --- |
|  |  |

and therefore is a Standard Normal and therefore is Chi-squared with one degree of freedom. **Eq. 40** now gives

|  |  |
| --- | --- |
|  | (41) |

Plugging **Eq. 39** and **Eq. 41** into **Eq. 38**, we find

|  |  |
| --- | --- |
|  | (42) |

We now turn to the second term of **Eq. 35**. First,

|  |  |
| --- | --- |
|  | (43) |

**Eq. 39** and **41** again give us

|  |  |
| --- | --- |
|  | (44) |

which then gives

|  |  |
| --- | --- |
|  | (45) |

Plugging **Eq. 42** and **Eq. 45** into **Eq. 35**, we get

|  |  |  |  |
| --- | --- | --- | --- |
|  | | | (46) |
|  |  |  | |

Finally, we estimate with the GWAS-derived point estimate of the effect and with its standard error, . Plugging back into **Eq. 34**, we get

|  |  |
| --- | --- |
|  | (47) |

To perform this estimation of A on the GWAS and FST data, we used paired FST and VGxSex points for all loci which passed all previous stages of filtering. Weights were set by **Eq. 33** and follow **Eq. 47** where and are the GWAS effect estimates as above, and and are the GWAS standard errors (SE) estimates for each locus’ effect size per trait.

To minimize the possibility of LD between sites used in the analysis as much as possible used the approximately independent LD blocks in Europeans (Berisa & Pickrell, 2016) as in REFX **Section “XXX”.** Namely, we subdivided the genome into 1703 LD blocks such that we expect sites from different LD blocks will have little to no LD between them. We iterated over the 1703 blocks and identified all post-filtering loci which fell within each block; we then randomly sampled one of these loci within each LD block and used this sample of (up to) 1703 sites to perform the weighted linear regression of on The slope and SE of this regression line was our estimate of A. We replicated this estimation process 1,000 times to generate 1,000 estimates of A. We then used the estimates of slope divided by the SE of these estimates to generate 1,000 Z-scores for A. The point estimate Z-score presented in REFX **Fig. 7B** is the mean of the 1,000 replicates, and the error bar is the middle 90% (i.e., the 5th to 95th quantile) of those replicates. In the main text, we focus on the results performed this estimation for Ashkenazi Jewish, Finnish, and Non-Finnish European populations as the other ancestry groups in gnomAD are more genetically diverged from the UKB White British sample and the GWAS estimates are expected to be less portable (Privé et al., 2022; Wang et al., 2020), similar to the UK Biobank dataset used to generate the GWAS data.

## Code and Data Availability

All code and sex-specific GWAS summary statistics are available at REFX XXX.

# References

Abbot, L., Bryant, S., Churchhouse, C., Ganna, A., Howrigan, D., Palmer, D., Neale, B., Walters, R., Carey, C., & The Hail Team. (n.d.). *UK Biobank -- Neale Lab*. Retrieved March 23, 2022, from http://www.nealelab.is/uk-biobank/

Arnqvist, G., & Rowe, L. (2005). *Sexual Conflict*. Princeton University Press. https://doi.org/10.1515/9781400850600

Auton, A., Abecasis, G. R., Altshuler, D. M., Durbin, R. M., Abecasis, G. R., Bentley, D. R., Chakravarti, A., Clark, A. G., Donnelly, P., Eichler, E. E., Flicek, P., Gabriel, S. B., Gibbs, R. A., Green, E. D., Hurles, M. E., Knoppers, B. M., Korbel, J. O., Lander, E. S., Lee, C., … Abecasis, G. R. (2015). A global reference for human genetic variation. *Nature*, *526*(7571), 68–74. https://doi.org/10.1038/nature15393

Barson, N. J., Aykanat, T., Hindar, K., Baranski, M., Bolstad, G. H., Fiske, P., Jacq, C., Jensen, A. J., Johnston, S. E., Karlsson, S., Kent, M., Moen, T., Niemelä, E., Nome, T., Næsje, T. F., Orell, P., Romakkaniemi, A., Sægrov, H., Urdal, K., … Primmer, C. R. (2015). Sex-dependent dominance at a single locus maintains variation in age at maturity in salmon. *Nature*, *528*(7582), 405–408. https://doi.org/10.1038/nature16062

Benonisdottir, S., & Kong, A. (2022). The Genetics of Participation: Method and Analysis. *BioRxiv*. https://doi.org/10.1101/2022.02.11.480067

Berisa, T., & Pickrell, J. K. (2016). Approximately independent linkage disequilibrium blocks in human populations. *Bioinformatics (Oxford, England)*, *32*(2), 283–285. https://doi.org/10.1093/bioinformatics/btv546

Bernabeu, E., Canela-Xandri, O., Rawlik, K., Talenti, A., Prendergast, J., & Tenesa, A. (2021). Sex differences in genetic architecture in the UK Biobank. *Nature Genetics*, *53*(9), 1283–1289. https://doi.org/10.1038/s41588-021-00912-0

Boyle, E. A., Li, Y. I., & Pritchard, J. K. (2017). An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell*, *169*(7), 1177–1186. https://doi.org/10.1016/j.cell.2017.05.038

Brown, B. C., Asian Genetic Epidemiology Network Type 2 Diabetes Consortium, Ye, C. J., Price, A. L., & Zaitlen, N. (2016). Transethnic Genetic-Correlation Estimates from Summary Statistics. *American Journal of Human Genetics*, *99*(1), 76–88. https://doi.org/10.1016/j.ajhg.2016.05.001

Bulik-Sullivan, B., Finucane, H. K., Anttila, V., Gusev, A., Day, F. R., Loh, P.-R., Duncan, L., Perry, J. R. B., Patterson, N., Robinson, E. B., Daly, M. J., Price, A. L., & Neale, B. M. (2015). An atlas of genetic correlations across human diseases and traits. *Nature Genetics*, *47*(11), 1236–1241. https://doi.org/10.1038/ng.3406

Bulik-Sullivan, B. K., Loh, P.-R., Finucane, H. K., Ripke, S., Yang, J., Patterson, N., Daly, M. J., Price, A. L., & Neale, B. M. (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics*, *47*(3), 291–295. https://doi.org/10.1038/ng.3211

Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O’Connell, J., Cortes, A., Welsh, S., Young, A., Effingham, M., McVean, G., Leslie, S., Allen, N., Donnelly, P., & Marchini, J. (2018). The UK Biobank resource with deep phenotyping and genomic data. *Nature*, *562*(7726), 203–209. https://doi.org/10.1038/s41586-018-0579-z

Carole Hooven. (2021). *T: The Story of Testosterone, the Hormone that Dominates and Divides Us*. Henry Holt and Co.

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, *4*(1), 7. https://doi.org/10.1186/s13742-015-0047-8

Cheng, C., & Kirkpatrick, M. (2016). Sex-Specific Selection and Sex-Biased Gene Expression in Humans and Flies. *PLOS Genetics*, *12*(9), e1006170. https://doi.org/10.1371/journal.pgen.1006170

Choi, S. W., Mak, T. S.-H., & O’Reilly, P. F. (2020). Tutorial: a guide to performing polygenic risk score analyses. *Nature Protocols*, *15*(9), 2759–2772. https://doi.org/10.1038/s41596-020-0353-1

Connallon, T., Cox, R. M., & Calsbeek, R. (2010). Fitness consequences of sex-specific selection. *Evolution; International Journal of Organic Evolution*, *64*(6), 1671–1682. https://doi.org/10.1111/j.1558-5646.2009.00934.x

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

des Marais, D. L., Hernandez, K. M., & Juenger, T. E. (2013). Genotype-by-environment interaction and plasticity: Exploring genomic responses of plants to the abiotic environment. In *Annual Review of Ecology, Evolution, and Systematics* (Vol. 44, pp. 5–29). Annual Reviews Inc. https://doi.org/10.1146/annurev-ecolsys-110512-135806

Duncan, L., Shen, H., Gelaye, B., Meijsen, J., Ressler, K., Feldman, M., Peterson, R., & Domingue, B. (2019). Analysis of polygenic risk score usage and performance in diverse human populations. *Nature Communications*, *10*(1), 3328. https://doi.org/10.1038/s41467-019-11112-0

Falconer, D. S., & Mackay, T. F. C. (1996). *Introduction to Quantitiative Genetics* (Addison Wesley Longman, Ed.; 4th ed.). Harlow.

Flynn, E., Tanigawa, Y., Rodriguez, F., Altman, R. B., Sinnott-Armstrong, N., & Rivas, M. A. (2021). Sex-specific genetic effects across biomarkers. *European Journal of Human Genetics : EJHG*, *29*(1), 154–163. https://doi.org/10.1038/s41431-020-00712-w

Galinsky, K. J., Reshef, Y. A., Finucane, H. K., Loh, P.-R., Zaitlen, N., Patterson, N. J., Brown, B. C., & Price, A. L. (2019). Estimating cross-population genetic correlations of causal effect sizes. *Genetic Epidemiology*, *43*(2), 180–188. https://doi.org/10.1002/gepi.22173

Ge, T., Chen, C.-Y., Neale, B. M., Sabuncu, M. R., & Smoller, J. W. (2017). Phenome-wide heritability analysis of the UK Biobank. *PLoS Genetics*, *13*(4), e1006711. https://doi.org/10.1371/journal.pgen.1006711

Gillespie, J. H. (2004). *Population Genetics: A Concise Guide* (2nd ed.). Johns Hopkins University Press.

Graffelman, J., & Moreno, V. (2013). The mid p-value in exact tests for Hardy-Weinberg equilibrium. *Statistical Applications in Genetics and Molecular Biology*, *12*(4), 433–448. https://doi.org/10.1515/sagmb-2012-0039

Graffelman, J., & Weir, B. S. (2016). Testing for Hardy-Weinberg equilibrium at biallelic genetic markers on the X chromosome. *Heredity*, *116*(6), 558–568. https://doi.org/10.1038/hdy.2016.20

Harpak, A., & Przeworski, M. (2021). The evolution of group differences in changing environments. *PLOS Biology*, *19*(1), e3001072. https://doi.org/10.1371/journal.pbio.3001072

Harper, J. A., Janicke, T., & Morrow, E. H. (2021). Systematic review reveals multiple sexually antagonistic polymorphisms affecting human disease and complex traits. *Evolution*, *75*(12), 3087–3097. https://doi.org/10.1111/evo.14394

Harrison, P. W., Wright, A. E., Zimmer, F., Dean, R., Montgomery, S. H., Pointer, M. A., & Mank, J. E. (2015). Sexual selection drives evolution and rapid turnover of male gene expression. *Proceedings of the National Academy of Sciences*, *112*(14), 4393–4398. https://doi.org/10.1073/pnas.1501339112

Huang, W., Campbell, T., Carbone, M. A., Jones, W. E., Unselt, D., Anholt, R. R. H., & Mackay, T. F. C. (2020). Context-dependent genetic architecture of Drosophila life span. *PLOS Biology*, *18*(3), e3000645. https://doi.org/10.1371/journal.pbio.3000645

Huang, W., Carbone, M. A., Lyman, R. F., Anholt, R. R. H., & Mackay, T. F. C. (2020). Genotype by environment interaction for gene expression in Drosophila melanogaster. *Nature Communications*, *11*(1), 5451. https://doi.org/10.1038/s41467-020-19131-y

Hudson, A. I., Odell, S. G., Dubreuil, P., Tixier, M.-H., Praud, S., Runcie, D. E., & Ross-Ibarra, J. (2022). Analysis of genotype-by-environment interactions in a maize mapping population. *G3 (Bethesda, Md.)*, *12*(3). https://doi.org/10.1093/g3journal/jkac013

Karczewski, K. J., Francioli, L. C., Tiao, G., Cummings, B. B., Alföldi, J., Wang, Q., Collins, R. L., Laricchia, K. M., Ganna, A., Birnbaum, D. P., Gauthier, L. D., Brand, H., Solomonson, M., Watts, N. A., Rhodes, D., Singer-Berk, M., England, E. M., Seaby, E. G., Kosmicki, J. A., … MacArthur, D. G. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*, *581*(7809), 434–443. https://doi.org/10.1038/s41586-020-2308-7

Kasimatis, K. R., Abraham, A., Ralph, P. L., Kern, A. D., Capra, J. A., & Phillips, P. C. (2021). Evaluating human autosomal loci for sexually antagonistic viability selection in two large biobanks. *Genetics*, *217*(1). https://doi.org/10.1093/genetics/iyaa015

Kidwell, J. F., Clegg, M. T., Stewart, F. M., & Prout, T. (1977). Regions of stable equilibria for models of differential selection in the two sexes under random mating. *Genetics*, *85*(1), 171–183. https://doi.org/10.1093/genetics/85.1.171

Liu, D., Sartor, M. A., Nader, G. A., Gutmann, L., Treutelaar, M. K., Pistilli, E. E., Iglayreger, H. B., Burant, C. F., Hoffman, E. P., & Gordon, P. M. (2010). Skeletal muscle gene expression in response to resistance exercise: sex specific regulation. *BMC Genomics*, *11*, 659. https://doi.org/10.1186/1471-2164-11-659

Long, Q., Rabanal, F. A., Meng, D., Huber, C. D., Farlow, A., Platzer, A., Zhang, Q., Vilhjálmsson, B. J., Korte, A., Nizhynska, V., Voronin, V., Korte, P., Sedman, L., Mandáková, T., Lysak, M. A., Seren, Ü., Hellmann, I., & Nordborg, M. (2013). Massive genomic variation and strong selection in Arabidopsis thaliana lines from Sweden. *Nature Genetics*, *45*(8), 884–890. https://doi.org/10.1038/ng.2678

Lumish, H. S., O’Reilly, M., & Reilly, M. P. (2020). Sex Differences in Genomic Drivers of Adipose Distribution and Related Cardiometabolic Disorders: Opportunities for Precision Medicine. In *Arteriosclerosis, thrombosis, and vascular biology* (Vol. 40, Issue 1, pp. 45–60). NLM (Medline). https://doi.org/10.1161/ATVBAHA.119.313154

Lynch, M., & Walsh, B. (1998). *Genetics and Analysis of Quantitative Traits* (1st ed.). Sinauer Associates.

MacQueen, A. H., Zhang, L., Bonette, J., Boe, A. R., Fay, P. A., Fritschi, F. B., Lowry, D. B., Mitchell, R. B., Rouquette, F. M., Wu, Y., & others. (2021). Mapping of genotype-by-environment interactions in phenology identifies two cues for flowering in switchgrass (Panicum virgatum). *BioRxiv*.

Mars, N., Koskela, J. T., Ripatti, P., Kiiskinen, T. T. J., Havulinna, A. S., Lindbohm, J. v, Ahola-Olli, A., Kurki, M., Karjalainen, J., Palta, P., FinnGen, Neale, B. M., Daly, M., Salomaa, V., Palotie, A., Widén, E., & Ripatti, S. (2020). Polygenic and clinical risk scores and their impact on age at onset and prediction of cardiometabolic diseases and common cancers. *Nature Medicine*, *26*(4), 549–557. https://doi.org/10.1038/s41591-020-0800-0

McLaren, W., Gil, L., Hunt, S. E., Riat, H. S., Ritchie, G. R. S., Thormann, A., Flicek, P., & Cunningham, F. (2016). The Ensembl Variant Effect Predictor. *Genome Biology*, *17*(1), 122. https://doi.org/10.1186/s13059-016-0974-4

Mills, M., Barban, N., & Tropf, F. C. (2020). *An introduction to Statistical Genetic Data Analysis*. The MIT Press.

Mokkonen, M., & Crespi, B. J. (2015). Genomic conflicts and sexual antagonism in human health: insights from oxytocin and testosterone. *Evolutionary Applications*, *8*(4), 307–325. https://doi.org/10.1111/eva.12244

Mostafavi, H., Harpak, A., Agarwal, I., Conley, D., Pritchard, J. K., & Przeworski, M. (2020). Variable prediction accuracy of polygenic scores within an ancestry group. *ELife*, *9*. https://doi.org/10.7554/eLife.48376

Nagpal, S., Tandon, R., & Gibson, G. (2022). Canalization of the Polygenic Risk for Common Diseases and Traits in the UK Biobank Cohort. *Molecular Biology and Evolution*. https://doi.org/10.1093/molbev/msac053

Ni, G., Moser, G., Schizophrenia Working Group of the Psychiatric Genomics Consortium, Wray, N. R., & Lee, S. H. (2018). Estimation of Genetic Correlation via Linkage Disequilibrium Score Regression and Genomic Restricted Maximum Likelihood. *American Journal of Human Genetics*, *102*(6), 1185–1194. https://doi.org/10.1016/j.ajhg.2018.03.021

Oliva, M., Muñoz-Aguirre, M., Kim-Hellmuth, S., Wucher, V., Gewirtz, A. D. H., Cotter, D. J., Parsana, P., Kasela, S., Balliu, B., Viñuela, A., Castel, S. E., Mohammadi, P., Aguet, F., Zou, Y., Khramtsova, E. A., Skol, A. D., Garrido-Martín, D., Reverter, F., Brown, A., … Stranger, B. E. (2020). The impact of sex on gene expression across human tissues. *Science*, *369*(6509), eaba3066. https://doi.org/10.1126/science.aba3066

Pasquali, R. (2006). Obesity and androgens: facts and perspectives. *Fertility and Sterility*, *85*(5), 1319–1340. https://doi.org/10.1016/j.fertnstert.2005.10.054

Patel, R. A., Musharoff, S. A., Spence, J. P., Pimentel, H., Tcheandjieu, C., Mostafavi, H., Sinnott-Armstrong, N., Clarke, S. L., Smith, C. J., Program, V. A. M. V., Durda, P. P., Taylor, K. D., Tracy, R., Liu, Y., Johnson, C. W., Aguet, F., Ardlie, K. G., Gabriel, S., Smith, J., … Pritchard, J. K. (2021). Effect sizes of causal variants for gene expression and complex traits differ between populations. *BioRxiv*. https://doi.org/10.1101/2021.12.06.471235

Pirastu, N., Cordioli, M., Nandakumar, P., Mignogna, G., Abdellaoui, A., Hollis, B., Kanai, M., Rajagopal, V. M., Parolo, P. D. B., Baya, N., Carey, C. E., Karjalainen, J., Als, T. D., van der Zee, M. D., Day, F. R., Ong, K. K., Morisaki, T., de Geus, E., Bellocco, R., … Ganna, A. (2021). Genetic analyses identify widespread sex-differential participation bias. *Nature Genetics*, *53*(5), 663–671. https://doi.org/10.1038/s41588-021-00846-7

Power, R. A., Kyaga, S., Uher, R., MacCabe, J. H., Långström, N., Landen, M., McGuffin, P., Lewis, C. M., Lichtenstein, P., & Svensson, A. C. (2013). Fecundity of Patients With Schizophrenia, Autism, Bipolar Disorder, Depression, Anorexia Nervosa, or Substance Abuse vs Their Unaffected Siblings. *JAMA Psychiatry*, *70*(1), 22. https://doi.org/10.1001/jamapsychiatry.2013.268

Privé, F., Aschard, H., Carmi, S., Folkersen, L., Hoggart, C., O’Reilly, P. F., & Vilhjálmsson, B. J. (2022). Portability of 245 polygenic scores when derived from the UK Biobank and applied to 9 ancestry groups from the same cohort. *The American Journal of Human Genetics*, *109*(1), 12–23. https://doi.org/10.1016/j.ajhg.2021.11.008

Purcell, S., & Chang, C. (n.d.). *Plink 2.00 alpha*. Retrieved March 23, 2022, from www.cog-genomics.org/plink/2.0/

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, *81*(3), 559–575. https://doi.org/10.1086/519795

Ruzicka, F., Dutoit, L., Czuppon, P., Jordan, C. Y., Li, X., Olito, C., Runemark, A., Svensson, E. I., Yazdi, H. P., & Connallon, T. (2020). The search for sexually antagonistic genes: Practical insights from studies of local adaptation and statistical genomics. *Evolution Letters*, *4*(5), 398–415. https://doi.org/10.1002/evl3.192

Ruzicka, F., Holman, L., & Connallon, T. (2021). Polygenic signals of sexually antagonistic selection in contemporary human genomes. *BioRxiv*. https://doi.org/10.1101/2021.09.20.461171

Schroderus, E., Jokinen, I., Koivula, M., Koskela, E., Mappes, T., Mills, S. C., Oksanen, T. A., Poikonen, T., Ketterson, A. E. E. D., & McPeek, E. M. A. (2010). Intra‐ and Intersexual Trade‐Offs between Testosterone and Immune System: Implications for Sexual and Sexually Antagonistic Selection. *The American Naturalist*, *176*(4), E90–E97. http://www.jstor.org/stable/10.1086/656264

Sella, G., & Barton, N. H. (2019). Thinking About the Evolution of Complex Traits in the Era of Genome-Wide Association Studies. *Annual Review of Genomics and Human Genetics*, *20*(1), 461–493. https://doi.org/10.1146/annurev-genom-083115-022316

Shi, H., Kichaev, G., & Pasaniuc, B. (2016). Contrasting the Genetic Architecture of 30 Complex Traits from Summary Association Data. *American Journal of Human Genetics*, *99*(1), 139–153. https://doi.org/10.1016/j.ajhg.2016.05.013

Shi, H., Mancuso, N., Spendlove, S., & Pasaniuc, B. (2017). Local Genetic Correlation Gives Insights into the Shared Genetic Architecture of Complex Traits. *The American Journal of Human Genetics*, *101*(5), 737–751. https://doi.org/10.1016/j.ajhg.2017.09.022

Sinnott-Armstrong, N., Naqvi, S., Rivas, M., & Pritchard, J. K. (2021). Gwas of three molecular traits highlights core genes and pathways alongside a highly polygenic background. *ELife*, *10*, 1–35. https://doi.org/10.7554/eLife.58615

Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., Liu, B., Matthews, P., Ong, G., Pell, J., Silman, A., Young, A., Sprosen, T., Peakman, T., & Collins, R. (2015). UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. *PLOS Medicine*, *12*(3), e1001779. https://doi.org/10.1371/journal.pmed.1001779

Urbut, S. M., Wang, G., Carbonetto, P., & Stephens, M. (2019). Flexible statistical methods for estimating and testing effects in genomic studies with multiple conditions. *Nature Genetics*, *51*(1), 187–195. https://doi.org/10.1038/s41588-018-0268-8

van Doorn, G. S. (2009). Intralocus sexual conflict. *Annals of the New York Academy of Sciences*, *1168*, 52–71. https://doi.org/10.1111/j.1749-6632.2009.04573.x

Wang, Y., Guo, J., Ni, G., Yang, J., Visscher, P. M., & Yengo, L. (2020). Theoretical and empirical quantification of the accuracy of polygenic scores in ancestry divergent populations. *Nature Communications*, *11*(1), 3865. https://doi.org/10.1038/s41467-020-17719-y

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-Statistics for the Analysis of Population Structure. *Evolution*, *38*(6), 1358. https://doi.org/10.2307/2408641

Wigginton, J. E., Cutler, D. J., & Abecasis, G. R. (2005). A note on exact tests of Hardy-Weinberg equilibrium. *American Journal of Human Genetics*, *76*(5), 887–893. https://doi.org/10.1086/429864

Wright, S. (1951). The genetical structure of populations. *Annals of Eugenics*, *15*(4), 323–354. https://doi.org/10.1111/j.1469-1809.1949.tb02451.x

Young, A. I., Benonisdottir, S., Przeworski, M., & Kong, A. (2019). Deconstructing the sources of genotype-phenotype associations in humans. *Science (New York, N.Y.)*, *365*(6460), 1396–1400. https://doi.org/10.1126/science.aax3710

Young, A. I., Wauthier, F. L., & Donnelly, P. (2018). Identifying loci affecting trait variability and detecting interactions in genome-wide association studies. *Nature Genetics*, *50*(11), 1608–1614. https://doi.org/10.1038/s41588-018-0225-6