

Cell Genomics, Volume 3

Supplemental information

**Amplification is the primary mode
of gene-by-sex interaction in complex human traits**

Carrie Zhu, Matthew J. Ming, Jared M. Cole, Michael D. Edge, Mark Kirkpatrick, and Arbel Harpak

Supplementary Materials for: Amplification is the Primary Mode of Gene-by-Sex Interaction in Complex Human Traits

Carrie Zhu^{1,2}, Matthew J. Ming^{1,2}, Jared M. Cole^{1,2}, Michael D. Edge³, Mark Kirkpatrick², Arbel Harpak^{1,2,+}

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

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

³ Department of Quantitative and Computational Biology, University of Southern California, Los Angeles, CA

+ Lead Contact: arbelharpak@utexas.edu

Table of Contents

Supplemental Methods.....	4
Text S1: Genotype data.....	4
Text S2: Phenotype data	4
Text S3: GWAS.....	4
Text S4: Miami plots	4
Text S5: Heritability and genetic correlation estimation using LD Score Regression	5
Text S6: Qualitative differences between our conclusions and approaches based on independent analysis of individual sex-heterogenous SNPs	5
Text S7: Posterior estimates of sex-specific effect sizes.....	6
Text S8: Improved utility of sex-specific models for complex phenotype prediction.	7
Text S9: Data Filtering for Sexually-Antagonistic Selection Analysis.	9
Text S10: Estimating Male-Female F_{ST} and $V_{G \times Sex}$	10
Text S11: Competing models for sex differences in trait variance	11
Supplementary Figures	13
Figure S1. Sex-specific Miami plots, related to Figure 3	13
Figure S2. All hypothesis covariance matrices, related to Figure 3	14
Figure S3 Inferring polygenic covariance structure, related to Figure 3	15
Figure S4 Testing p-value thresholds for input into <i> mash </i> , related to Figure 3	17
Figure S5: Simulation of sex-invariant genetic effects but with variable environmental variance, related to Figure 3.....	19
Figure S6: Simulation of covariance structure, related to Figure 3	21
Figure S7: Proportion of non-trivial weights on hypothesis matrices, related to Figure 3 and Figure 4	22
Figure S8: Across traits, phenotypic variance sex ratio correlates with the phenotypic mean sex ratio, related to Figure 4.....	23
Figure S9: Process of generating posterior estimates from <i> mash </i> , related to Figure 3	24

Figure S10: Process of comparing models for complex phenotype prediction, related to Figure 3	25
Figure S11: Evaluating evidence for systematic amplification, related to Figure 2.....	27
Figure S12: The predictive utility of G-by-Sex aware polygenic scores, related to Figure 3.....	29
Figure S13: Genetic effect and testosterone levels – using additive model PGS, related to Figure 5....	31
Figure S14: Genetic effect and testosterone levels – using sex-specific model PGS, related to Figure 5	33
Figure S15: Genetic effect and PGS for testosterone, related to Figure 5	34
Figure S16: Genetic effect residualized for age and testosterone levels, related to Figure 5.....	36
Figure S17: Example of shared polygenic and environmental amplification in muscle, related to Figure 6	37
Figure S18: Competing models for sex differences in trait variance, related to Figure 6	38
Figure S19: Z-scores for strength of sexually-antagonistic selection, related to Figure 7.....	40
Figure S20: Z-scores using UK Biobank data, related to Figure 7	41
Supplementary Tables	42
Table S1: Filtering site with small sample sizes, related to Figure 7.....	42
Table S2. Filtering monoallelic sites, related to Figure 7	43
Table S3: Filtering sites that were not overlapping between the gnomAD dataset and the GWAS dataset, related to Figure 7.....	44
Table S4: Filtering by p-value threshold, related to Figure 7.....	45
Table S5: Identification of SNPs susceptible to mis-mapping to sex chromosomes— Kasimatis et al., related to Figure 7.....	46
Table S6: Characterizing GxSex based on independent analysis of individual sex-heterogenous SNPs, related to Figure 3.....	47
Supplementary Data	48
Data S1-S27: Mixture weights on hypothesis covariance matrices, related to Figure 3	48
Data S1: Albumin.....	49
Data S2: Arm fat-free mass (L)	50
Data S3: Arm fat-free mass (R)	51
Data S4: BMI.....	52
Data S5: Calcium	53
Data S6: Creatinine	54
Data S7: Diastolic blood pressure	55
Data S8: Eosinophil percentage	56
Data S9: Forced vital capacity	57

Data S10: HbA1c	58
Data S11: Height	59
Data S12: Hip circumference	60
Data S13: IGF-1	61
Data S14: Lymphocyte percentage	62
Data S15: Total protein	63
Data S16: Pulse rate	64
Data S17: Red blood cell count	65
Data S18: SHBG	66
Data S19: Systolic blood pressure.....	67
Data S20: Testosterone.....	68
Data S21: Urate	69
Data S22: Urea	70
Data S23: Waist circumference.....	71
Data S24: Waist to hip ratio	72
Data S25: Weight	73
Data S26: Whole body fat mass	74
Data S27: Waist:hip (BMI adjusted).....	75
References	76

Supplemental Methods

Text S1: Genotype data

We focused on autosomal bi-allelic SNPs with an INFO score greater than 0.8. Using *plink 2.0 alpha*¹⁻⁶, we further retained variants with calling rate > 0.95, Hardy-Weinberg equilibrium test p-value > 10⁻⁹, minor allele frequency (MAF) > 0.001 and MAF < 0.999, following a QC procedure used by the Neale Lab on version 2 of UK Biobank⁷. These were computed on the aforementioned quality-controlled sample set (**Methods “UK Biobank sample characteristics”**) and resulted in 9,607,691 SNPs.

Text S2: Phenotype data

Our analysis consisted of 27 continuous traits for their relatively high SNP heritability estimates, based on LD Score regression^{7,8}. BMI-adjusted waist:hip ratio (WHR) was calculated by regressing WHR on BMI and obtaining the residuals, as obtained using the following commands in R:

```
model <- lm(<WHR>~<BMI>, data=<dataframe>)  
residuals <- summary(model)$residuals
```

Text S3: 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 applying the same regression model for each sex independently. Any variants with missing values in the summary statistics were removed from further analysis.

Text S4: Miami plots

We used the *Ensembl Variant Effect Predictor (VEP)*⁹ based on the GRCh38 genome build to annotate SNPs with p-value < 5x10⁻⁸ in Miami plots (**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.

Text S5: 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^{8,10}. 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.⁸.

Text S6: Qualitative differences between our conclusions and approaches based on independent analysis of individual sex-heterogeneous SNPs

A common approach for detecting and characterizing GxSex based on sex-stratified GWAS data is to test the hypothesis of sex differences in genetic effects at each site independently. As an example, Traglia et al.¹¹ meta-analyzed sex-stratified GWAS from various sources, many of which standardized effects within-sex. Traglia et al. defined sex-heterogeneous (“sex-het”) SNPs as ones where a t-test testing a null hypothesis of equal effects in males and females was significant at a level of 0.05. They then characterized the pervasiveness of different modes of GxSex focusing on this subset. In particular, they categorized sex-het SNPs where the marginal association p-value was significant at a level of 0.05 as “having an effect in only one sex”. They then categorized the remaining sex-het SNPs as having opposite signs of effects in the two sexes or the same sign and different magnitudes. With these categorizations at hand, they argued that the vast majority of sex-het SNPs have an effect in only one sex.

In this study, we analyzed the polygenic covariance of genetic effects genome-wide, rather than focusing on significant individual SNPs. We proposed that sex differences were largely due to differences in magnitude rather than opposite or sex-private effects. Here, we show that the seemingly discrepant arguments may be a direct result of the different analysis approaches. We show that when we generate data from a generative model adhering to pervasive amplification and use Traglia’s et al.’s approach to characterize GxSex based on the simulated data, their classification suggests the same qualitative result—that GxSex manifests primarily via effects that are private to one of the sexes.

We performed a simulation study following a procedure similar to that in the section “**Simulating equal genetic effects and heterogeneous estimation noise among sexes**” of the main text. Here, however, we sample genetic effects using the covariance mixtures estimated for five traits – height, BMI, creatinine, IGF-1 and systolic blood pressure (**Data S11, 4, 6, 13, and 19**

respectively). We set the female to male environmental variance ratio as 1.2 and the heritability as 0.05. We simulated genotypes and phenotypic values, and then standardized phenotypic values within-sex and performed a sex-stratified GWAS. Using the effect estimates and corresponding standard errors for males and females, we calculated a t-statistic for each SNP, where

$$t = \frac{\hat{\beta}_{female} - \hat{\beta}_{male}}{\sqrt{SE_{female}^2 + SE_{male}^2}},$$

and identified SNPs with p-value < 0.05 as sex-het SNPs. We further categorized sex-het SNPs as having an effect in only one sex if the effect was only significant in one sex (t-test p-value < 0.05). For SNPs where the sex-specific effect was significant for both sexes or for neither, the SNPs were categorized based on if the effects were of opposite or same sign.

For data generated based on the covariance structures, we indeed find that the classification of sex-het SNPs qualitatively mirrors that of Traglia et al., with the majority of the sex het SNPs categorized as having an effect in just one sex (**Table S6**). Importantly, this is despite having simulated traits with pervasive differences in the magnitude of effects. We therefore conclude that such an approach, including the within-sex standardization of effect sizes (which heavily dilutes signals of pervasive magnitude differences), the ascertainment bias resulting from a focus on individually-significant sex-heterogenous SNPs, and the use of the above-described classification result in mis-characterization of the mode of GxSex, can lead to the mischaracterization of GxSex in complex traits.

Text S7: 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 lesser variation in effect sizes between males and females. Therefore, using the average of the fitted mixture model over 100 repetitions (**Methods “Mixture weights for covariance structure between male and female effects”**), we computed posterior estimates for each trait to be used in further analysis (**Text S8**). *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 a p-value of the same rank (**Fig. S10**).

Text S8: Improved utility of sex-specific models for complex phenotype prediction.

The pervasiveness of GxSex that we infer, alongside the mixture 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 four polygenic scores (PGS) for various traits, together with covariates (**Fig. S12**):

1. An additive PGS, assuming no GxSex, based on both-sex GWAS summary statistics
2. An additive PGS, standardized by sex, based on both-sex GWAS summary statistics
3. A sex-specific, but polygenic covariance-naïve PGS, based on stratified GWAS, fit independently for each sex.
4. A sex-specific, covariance-aware PGS based on posterior sex-specific effect estimates (**Text S7**)

The estimation, prediction and evaluation pipeline are illustrated in **Fig. S10**, and, for models (a) and (b), parallels the procedure described by Choi¹¹.

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 (**Text S3**), generating both-sex, female-specific, and male-specific summary statistics. For the additive model standardized by sex, we used phenotype values standardized within-sex before performing GWAS to get both-sex summary statistics. Third, to obtain posterior estimates, we first input the effect sizes and standard errors from the male-specific and female-specific GWAS, and took the average of 100 resampling estimates of mixture proportion vectors estimated from a random subset of SNPs with p-value < 1e-5 available in each of 1703 LD blocks¹². We used the 1e-5 p-value threshold to create the random subset to provide stronger signals and patterns for *mash* to learn from. We then feed estimated mixture proportion vectors into *mash* to perform the refined, covariance-aware estimation of effects. The output from *mash* included posterior mean and a local false sign rate (*lfsr*) for each SNP. We sorted SNPs by *lfsr* and matched it to a ranked list of p-values from the additive GWAS summary statistics to create a “pseudo p value” which we use later for thresholding.

Afterwards, we performed clumping separately for the additive both-sex, additive both-sex standardized by sex, male-specific additive, female-specific additive, male covariance-aware and female covariance aware models to produce a subset of significant SNPs that are approximately independent using *plink 1.9 beta's* --clump command, removing SNPs with pairwise LD threshold $r^2 > 0.1$ or within 250kb. To estimate pairwise LD values, we used a sample of 187 unrelated individuals (population codes GBR and CEU) from 1000 Genomes phase 3¹³.

Using the resulting subset of SNPs, we estimated 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 used *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 24 PGS runs were performed for each trait over the combination of six models and four p-value thresholds (**Fig. S10**).

Finally, to assess the prediction accuracy, we computed R^2 for the following,

$$y \sim \text{birth year} + \text{sex} + \text{principal componenets} + (\text{sex} \times M) + ((1 - \text{sex}) \times F)$$

$$\text{sex}: \{0(\text{female}), 1(\text{male})\}$$

$$M: \text{male PGS} \quad F: \text{female PGS}$$

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 four predictors—additive model, additive standardized by sex model, sex-specific additive model, and sex-specific polygenic covariance-aware model—we selected the p-value threshold with the greatest R^2 . We also calculated the incremental R^2 , which is the increment in R^2 after adding the PGS to a null model with only the covariates (**Fig. S10**).

We performed 20-fold cross validation for our PGS procedure. Each prediction was performed on a different randomly sampled 5K female and 5K male test set, with the remaining used as the training set. For each model, we averaged the greatest incremental R^2 from the p-value thresholds over the twenty folds for comparison across all models.

The sex-specific, covariance-aware PGS model outperformed the additive model (including model standardized by sex) for 20/27 traits (**Fig. S12**). The sex-specific, covariance-aware PGS model outperformed the sex-specific, covariance-naïve model for all traits. The additive model also outperformed the sex-specific covariance-naïve for all traits but testosterone.

The additive model using phenotype values standardized within sex only outperformed the base additive model for 11/27 traits.

The increase of prediction accuracy in the additive model over the sex-specific covariance-naïve model may be due the additive model being estimated on nearly twice the sample size as the sex-specific models. We chose not to equalize the sample sizes to align the comparison to what would be used in practice. However, despite the disadvantage in decreased sample size, the covariance-aware PGS model outperformed the additive model in approximately 3/4 of the total traits. Interestingly, the additive model standardized by sex did not outperform the base additive model for many traits, including those showing evident sexual dimorphism such as testosterone. These findings support the argument for characterizing GxSex using multiple covariance structures rather than broadly applying within-sex standardization, which can weaken amplification signals and sex differences in the data.

Polygenic scores used in **Methods “Testosterone as an amplifier”** and **Fig. 2** were based on a 5-fold cross validation with 50K individuals in the test set (25K males and 25K females). The polygenic scores were gathered from the first fold for each trait.

Text S9: Data Filtering for Sexually-Antagonistic Selection Analysis.

We downloaded allele frequency data from the gnomAD dataset¹⁴. Specifically, we used gnomAD v3.1.2, which consists of 76,156 whole genome sequences mapped to the GRCh38 reference. In order to allow this assembly to conform to the GRCh37 build used in the UK Biobank GWAS, we used the UCSC in-browser tool LiftOver¹⁵ (<https://genome.ucsc.edu/cgi-bin/hgLiftOver>) to identify and convert the GRCh38 sites of interest to their corresponding GRCh37 positions. Allele count summaries are available for the total sample, as well as stratified by sex chromosomes karyotype and genetic ancestry groupings: 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 are designated Other (“oth”). Aneuploid individuals (e.g., X or XXY) are not included in the dataset. For the purposes of this study, we again 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>).¹⁴

We downloaded gnomAD VCF files from the gnomAD [browser](#) for all autosomes and used VCFTools¹⁶ to parse the file. We filtered the data to exclude insertions or deletions, and only kept

bi-allelic SNPs. We further removed missing data (3,698 sites). These filtering steps resulted in 2,285,169 remaining 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 sites with less than 1,000 alleles in each ancestry group independently. The number of sites we removed at this step depended on the sample size of each group (**Table S1**). This step resulted in the complete removal of the Amish and Middle Eastern subsamples because their low sample sizes.

Finally, we filtered out sites where difference in male and female allele counts may be partly or fully driven by the mismapping of autosomal reads to sex chromosomes or vice-versa^{17,18}. We follow a similar approach to Kasimatis et al. to identify such sites¹⁸. In particular, for every SNP, we extract the 301bp sequence surrounding the SNP, with the SNP's position at the center, from the GRCh37 genome assembly¹⁹. We further shorten the 301bp sequence into three 150bp-long subsequences with the SNP's position at the center, start or end of the sequence. We then use Mega-BLAST through NCBI's command-line BLAST tool²⁰ to search for regions of high sequence homology to either of the three subsequences. If any of the three was found to have a 90% or greater sequence identity to a sequence on a sex chromosome, we filtered out the site. This filtering was performed agnostic of ancestry.

We qualitatively compared our list of filtered SNPs to Kasimatis et al.¹⁸. We based our comparison on sites considered in both works. In particular, we limited the comparison to the genotype array sites, as they did not consider UKB imputed genotypes, and further only to sites we had not removed in a previous filtering step. In terms of parameters, our 90% homology threshold is the same as Kasimatis et al., but we diverge from their algorithm in testing three different sequences per SNP; and also in using shorter sequences (150bp) to model for the mapping of short reads rather than the hybridization of probes. In general, our approach tended to identify more sites than Kasimatis et al. as invalid for analysis (**Table S5**).

Text S10: Estimating Male-Female F_{ST} and $V_{G \times Sex}$

We estimated Male-Female F_{ST} for every site remaining in our dataset based on the new estimator we propose in **Eq. 51**, based on the sample allele frequencies. We removed sites for which there were no alternative allele calls for either males or females. This step resulted in a significant

reduction in sample size by one or two orders of magnitude, depending on the ancestry (**Table S2**). We further removed sites that were not found in both the gnomAD dataset and the UKB GWAS dataset. The number of sites removed through this step varied greatly across ancestry groups (**Table S3**).

We used the point estimates and the standard errors of the sex-stratified GWAS (**Table S6**) for 27 physiological or physical traits. We further filtered sites by GWAS p-value. We used four different p-value thresholds at 10^{-3} , 10^{-5} , 10^{-8} and 1 (i.e., all SNPs; see **Table S4** for the number of sites remaining for each p-value threshold). In the main text, we focus on the 10^{-5} threshold, reasoning that it strikes a reasonable, albeit arbitrary, middle ground between sample size and noise. Results for other p-value thresholds are shown in **Fig. S19**.

Finally, we obtained an estimate of V_{GxSex} using **Eq. 19**, where β_m^2 and β_f^2 were estimated using the squared GWAS effect estimates, and p is the total alternate allele frequency as above. For detail on the estimation of the sampling error of V_{GxSex} , see “**Estimating the potential for sexually-antagonistic selection on standing variation (A)**” in the main text.

Text S11: Competing models for sex differences in trait variance

In the main text section “**Are polygenic and environmental effects jointly amplified**”, we show that under a model where amplification is pervasive and shared between environmental and genetic effects, we expect the male-female ratio of environmental variance to equal the male-female ratio of genetic variance. Here, we compare the expectation under the pervasive, joint amplification model with expectations under two other longstanding models.

As recently discussed by Zajitschek et al.²¹, the “estrus-mediated variability” hypothesis predicts that females will display higher trait variability in traits affected by the estrous cycle. If we interpret estrus-mediated effects as environmental, this hypothesis suggests equal genetic variance but larger environmental variance in females (**Fig. S18A**; orange line in **Fig. S18B**).

The “greater male variability” hypothesis predicts that males will display higher trait variability. There are multiple rationales offered for this hypothesis, such as stronger sexual selection on males that leads to higher variability via group selection²²⁻²⁴. A more widely applicable rationale for this prediction is that mammalian males, as the heterogametic sex, will experience more variable X chromosome effects, whereas these effects will be “averaged out” through heterogeneous X inactivation in females²³. This hypothesis therefore predicts greater genetic

variance in males and equal environmental variance between males and females (**Fig. S18A**;
green line in **Fig. S18B**).

Supplementary Figures

Figure S1. Sex-specific Miami plots, related to Figure 3

“Miami plots”, contrasting statistical significance of marginal effects in a female-specific GWAS (top) and a male-specific GWAS (bottom), for each of the 27 traits analyzed. Data points are thinned out by random selection over multiple levels of p-value thresholds. For larger p-values, more points are not shown.

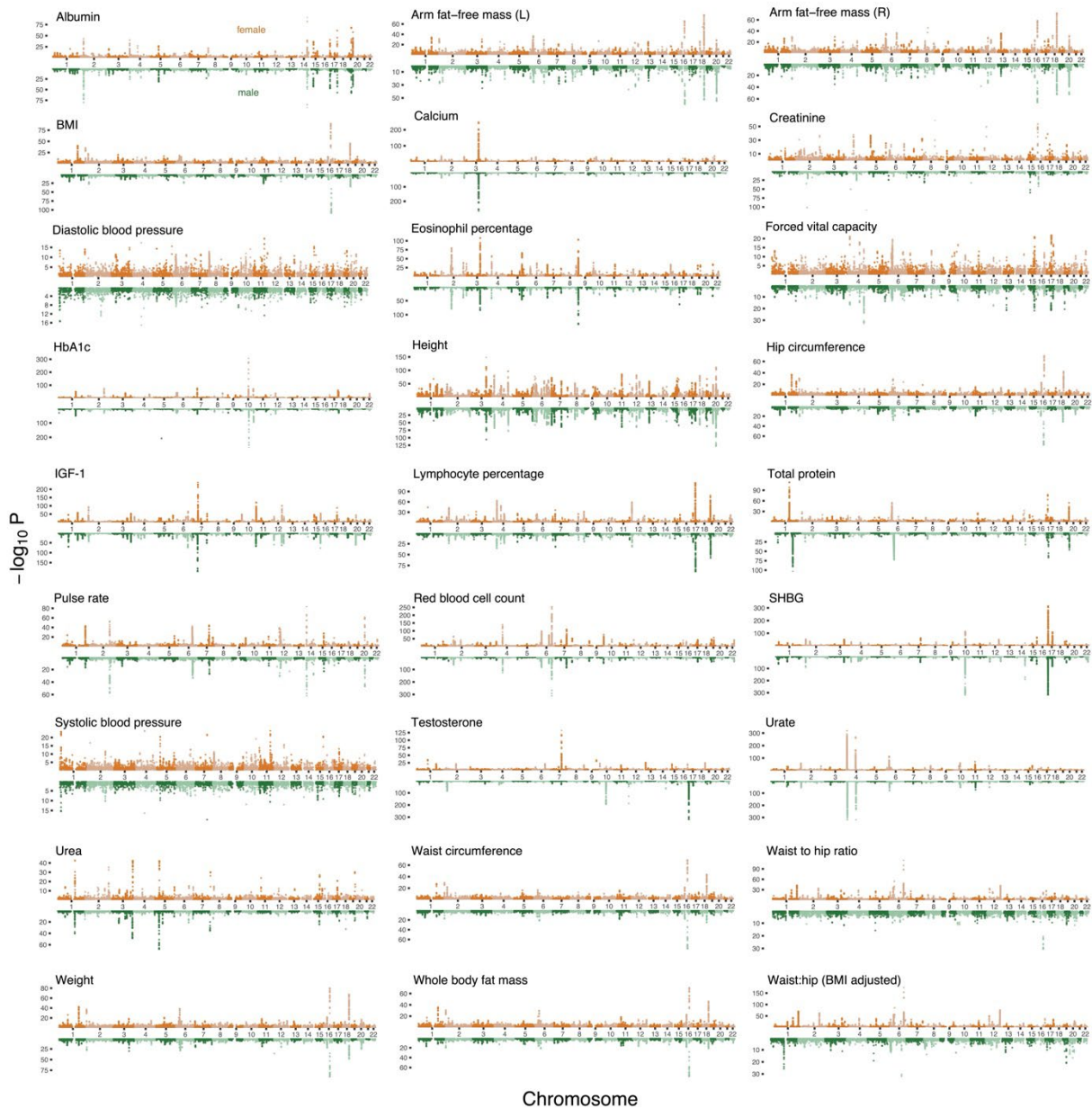


Figure S2. All hypothesis covariance matrices, related to Figure 3

All hypothesis matrices we inputted into *mash* are shown here organized by correlation and magnitude of effect size. The placement of the matrices parallels in the weights in **Data S1-27**.

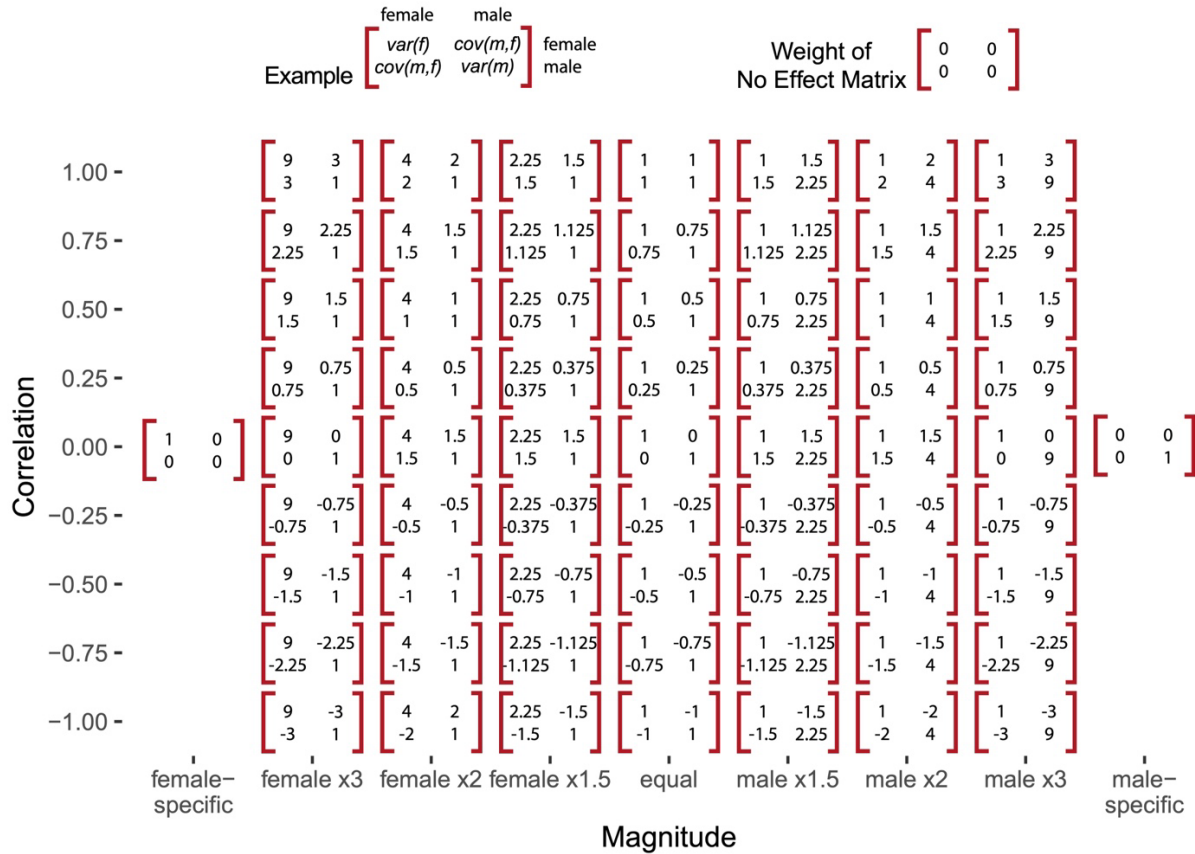
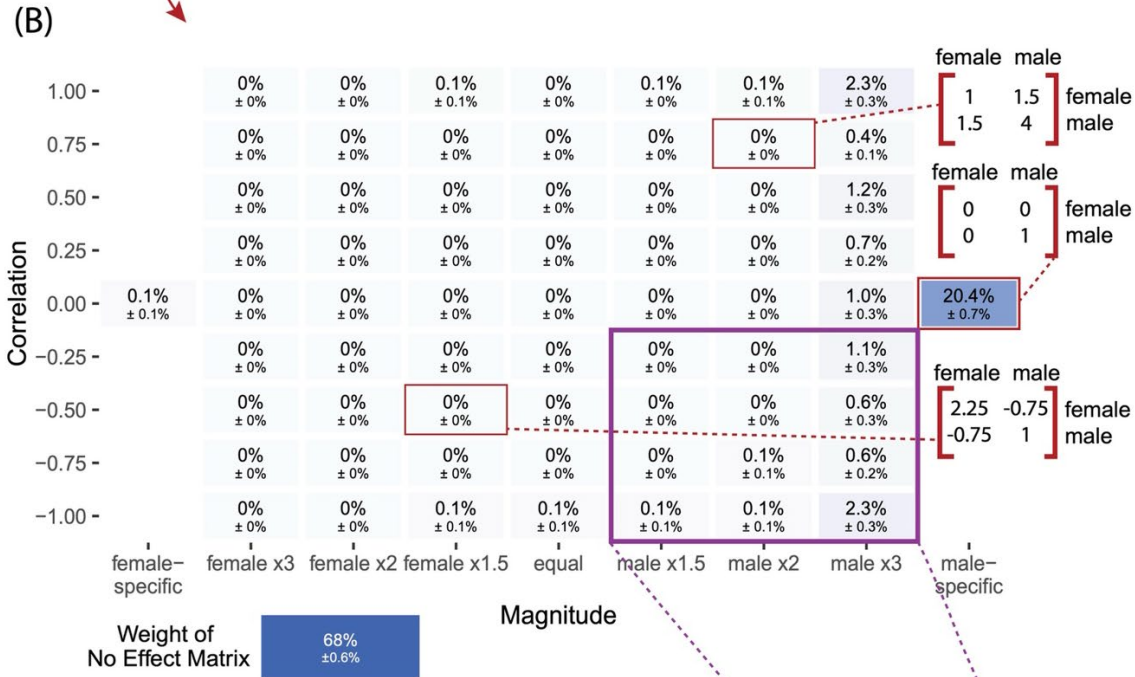
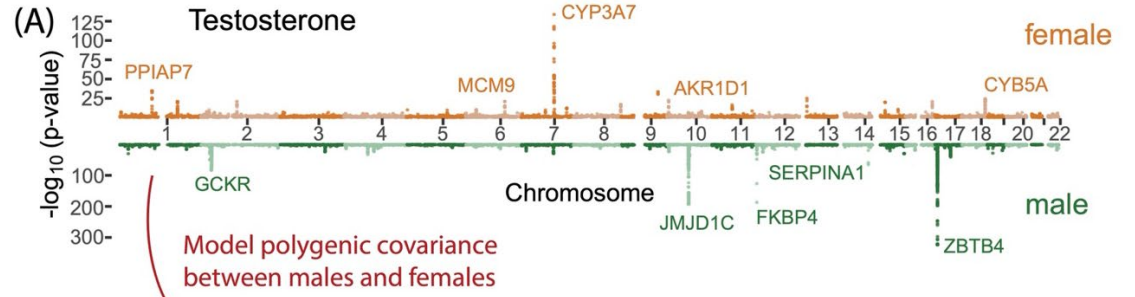


Figure S3 Inferring polygenic covariance structure, related to Figure 3

(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 that are less than 5 kbp away from a transcription start site were annotated with the gene corresponding to the closest one. **(B)** We modelled the sex-stratified GWAS estimates as sampled with error from true effects arising from a mixture of pre-specified hypothesis covariance relationships between female and male genetic effects; see examples in red frames. Each box specifies a mixture weight (\pm SE) that we infer for one hypothesis matrix. The weight, also indicated by the shade, corresponds to the relative frequency that the specified hypothesis matrix is represented by the variants. The axes state the relative magnitude (amplification) and correlation between males and females, which jointly make up the covariance relationship. **(C)** The x and y axes are a condensed version of the x and y axes from (B) for testosterone. The weights are a proportion of the non-null weights, i.e., the weight divided by sum of all weights except for the weight on the no effect matrix, corresponding to no effects 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 effects in females, 5.1%, divided by the total weight on matrices with nonzero effects, 32%.

Inferring Polygenic Covariance Structure between Males and Females



(C) Covariance of Genetic Effects: Compact Representation

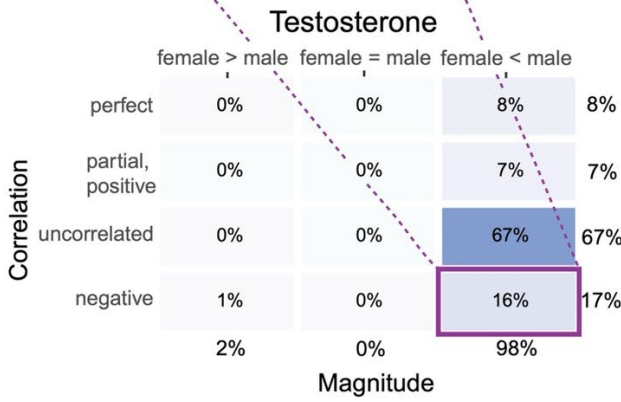


Figure S4 Testing p-value thresholds for input into *mash*, related to Figure 3

We tested the effect of inputting random subsets taken from various p-value thresholds [1, 5e-2, 1e-5, 5e-8] in the *mash* fitting step to estimate mixture proportions. The random subsets were drawn by sampling from the 1703 LD blocks (**Methods “Mixture weights for covariance structure between male and female effects”**). **(A)** Both depict the weight on the no-effect matrix per p-value threshold for four different traits. The left plot shows results from sampling once from each available LD block as long as there were still p-values below the specified threshold. For the right plot, we keep sampling from LD blocks until we reach a subset of 1703 SNPs, ensuring equal sample sizes across the thresholds. **(B)** The percentage of weight by correlation of effects are depicted across the thresholds. Plots in the box show results from equal sample sizes. **(C)** The percentage of weight on different types of magnitude of effects are shown for four different traits across the threshold. Plots in boxes also show results from equal sample sizes.

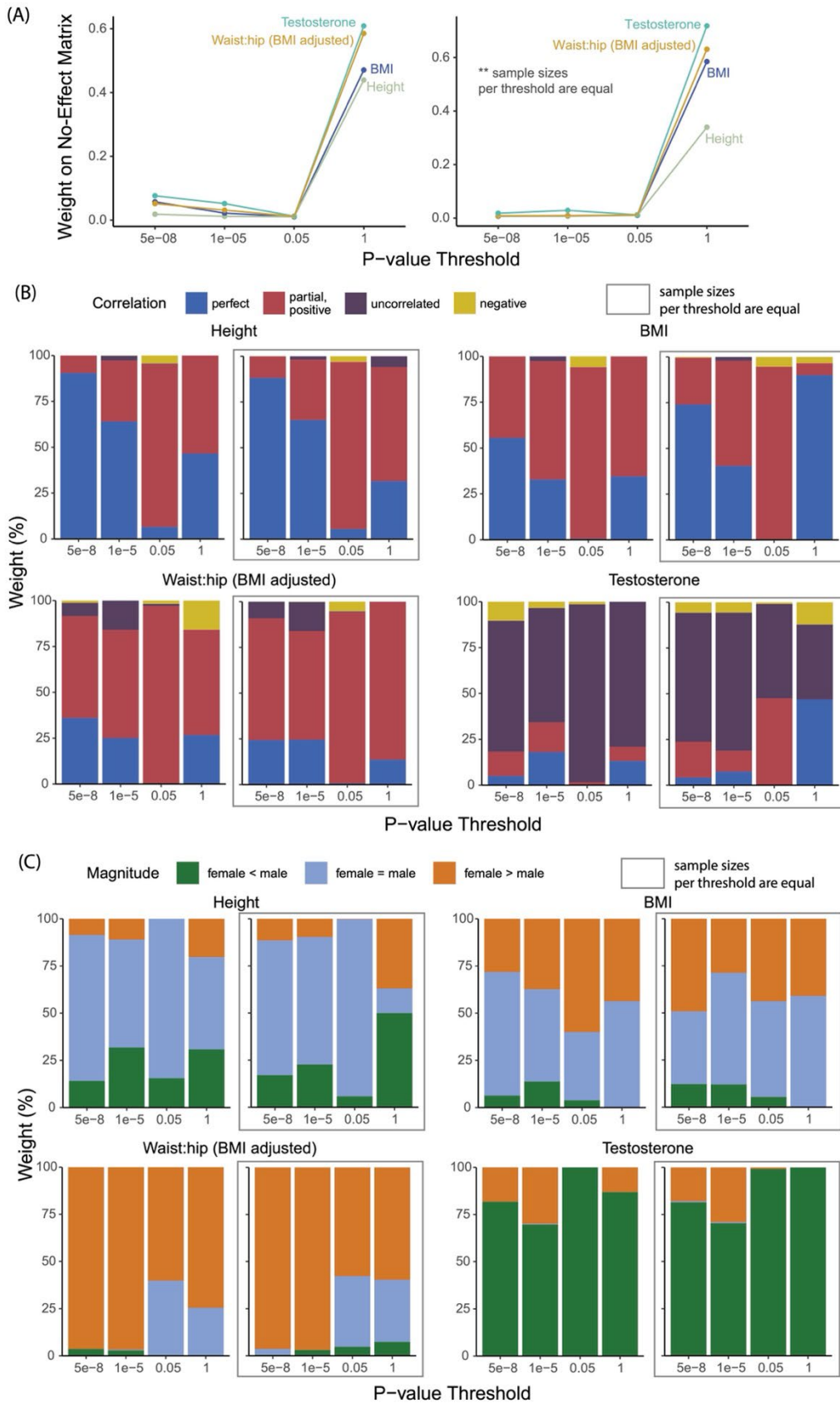
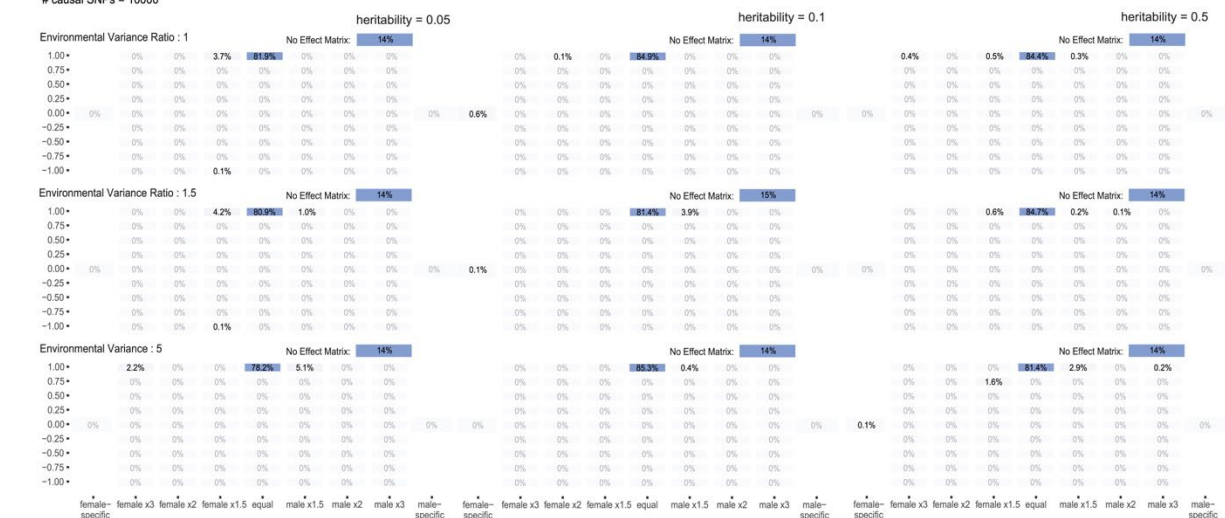
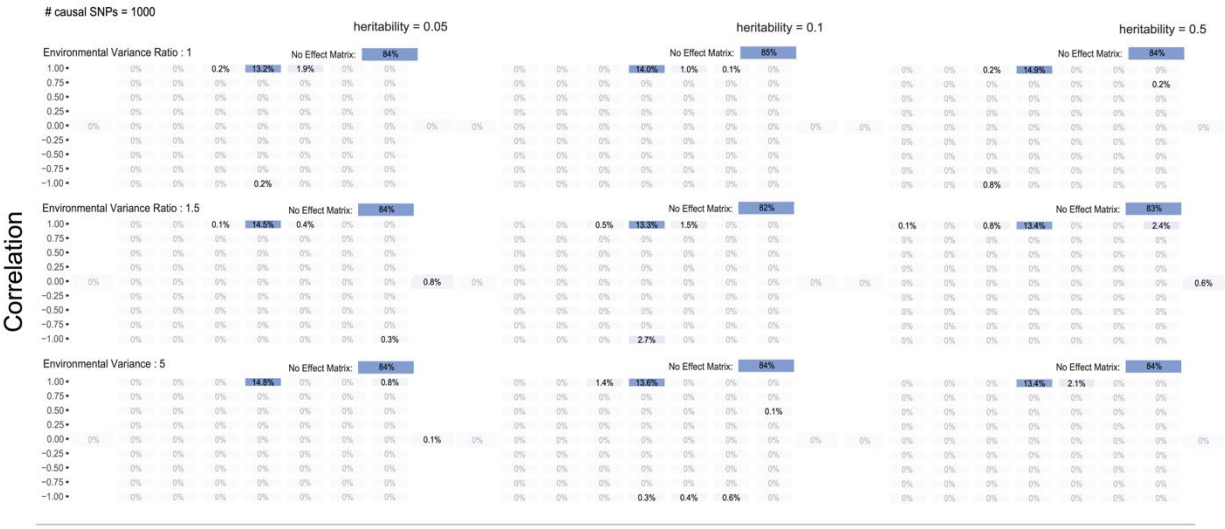


Figure S5: Simulation of sex-invariant genetic effects but with variable environmental variance, related to Figure 3

Mixture weights based on various parameters for number of causal SNPs, male-specific heritability, and female to male environmental variance ratio are shown for all hypothesis covariance matrices represented by the axes. Results are generated from a simulation study based on **Eq. 2**, in which environmental variance differs between the sexes while genetic effects stay the same to test whether *mash* interprets the environmental variance difference as GxSex (**Methods “Environmental variance simulation for *mash*”**). We tested three parameters for female to male environmental variance ratio [1, 1.5, 5]. Parameters for the number of causal SNPs [100, 1K, 10K] and male heritability [0.01, 0.1, 0.5] are listed on the top right.



Magnitude

Figure S6: Simulation of covariance structure, related to Figure 3

We compared the distribution of mixture weights estimated in *mash* when inputting effect estimates and standard errors are sampled from a pre-specified variance-covariance matrix. The matrices and associated proportions sampled from a particular matrix is shown on the left. We set the female to male environmental variance ratio as 1.2:1 and the heritability to 0.5. **(A)** We simulate a null model in which all effect estimates are drawn from a matrix representing equal effects between males and females. **(B)** 86% of the true effects have equal effects in males and females and 14% are sampled from a relationship in which effects in females are perfectly correlated but have twice the variance of that in males. **(C)** 86% of the true effects have equal effects in males and females and 14% are sampled from a matrix corresponding to female-private effects.

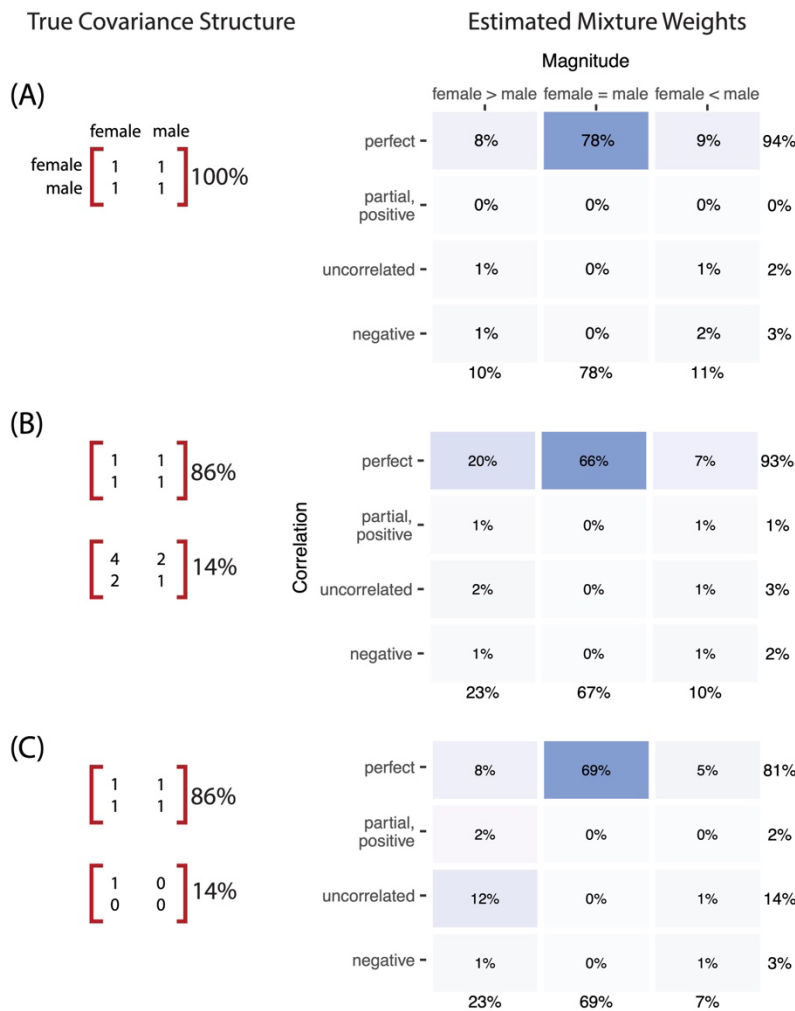


Figure S7: Proportion of non-trivial weights on hypothesis matrices, related to Figure 3 and Figure 4

We examine the proportion of weight on non-trivial hypothesis matrices across 27 complex traits. Here, we define non-trivial as matrices with correlation < 1 or female \neq male magnitude. The heritability ratio represented by point size is estimated by taking the larger of the two sex-specific heritabilities and dividing it by the smaller. We refer to the 1:1 line to see if more traits are represented more by unequal magnitude or imperfect correlation.

Weight on Nontrivial Correlation vs Nontrivial Magnitude Matrices

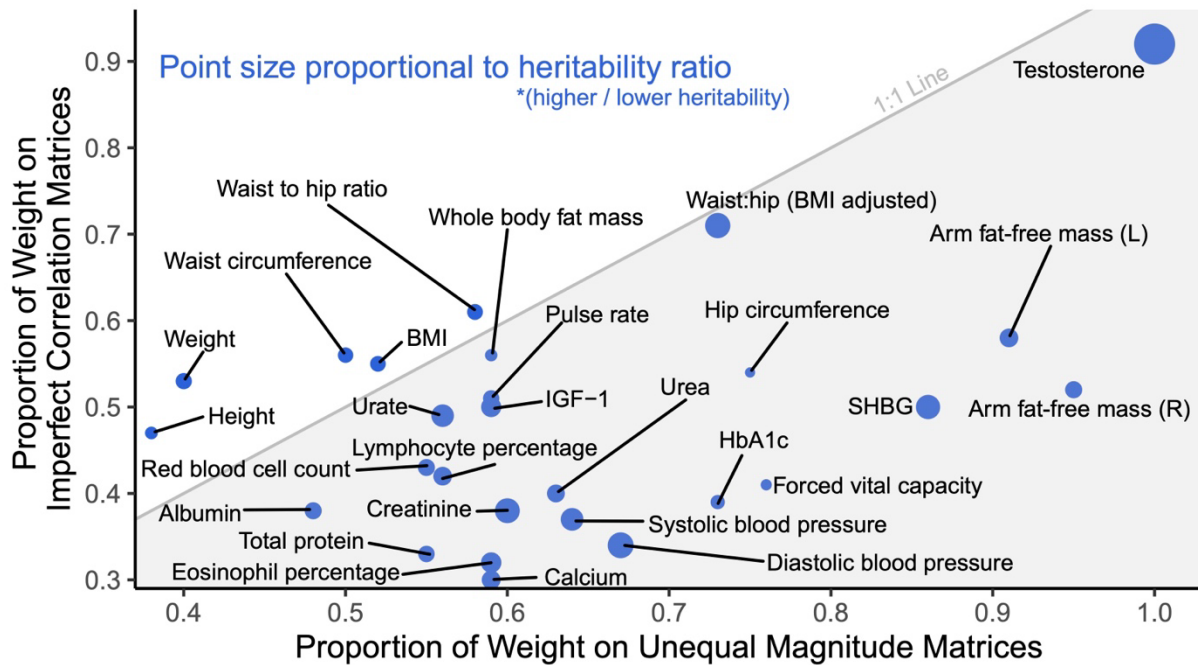


Figure S8: Across traits, phenotypic variance sex ratio correlates with the phenotypic mean sex ratio, related to Figure 4

The solid gray line shows a linear fit, excluding two outliers: testosterone and waist:hip ratio adjusted for BMI.

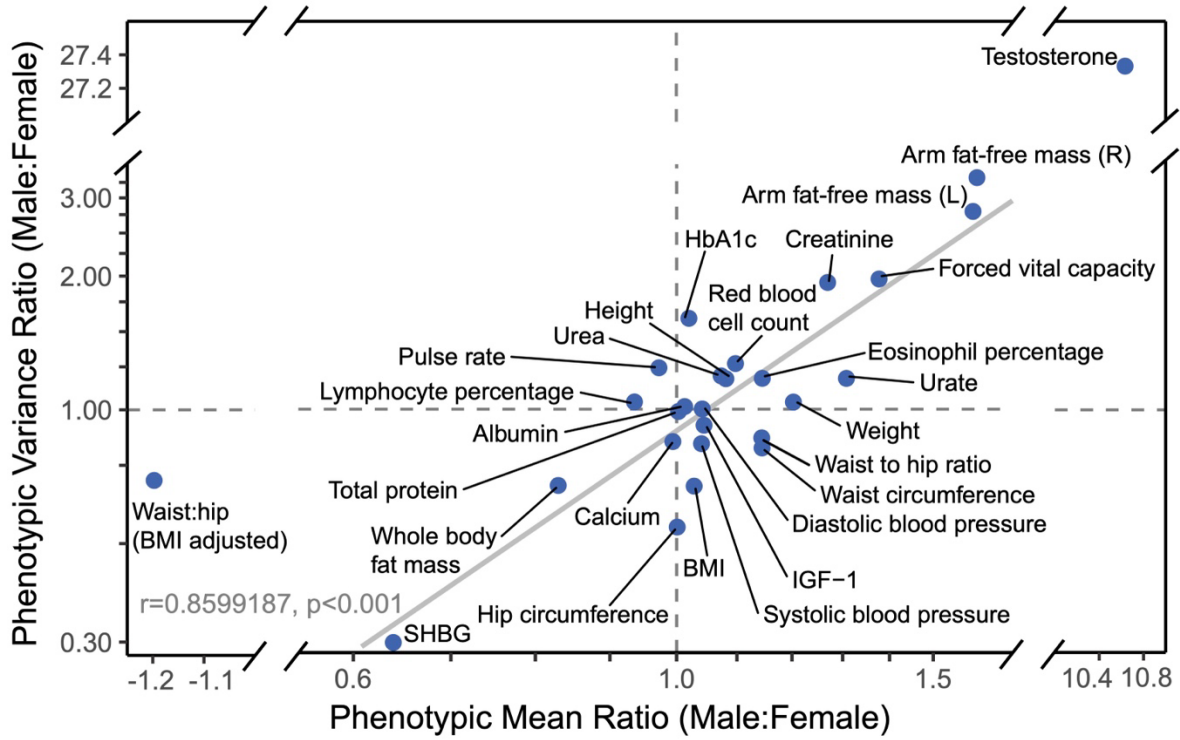


Figure S9: Process of generating posterior estimates from *mash*, related to Figure 3

Posterior mean and local false sign rate (lfsr) estimates are generated using *mash* from sex-specific summary statistics. We fit the model on the average of 100 mixture weights. The process is described in the **Text S7** section.

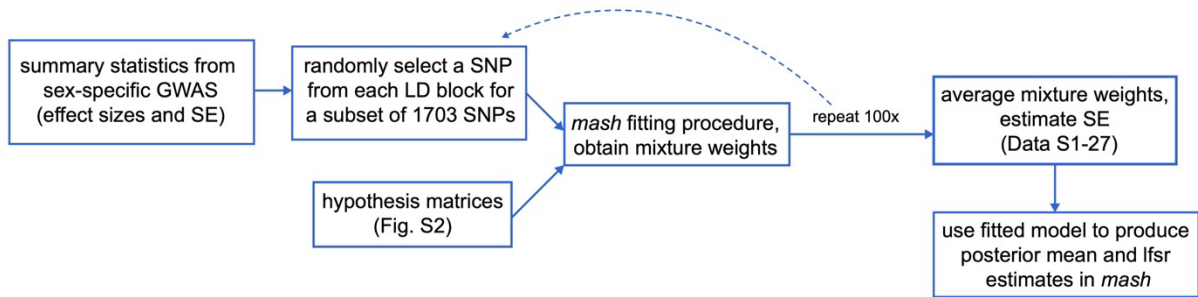


Figure S10: Process of comparing models for complex phenotype prediction, related to Figure 3

The process is described in **Text S8**. **(A)** The sample test set is obtained by sampling 25k males and 25k females. We then create the training set, by removing the test set from the original list of sample IDs. **(B)** After obtaining summary statistics from GWAS, we perform clumping to adjust for LD. Covariance-aware effect estimates are acquired by inputting the sex-specific summary statistics through *mash* to generate posterior mean and lfsr estimates **Fig. S9**. lfsr estimates are converted into pseudo p-values by matching to original p-values after ordering both from smallest to largest. We also conduct PGS analysis over four p-value thresholds on each of the five models, thus obtaining 20 PGS. **(C)** Prediction accuracy is measured based on R^2 , from the linear regression of phenotype value on covariates and the polygenic scores. The incremental R^2 is also estimated by subtracting R^2 from the null model R^2 , based on the regression of the phenotype value on just the covariates.

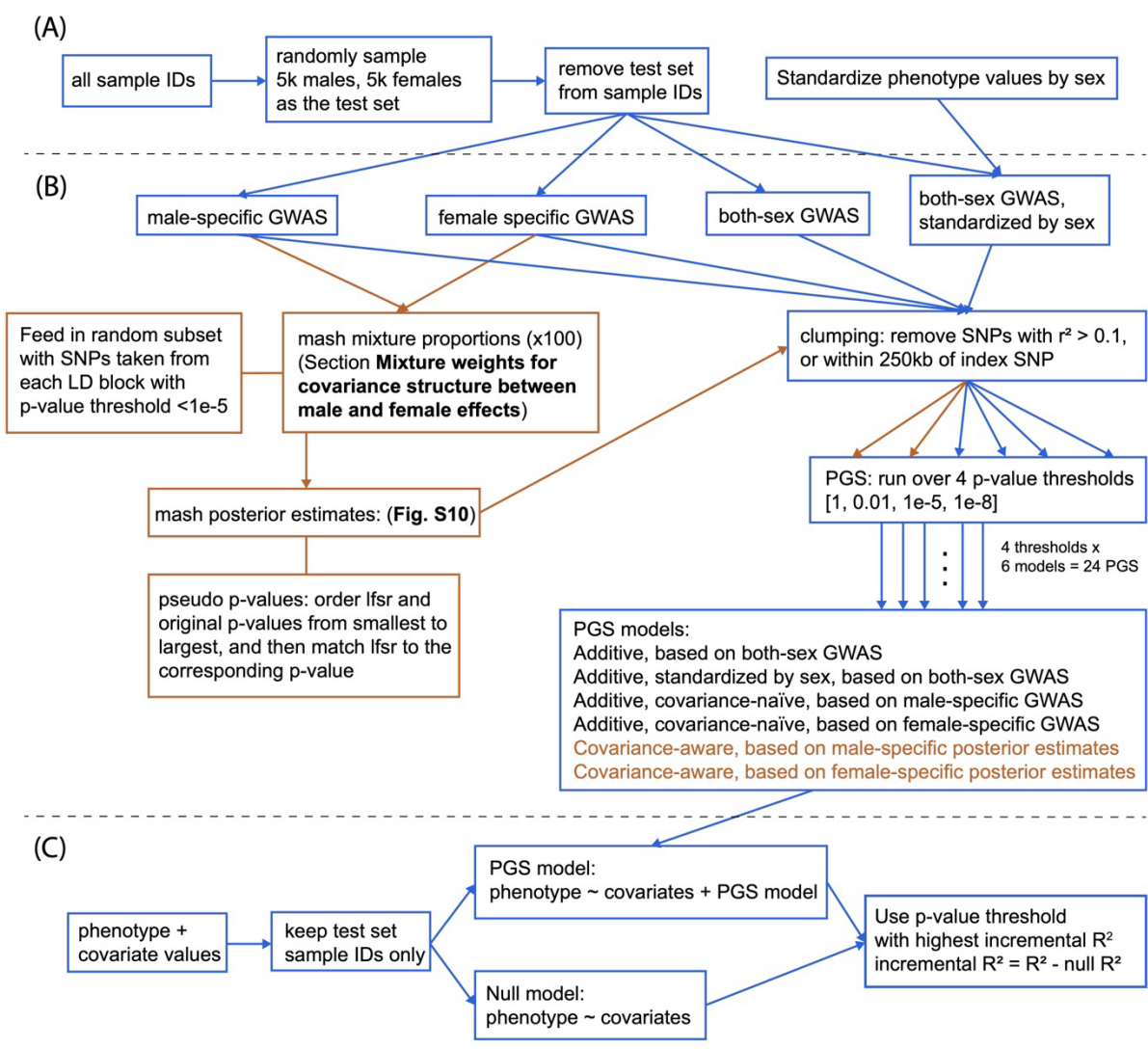


Figure S11: Evaluating evidence for systematic amplification, related to Figure 2

We regressed the male (green) and female (orange) trait values to polygenic scores estimated in an independent sample of males (top) and females (bottom) separately. Each line corresponds to a separate regression of trait values in one sex to polygenic scores estimated in one sex. Each point represents the mean value in one decile of the polygenic score. The fitted line and associated R^2 are estimated from the raw, non-binned data.

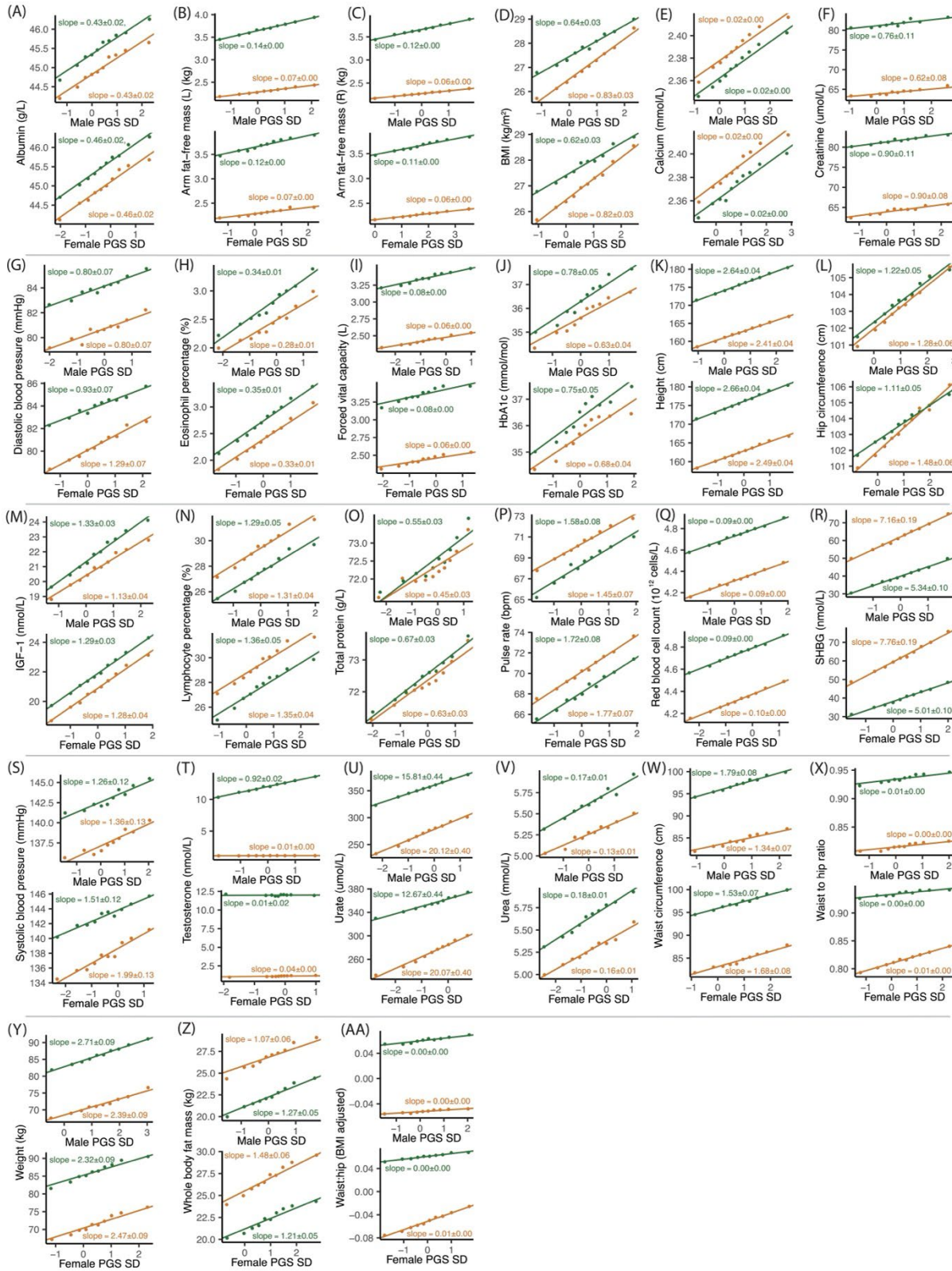


Figure S12: The predictive utility of G-by-Sex aware polygenic scores, related to Figure 3

We compared the incremental squared correlation (incremental R^2) of the phenotypic values and a predictor for four predictors. Incremental R^2 was obtained by taking the difference in R^2 between a prediction with covariates and the polygenic score and a prediction with only covariates. For each of the three models, we chose the highest incremental R^2 across four p-value thresholds (**Text S8**). The additive (purple), additive standardized by sex (yellow), sex-specific covariance-naïve (blue), and sex-specific covariance aware (red) models were calculated from effect estimates generated in GWAS, with the latter two in sex-stratified samples. The sex-specific, covariance aware model (red) 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 size of around 123K males and 147K females, whereas the additive model retained the full training set of around 270K individuals. Incremental R^2 values and ± 3 SE-wide error bars were computed using 20-fold cross-validation.

Performance of GxSex Aware Polygenic Scores

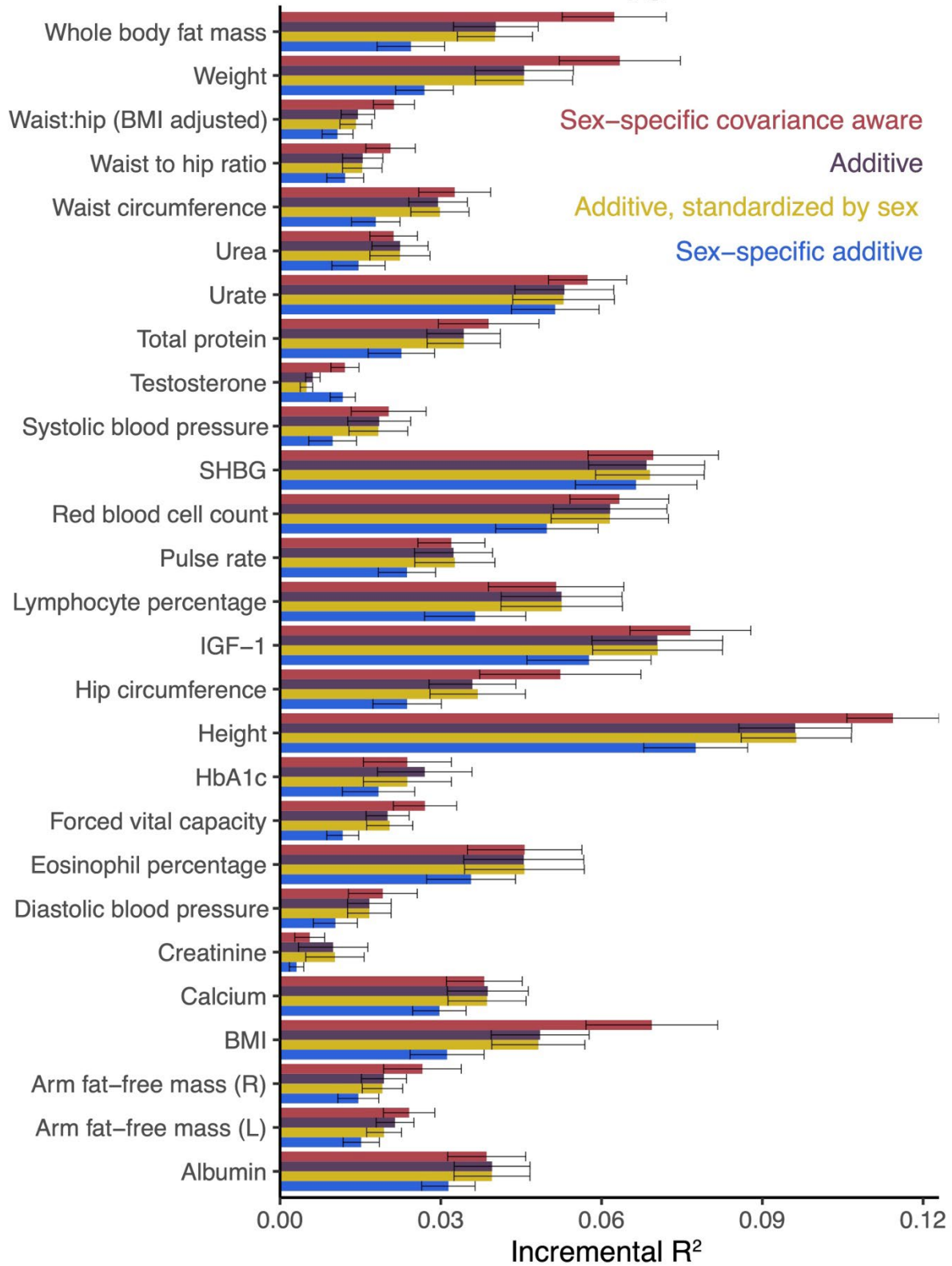


Figure S13: Genetic effect and testosterone levels – using additive model PGS, related to Figure 5

The relationship between genetic effect and testosterone level bins, separated by sex, is depicted for all traits. Genetic effect, with ± 1 SE error bars, is estimated from the slope of the regression of phenotypic values and polygenic scores from the additive model, multiplied by the polygenic score standard deviation (**Fig. S12**). The hollow data point represents a testosterone bin with overlapping range between males and females.

Effect of 1 PGS SD on Phenotype

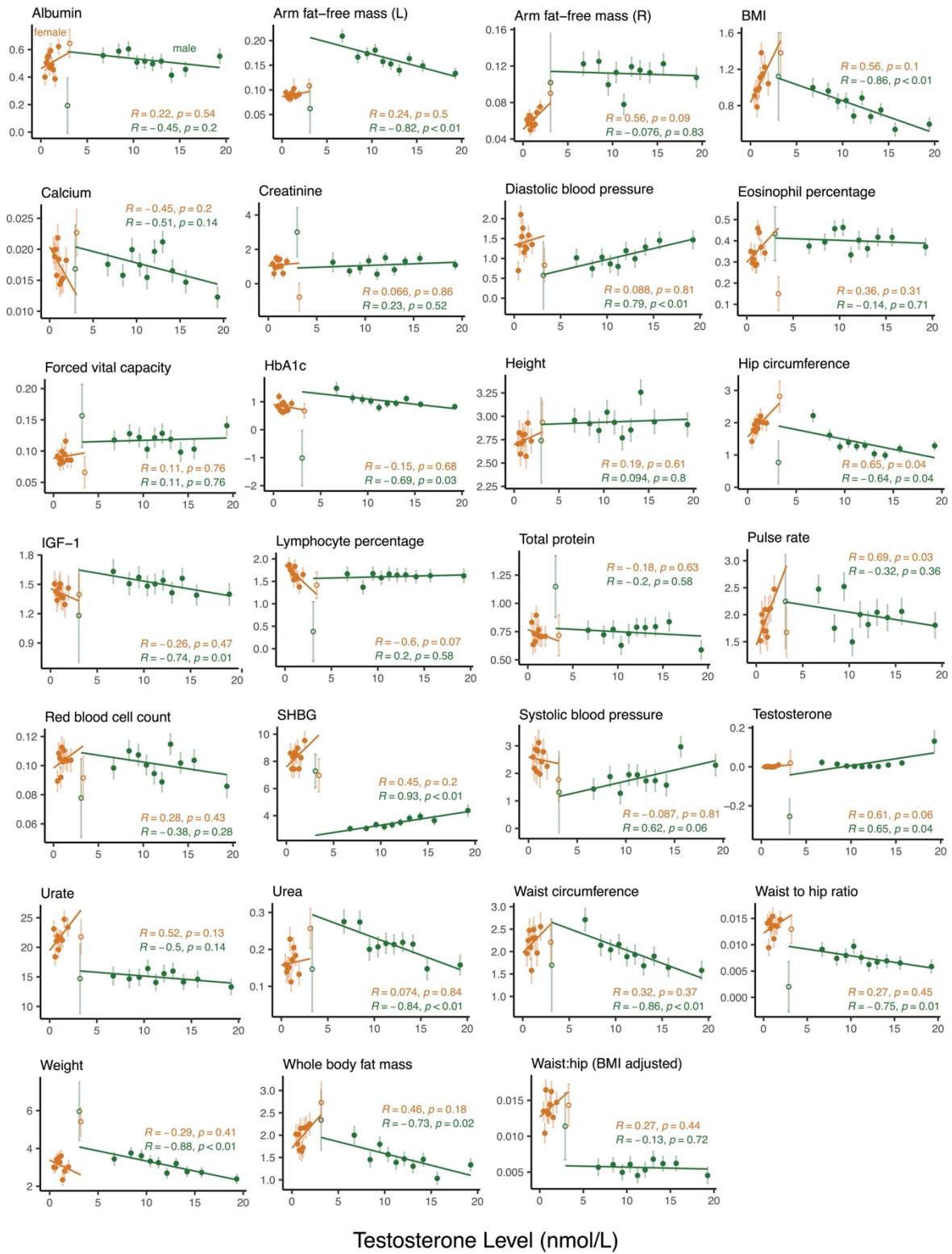


Figure S14: Genetic effect and testosterone levels – using sex-specific model PGS, related to Figure 5

The same type of figures as **Fig. S13** are displayed with the exception of utilizing PGS from the sex-specific, covariance-naïve model for estimation of genetic effect.

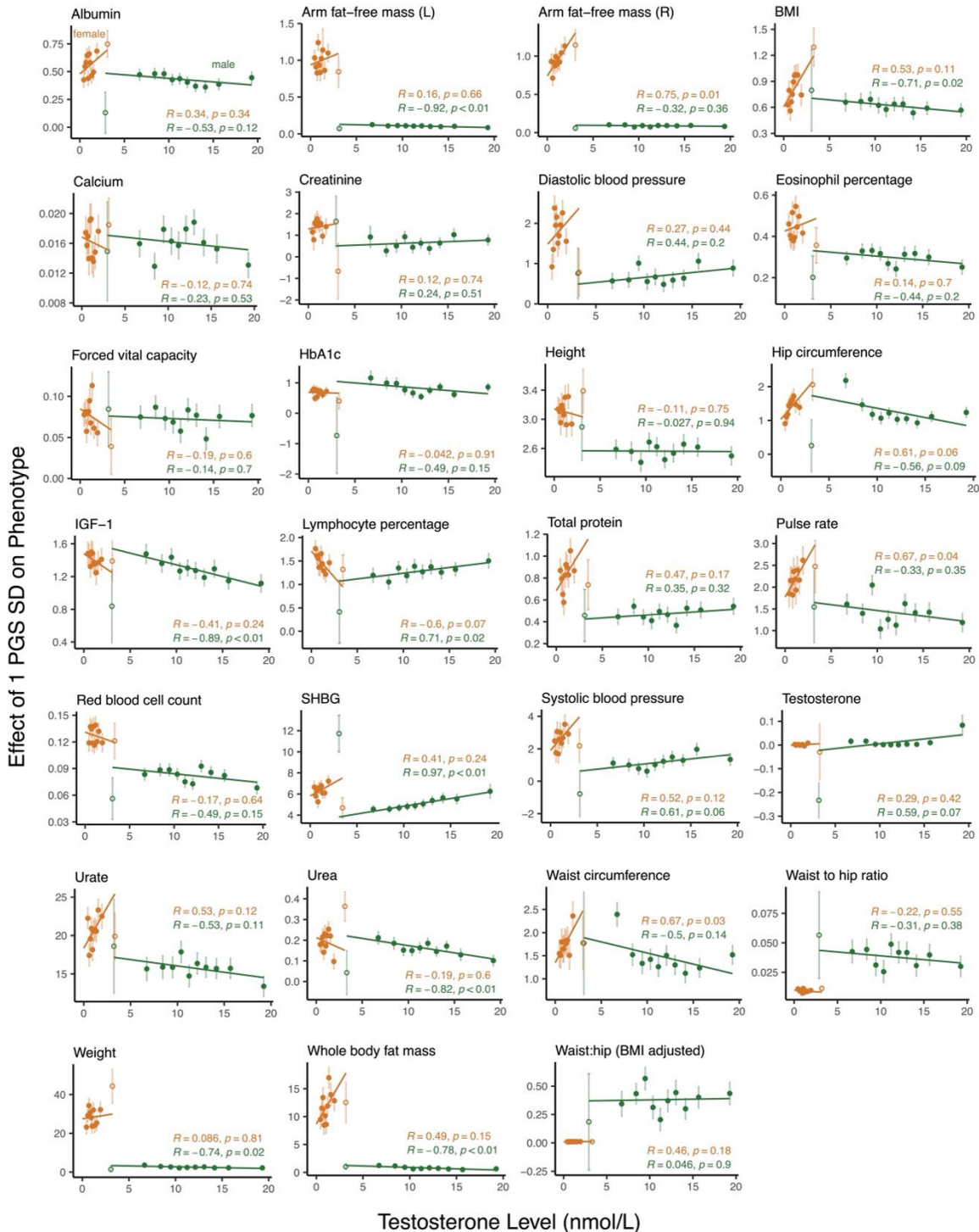


Figure S15: Genetic effect and PGS for testosterone, related to Figure 5

We used mendelian randomization to assess the causal relationship between genetic effect and testosterone levels, similar to **Fig. S13**. **(A)** The plots illustrate the relationship between amplification of total genetic effect of each trait and PGS based on sex-specific summary statistics for testosterone. Genetic effect, with ± 1 SE error bars, is estimated from the slope of the regression of phenotype values on the PGS based on additive summary statistics in each PGS bin for testosterone, multiplied by the PGS standard deviation. The x-axis takes the mean PGS for testosterone from each bin. **(B)** The correlation for each sex (90% CI) and all traits between genetic effect and PGS for testosterone across the bins is shown, with traits in descending order by the difference of correlation estimates between the sexes.

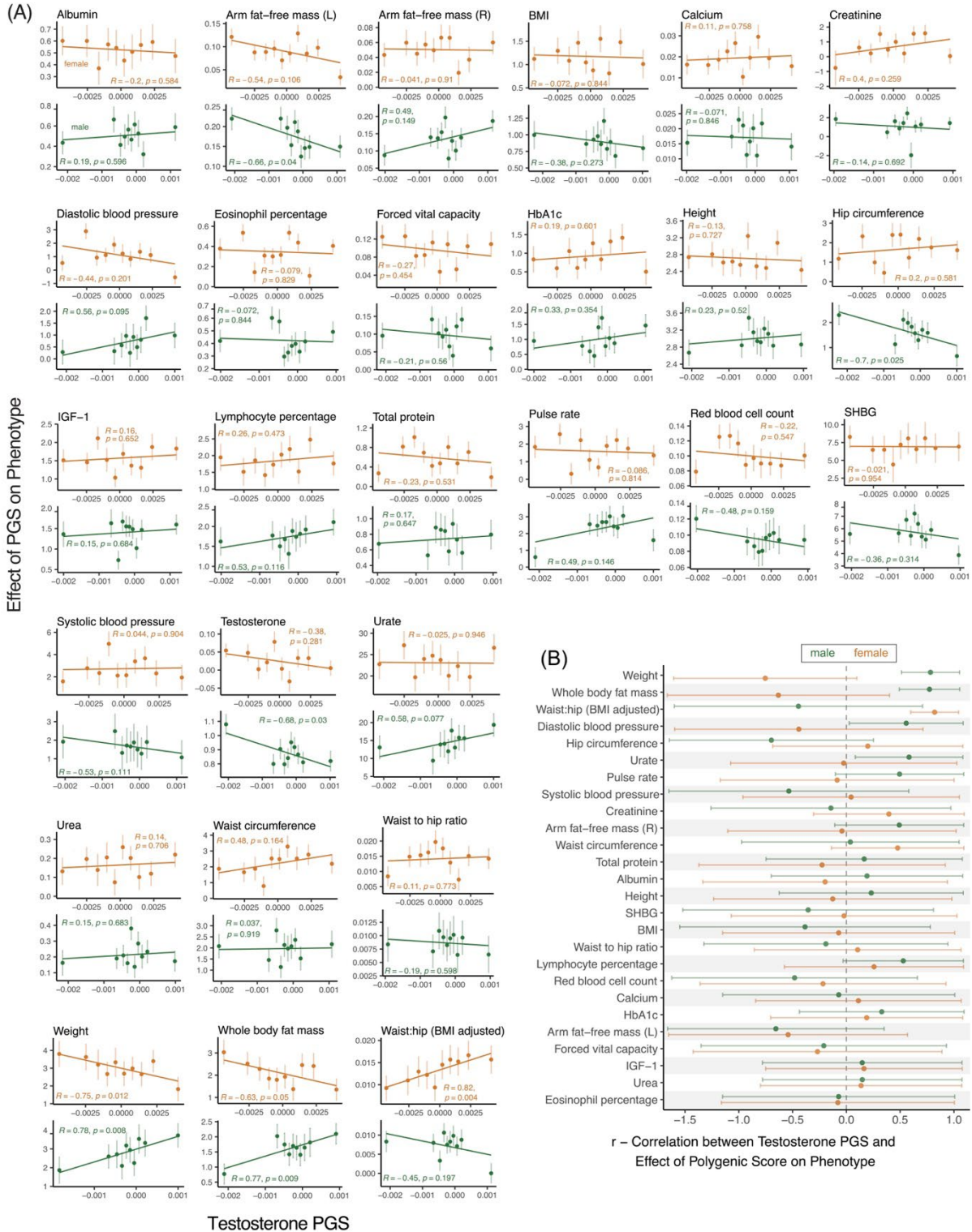


Figure S16: Genetic effect residualized for age and testosterone levels, related to Figure 5

In each bin, we examined the relationship between the polygenic effect residualized for mean age and the mean testosterone level. This plot shows the correlation (90% CI) between the two across ten bins.

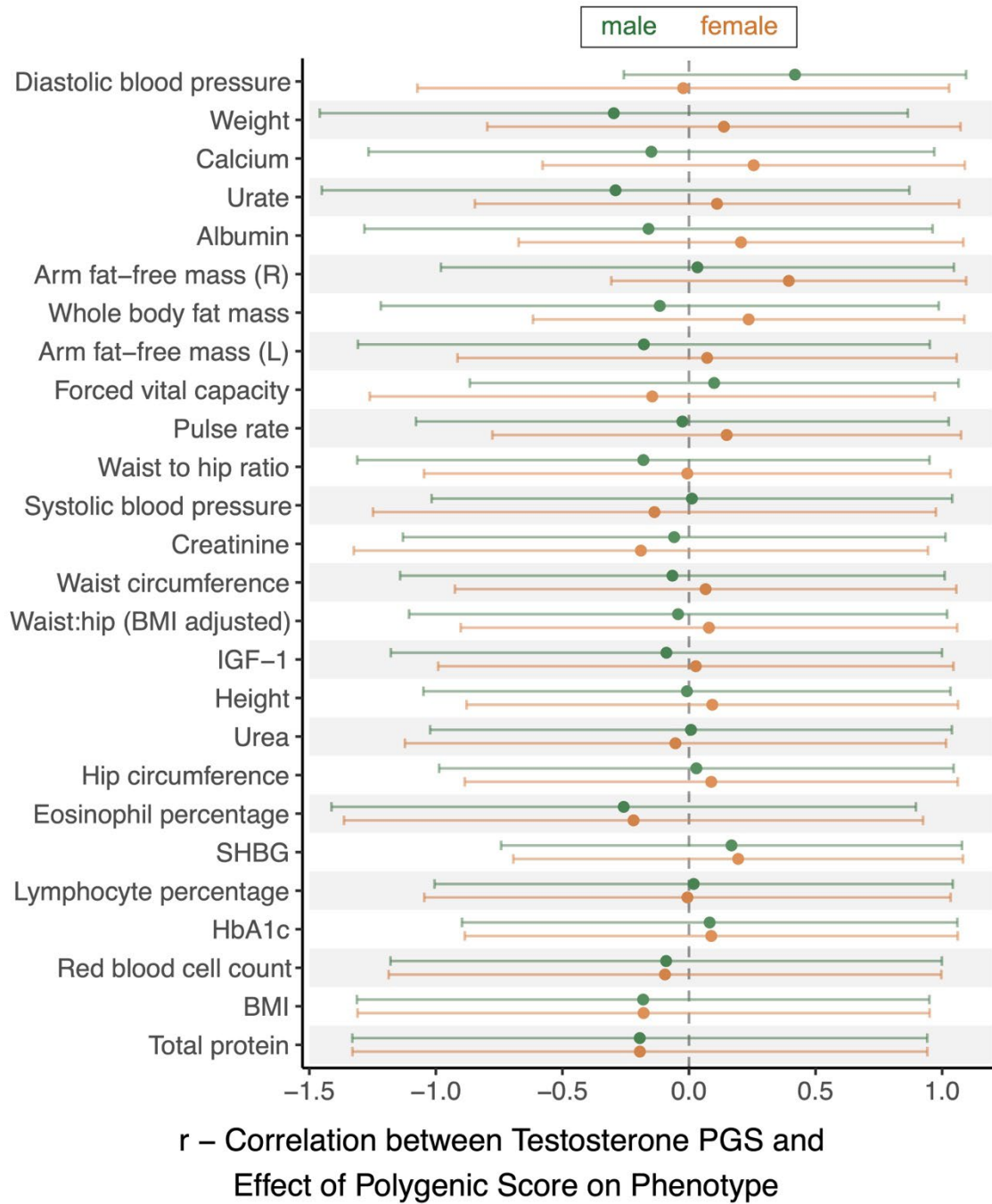


Figure S17: Example of shared polygenic and environmental amplification in muscle, related to Figure 6

An example of equal amplification of genetic and environmental effect (“**Are polygenic and environmental effects jointly amplified?**” in main text) is depicted in muscle after resistance training. The example follows the model in **Fig. 6A**, in which the genetic effect (genetic regulation) and environmental effect (resistance exercise), both affect a core pathway (involving IGF-1) for muscle. The effect size of the core pathway is then amplified by a modulator such as testosterone, producing the sexual-dimorphism shown in muscle.

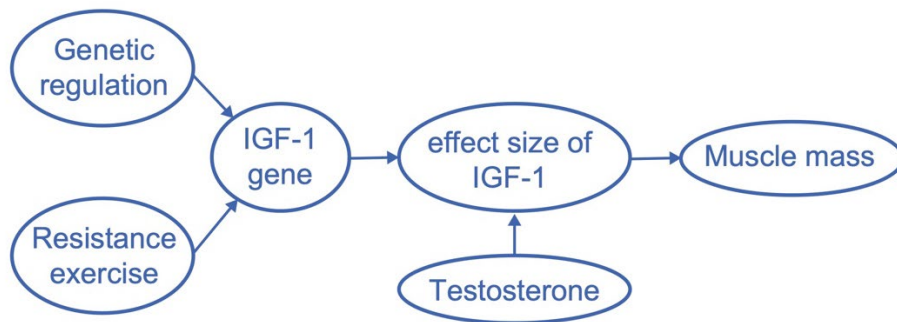
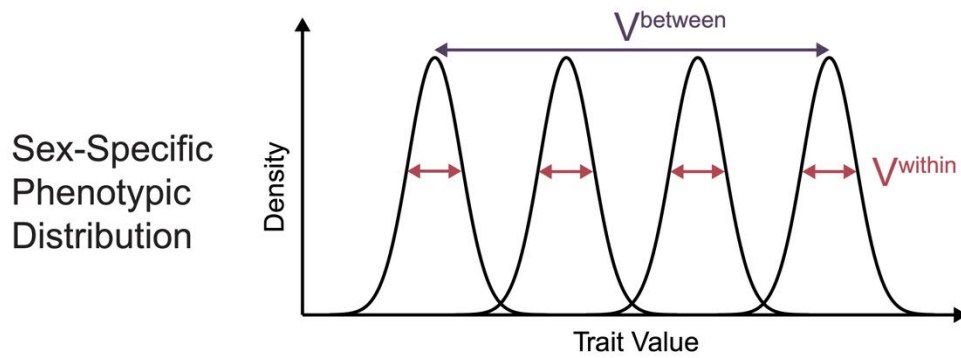


Figure S18: Competing models for sex differences in trait variance, related to Figure 6

We compare two different models for sex differences in trait variance, greater-male and estrus-mediated variability as discussed in Zajitschek et al.²¹ and **Text S11**, with our model of pervasive amplification linking genetic and environmental effects (“**Are polygenic and environmental effects jointly amplified?**” in main text). **(A)** We illustrate the difference between the variability between (V_{between}) and the variability within (V_{within}) using individual distributions. We define the three models according to the relationship between male and female V_{between} and V_{within} . **(B)** The three models are superimposed as different colored lines on **Fig. 6B** to compare the distribution of traits across the models. Traits in blue, green, and yellow are consistent (within 90% CI) with the model of pervasive amplification, greater-male variability, and estrus-mediated variability alone, respectively. Traits in grey are consistent with more than one model. Traits in black are inconsistent with either model.

(A)



	within genotype	between genotype
Greater Male Variability	$V_m^{within} = V_f^{within}$	$V_m^{between} > V_f^{between}$
Estrus-mediated Variability	$V_m^{within} < V_f^{within}$	$V_m^{between} = V_f^{between}$
Pervasive Joint Amplification	$V_m^{within} < V_f^{within}$ if and only if	$V_m^{between} < V_f^{between}$

(B)

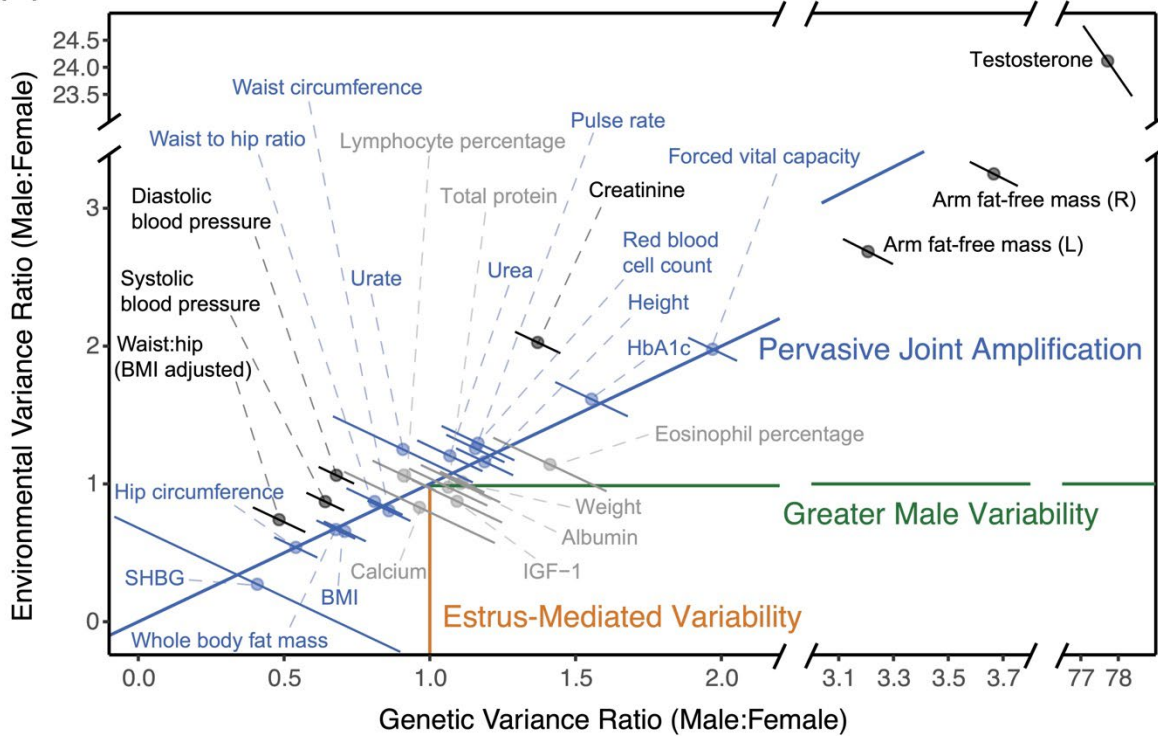


Figure S19: Z-scores for strength of sexually-antagonistic selection, related to Figure 7

Shown are Z-scores for the strength of sexually-antagonistic selection, A in Eq 1, computed for various gnomAD subsamples (categorized by ancestry group). Different panels show results obtained with different p-value thresholds on marginal significance of sex-specific association in the GWAS (in either sex), including 10^{-8} (A), 10^{-5} (B), and 10^{-3} (C). Bars show 90% resampling confidence intervals.

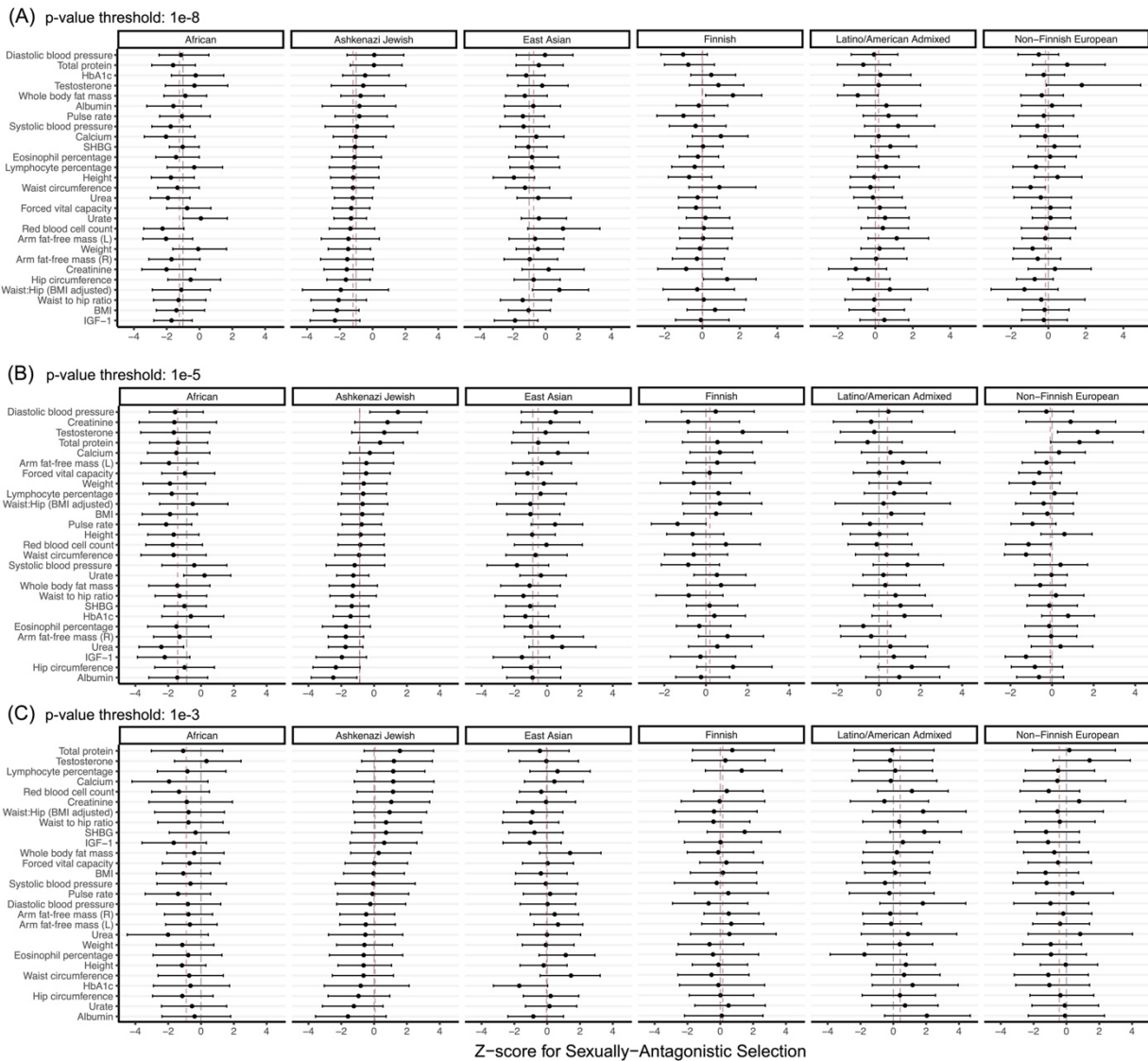
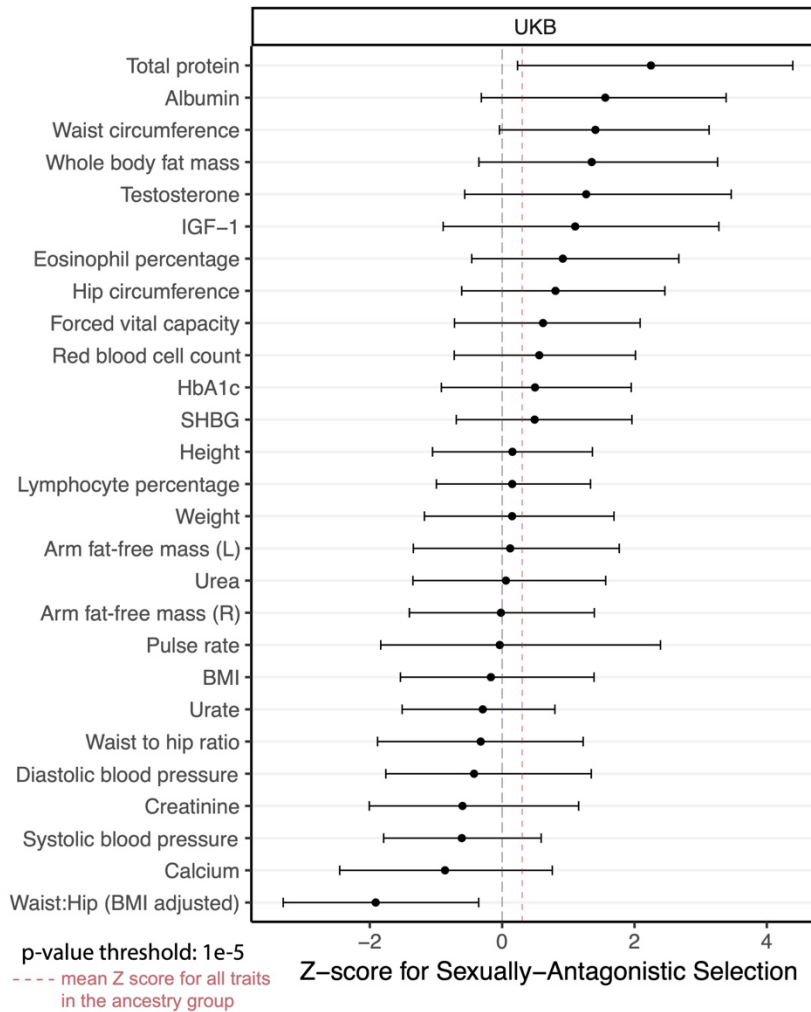


Figure S20: Z-scores using UK Biobank data, related to Figure 7

This figure parallels **Fig. 7** and **Fig. S19** but uses male-female allele frequency differentiation data from the UK BioBank. Here, the statistic L_{ST} replaces F_{ST} , as explained in the section “**Estimating the potential for sexually-antagonistic selection on standing variation (A)**” of the main text. The p-value cutoff for marginal sex-specific GWAS associations included in this analysis was set to 10^{-5} . Bars show 90% resampling confidence intervals.



Supplementary Tables

Table S1: Filtering site with small sample sizes, related to Figure 7

For the sexually-antagonistic selection analysis, we removed all sites with less than 1,000 alleles in the sample in order to narrow our research on sites with sufficiently large sample sizes. There were 2,285,169 sites before filtering.

Ancestry	# Sites After Filtering
African/African American	2,282,394
Amish	0
Latino/American Admixed	2,282,394
Ashkenazi Jewish	2,247,203
East Asian	2,255,778
Finnish	2,249,604
Middle Eastern	0
Non-Finnish European	2,283,952
South Asian	2,167,009
Other	2,080,336

Table S2. Filtering monoallelic sites, related to Figure 7

For the sexually-antagonistic selection analysis, and for each ancestry, we removed sites which were monoallelic in one or both sexes.

Ancestry	# Sites after filtering
African/African American	924,946
Latino/American Admixed	519,496
Ashkenazi Jewish	161,271
East Asian	300,323
Finnish	207,162
Non-Finnish European	1,104,008
South Asian	330,518
Other	267,812

Table S3: Filtering sites that were not overlapping between the gnomAD dataset and the GWAS dataset, related to Figure 7

For the sexually-antagonistic selection analysis, any sites which did not occur in both the filtered GWAS data from UKB and the filtered gnomAD data were excluded.

Ancestry	# Sites before filtering by GWAS-gnomAD overlap	# Sites after filtering	# Sites removed
African/African American	924,946	886,467	38,479
Latino/American Admixed	519,496	506,760	12,736
Ashkenazi Jewish	161,271	160,358	913
East Asian	300,323	295,801	4,522
Finnish	207,162	204,914	2,248
Non-Finnish European	1,104,008	1,051,015	52,993
South Asian	330,518	325,444	5,074
Other	267,812	264,682	3,130

Table S4: Filtering by p-value threshold, related to Figure 7

For the sexually-antagonistic selection analysis, we used three different thresholds of GWAS p-values to further restrict which sites were used for analysis by focusing on sites with strong correlation with the trait of interest. Shown are the results of filtering for height, as an example. Results for other traits are highly similar.

Ancestry	p-value threshold	# Sites before this filtering step	# Sites after filtering
Ashkenazi Jewish	1e-3	160,358	2,749
Ashkenazi Jewish	1e-5	160,358	578
Ashkenazi Jewish	1e-8	160,358	298
Finnish	1e-3	204,914	3,975
Finnish	1e-5	204,914	852
Finnish	1e-8	204,914	373

Table S5: Identification of SNPs susceptible to mis-mapping to sex chromosomes—Kasimatis et al., related to Figure 7

We compare our results for excluding sites due to mis-mapping to the results from Kasimatis et al.¹⁸, who removed from their analysis sites in regions with 90% identity or higher with a region in a sex chromosome (as we have, based on BLAT²⁵ matches), as well as sites overlapping probes of length 40bp or higher.

90% BLAT identity threshold	This study, susceptible to mis-mapping	This study, not identified as susceptible
Kasimatis et al., susceptible to mis-hybridization	15,565 (2.51%)	4,963 (0.80%)
Kasimatis et al., not identified as susceptible	469,864 (75.78%)	129,648 (20.91%)

Table S6: Characterizing GxSex based on independent analysis of individual sex-heterogenous SNPs, related to Figure 3

We simulated GWAS data study arising from the covariance structure inferred for five traits. We then followed the method described in **Text S6** for calling SNPs as sex-heterogenous SNPs and categorizing them.

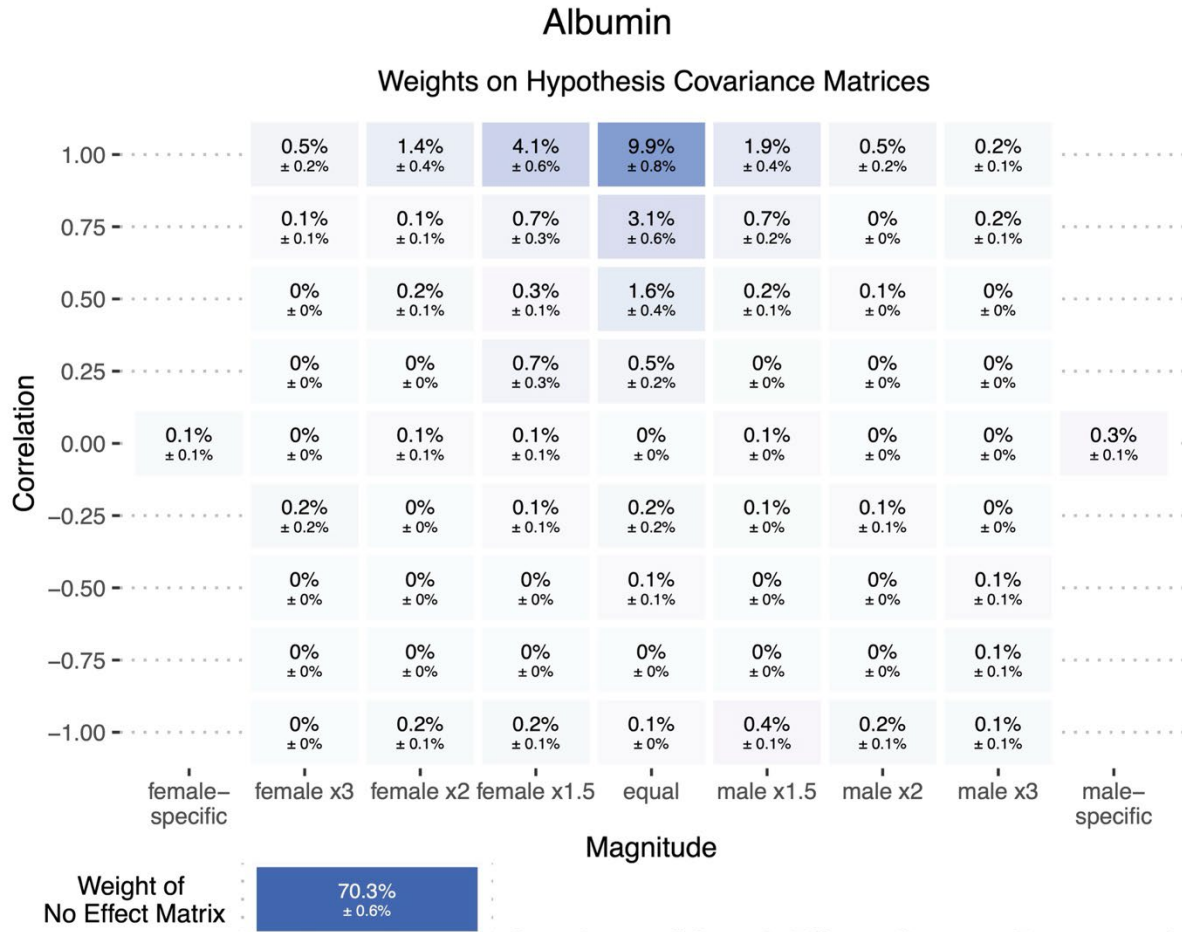
Generative covariance structure	# sex-heterogenous SNPs	% sex-specific effects	% effects in the opposite direction	% effects in the same directions
Height	994.7 (13.72)	62.3% (0.6%)	37.7% (0.6%)	0.0% (0.0%)
BMI	1527.7 (50.74)	52.1% (2.7%)	26.8% (0.9%)	21.1% (0.9%)
Creatinine	1243.6 (7.67)	52.8% (0.4%)	30.7% (0.4%)	16.4% (0.4%)
IGF-1	1392.3 (10.82)	52.2% (0.5%)	29.1% (0.7%)	18.7% (0.4%)
Systolic blood pressure	1349.2 (15.68)	51.7% (0.3%)	28.4% (0.3%)	20.0% (0.4%)

Supplementary Data

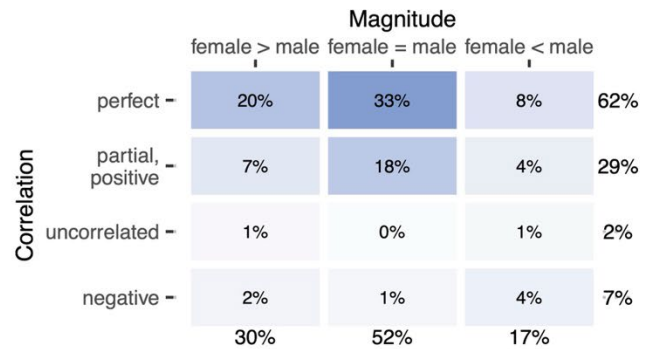
Data S1-S27: Mixture weights on hypothesis covariance matrices, related to Figure 3

Figures similar to **Fig. S3B, C** for all traits. Mixture weights (\pm SE) estimated from a sample of SNP effects using *mash* are shown with the corresponding hypothesis covariance matrix represented by the axes. Covariance matrices span all hypothesis matrices shown in **Fig. S2**. The “Weight of No Effect Matrix” box represents the weight on a matrix signifying no effects in both sexes. The bottom plot for each traits represents a consolidated version of the top plot. Weights on the bottom plot are shown as percentages of non-null weights.

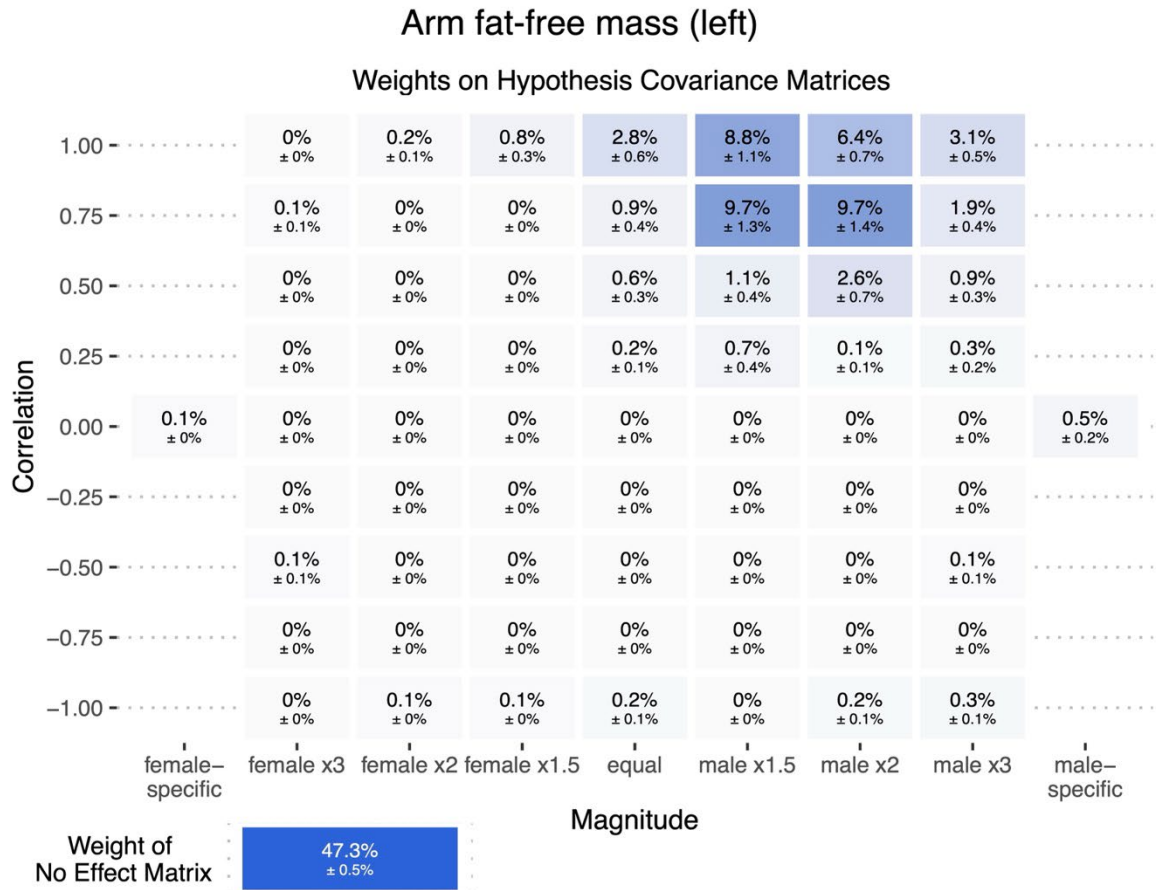
Data S1: Albumin



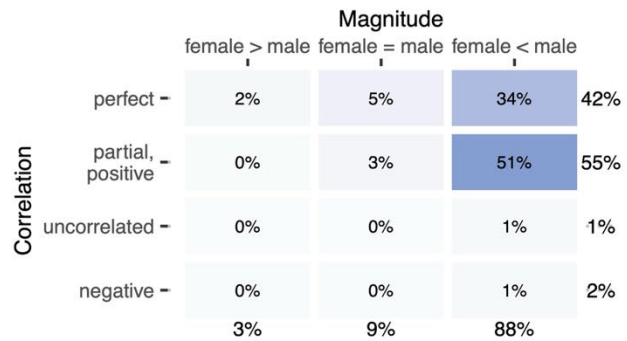
Covariance of Genetic Effects: Compact Representation



Data S2: Arm fat-free mass (L)



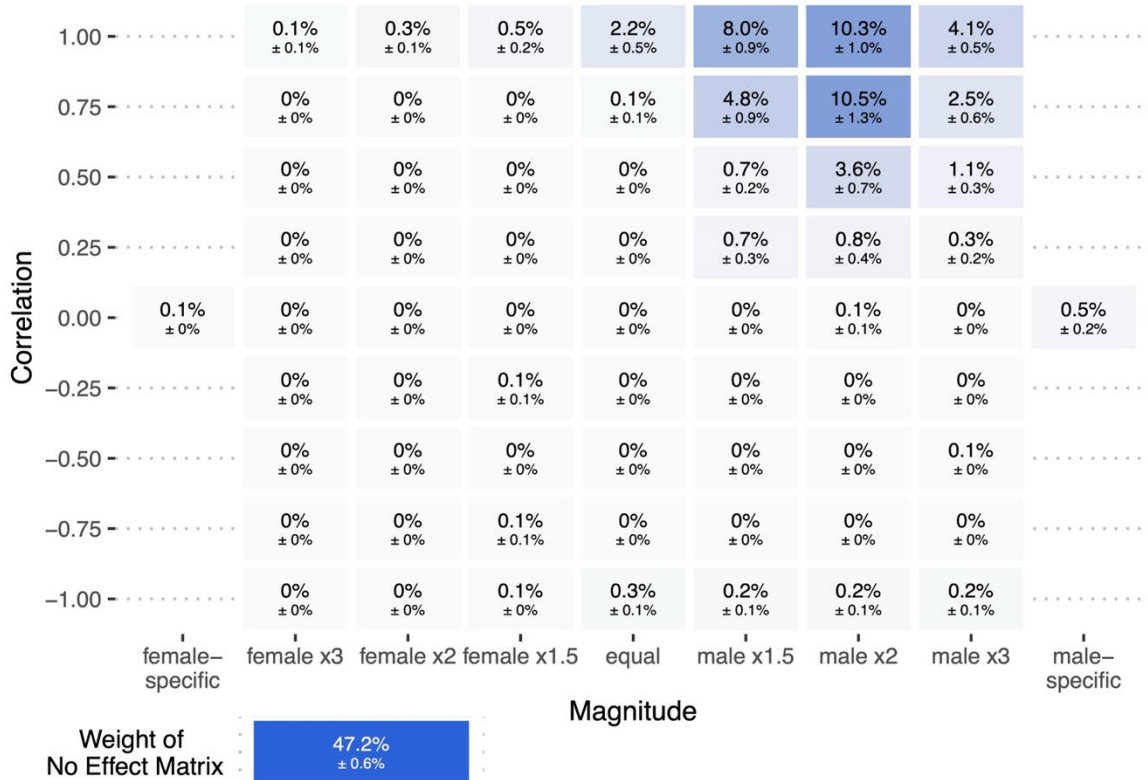
Covariance of Genetic Effects: Compact Representation



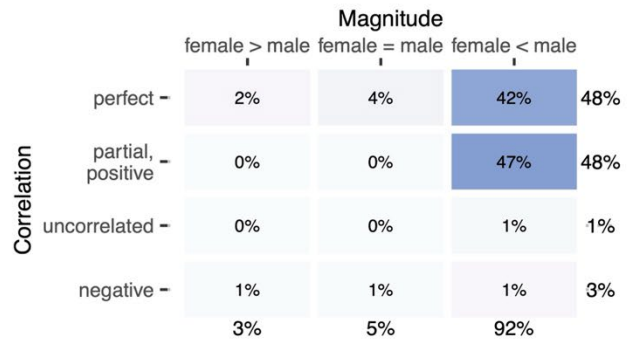
Data S3: Arm fat-free mass (R)

Arm fat-free mass (right)

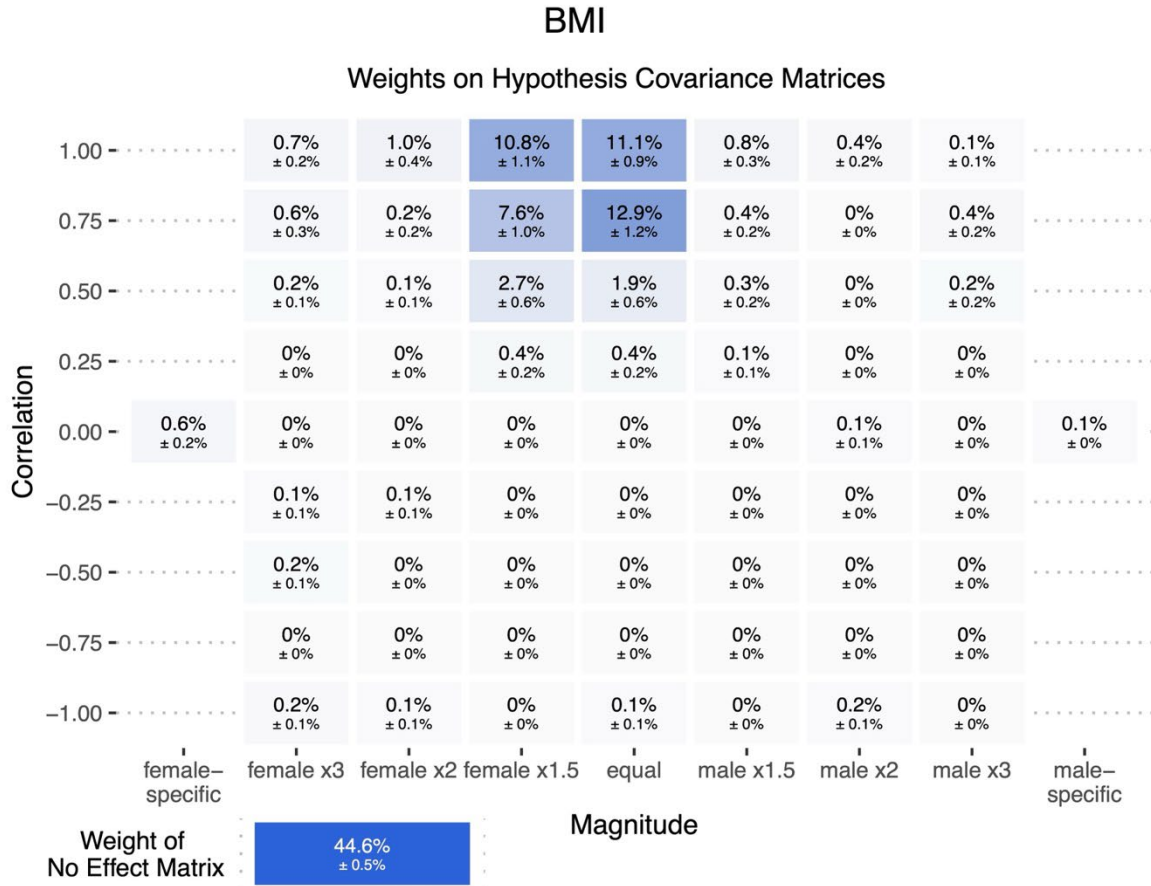
Weights on Hypothesis Covariance Matrices



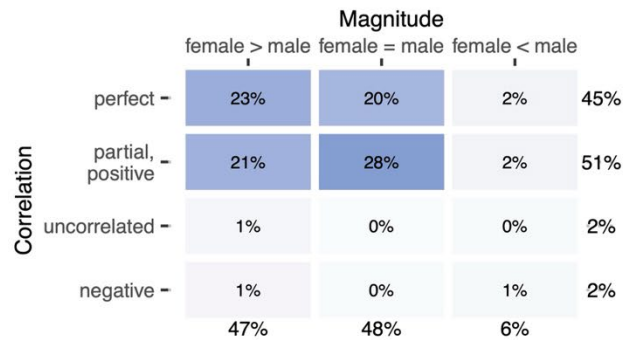
Covariance of Genetic Effects: Compact Representation



Data S4: BMI



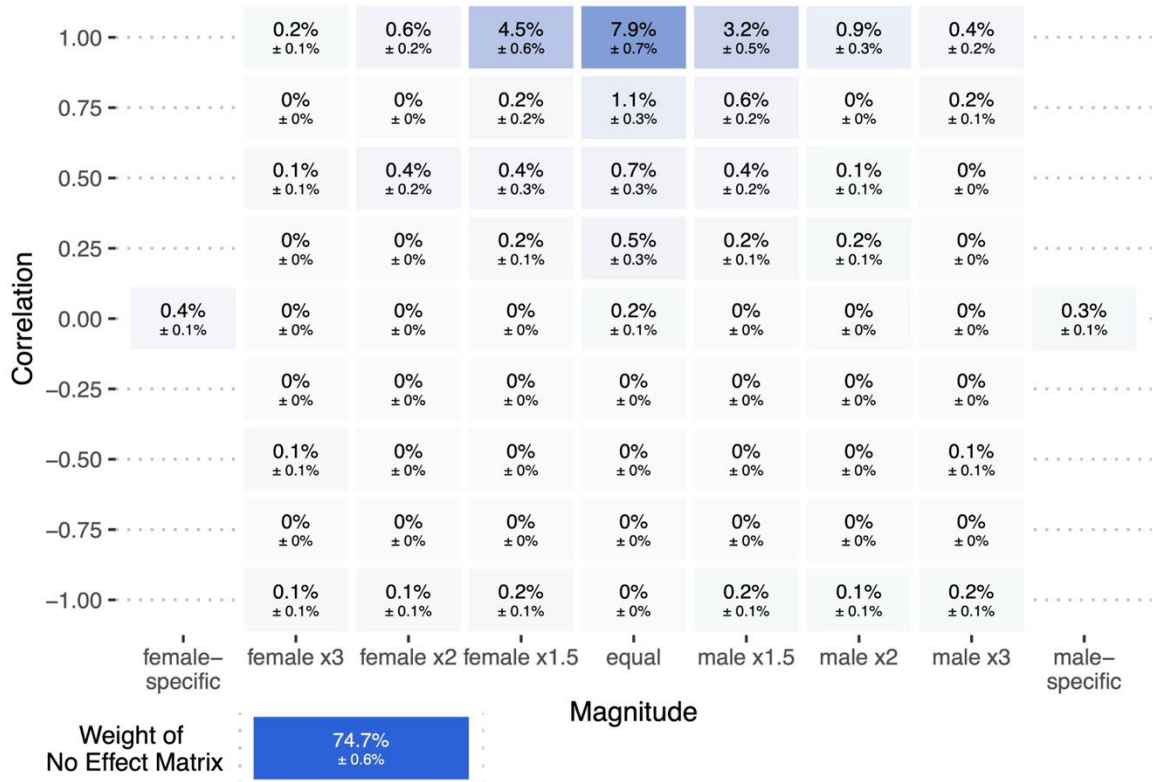
Covariance of Genetic Effects: Compact Representation



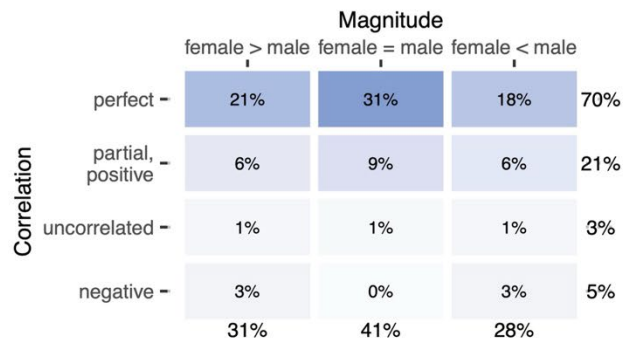
Data S5: Calcium

Calcium

Weights on Hypothesis Covariance Matrices



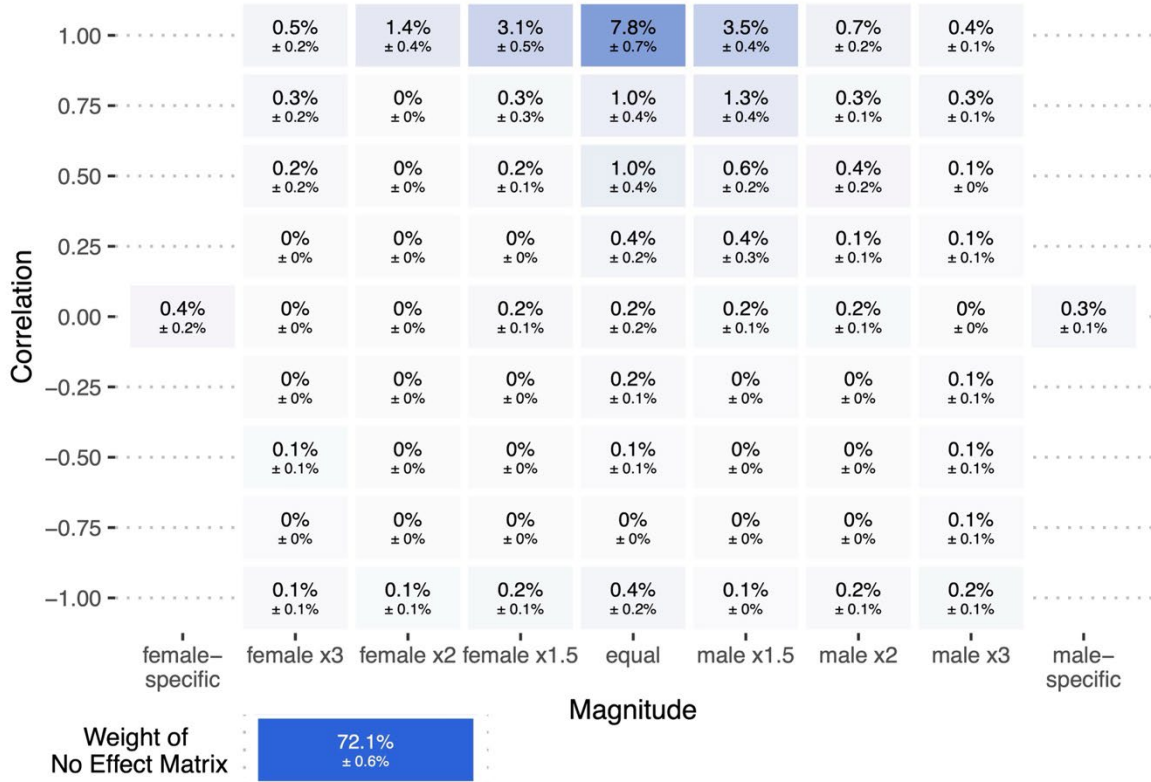
Covariance of Genetic Effects: Compact Representation



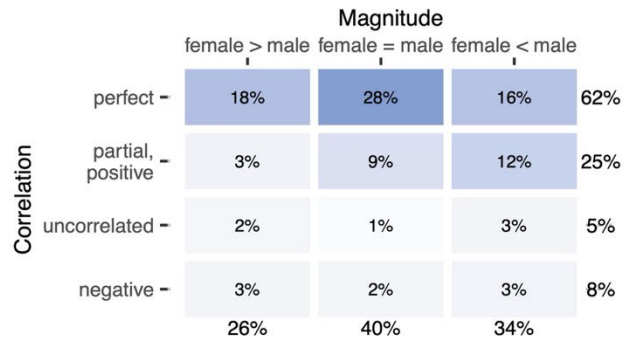
Data S6: Creatinine

Creatinine

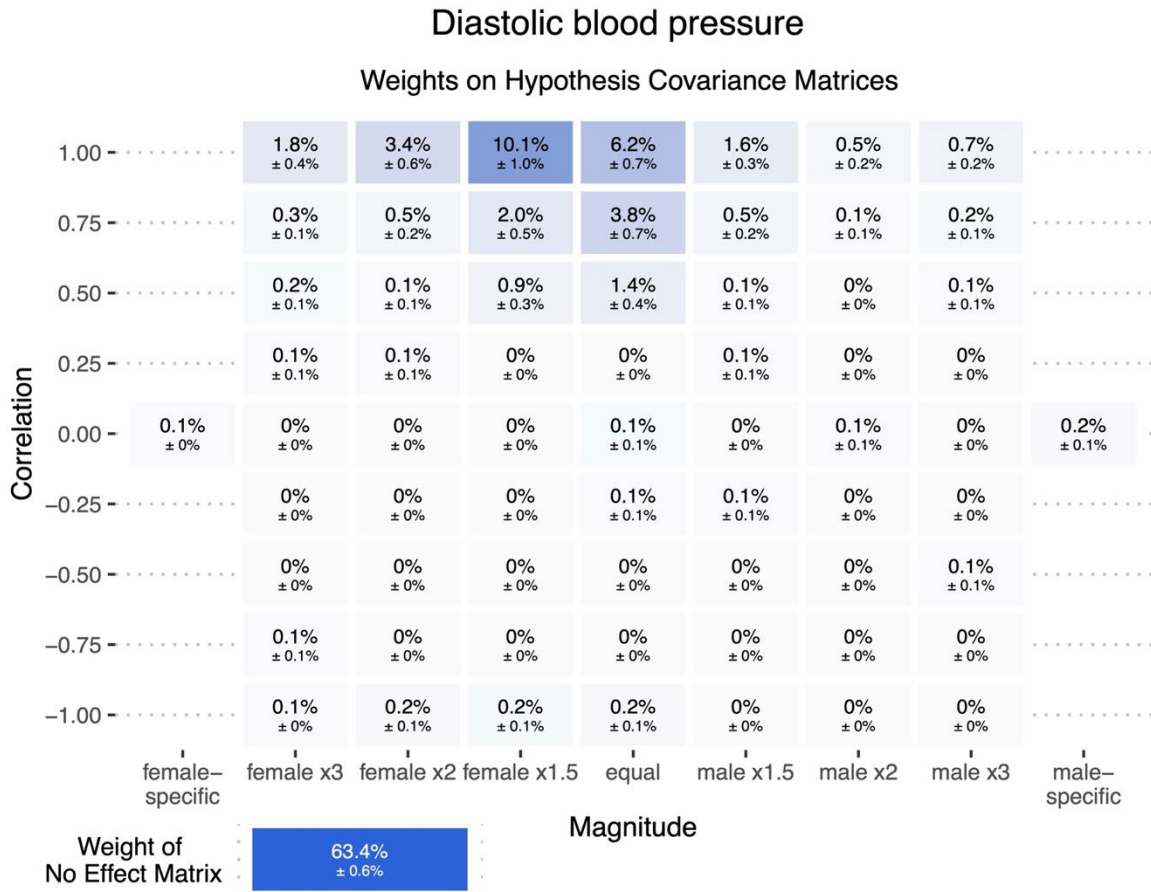
Weights on Hypothesis Covariance Matrices



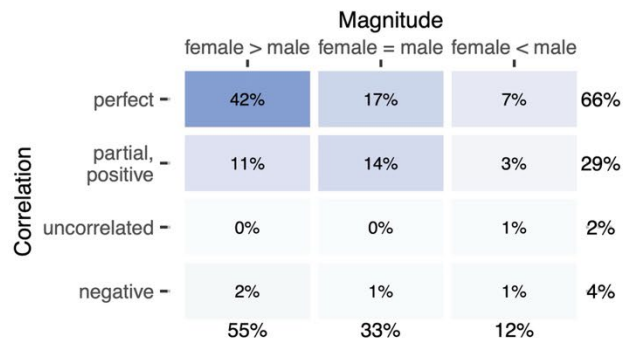
Covariance of Genetic Effects: Compact Representation



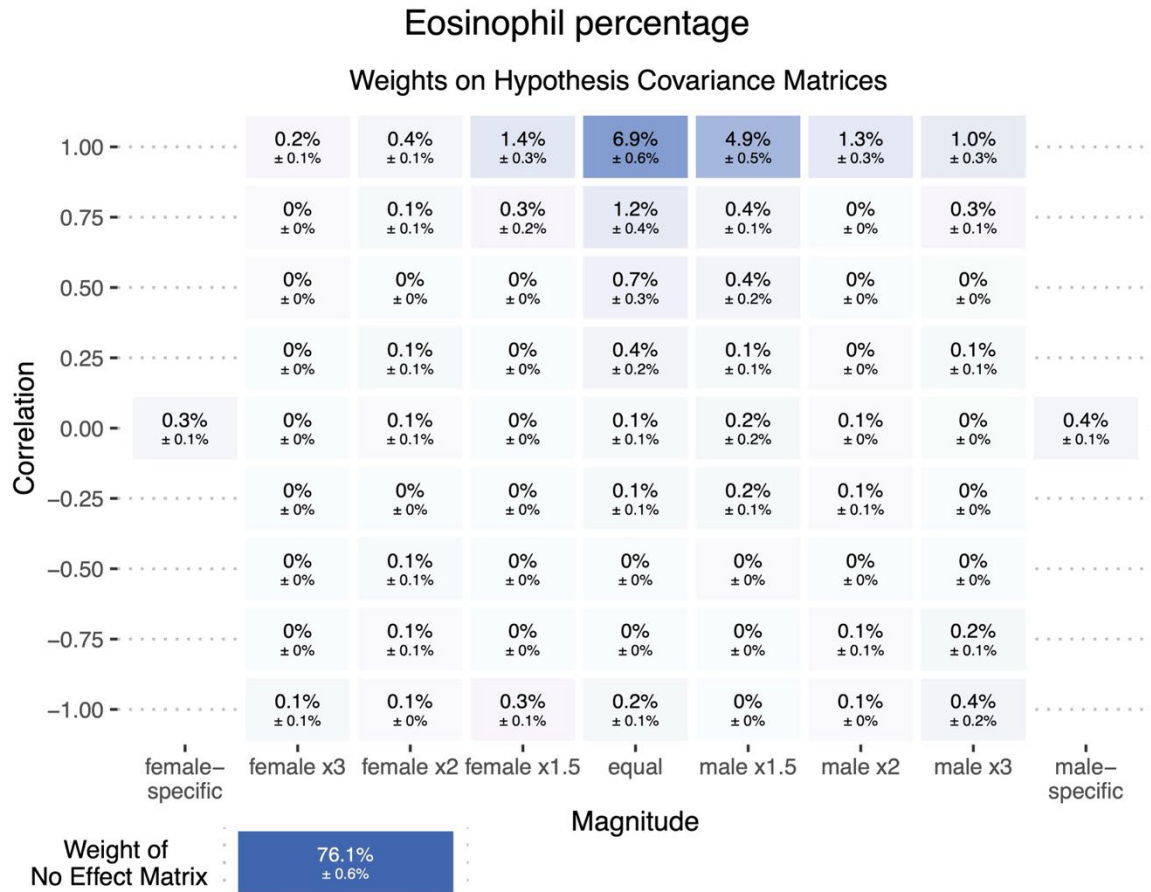
Data S7: Diastolic blood pressure



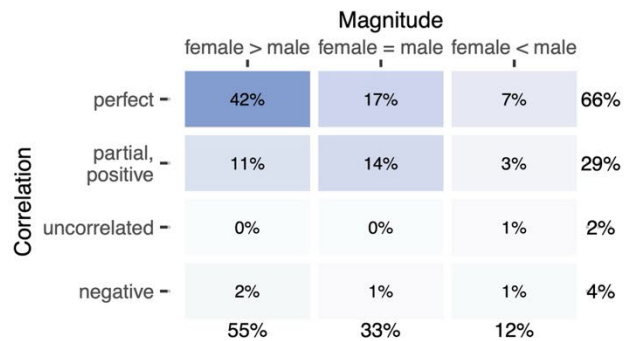
Covariance of Genetic Effects: Compact Representation



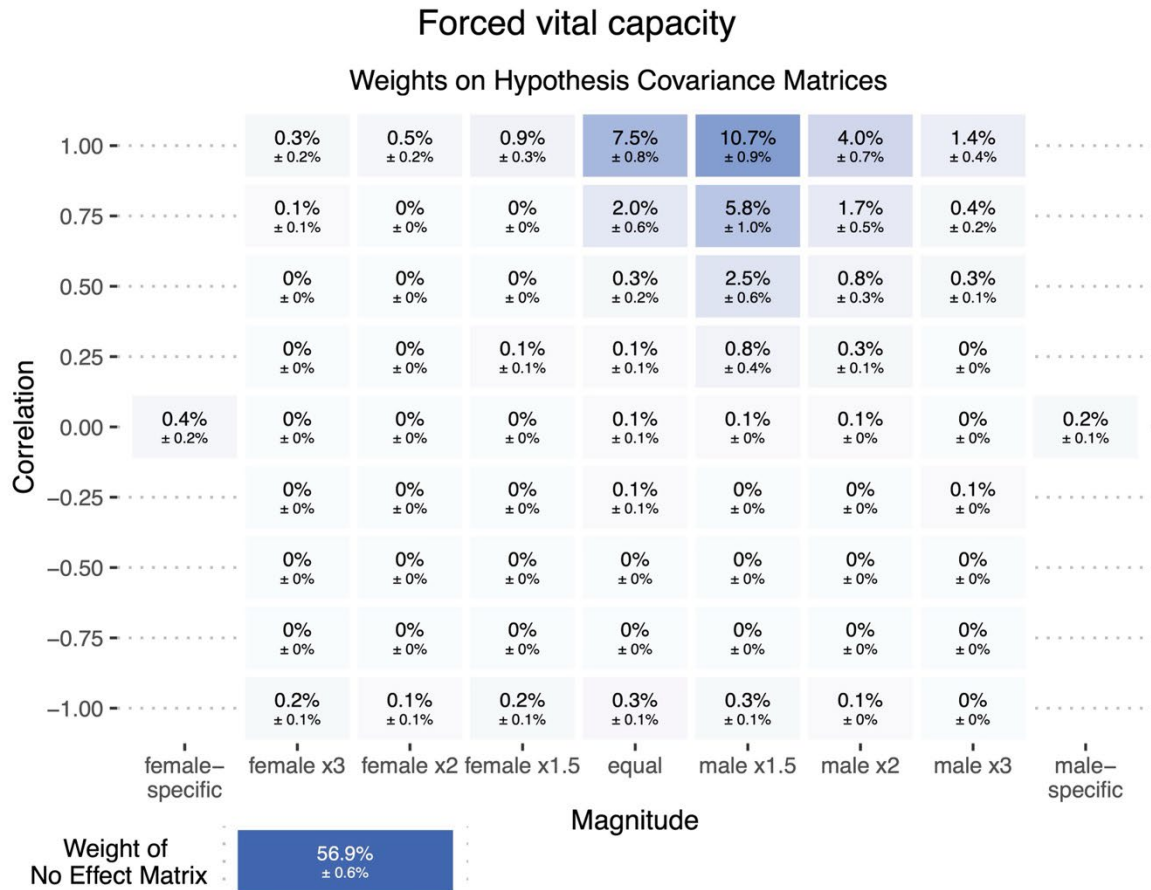
Data S8: Eosinophil percentage



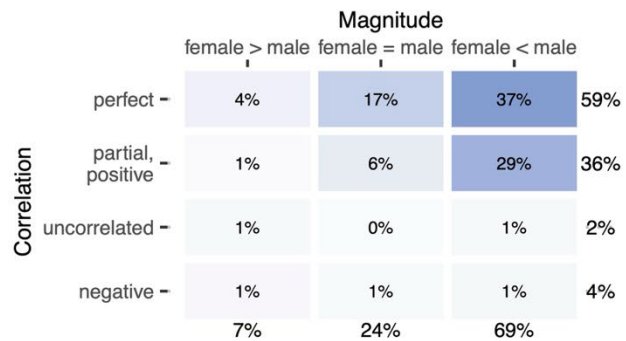
Covariance of Genetic Effects: Compact Representation



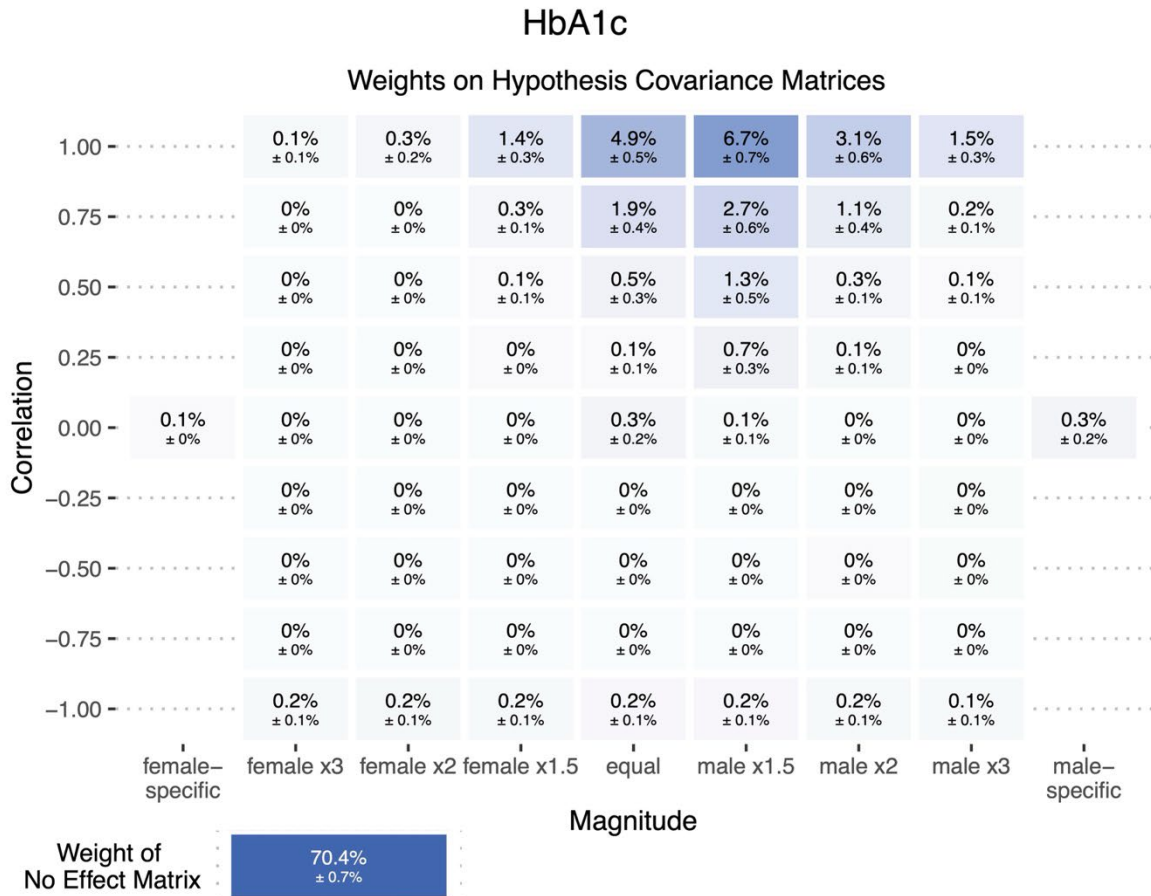
Data S9: Forced vital capacity



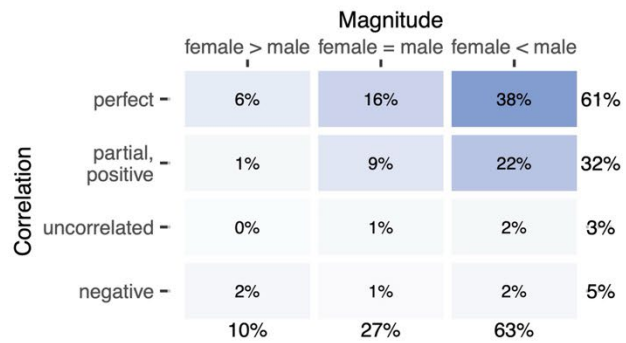
Covariance of Genetic Effects: Compact Representation



Data S10: HbA1c



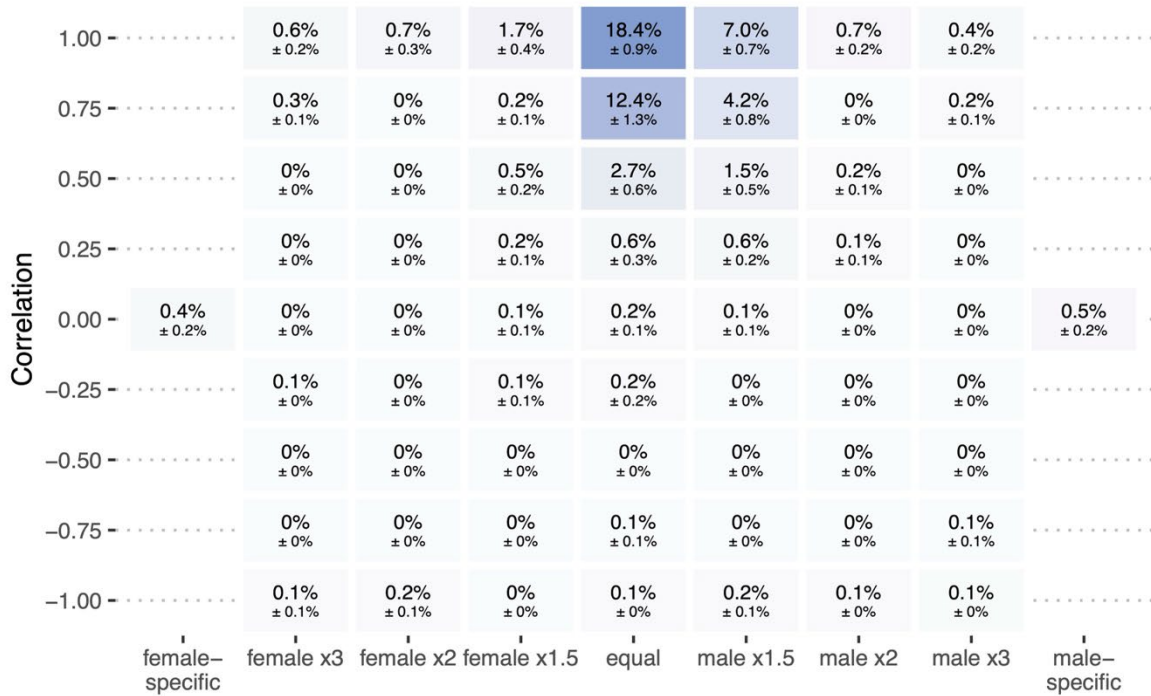
Covariance of Genetic Effects: Compact Representation



Data S11: Height

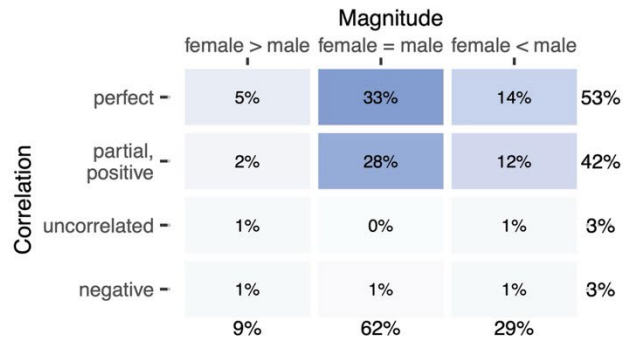
Height

Weights on Hypothesis Covariance Matrices

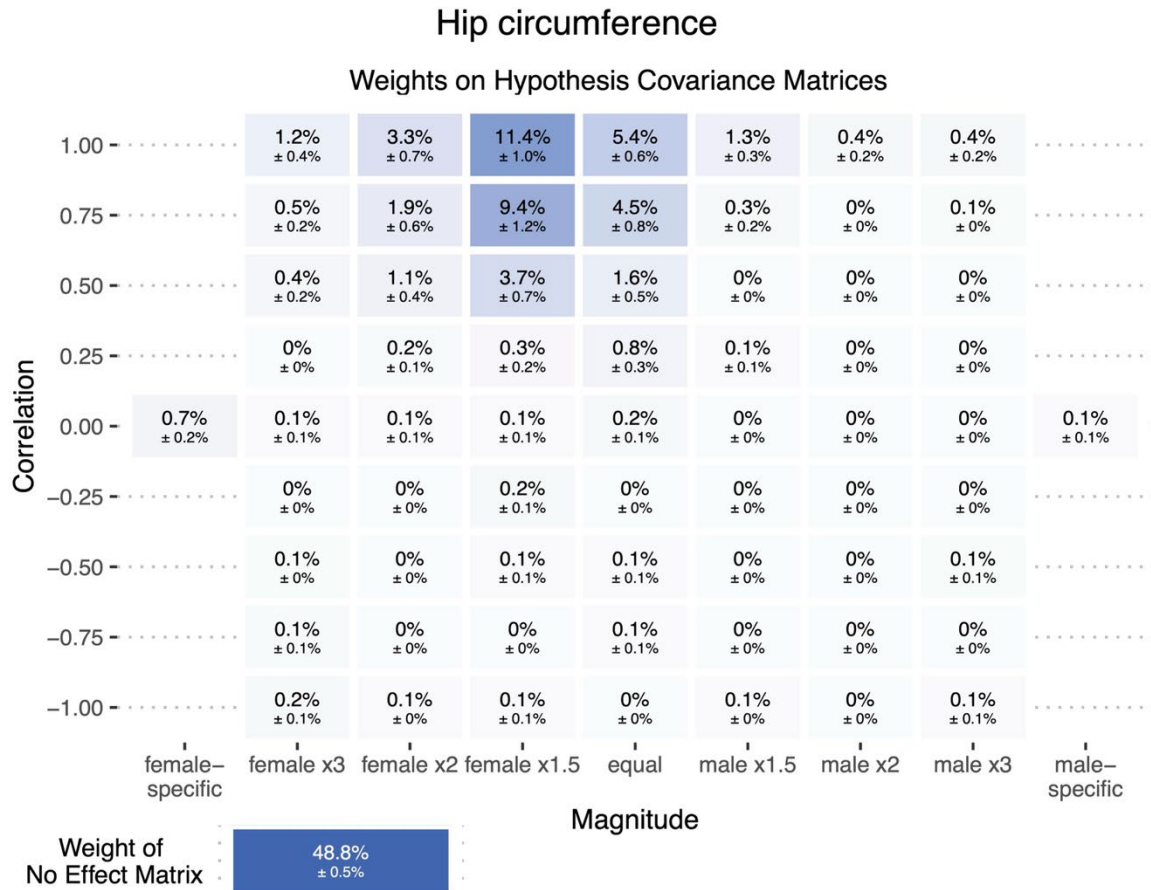


Magnitude

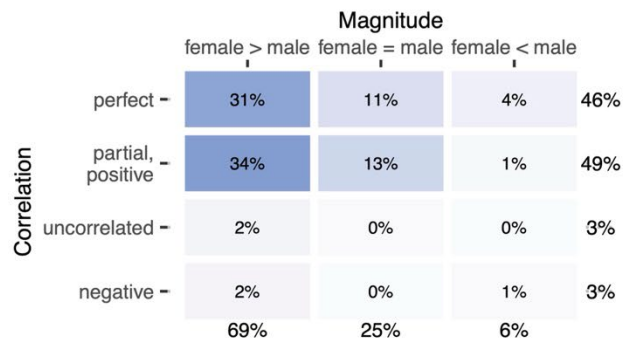
Covariance of Genetic Effects: Compact Representation



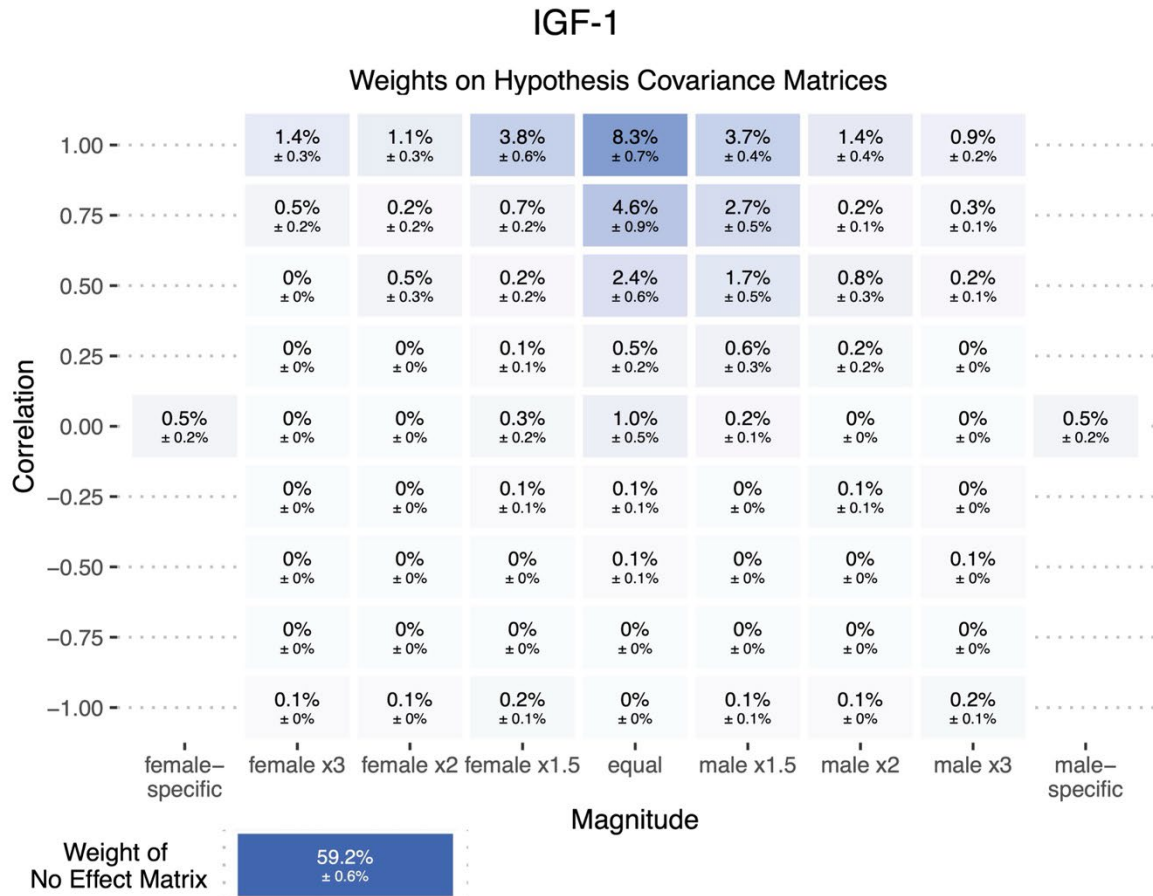
Data S12: Hip circumference



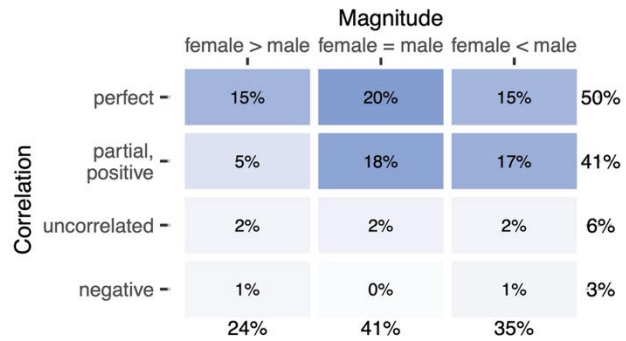
Covariance of Genetic Effects: Compact Representation



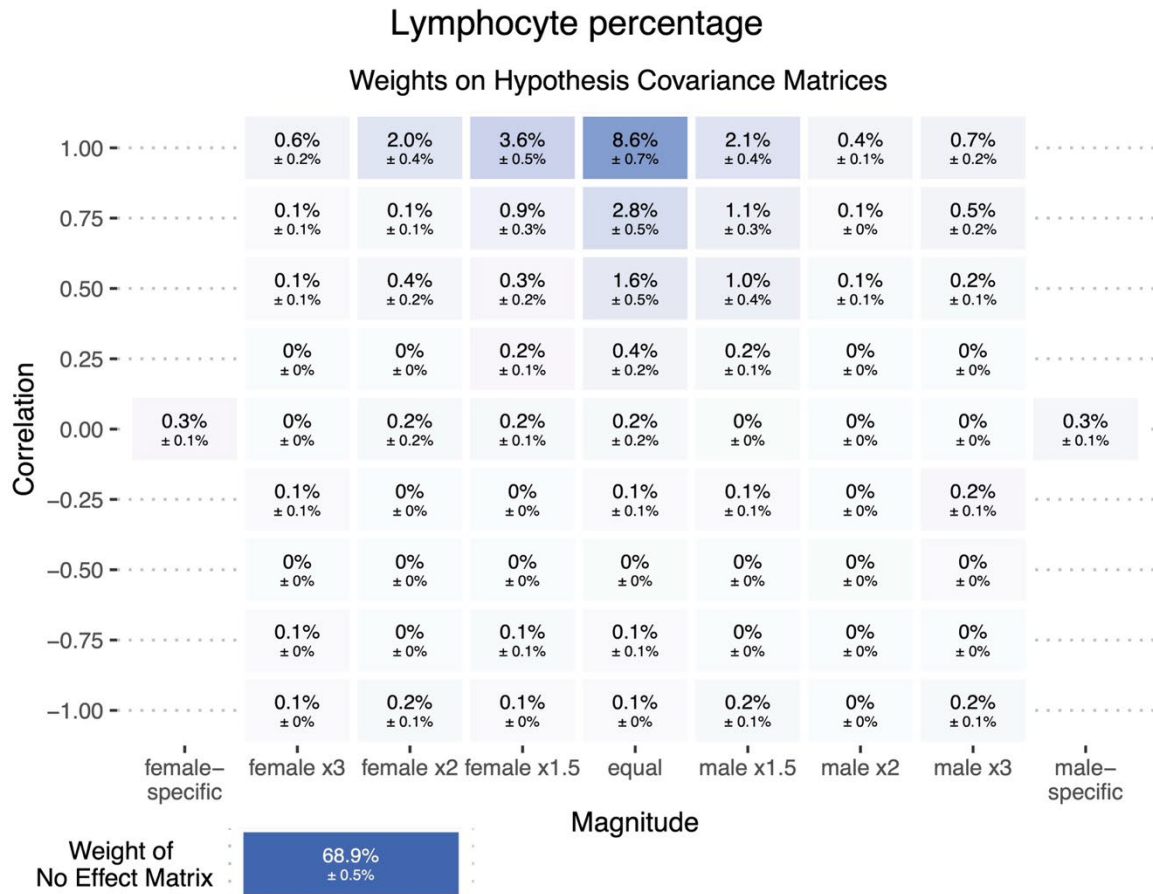
Data S13: IGF-1



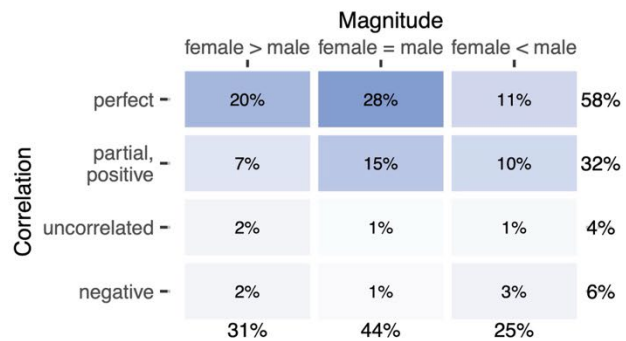
Covariance of Genetic Effects: Compact Representation



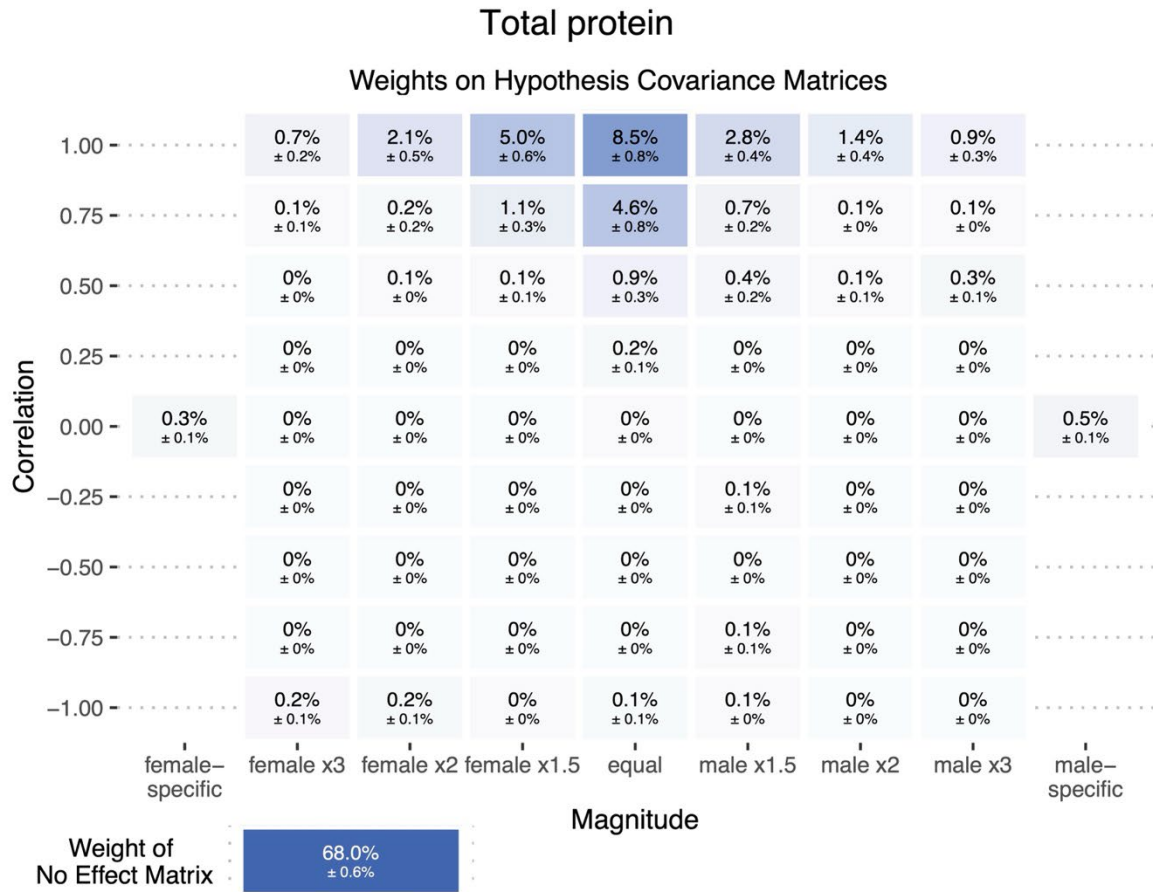
Data S14: Lymphocyte percentage



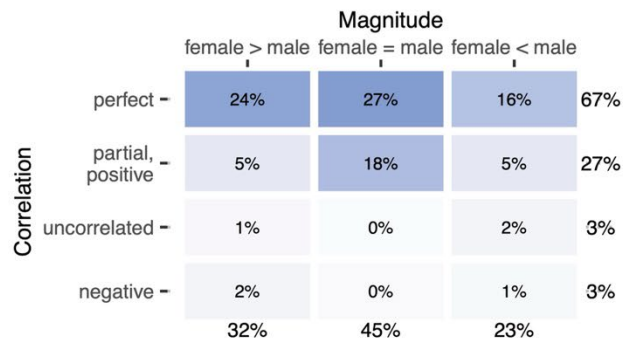
Covariance of Genetic Effects: Compact Representation



Data S15: Total protein



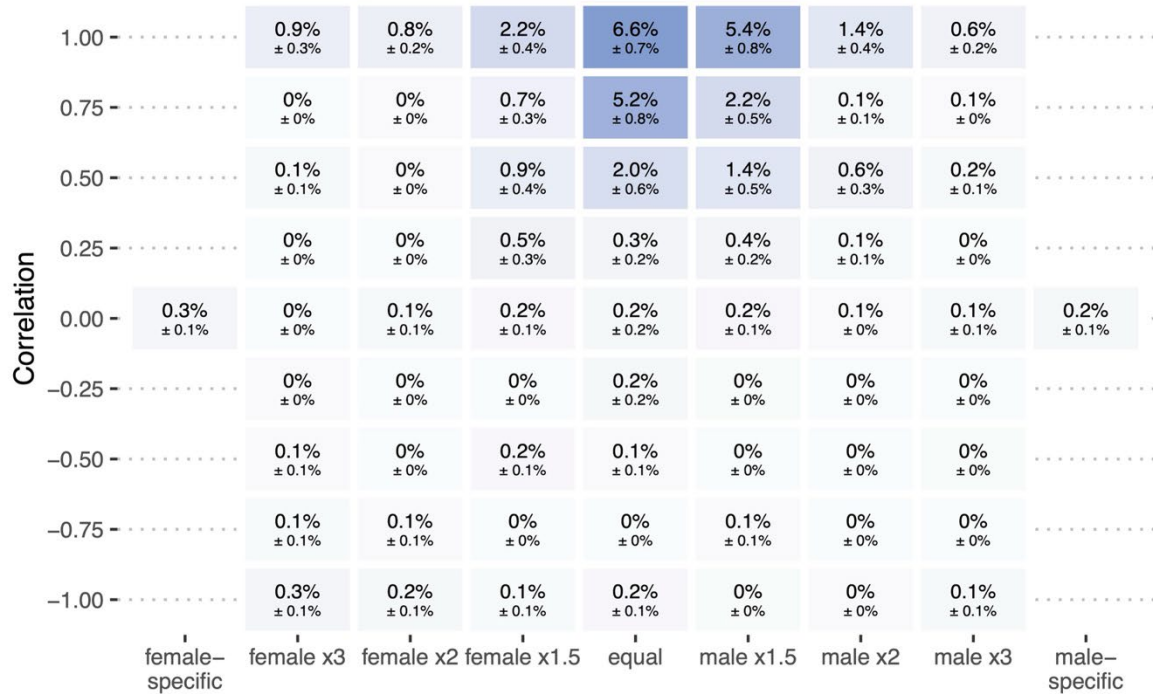
Covariance of Genetic Effects: Compact Representation



Data S16: Pulse rate

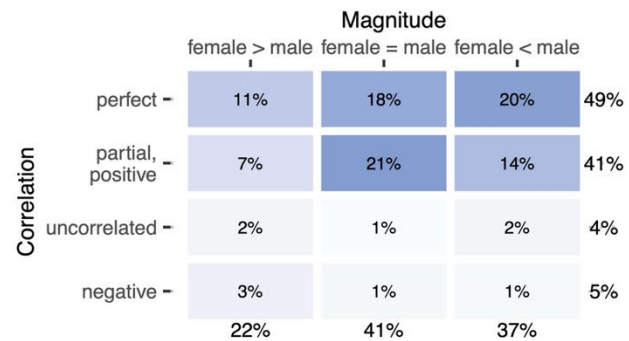
Pulse rate

Weights on Hypothesis Covariance Matrices

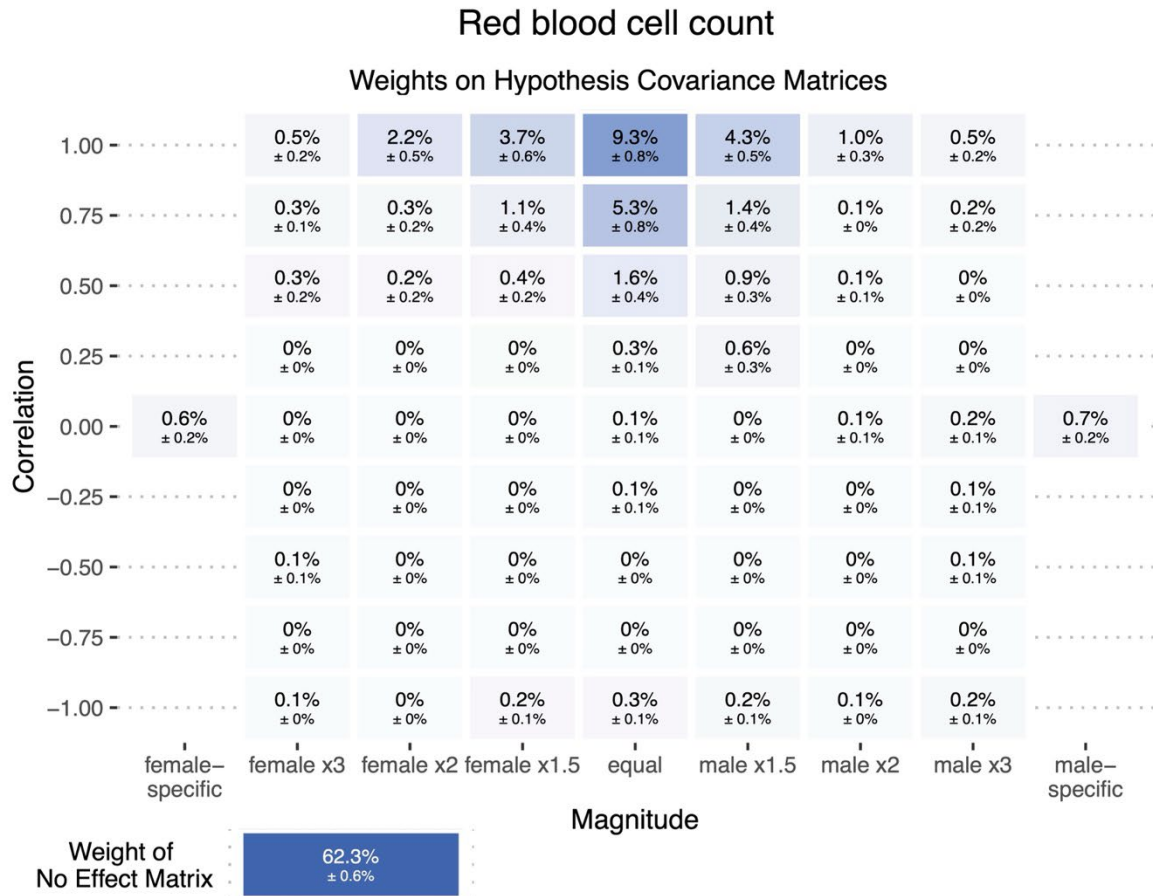


Weight of No Effect Matrix: 63.7% ± 0.6%

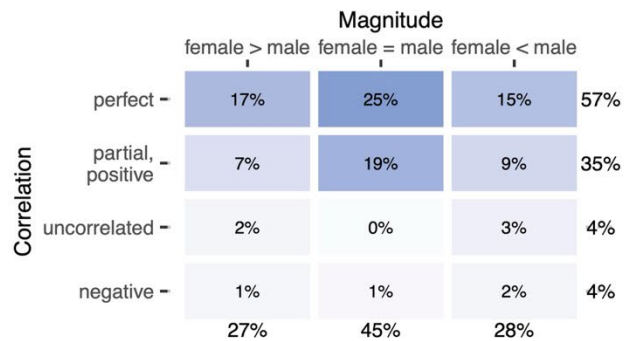
Covariance of Genetic Effects: Compact Representation



Data S17: Red blood cell count



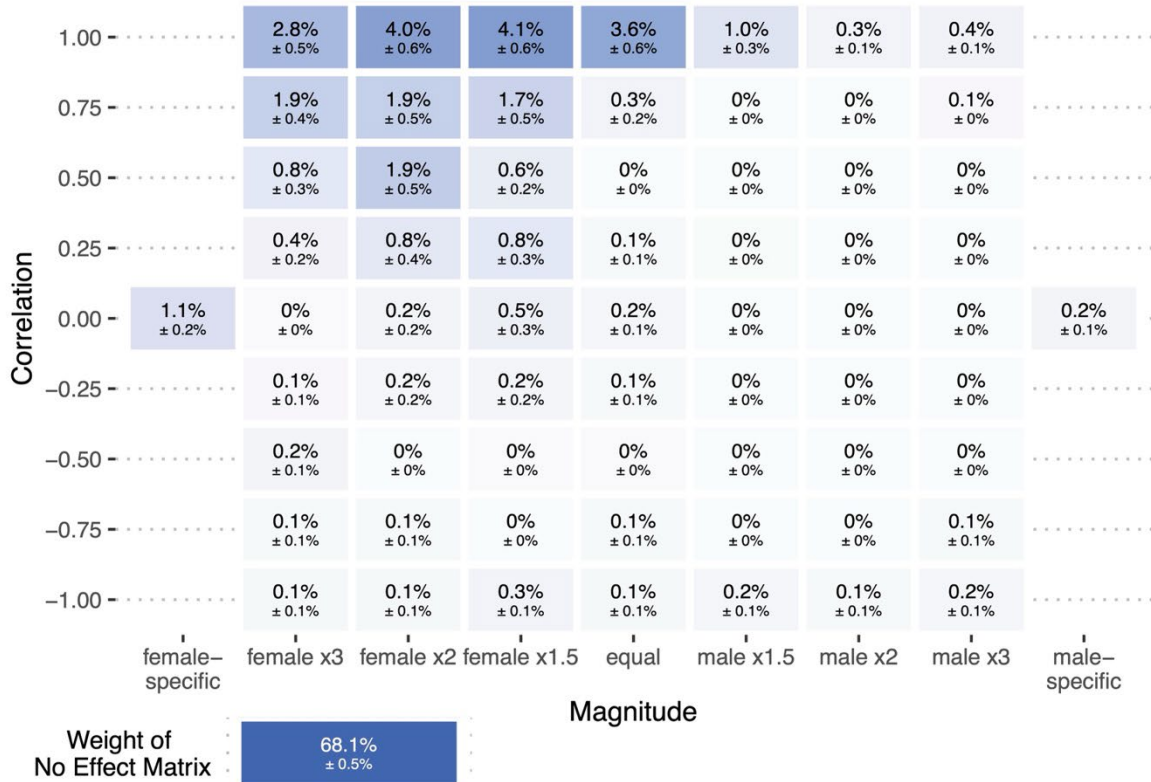
Covariance of Genetic Effects: Compact Representation



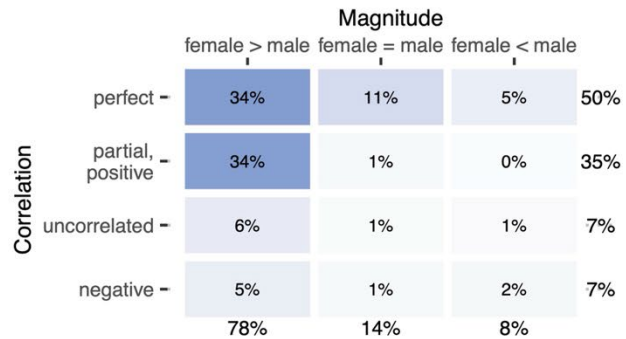
Data S18: SHBG

SHBG

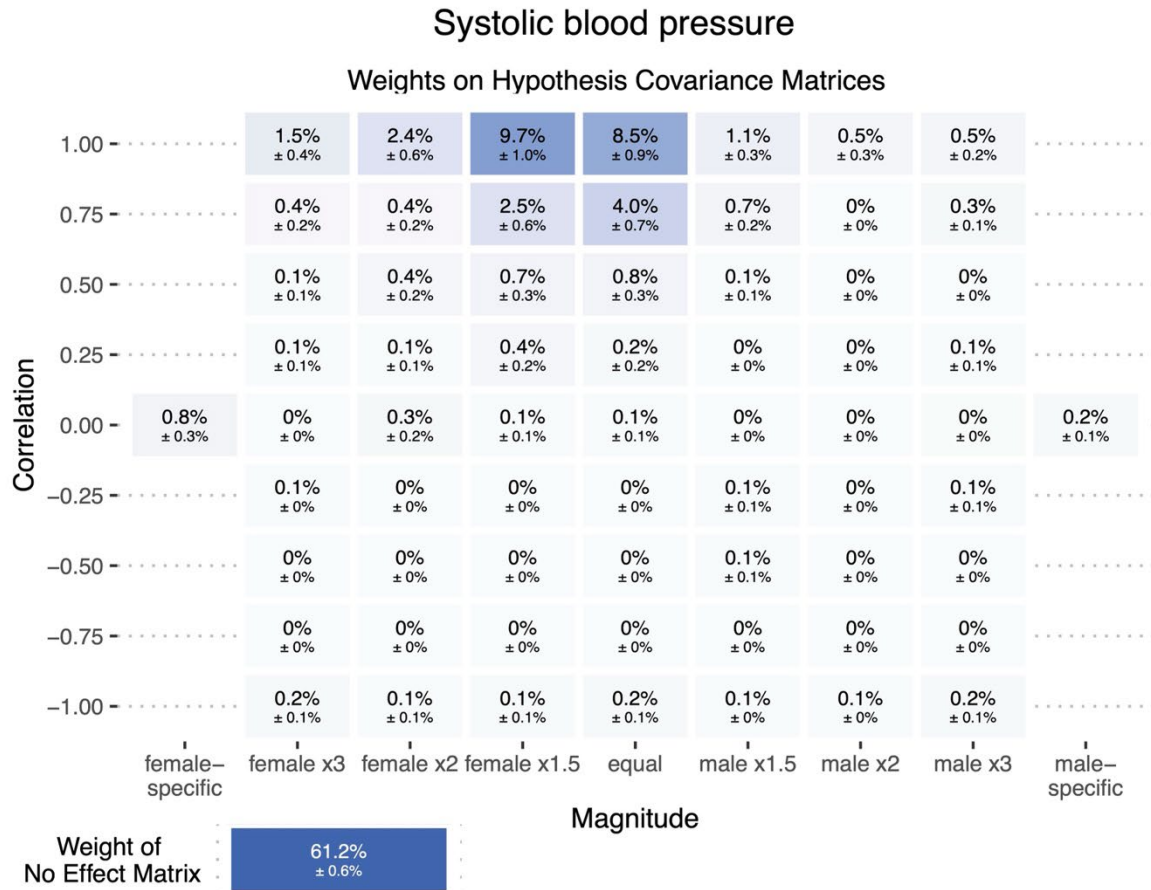
Weights on Hypothesis Covariance Matrices



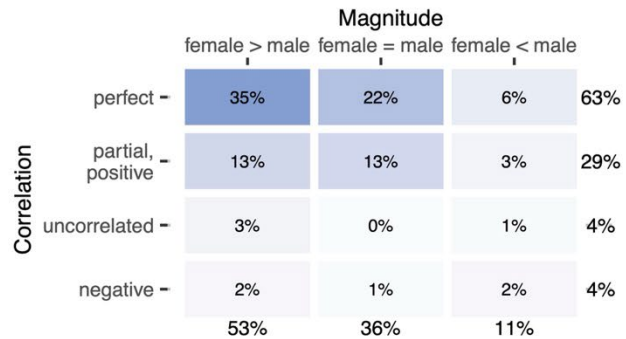
Covariance of Genetic Effects: Compact Representation



Data S19: Systolic blood pressure



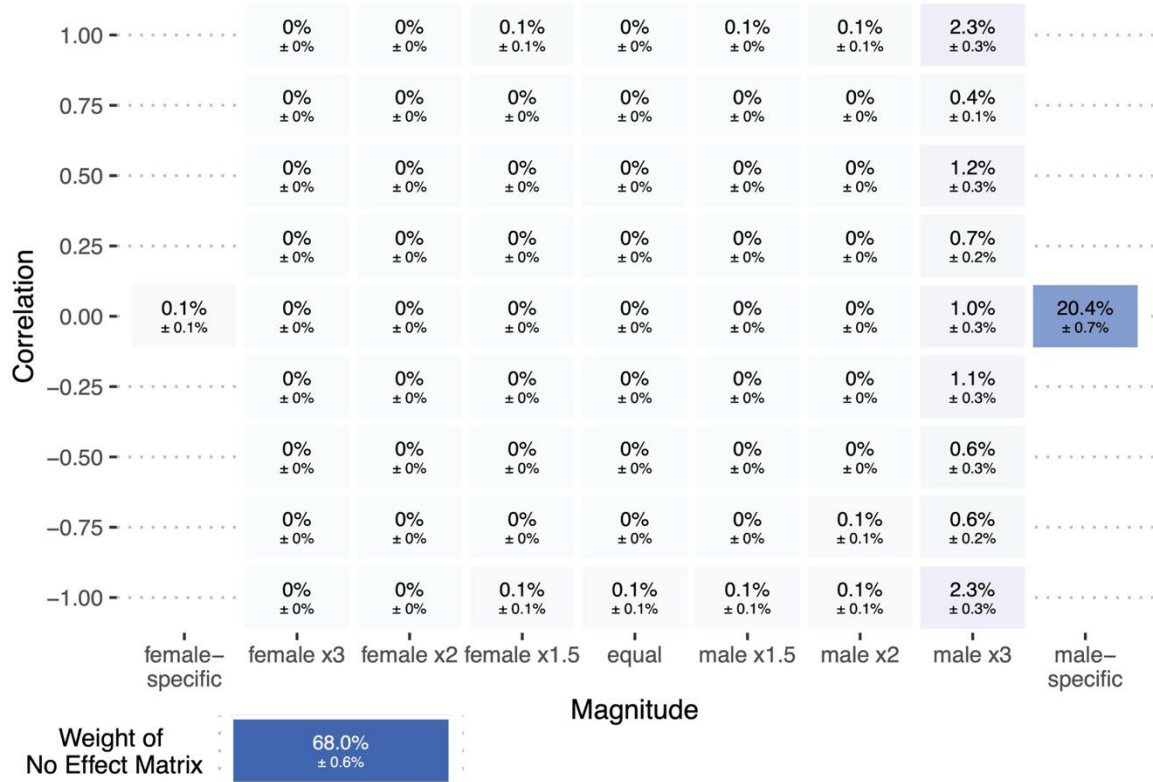
Covariance of Genetic Effects: Compact Representation



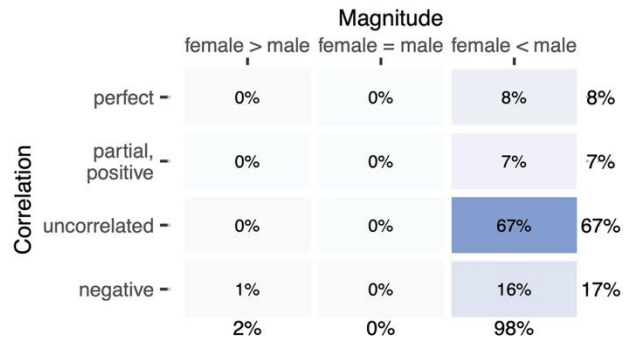
Data S20: Testosterone

Testosterone

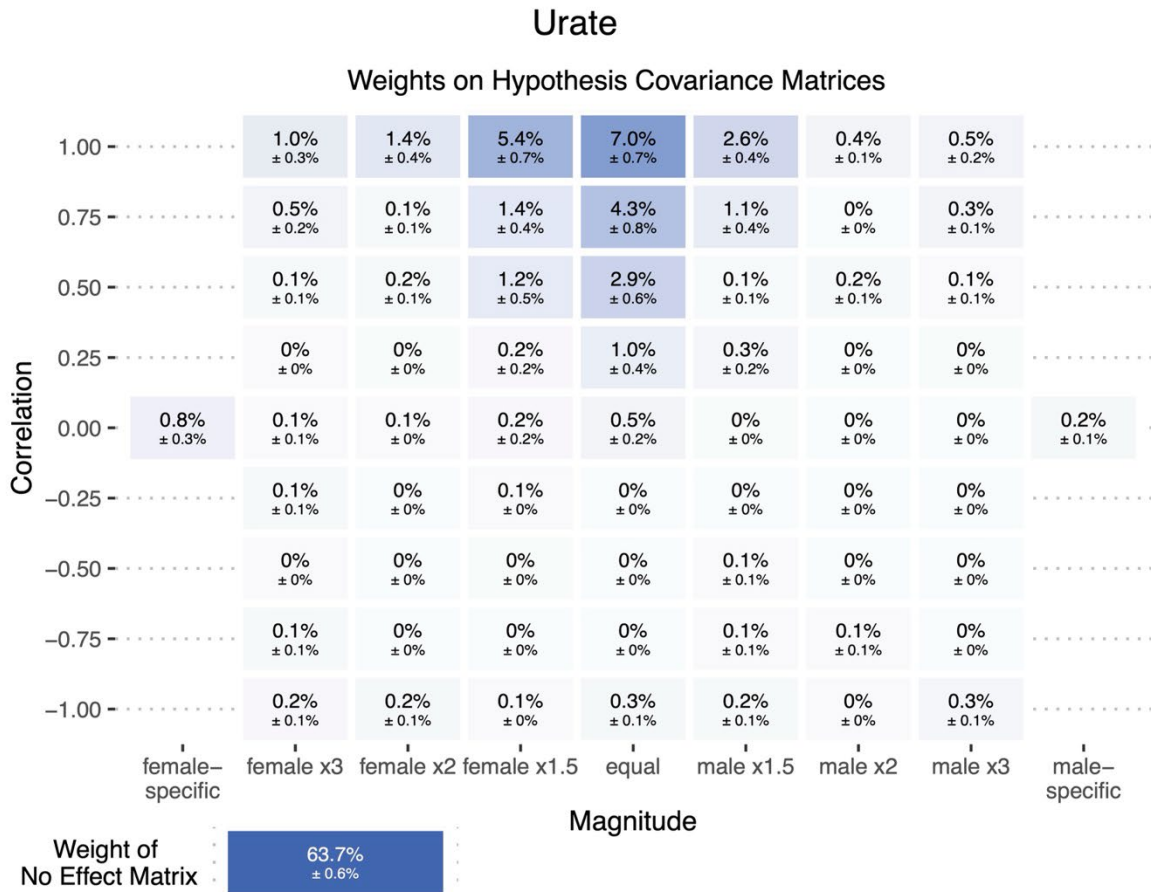
Weights on Hypothesis Covariance Matrices



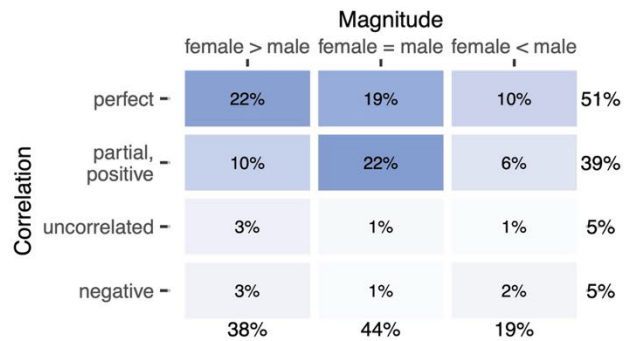
Covariance of Genetic Effects: Compact Representation



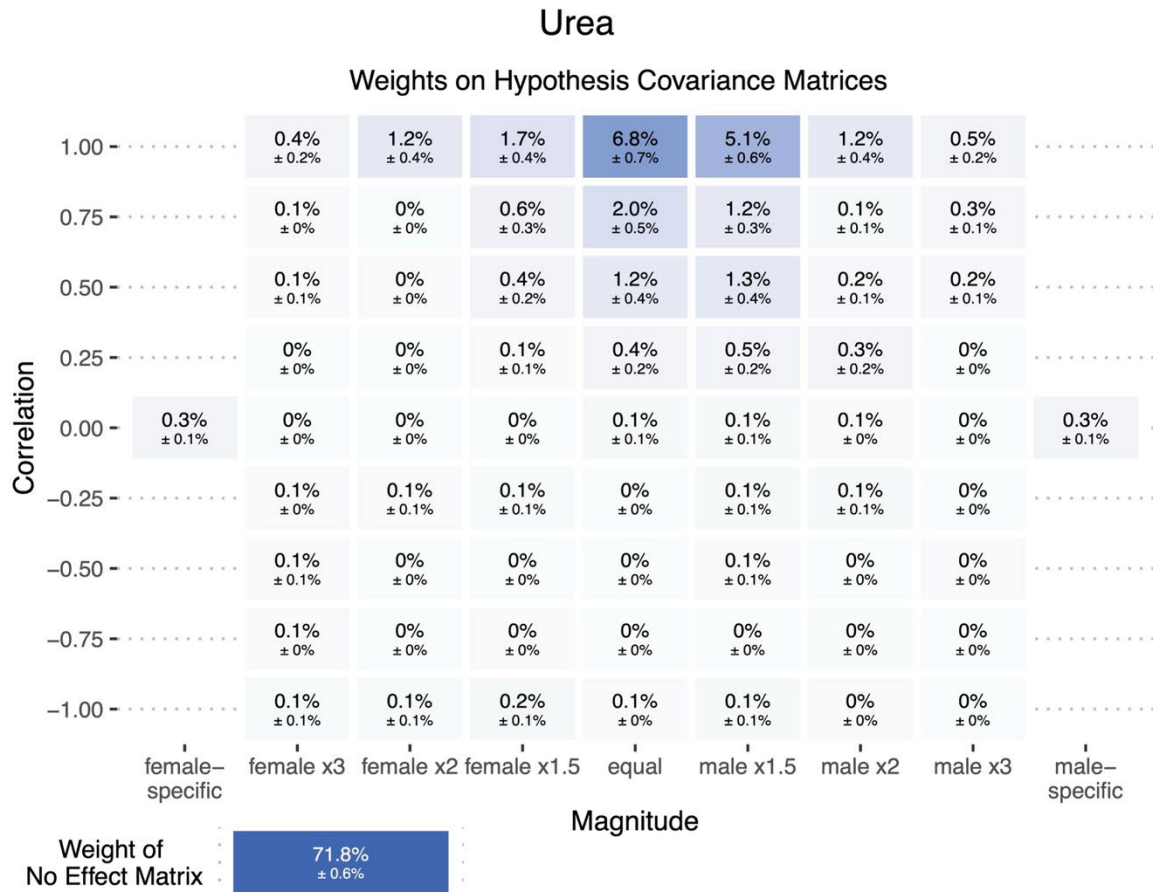
Data S21: Urate



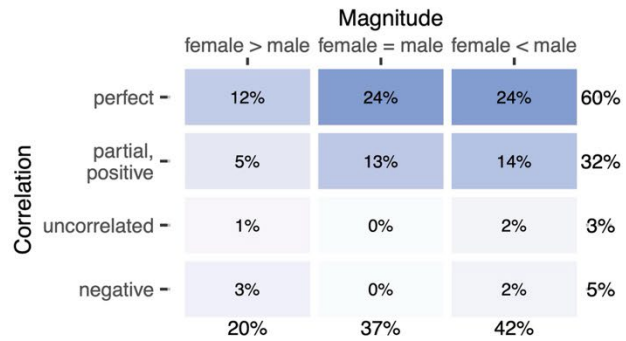
Covariance of Genetic Effects: Compact Representation



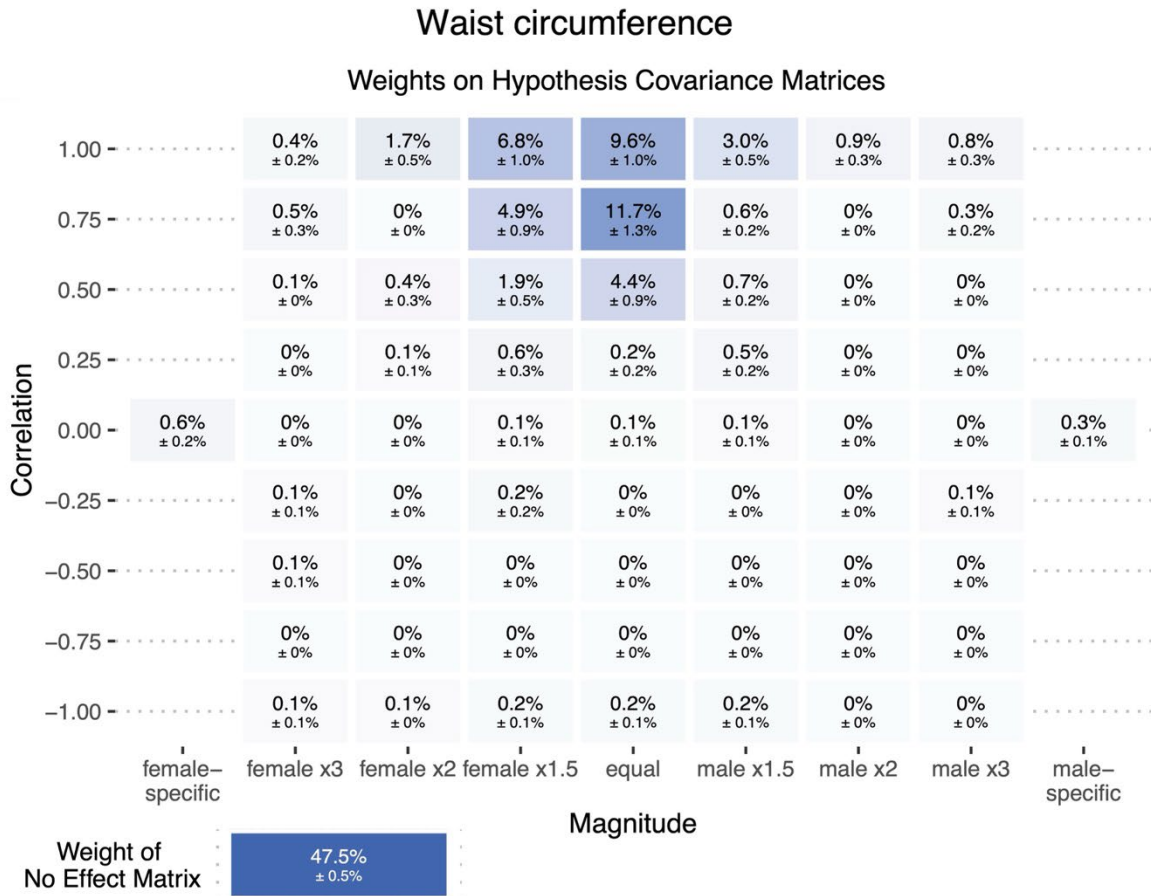
Data S22: Urea



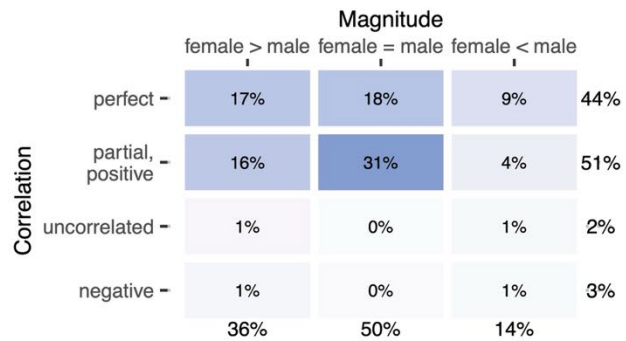
Covariance of Genetic Effects: Compact Representation



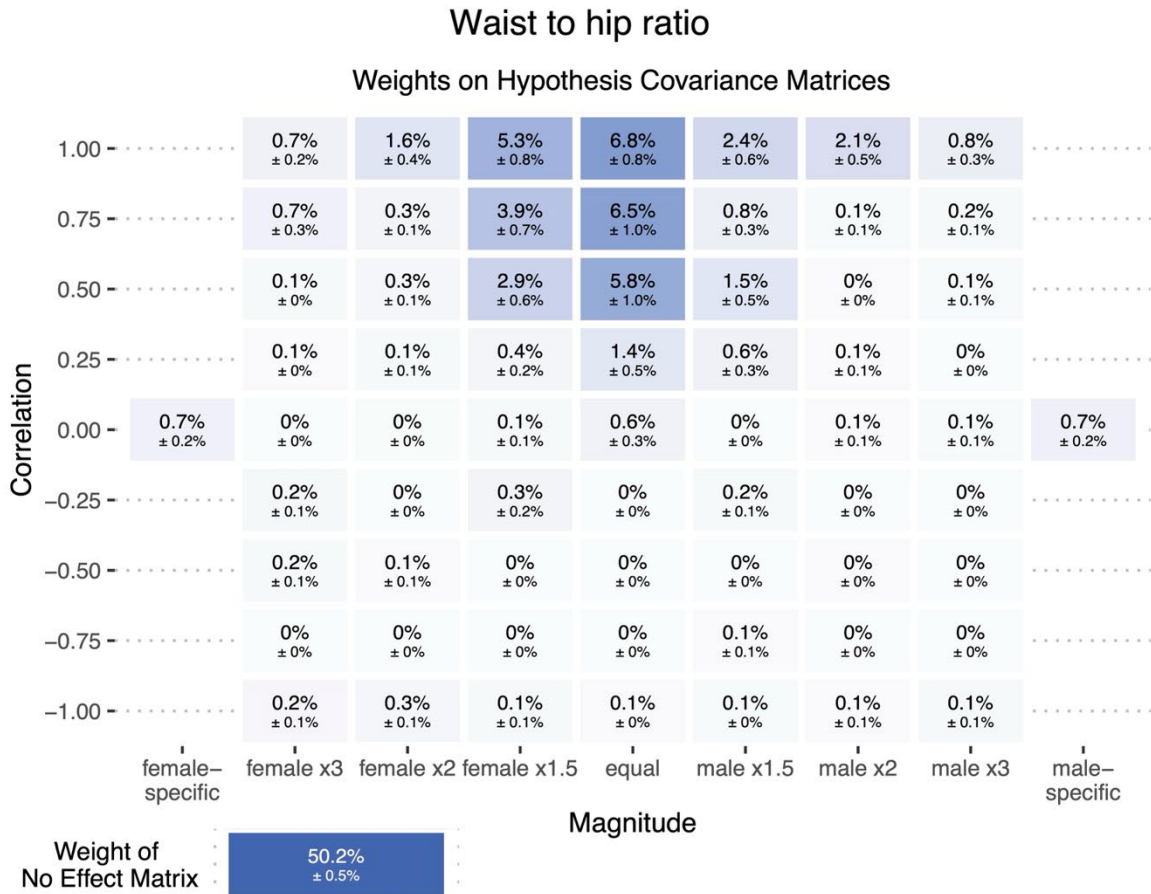
Data S23: Waist circumference



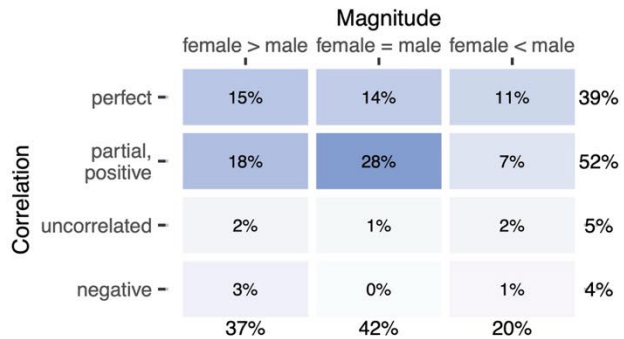
Covariance of Genetic Effects: Compact Representation



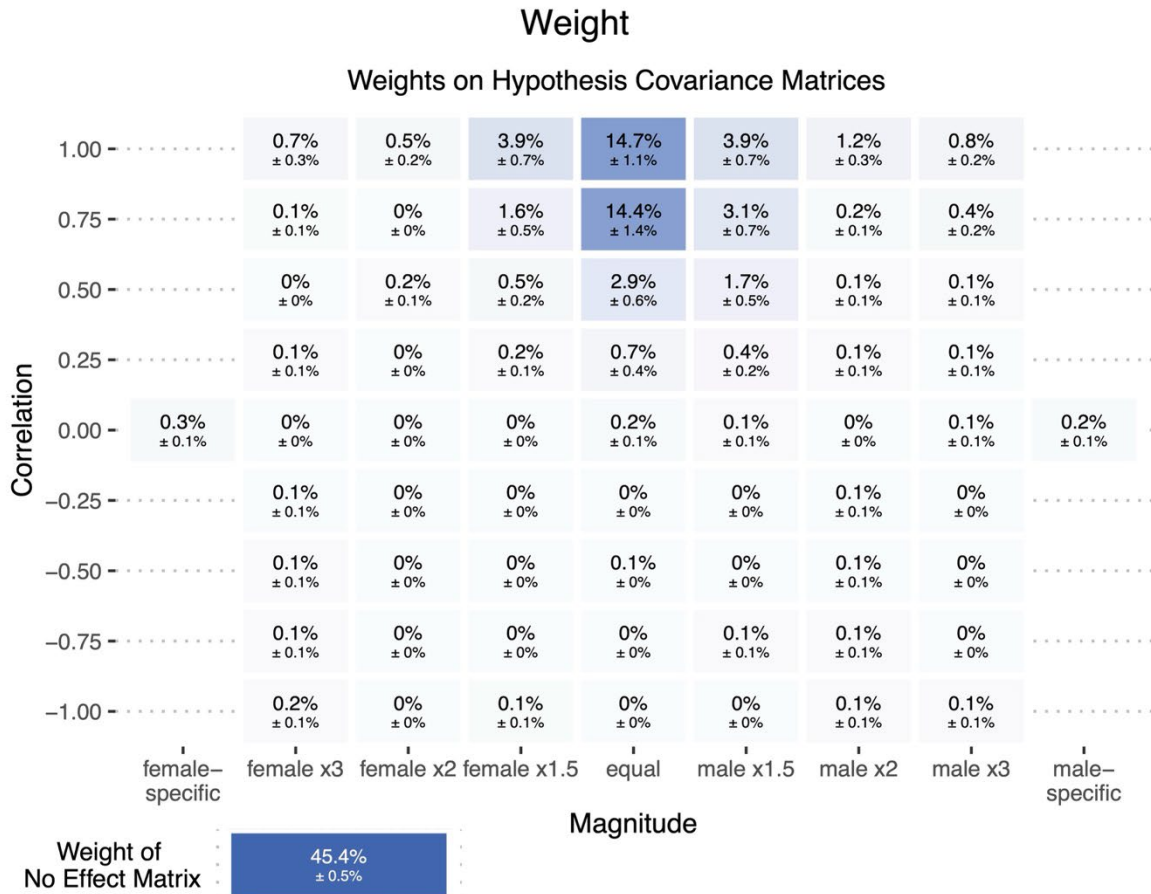
Data S24: Waist to hip ratio



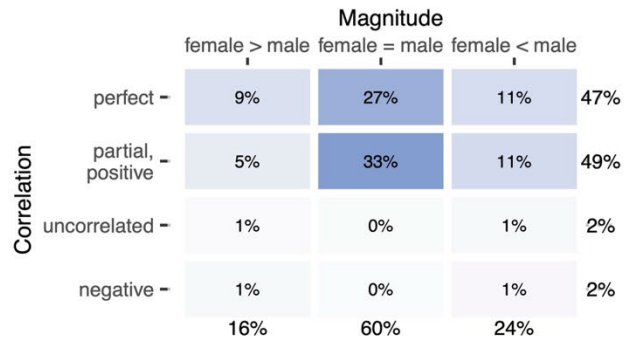
Covariance of Genetic Effects: Compact Representation



Data S25: Weight



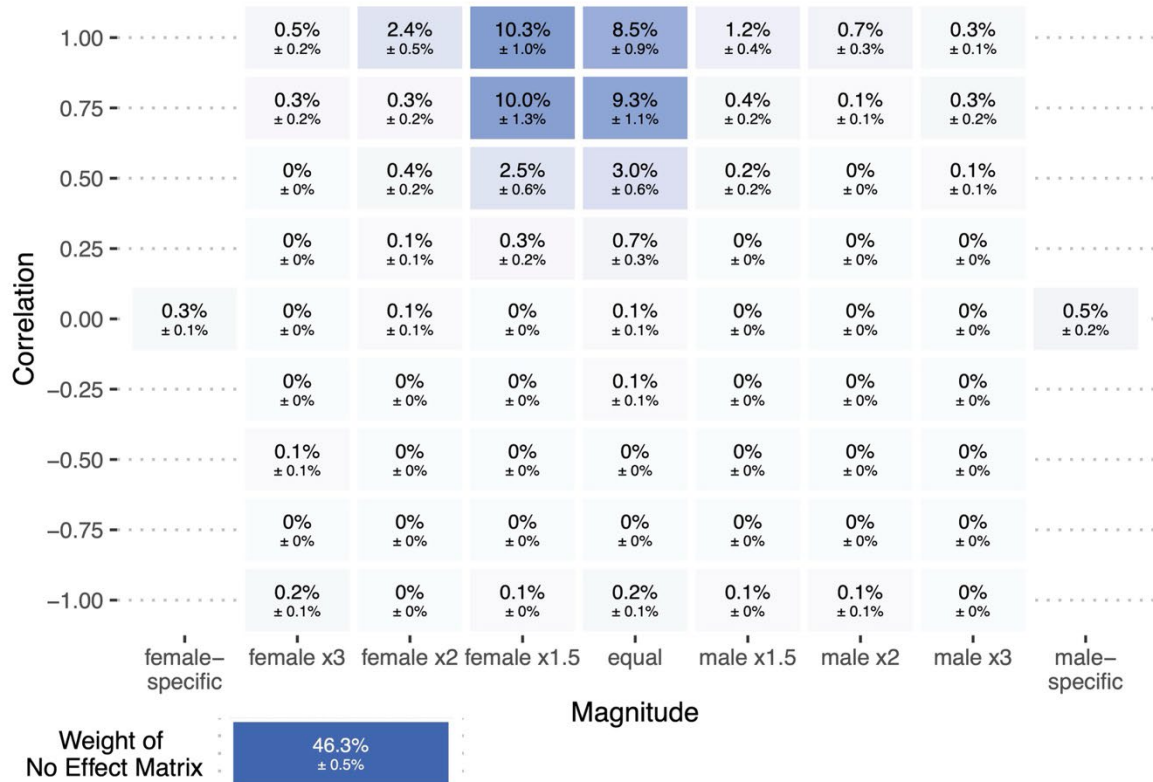
Covariance of Genetic Effects: Compact Representation



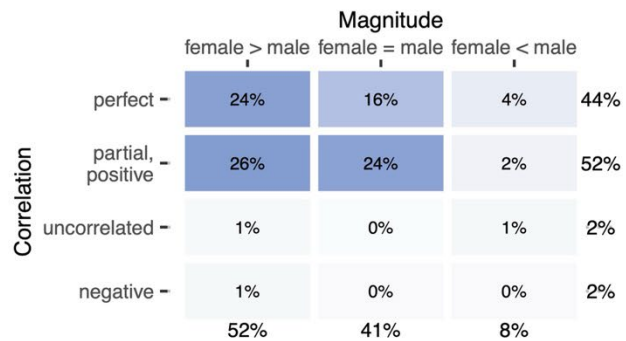
Data S26: Whole body fat mass

Whole body fat mass

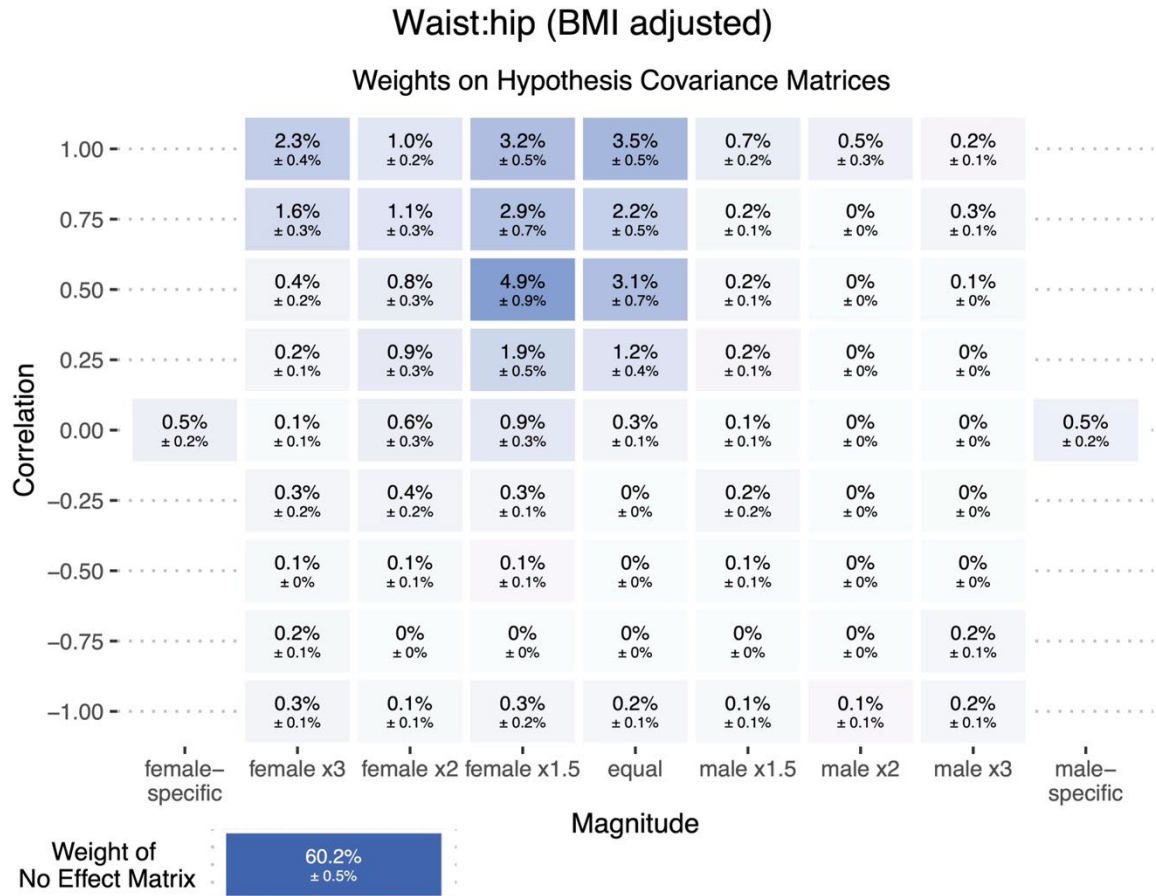
Weights on Hypothesis Covariance Matrices



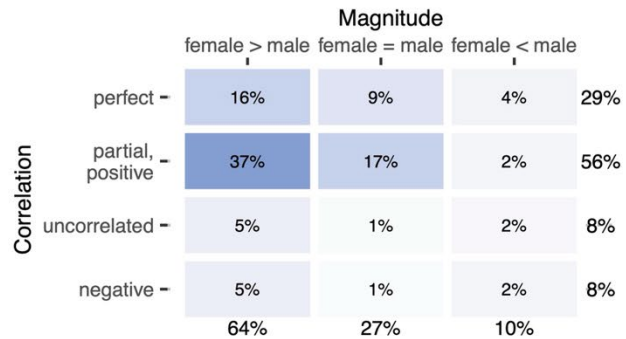
Covariance of Genetic Effects: Compact Representation



Data S27: Waist:hip (BMI adjusted)



Covariance of Genetic Effects: Compact Representation



References

1. Purcell, S., and Chang, C. (2020). PLINK 2.00 alpha.
2. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4, 7. 10.1186/s13742-015-0047-8.
3. Graffelman, J., and Moreno, V. (2013). The mid p-value in exact tests for Hardy-Weinberg equilibrium. *Stat Appl Genet Mol Biol* 12, 433–448. 10.1515/sagmb-2012-0039.
4. Graffelman, J., and Weir, B.S. (2016). Testing for Hardy-Weinberg equilibrium at biallelic genetic markers on the X chromosome. *Heredity (Edinb)* 116, 558–568. 10.1038/hdy.2016.20.
5. Wigginton, J.E., Cutler, D.J., and Abecasis, G.R. (2005). A note on exact tests of Hardy-Weinberg equilibrium. *Am J Hum Genet* 76, 887–893. 10.1086/429864.
6. 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., et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81, 559–575. 10.1086/519795.
7. Abbot, L., Bryant, S., Churchhouse, C., Ganna, A., Howrigan, D., Palmer, D., Neale, B., Walters, R., Carey, C., and The Hail Team (2017). GWAS Results. <http://www.nealelab.is/uk-biobank/>.
8. Bulik-Sullivan, B.K., Loh, P.-R., Finucane, H.K., Ripke, S., Yang, J., Patterson, N., Daly, M.J., Price, A.L., and Neale, B.M. (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* 47, 291–295. 10.1038/ng.3211.
9. McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R.S., Thormann, A., Flicek, P., and Cunningham, F. (2016). The Ensembl Variant Effect Predictor. *Genome Biol* 17, 122. 10.1186/s13059-016-0974-4.
10. Bulik-Sullivan, B.K., Finucane, H.K., Anttila, V., Gusev, A., Day, F.R., Loh, P.-R., Duncan, L., Perry, J.R.B., Patterson, N., Robinson, E.B., et al. (2015). An atlas of genetic correlations across human diseases and traits. *Nat Genet* 47, 1236–1241. 10.1038/ng.3406.
11. Choi, S.W., Mak, T.S.-H., and O'Reilly, P.F. (2020). Tutorial: a guide to performing polygenic risk score analyses. *Nat Protoc* 15, 2759–2772. 10.1038/s41596-020-0353-1.
12. Berisa, T., and Pickrell, J.K. (2016). Approximately independent linkage disequilibrium blocks in human populations. *Bioinformatics* 32, 283–285. 10.1093/bioinformatics/btv546.

13. 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., et al. (2015). A global reference for human genetic variation. *Nature* 526, 68–74. 10.1038/nature15393.
14. 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., et al. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 581, 434–443. 10.1038/s41586-020-2308-7.
15. Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, and D. (2002). The Human Genome Browser at UCSC. *Genome Res* 12, 996–1006. 10.1101/gr.229102.
16. 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., et al. (2011). The variant call format and VCFtools. *Bioinformatics* 27, 2156–2158. 10.1093/bioinformatics/btr330.
17. Pirastu, N., Cordioli, M., Nandakumar, P., Mignogna, G., Abdellaoui, A., Hollis, B., Kanai, M., Rajagopal, V.M., Parolo, P.D.B., Baya, N., et al. (2021). Genetic analyses identify widespread sex-differential participation bias. *Nat Genet* 53, 663–671. 10.1038/s41588-021-00846-7.
18. Kasimatis, K.R., Abraham, A., Ralph, P.L., Kern, A.D., Capra, J.A., and Phillips, P.C. (2021). Evaluating human autosomal loci for sexually antagonistic viability selection in two large biobanks. *Genetics* 217. 10.1093/genetics/iyaa015.
19. Schneider, V.A., Graves-Lindsay, T., Howe, K., Bouk, N., Chen, H.-C., Kitts, P.A., Murphy, T.D., Pruitt, K.D., Thibaud-Nissen, F., Albracht, D., et al. (2017). Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly. *Genome Res* 27, 849–864. 10.1101/gr.213611.116.
20. Morgulis, A., Coulouris, G., Raytselis, Y., Madden, T.L., Agarwala, R., and Schäffer, A.A. (2008). Database indexing for production MegaBLAST searches. *Bioinformatics* 24, 1757–1764. 10.1093/bioinformatics/btn322.
21. Zajitschek, S.R., Zajitschek, F., Bonduriansky, R., Brooks, R.C., Cornwell, W., Falster, D.S., Lagisz, M., Mason, J., Senior, A.M., Noble, D.W., et al. (2020). Sexual dimorphism in trait variability and its eco-evolutionary and statistical implications. *Elife* 9. 10.7554/eLife.63170.
22. CUERVO, J.J., and MØLLER, A.P. (1999). Phenotypic variation and fluctuating asymmetry in sexually dimorphic feather ornaments in relation to sex and mating system. *Biological Journal of the Linnean Society* 68, 505–529. 10.1111/j.1095-8312.1999.tb01186.x.
23. Reinhold, K., and Engqvist, L. (2013). THE VARIABILITY IS IN THE SEX CHROMOSOMES. *Evolution (N Y)* 67, 3662–3668. 10.1111/evo.12224.
24. Pomiankowski, A., and Møller, A.P. (1995). A resolution of the lek paradox. *Proc R Soc Lond B Biol Sci* 260, 21–29. 10.1098/rspb.1995.0054.

25. Kent, W.J. (2002). BLAT — The BLAST-Like Alignment Tool. *Genome Res* 12, 656–664. 10.1101/gr.229202.