A study of analytical strategies to include X-chromosome in variance heterogeneity analysis: evidence for trait–specific polygenic variance structure

Wei Q. Deng[1], Anette Kalnapenkis[2,3], Shihong Mao[4], Tõnu Esko[2,5], Reedik Mägi[2], Guillaume Paré[4,6], Lei Sun[†] [1,7]

[1]Department of Statistical Sciences, Faculty of Arts and Science, University of Toronto, Toronto, Ontario, M5S 3G3, Canada;

[2]Estonian Genome Center, University of Tartu, Tartu 51010, Estonia;

[3]Institute of Molecular and Cell Biology, University of Tartu, Tartu 51010, Estonia;

[4]Population Health Research Institute, Hamilton Health Sciences and McMaster University, Hamilton, Ontario, L8L 2X2, Canada;

[5]Program in Medical and Population Genetics, Broad Institute, Cambridge, USA

[6]Department of Pathology and Molecular Medicine, McMaster University, Hamilton, Ontario, L8N 4A6, Canada;

[7]Division of Biostatistics, Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, M5T 3M7, Canada.

[†]*Corresponding author*:

Lei Sun, Department of Statistical Sciences, 100 St George Street, University of Toronto, Toronto, Ontario, M5S 3G3, Canada. E-mail: sun@utstat.toronto.edu.

ABSTRACT

Genetic markers associated with *variance* of quantitative traits are considered promising candidates for follow-up including interaction analyses. However, as in studies of main effects, X-chromosome is routinely excluded from ‘whole-genome’ scans due to analytical challenges. Specifically, as males carry only one copy of X-chromosome, the inherent sex-genotype dependency could bias the trait-genotype association, through sexual dimorphism in quantitative traits with sex-specific means or variances. Here we investigate phenotypic variance heterogeneity associated with SNPs on X-chromosome and propose robust strategies. Among those, a generalized Levene’s test, adjusting for sex and sex-genotype interaction effects, has adequate power and remains robust to sexual dimorphism. An alternative sex-stratified approach via Fisher’s method is the most robust at the cost of slightly reduced power. We applied both methods to an Estonian study of gene expression quantitative trait loci (eQTL; n=841), and two complex trait studies of height, hip and waist circumferences, and body mass index (BMI) collected on Caucasians in UK Biobank (UKB; n=132,968) and multi-ethnic study of atherosclerosis (MESA; n=2,073). Consistent with previous eQTL findings on mean, we found some but not conclusive evidence for *cis* regulators being enriched for variance association. Individual SNP rs148191803 was X-chromosome-wide significant for waist circumference (*p*=2.4E-6) and suggestive for BMI (*p*=1.2E-5) in UKB but not in MESA. However, a permutation study based on MESA showed a trait-specific polygenic model whereby multiple X-chromosome loci collectively influence variance of height (λ_{GC}=1.14, *p*<1/100), calling for developments of methods to examine broad-sense heritability by incorporating variance loci and quantifying their sex-specific contributions.

INTRODUCTION

Several recent reports have examined autosomal genetic loci contributing to phenotypic variance (as opposed to mean) for a wide range of complex traits[1–3], and corresponding methodology development remains an active area of research[4–11]. One possible reason for such phenotypic variance and SNP genotype association, or variance heterogeneity, is that genotype-stratified variances of a trait differ in the presence of gene-gene (*G*x*G*) or gene-environment (*G*x*E*) interactions; both referred to as *G*x*E* hereinafter. For example, rs1358030 (*SORCS1*) was shown to interact with treatment type affecting HbA1c levels in Type 1 Diabetes subjects[12]. And indeed, in a proof-of-principle study where the treatment information was intentionally masked, the SNP was then demonstrated to be associated with variance of HbA1c[10]. Conversely, because direct *G*x*E* modeling may not be feasible in an initial whole-genome scan, the question was then raised as to whether SNPs having effects on the variance of a trait make good candidates for follow-up interaction testing[3]. For instance, rs7202116 (*FTO* as the nearest gene) was significantly associated with variance of BMI[2], and at the same locus, rs1121980 (*FTO)* showed evidence for a statistical interaction with physical activity influencing the mean of BMI[13,14]; it is worth noting that un-modeled interaction induces variance heterogeneity, but the causes of variance heterogeneity are multifaceted[8,10,15–18]. In practice, although it is possible that an interacting SNP has a stronger effect on variance than on mean, as in the case of rs12753193 (*LEPR*) interacting with BMI in the prediction of CRP levels in the absence of detectable main effect[1], a more powerful approach to selecting association candidates is to jointly evaluate their mean and variance effects[8,10,19,20].

Despite enthusiasm to discover SNPs with variance effects and the availability of statistical tests, variance heterogeneity has not been formally explored for SNPs on X-chromosome (XCHR). As in the conventional 'genome-wide' (mean) association studies[21], the reluctance to include XCHR is due to analytical challenges[21,22]. They range from technical difficulties in genotype calling to statistical complexities in imputation and association (e.g. model uncertainty involving random or skewed X-inactivation[23–26] and sex as a potential confounder). Solutions to overcome some of these challenges had been provided, but all in the context of genetic association analysis of main effects[25,27–30].

Here we focus on understanding the impact of the inherent sex-genotype dependency on variance heterogeneity association analysis, and when the trait of interest has sex-specific mean or variance values for males and females. In practice, sexual dimorphism is consistently observed. For example, based on the UK Biobank (UKB)[31] and Multi-Ethnic Study of Atherosclerosis[32] (MESA) data, height displays a sex-specific difference in mean, hip circumference differs in variance, while body mass index (BMI) and waist circumference contrast in both mean and variance between males and females (Figure 1). These empirical patterns of sexual dimorphism vary according to the underlying physiology of the trait, which might or might not be related to genes. Thus, association analyses of phenotypic mean or variance with XCHR SNPs could be biased if these potential sex-specific main or variance effects were not appropriately accounted for.

Figure 1 here

For an autosomal SNP, evaluating differences in phenotypic variance across the three genotype groups can be readily achieved by the classical Levene's test for variance heterogeneity[33]. SNPs with significant variance association *p*-values are then selected as likely candidates for follow-up interaction studies. However, the same strategy to prioritize SNPs on XCHR can be problematic, because sex-specific mean and variance differences could create spurious variance heterogeneity unrelated to the putative *G*x*E* interactions. Thus, the correct formulation of variance test is dependent on a proper formulation of sex effect with respect to both mean and variance.

In this paper, we explicitly model the possible sources of confounding related to sex, and propose two general testing strategies that strike a balance between power and robustness. Using extensive simulations, we demonstrate the danger of directly applying autosomal methods to XCHR that would otherwise be suitable for testing variance heterogeneity, and we conclude that special consideration for sex-genotype dependence must be made for XCHR to maintain correct type I error rates. Application studies include identifying SNPs associated with variances of height, BMI, hip and waist circumference using the UK Biobank and MESA data, as well as detecting loci associated with variance of expression quantitative traits using data from the Estonian Genome Center at the University of Tartu (EGCUT) cohort[34–36].

MATERIAL AND METHODS

Notation and model setup

Of interest is a quantitative trait *Y*, assumed to be (approximately) normally distributed or had been inversely transformed to resemble a normal distribution. Without loss of generality, consider the following linear model for the '*true'* association relationship between *Y* and a SNP,□

$$Y \; = \; \beta_0 + \beta_G G + \beta_s S + \beta_{GS} GS + \beta_E E + \beta_{GE} GE + \beta_{SE} SE + \beta_{GSE} GES + \varepsilon, \qquad (1)$$

where *G* denotes the SNP genotype (coded additively with respect to the number of the minor allele as 0, 1 and 2 for *bb, Bb* and *BB* as in convention[37]), *S* is the sex indicator variable (female

= 0 and male = 1), $E \sim N(0, 1)$ is a standardized continuous covariate following the classical *G-E* independence assumption[38], and the error term ε □ $N(0, 1)$ is independent of *G*, *S* and *E*. The minor allele frequency (MAF) of *G* is assumed to be the same for male and females; sex-specific MAF affects the naïve methods and we will return to this point in the Discussion section.

Under these assumptions, it is possible to identify autosomal SNPs potentially involved in *G*x*E* or high-order interactions, without having to measure *E* directly, through detecting phenotypic variance associated with *G* via the *working* model of $Y \sim G$. Note that the analytical context here is that direct *G*x*E* (or *G*x*G*) modeling may not be possible (e.g. *E* may not be known or measured precisely) or desirable (e.g. due to computational or multiple hypothesis testing concerns for whole-genome *G*x*G* scans). To see the rationale behind the working model, with the additional assumption of conditional independence between *E* and *S* conditional on *G*, one can show that the conditional variance of *Y* on *G* is,

$$\mathrm{Var}(Y|G = g) = (\beta_E + \beta_{GE}g)^2 + (\beta_S + \beta_{GS}g)^2\mathrm{Var}(S|G = g) + [(\beta_{SE} + \beta_{GSE}g)^2 + 2(\beta_E + \beta_{GE}g)(\beta_{SE} + \beta_{GSE}g)]E(S|G = g) + 1. \tag{2}$$

Since sex *S* is independent of *G* for an autosomal SNP, Pr(*S*|*G* = *g*) is constant across *g* = 0, 1 and 2, so are *E(S|G=g)* and *Var(S|G=g)*. Thus, if $\beta_{GS} = \beta_{GE} = \beta_{GSE} = 0$, expression (2) can be reduced to a constant with respect to *G*:

$$\mathrm{Var}(Y|G = g) = {\beta_E}^2 + {\beta_S}^2\mathrm{Var}(S|G = g) + [{\beta_{SE}}^2 + 2\beta_E\beta_{SE}]E(S|G = g) + 1 \tag{3}$$

$$= {\beta_E}^2 + {\beta_S}^2\mathrm{Var}(S) + [{\beta_{SE}}^2 + 2\beta_E\beta_{SE}]E(S) + 1. \tag{4}$$

Conversely, variation in Var(*Y*|*G*=*g*) across *G* suggests that at least some of the (un-modeled) interaction terms involving *G* (i.e. β_{GS} , β_{GE} and β_{GSE}) are non-zero. This was precisely the motivation behind the original idea[1] of using Levene's test to identify variance heterogeneity induced by the underlying but un-modeled *G*x*E* interaction.

X-chromosome (XCHR) specific challenges for variance tests

The same approach to draw similar conclusions for XCHR SNPs, however, is questionable, because Pr($S|G = g$) is no longer constant in G and expression (3) cannot be further reduced to (4). For example, under the X-inactivation coding of 0, 1 and 2 for the *bb*, *Bb* and *BB* genotypes in females and 0 and 2 for the *b* and *B* genotypes in males, the $G=1$ group contains only females. Similarly, under the no X-inactivation coding of 0, 1 and 2 for females and 0 and 1 for males, the $G=2$ group then contains only females. Thus, omitting the sex indicator S from the covariates can bias the conclusion through sexual dimorphism as seen in Figure 1.

Consider the simplest case of no interaction effects at all ($\beta_{GS} = \beta_{GE} = \beta_{SE} = \beta_{GSE} = 0$) nor environmental main effect ($\beta_E = 0$), but there is a sex main effect ($\beta_S \neq 0$, i.e. the sex-stratified phenotypic means differ between males and females), then expression (3) is reduced to

$$\mathrm{Var}(Y|G = g) = {\beta_S}^2\mathrm{Var}(S|G = g) + 1. \qquad (5)$$

Thus, in the absence of any interactions that involve G, there is a spurious phenotypic variance heterogeneity across levels of G through a non-zero sex main effect (β_S), or through a sex-environment interaction effect (β_{SE}) if present as in equation (3). Severity of the confounding depends on the discrepancy between the two sex-stratified trait distributions (real data in Figure 1 and conceptual data in Figures 2 A-D), as well as on the strength of correlation between sex and the observed genotype, which in turn depends on the MAF and proportions of males and females in a sample (details in Supplemental Data).

Figure 2 here

To avoid spurious variance heterogeneity signals, alternative approaches are needed to quantify variance differences induced by GxE or higher order interactions involving G. To this end, it is important to appropriately define the null hypothesis of variance homogeneity that corresponds to an absence of phenotypic variance associated with genotype while allowing for variance (and mean) to differ between males and females (Figures 2 A-D).

X-chromosome (XCHR) variance heterogeneity tests

Here we consider various analytical strategies to assess phenotypic variance associated with genotypes of XCHR SNPs, including naïve methods that directly apply the original Levene's test to different genotype groups, and alternative approaches that utilize a generalized Levene's test derived from a two-stage regression framework[20,39].

Naïve methods: apply Levene's test to three or five genotype groups

The original Levene's test for variance heterogeneity treats an autosomal genotype G as a categorical variable[33,39] and examines any variance difference in trait Y amongst the three possible genotype groups. A direct application to XCHR, however, is problematic. Because sex S is inherently correlated with XCHR G, so any potential correlation between S and Y (e.g. as observed in human height) would create the classic case of confounding. To see this, consider the null situation where G is *not* associated with *variance* of Y as in the top panel of Figure 2. Assume the X-inactivation coding of G was used (same conclusion for the no X-inactivation coding), the Bb group contains only females and its variance would be the same as σ_f^2, reflected by variance of the orange curve in the figure. In contrast, the other two groups (bb+b and BB+B) contain both males and females, and their respective variance values involve both the orange and blue curves depending on sex-specific means (μ_m and μ_f) and variances (σ_m^2 and σ_f^2), as well the proportion of males in each group. Thus, in the presence of sexual dimorphism (either in mean-Figure 2B or variance-Figure 2C or both-Figure 2D), there would be spurious variance heterogeneity resulting in increased false positive rates (also confirmed by empirical results).

As an alternative, one may be tempted to treat each genotype and sex combination as one group, resulting in a total of 5 groups. Indeed, this five-group strategy does not induce spurious association in the presence of sex-specific mean effect ($\mu_f \neq \mu_m$ as in Figure 2B). However, it is not difficult to see that the problem remains when there is a sex-specific variance effect ($\sigma_m^2 \neq \sigma_f^2$ as in Figures 2C or 2D).

Fisher's method: combine sex-stratified Levene's test

Sex-stratified analysis provides a practical strategy whereby variance heterogeneity is assessed separately in males (two-group Levene's test) and females (three-group Levene's test). Fisher's method can then be used to combine the two *p*-values. Note that Leven's test statistic is asymptotically $\chi^2_{\#groups-1}$ distributed without apparent 'direction of effect', so the traditional meta-analysis that combines the weighted (directional) Z-values for testing mean effect is not applicable here. Though sex-stratified analysis does not allow direct *G*x*S* modeling, it is robust to various forms of sexual dimorphism as seen in Figures 2B-D, and it does not require the knowledge of X-inactivation status.

Model-based generalized Levene's test: account for sex-specific mean and variance effects via two-stage regression models.

Since the null hypothesis is defined in terms of phenotypic variance heterogeneity induced by (un-modeled) *G*x*E* interactions while allowing for sexual dimorphism (Figures 2B-D), a preferred method should explicitly account for the effect of sex on the phenotype of interest. To this end, we consider the generalized Levene's test that established a flexible two-stage regression framework[20,39]. In essence, stage one regresses *Y* on *G* and obtains the absolute residual *d* (i.e. absolute of deviation between observed and model fitted *Y* values). Stage two regresses *d* on *G* again, and testing the slope was shown to be equivalent to evaluating variance heterogeneity in *Y* associated with *G* because the expectation of *d* linearly depends on variance of *Y* [20,39].

The generalized Levene's test has been used to study autosomal SNPs with more complex data structures including genotype group uncertainty (e.g. imputed SNPs) or sample dependency (e.g. correlated family members)[20]. For XCHR analysis, the implementation requires additional care. For example, it is not immediately clear if *S* (or *G*x*S*) should be included in both stages. For a comprehensive evaluation, we consider all combinations of the following two-stage models:

Stage One: Mean models,

$$Y{\sim}\alpha_0+\alpha_G G\ (\mathbf{M1}),$$

$$Y{\sim}\alpha_0+\alpha_G G+\alpha_s S\ \ (\mathbf{M2}),$$

$$Y{\sim}\alpha_0+\alpha_G G+\alpha_s S+\alpha_{GS} GS\ (\mathbf{M3}).$$

Stage Two: Variance models,

$$d{\sim}\gamma_0+\gamma_G G\ (\mathbf{V1}),$$

$$d{\sim}\gamma_0+\gamma_G G+\gamma_s S\ (\mathbf{V2}),$$

$$d{\sim}\gamma_0+\gamma_G G+\gamma_s S+\gamma_{GS} GS\ (\mathbf{V3}).$$

The models in stage one are only used to calculate residuals, using either the traditional ordinary least squares (OLS) or the recommended least absolute deviations (LAD); LAD is more robust to data with asymmetric distributions or low genotype counts in a specific group[20,40]. The goal of this stage is to remove any *Mean* effects associated with the covariates included in the model (i.e. *G*, *S* or *G*x*S*), thus denoted as M1, M2 or M3.

Test for *Variance* heterogeneity is achieved in stage two (V1, V2 or V3), by testing $H_o: \gamma_G = 0$ or $H_o: \gamma_G = \gamma_{GS} = 0$ via the standard regression *F*-test, where model is fitted using OLS for independent samples or generalized least square for dependent samples.

The model-based regression approach includes a total of nine M+V two-stage models, and V3 also allows a two degrees of freedom (d.f.) test (Table S1). Based on the earlier discussion, it is expected that mean modeling strategies omitting *S* (i.e. M1) would be sensitive to sex-specific mean effect (e.g. Figures 2B or 2D). Meanwhile, variance testing strategies omitting *S* (i.e. V1) are anticipated to be sensitive to sex-specific variance effect (e.g. Figures 2C or 2D). For completeness of our empirical validation, we first examined all 12 testing strategies in simulation studies then focused on the robust approaches in applications.

Simulation studies

A sample of 2,500 females and 2,500 males were simulated, and the MAF was fixed at 0.2; other sample sizes and MAFs led to qualitatively similar results. Note that although *G* could be coded assuming X-inactivation or no X-inactivation, the two types of coding are generally highly correlated leading to similar association results[29]. Thus, for a more focused study here the genotype was simulated assumed no X-inactivation. The number of simulated replicates was 10,000 so that estimates of the empirical Type I Error (T1E) rates within ±0.5% of the nominal rate of 5% were considered satisfactory.

A joint mean and variance test can be more powerful than testing for variance heterogeneity alone, but the power of the joint test depends on the individual components[10]. Therefore, here we focus on comparing the different variance-testing strategies as outlined above, recommending the most robust yet powerful method that is also suitable for the joint location-scale test.

Simulations for T1E evaluation - design I based on model (1)

The genotype-phenotype relationship was generated according to model (1), where the environmental variable $E \sim N(0, 1)$ was used in generating observed phenotypic values but assumed not being available for the actual association analysis. The null scenarios were defined by the absence of interaction effects for *G*x*E* and *G*x*E*x*S*, so the quantitative trait for each null scenario was generated assuming $\beta_{GE} = \beta_{GES} = 0$ in model (1). A SNP could have a *G* main effect, but it does not affect the phenotypic variance of interest which is induced by un-modeled β_{GE} and β_{GES} in the working model, so $\beta_G = 0$ without loss of generality. Note that the naïve variance methods could also pick up a non-zero *G*x*S* interaction effect if $\beta_{GS} \neq 0$, but β_{GS} itself in fact can be directly tested because gender information is routinely collected (or robustly inferred from the available genotype data). Thus, β_{GS} is not related to the variance heterogeneity of interest here and was set to be zero. For the remaining parameters, without loss of generality, β_0= 0, β_E = 0 or 0.5, β_S= 0 or 0.5, and β_{SE} = 0, -0.25 or 0.25, giving a total of 12 scenarios. They roughly fall into four categories, corresponding to the four conceptual sex-stratified distributions as shown in Figures 2A-D. For example, sexual dimorphism was introduced via β_S and β_{SE}, where a none-zero β_S allows for *sex*-specific mean effect (Figures 2B and 2D) while a non-zero β_{SE} allows for *sex*-specific variance effect (Figures 2C and 2D). Note that both β_S and

β_{SE} are independent of the *genotype*-specific variance effect to be identified, which is absent in the null cases.

Simulations for T1E evaluation – design II based on sex-stratified mean and variance

The null scenarios based on model (1) may not fully capture the extremes of sexual dimorphism, thus we further simulated trait values directly according to sex-specific distributions using means (μ_m and μ_f) and variances (σ_m^2 and σ_f^2) that mimic the values observed in inverse-normally transformed BMI, height, hip and waist circumference from MESA (Table S2). The simulated traits, generated independent of any genotypes, were then tested for variance association with genotypes of 12,206 XCHR SNPs from the MESA dataset, after filtering by a minimum count of 30 observations in the five sex-genotype stratified groups as variance test is sensitive to small group size.

Simulations for power study

Only strategies with satisfactory T1E control were considered for power evaluation. We focused on model-based design I where power directly depends on the size of *G*x*E* and *G*x*E*x*S* interaction effects and has a clearer genetic interpretation than design II. Under model (1), β_{GE} was varied from 0 to 0.2 with a 0.025 incremental increase, and combined with a possible three-way interaction β_{GES} of 0 or 0.1. Other parameter values were the same as in the null case with the exception that β_S = β_E = 0 were not considered for a more focused study of power. That is, β_0= 0, β_G = β_{GS} = 0, β_S = β_E = 0.5, and β_{SE} = 0, -0.25 or 0.25. In total, there were 54 scenarios.

Applications

Robust variance testing strategies for X-chromosome SNPs that also had reasonable power performance were then applied to real data. Only reportedly unrelated and ethnically Caucasian individuals were included, and diabetic individuals were excluded based on electronic medical records in the UK Biobank[31], and based on blood glucose level greater than 7 mmol/L in MESA[32]. All quantitative traits were quantile-normally transformed to avoid 'scale-effect' where the variance values tend to be proportional to mean values[1,2].

The significance level for discovery was set at a nominal level of 5% with Bonferroni correction, depending on the total number of XCHR SNPs examined in each application. Further, the proportion of truly variance-associated variants was estimated using the method proposed by Storey and Tibshirani[41].

The UK Biobank (UKB) data[31]

Available genotyped XCHR SNPs were filtered based on whether they were in pseudo autosomal region and a minimal sample count of 30 across the five sex-genotype groups. In total, 7,344 XCHR SNPs on the Caucasian sample (71,452 females and 61,516 males) were analyzed, and the XCHR-wide significance level was 6.8E-6.

The Multi-Ethnic Study of Atherosclerosis (*MESA*) *data*[32].

The genotype data in MESA, available from dbGap (Study accession: phs000209.v10.p2), were filtered similarly as the UKB data. In total, 12,206 XCHR SNPs on the Caucasian sample (1,003 females and 1,070 males) were analyzed, and the XCHR-wide significance level was 4.1E-6. We did not perform a multi-ethnic analysis with all ethnicities combined. Instead, we focused on the Caucasian subset and used it to corroborate findings from the UKB data.

Estonian Genome Center at the University of Tartu (EGCUT) cohort[34–36]

We sought to discover XCHR SNPs influencing the variance of expression traits, as variability of gene expression has been suggested to be associated with genetic variants on autosomes[7]. The recommended strategies were applied to a sample of 413 male and 421 female Estonians across 648 gene expression traits that had gone through standard quality control procedures and further inversely normal transformed. After filtering using the same criteria as the UKB data, 4,034 XCHR SNPs were analyzed for variance association with each of the 648 gene expression traits, resulting in a total number of 2,614,032 tests and a global significance level of 1.9E-8.

RESULTS

Simulation studies

As expected, the naïve Levene's test with either a three-level factor *G* factor or a five-level *GxS* factor grossly overestimated the number of false positives in almost all scenarios except for when β_{E}, β_{GS} and β_{SE} were all set to zero or the in the absence of any sexual dimorphism (Table 1).

For generalized two-stage Levene's tests, good choices of the mean model in stage one and variance test in stage two should remove any effect from sex to avoid inflating the test statistics. Thus, as expected, any strategies involving M1 or V1 had T1E issues, where the degrees of departure from the nominal T1E rate varied according to sizes of the unadjusted sex mean or variance effects (Table 1). The remaining strategies appear to have reasonably controlled T1E rates and are underlined in Table 1. However, when considering design II where sexual dimorphism was more extreme, only M2V3.2 and M3V3.2 have good T1E control and behave quite similarly (Table S3).

The sex-stratified approach, as expected, gave correct T1E rates in females and males separately, and subsequently in the combined sample via Fisher's methods, under both design I (Table 1) and design II (Table S3).

In terms of statistical power among testing strategies with reasonable T1E control (M2V3.2, M3V3.2 and Fisher), M2V3.2 and M3V3.2 are nearly identical and had slightly better power than Fisher's method across most of the scenarios considered (Figures S1).

The reason for performance similarity between M2V3.2 and M3V3.2 is because the model was generated (and correctly modeled) under the assumption of no X-inactivation and $\beta_G = 0$. Interestingly, when there is a strong genotypic main effect ($\beta_G \neq 0$), an increased variance in the female homozygote *Bb* group could be observed as a result of unknown X-inactivation[42]. Indeed, additional simulation studies confirmed that variance heterogeneity *p*-values given by M2V3.2 were influenced by X-inactivation while Fisher's method and M3V3.2 remained consistent (Figure S2, Table S4). Thus, we recommend the M3V3.2 model-based approach and the complementary sex-stratified Fisher's method, which were then applied to the three application datasets.

Applications

XCHR-wide analysis of waist circumference showed that the recommended M3V3.2 and Fisher's tests indeed have good T1E control (Figure 3); similar conclusions were drawn based on results of other traits (Figures S3-5, respectively, for BMI, height and hip circumference). In UKB (Table S5), rs148191803 (*MED14*; in a region known to escape X-inactivation) was found to be associated with waist circumference, based on the M3V3.2 test (p = 2.08E-06) and Fisher's method (p = 2.4E-06), at the XCHR-wide significance level (p < 6.8E-06), and the same SNP was associated with variance of BMI (M3V3.2 p = 7.01E-06; Fisher p = 1.20E-05; Table S6). However, no SNPs in the *MED14* locus had waist circumference or BMI variance association p-values less than 0.05 in MESA.

The association of rs2023750 at *TBL1X* with BMI was suggestive (M3V3.2 p = 5.30E-04; Fisher p = 2.61E-04, Table S5) in UKB, and at the same gene locus, rs2521580 was also suggestively associated with BMI (M3V3.2 p = 2.85E-03; Fisher p = 9.04E-04) and waist circumference (M3V3.2 p = 9.77E-03; Fisher p = 5.75E-04) in MESA (Table S6).

Although there were no additional X-chromosome-wide significant SNPs in UKB or MESA, the overall distributions of the p-values suggest enrichment of associated variants for some of the traits. In Figure S4 for example, the estimated genomic lambda λ_{GC} based on the M3V3.2 variance test for height was 1.028 and 1.194, respectively in UKB and MESA, and it was 1.014 and 1.186 based on Fisher's method. The estimated proportion of truly associated SNPs also suggested signal enrichment (Table S7).

To benchmark the observed estimates, a permutation analysis was conducted using the individual-level MESA data available to us. Each quantitative trait under the study was permutated within the two sex strata, independently, 100 times. For each permutated dataset, Fisher's method and M3V3.2 were applied and the corresponding λ_{GC}values were then calculated (Figure 4). For height, the permuted values, as expected, centered around λ_{GC}= 1. The observed λ_{GC}value, based on either the M3V3.2 or Fisher's method was larger than all 100 null values (Figure 4), supporting the apparent enrichment of XCHR SNPs associated with variance of height. The same conclusion holds, but to a lesser extent for waist circumference. However, for hip circumference and BMI, the observed estimates were not visibly different from the null estimates obtained from the permuted datasets. Similar observations were made based on the estimated proportion of truly associated SNPs (Figure S6).

For the eQTL analysis, we observed various forms of sexual dimorphism in expression traits. In total, 182 out of the 648 expression traits had p <0.05 based on either a t-test of equality of means or an F-test for equality of variance between the two sexes (Figure S7). Among the eQTLs, the top five variance-associated SNPs belonged to three genes, *FTX*, *PLAC1* and *TEX11* (Figures S8-10), but no SNPs passed the strict Bonferroni correction at p < 1.9E-8 (Table S8). There was no apparent enrichment of association globally over all SNP-expression 2,614,032 (= 648*4,034) p-values (Figures S11-12). However, upon further investigation based on stratifying SNPs and gene expression pairs according to whether they were *cis* or *trans* acting (using a physical distance of 5Mbps from the start and the end of the gene for each expression trait), we found that the estimated proportion of truly associated SNP-expression pairs appear to be slightly higher for SNPs in *cis* (0.006 and 0.013 based on p-values of Fisher's and M3V3.2 methods, respectively), as compared to those in *trans* (0 based on either). The estimated lambda control was 0.996 and 1.003 for *cis*-acting pairs using Fisher's and M3V3.2 methods, respectively, and both at 0.99 for SNP-expression pairs in *trans*, suggesting additional studies are needed to establish convincing evidence for enrichment of variance-associated eQTLs.

DISCUSSION

This work was motivated by the recent call to include X-chromosome (XCHR) in 'whole-genome' scans[21], as well as the recent development of identifying autosomal SNPs associated with phenotypic variance[1,20]. To pave the way for future XCHR-wide study of variance heterogeneity and subsequent joint location-scale test[10], we examined a catalogue of analytical strategies and recommended two robust and power approaches. We emphasize the importance of recognizing sex as an inherent confounder in analyzing XCHR variants that contribute to phenotypic variance heterogeneity, particularly for traits displaying sexual dimorphism with either sex-specific means or variances, or both; this also holds for the traditional association analysis of XCHR variants studying their effects on phenotypic mean[22].

Between the two recommended strategies, Fisher's method to combine sex-specific Levene's *p*-values is intuitive and the most robust, but at the cost of slightly reduced power. Through exploiting the recently proposed generalized Levene's test based on a two-stage regression approach, the model-based M3V3.2 test can directly account for sex main effect as well as *G*x*S* interaction effect. The model-based regression testing strategy also allows flexible adjustments for other potential confounder such as principal components[43]. Thus, we recommend in practice to apply both methods to complement each other.

The naïve strategies that directly test for variance heterogeneity across either the classical three genotype groups or the sex-stratified five groups are inadequate with grossed inflated T1E rates in the presence of any sexual dimorphism (Table 1). Note that if the status of X-inactivation were known *a priori*, a non-additive variance model (VNA) in stage two may be considered:

$$d \sim \gamma_0 + \gamma_{G1} G1 + \gamma_{G2} G2 \ (\textbf{NAV1}),$$

$$d \sim \gamma_0 + \gamma_{G1} G1 + \gamma_{G2} G2 \ + \gamma_s S \ (\textbf{NAV2}),$$

$$d \sim \gamma_0 + \gamma_{G1} G1 + \gamma_{G2} G2 \ + \gamma_s S + \gamma_{G1S} G1S \ (\textbf{NAV3}),$$

where $G1$ and $G2$ are indicator variables for the *Bb* and *BB*+*B* groups under X-inactivation, respectively, or alternatively for the *Bb*+*B* and *BB* groups without X-inactivation. Under this representation, Lev3 is equivalent to M1VNA1 in which the main effect of sex is not account for in stage one; while Lev5 is equivalent to the M3VNA3 model-based testing of $\gamma_{G1} = \gamma_{G2} = \gamma_S = \gamma_{G1S} = 0$ in which variance heterogeneity due to sex γ_S is erroneously being tested. These observations clearly revealed the source of bias inherent in the naïve methods.

The additive coding for *G* in stage one is believed to sufficiently capture the genetic main effect[37], while it may not be the case for analysis of variance. The ambiguous genotype grouping under unknown X-inactivation status adds another layer of complexity for non-additive variance models (Figure S13). In fact, the choice of reference allele coding matters in XCHR when the mean association model does not include the *G*x*S* interaction terms as shown recently[29]. This has direct consequences for XCHR variance testing since a variance difference in the female homozygote group could be observed as a result of a strong marginal effect coupled with unknown X-inactivation[42]. Additional simulation results under model (1) with a non-zero genetic main effect suggested that *p*-values derived from M2V3.2 varied according to the underlying X-inactivation status (Table S4). Though in applications, the genetic main effect would have to be extremely large for the M2V3.2 *p*-values to lead to different conclusions. We also note that X-inactivation and no X-inactivation lead to similar (mean) association results if the *G*x*S* interaction term is included in the model, which explains the consistent performance of M3V3.2 irrespective of the X-inactivation status.

Since variance testing requires larger sample size than mean testing, detecting individual variance signals that are significant at the XCHR-wide or genome-wide level requires studies of very large size that might only be viable through meta-analysis. Meta-analyses of variance heterogeneity[4] for XCHR variants can be conducted in parallel to that of a single study incorporating the analytical strategies proposed for autosomal variants[4].

Similar to a polygenic model proposed for association studies of main effects, it is possible that a large proportion of genetic variants, though not individually detectable, could collectively contribute to variance heterogeneity in certain complex traits[44,45]. The permutated genomic control values showed a clear enrichment and point to a possible XCHR polygenic inheritance model for variance of height, which suggests that height could be potentially enriched for gene-environment interactions. Some have suggested that X-linked genes contribute to the sex-specific architecture of complex traits[46], yet the amount of contribution from XCHR SNPs involved in possible *G*x*E* or higher-order interactions is unclear. On that note, it is interest to point out that a sex-stratified approach could reveal sex-specific enrichment of XCHR SNPs associated with phenotypic variance of, for example, waist circumference in males (Figure 4). It has been reported genetic loci can have gender specific marginal effects on traits such as height[47] and HDL[48], i.e. *G*x*S* interactions. Results from this study call for new developments of the broad-sense heritability estimation methods that can incorporate variance loci as well as quantify their contributions to sex-specific heritability.

SUPPLEMENTAL DATA

Supplemental data include 13 figures, 8 tables, and theoretical derivations.

ACKNOWLEDGEMENTS

The authors thank Professor Radu V. Craiu and Mr. Bo Chen for helpful suggestions and critical reading of the original version of the paper. The authors would like to thank Silva Kasela for her help on optimizing the eQTL analysis. We are thankful to all the participants of the Multi-Ethnic Study of Atherosclerosis, UK Biobank, Estonian Genome Center at the University of Tartu cohort. This research was funded by the Canadian Institutes of Health Research (CIHR, 201309MOP-310732-G-CEAA-117978) and the Natural Sciences and Engineering Research Council of Canada (NSERC, 250053-2013) to LS. WQD is supported by NSERC Alexander Graham Bell Canada Graduate Scholarship and Ontario Graduate Scholarship.

WEB RESOURCES

An implementation of our methods is provided as an open-source and user-friendly R package available on github (https://github.com/WeiAkaneDeng/Xvarhet).

REFERENCES

1. Paré, G., Cook, N.R., Ridker, P.M., and Chasman, D.I. (2010). On the use of variance per genotype as a tool to identify quantitative trait interaction effects: a report from the Women's Genome Health Study. PLoS Genet. *6*, e1000981.

2. Yang, J., Loos, R.J.F., Powell, J.E., Medland, S.E., Speliotes, E.K., Chasman, D.I., Rose, L.M., Thorleifsson, G., Steinthorsdottir, V., Mägi, R., et al. (2012). FTO genotype is associated with phenotypic variability of body mass index. Nature.

3. Shungin, D., Deng, W.Q., Varga, T. V, Luan, J., Mihailov, E., Metspalu, A., Morris, A.P., Forouhi, N.G., Lindgren, C., Magnusson, P.K.E., et al. (2017). Ranking and characterization of established BMI and lipid associated loci as candidates for gene-environment interactions. PLOS Genet. *13*, e1006812.

4. Struchalin, M. V, Dehghan, A., Witteman, J.C., van Duijn, C., and Aulchenko, Y.S. (2010). Variance heterogeneity analysis for detection of potentially interacting genetic loci: method and its limitations. BMC Genet. *11*, 92.

5. Deng, W.Q., and Paré, G. (2011). A fast algorithm to optimize SNP prioritization for gene-gene and gene-environment interactions. Genet. Epidemiol. 1–10.

6. Aschard, H., Chen, J., Cornelis, M.C., Chibnik, L.B., Karlson, E.W., and Kraft, P. (2012). Inclusion of gene-gene and gene-environment interactions unlikely to dramatically improve risk prediction for complex diseases. Am. J. Hum. Genet. *90*, 962–972.

7. Hulse, A.M., and Cai, J.J. (2013). Genetic variants contribute to gene expression variability in humans. Genetics *193*, 95–108.

8. Cao, Y., Wei, P., Bailey, M., Kauwe, J.S.K., and Maxwell, T.J. (2014). A versatile omnibus test for detecting mean and variance heterogeneity. Genet. Epidemiol. *38*, 51–59.

9. Deng, W.Q., Asma, S., and Paré, G. (2014). Meta-analysis of SNPs involved in variance heterogeneity using Levene's test for equal variances. Eur. J. Hum. Genet. *22*, 427–430.

10. Soave, D., Corvol, H., Panjwani, N., Gong, J., Li, W., Boëlle, P.Y., Durie, P.R., Paterson, A.D., Rommens, J.M., Strug, L.J., et al. (2015). A Joint Location-Scale Test Improves Power to Detect Associated SNPs, Gene Sets, and Pathways. Am. J. Hum. Genet. *97*, 125–138.

11. Hong, C., Ning, Y., Wei, P., Cao, Y., and Chen, Y. (2016). A semiparametric model for vQTL mapping. Biometrics.

12. Paterson, A.D., Waggott, D., Boright, A.P., Hosseini, S.M., Shen, E., Sylvestre, M.P., Wong, I., Bharaj, B., Cleary, P.A., Lachin, J.M., et al. (2010). A genome-wide association study identifies a novel major locus for glycemic control in type 1 diabetes, as measured by both A1C and glucose. Diabetes *59*, 539–549.

13. Qi, L., Brage, S., Sharp, S.J., Sonestedt, E., Demerath, E., Ahmad, T., Mora, S., Kaakinen, M., Sandholt, C.H., Holzapfel, C., et al. (2011). Physical Activity Attenuates the Influence of FTO Variants on Obesity Risk□: A Meta-Analysis of 218 , 166. PLoS Med. *8*,.

14. Ahmad, S., Rukh, G., Varga, T. V., Ali, A., Kurbasic, A., Shungin, D., Ericson, U., Koivula, R.W., Chu, A.Y., Rose, L.M., et al. (2013). Gene x Physical Activity Interactions in Obesity: Combined Analysis of 111,421 Individuals of European Ancestry. PLoS Genet. *9*, 1–9.

15. Paré, G., Mehta, S.R., Yusuf, S., Anand, S.S., Connolly, S.J., Hirsh, J., Simonsen, K., Bhatt, D.L., Fox, K. a a, and Eikelboom, J.W. (2010). Effects of CYP2C19 genotype on outcomes of clopidogrel treatment. N. Engl. J. Med. *363*, 1704–1714.

16. Sun, X., Elston, R., Morris, N., and Zhu, X. (2013). What is the significance of difference in phenotypic variability across SNP genotypes? Am. J. Hum. Genet. *93*, 390–397.

17. Dudbridge, F., and Fletcher, O. (2014). Gene-Environment Dependence Creates Spurious Gene-Environment Interaction. Am. J. Hum. Genet. *95*, 301–307.

18. Wood, A.R., Tuke, M.A., Nalls, M.A., Hernandez, D.G., Bandinelli, S., Singleton, A.B., Melzer, D., Ferrucci, L., Frayling, T.M., and Weedon, M.N. (2014). Another explanation for apparent epistasis. Nature *514*, E3–E5.

19. Aschard, H., Hancock, D.B., London, S.J., and Kraft, P. (2011). Genome-wide meta-analysis of joint tests for genetic and gene-environment interaction effects. Hum. Hered. *70*, 292–300.

20. Soave, D., and Sun, L. (2017). A generalized Levene's scale test for variance heterogeneity in the presence of sample correlation and group uncertainty. Biometrics *73*, 960–971.

21. Wise, A.L., Gyi, L., and Manolio, T.A. (2013). EXclusion: Toward integrating the X chromosome in genome-wide association analyses. Am. J. Hum. Genet. *92*, 643–647.

22. König, I.R., Loley, C., Erdmann, J., and Ziegler, A. (2014). How to Include Chromosome X in Your Genome-Wide Association Study. Genet. Epidemiol. *38*, 97–103.

23. Carrel, L., and Willard, H.F. (2005). X-inactivation profile reveals extensive variability in X-linked gene expression in females. Nature *434*, 400–404.

24. Ross, M.T., Grafham, D. V, Coffey, A.J., Scherer, S., McLay, K., Muzny, D., Platzer, M., Howell, G.R., Burrows, C., Bird, C.P., et al. (2005). The DNA sequence of the human X chromosome. Nature *434*, 325–337.

25. Wang, J., Yu, R., and Shete, S. (2014). X-Chromosome Genetic Association Test Accounting for X-Inactivation, Skewed X-Inactivation, and Escape from X-Inactivation. Genet. Epidemiol. *38*, 483–493.

26. Tukiainen, T., Villani, A.-C., Yen, A., Rivas, M.A., Marshall, J.L., Satija, R., Aguirre, M., Gauthier, L., Fleharty, M., Kirby, A., et al. (2017). Landscape of X chromosome inactivation across human tissues. Nature *550*, 244–248.

27. Clayton, D.G. (2009). Sex chromosomes and genetic association studies. Genome Med. *1*, 110.

28. Clayton, D. (2008). Testing for association on the X chromosome. Biostatistics *9*, 593–600.

29. Chen, B., Radu, Craiu, V., and Sun, L. (2017). Bayesian Model Averaging for the X-Chromosome Inactivation Dilemma in Genetic Association Study. arXiv.

30. Peter F Hickey, and Bahlo, M. (2011). X Chromosome Association Testing in Genome Wide Association Studies. Genet. Epidemiol. *670*, 664–670.

31. Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., et al. (2015). UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. PLoS Med. *12*, 1–10.

32. Bild, D.E., Bluemke, D.A., Burke, G.L., Detrano, R., Diez Roux, A. V., Folsom, A.R., Greenland, P., Jacobs, D.R., Kronmal, R., Liu, K., et al. (2002). Multi-Ethnic Study of Atherosclerosis: Objectives and design. Am. J. Epidemiol. *156*, 871–881.

33. Levene, H. (1960). Robust tests for equality of variances. In Contributions to Probability and

Statistics: Essays in Honor of Harold Hotelling. In In Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling., I. Olkin, ed. (Stanford University Press), pp. 278–292.

34. Leitsalu, L., Haller, T., Esko, T., Tammesoo, M.L., Alavere, H., Snieder, H., Perola, M., Ng, P.C., Mägi, R., Milani, L., et al. (2015). Cohort profile: Estonian biobank of the Estonian genome center, university of Tartu. Int. J. Epidemiol. *44*, 1137–1147.

35. Westra, H.J., Peters, M.J., Esko, T., Yaghootkar, H., Schurmann, C., Kettunen, J., Christiansen, M.W., Fairfax, B.P., Schramm, K., Powell, J.E., et al. (2013). Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat. Genet. *45*, 1238–1243.

36. Metspalu, A. (2004). The Estonian Genome Project. Drug Dev. Res. *62*, 97–101.

37. Hill, W.G., Goddard, M.E., and Visscher, P.M. (2008). Data and theory point to mainly additive genetic variance for complex traits. PLoS Genet. *4*,.

38. Lindström, S., Yen, Y.-C., Spiegelman, D., and Kraft, P. (2009). The impact of gene-environment dependence and misclassification in genetic association studies incorporating gene-environment interactions. Hum. Hered. *68*, 171–181.

39. Gastwirth, J.L., Gel, Y.R., and Miao, W. (2009). The Impact of Levene’s Test of Equality of Variances on Statistical Theory and Practice. Stat. Sci. *24*, 343–360.

40. Hines, W.G., and Hines, R.J. (2000). Increased power with modified forms of the Levene (Med) test for heterogeneity of variance. Biometrics *56*, 451–454.

41. Storey, J.D., and Tibshirani, R. (2003). Statistical significance for genomewide studies. Proc. Natl. Acad. Sci. U. S. A. *100*, 9440–9445.

42. Ma, L., Hoffman, G., and Keinan, A. (2015). X-inactivation informs variance-based testing for X-linked association of a quantitative trait. BMC Genomics *16*, 241.

43. Price, A.L., Patterson, N.J., Plenge, R.M., Weinblatt, M.E., Shadick, N. a, and Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. Nat. Genet. *38*, 904–909.

44. Yang, J., Weedon, M.N., Purcell, S., Lettre, G., Estrada, K., Willer, C.J., Smith, A. V, Ingelsson, E., Connell, J.R.O., Mangino, M., et al. (2011). Genomic inflation factors under polygenic inheritance. Eur. J. Hum. Genet. 807–812.

45. Purcell, S.M., Wray, N.R., Stone, J.L., Visscher, P.M., O'Donovan, M.C., Sullivan, P.F., Ruderfer, D.M., McQuillin, A., Morris, D.W., Oĝdushlaine, C.T., et al. (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature *460*, 748–752.

46. Weiss, L.A., Pan, L., Abney, M., and Ober, C. (2006). The sex-specific genetic architecture of quantitative traits in humans. Nat Genet *38*, 218–222.

47. Randall, J.C., Winkler, T.W., Kutalik, Z., Berndt, S.I., Jackson, A.U., Monda, K.L., Kilpelainen, T.O., Esko, T., Magi, R., Li, S., et al. (2013). Sex-stratified Genome-wide Association Studies Including 270,000 Individuals Show Sexual Dimorphism in Genetic Loci for Anthropometric Traits. PLoS Genet. *9*,.

48. Ober, C., Loisel, D. a, and Gilad, Y. (2008). Sex-specific genetic architecture of human disease. Nat. Genet. *9*, 911–922.

Table 1. Empirical type I error (T1E) rates of X-chromosome variance heterogeneity tests under Simulation Design I.

A quantitative trait was simulated according to simulation design I based on linear regression model (1) with coefficient values specified above such that CONDITION 1 captures the null scenario of no sexual dimorphism, i.e. no sex-specific mean nor variance differences as depicted in Figure 2A; CONDITION 2 corresponds to the conceptual null scenario in Figure 2B with the presence of sex-specific means via a non-zero β_S; CONDITIONS 3A and 3B correspond to Figure 2C, representing a sex-specific variance difference through a non-zero β_{SE}, the *SxE* interaction effect, where the environmental effect β_E takes a value of either 0 (CONDITION 3A) or 0.5 (CONDITION 3B). Similarly, CONDITIONS 4A and 4B correspond to Figure 2D with sexual dimorphism in both means and variances, with the absence and presence of environmental effect β_E, respectively. The total sample size was 5,000 with 2,500 females and 2,500 males, and the MAF was 0.2. The nominal T1E rate was set to 0.05 and the empirical T1E rates were calculated based on 10,000 simulated replicates. Those empirical T1E rates exceeding 5%±0.5% were in bold. Testing strategies that showed satisfactory T1E controls were underlined, and details of the testing strategies are provided in the text and summarized in Table S1.

	CONDITION 1		CONDITION 2		CONDITION 3A		CONDITION 3B		CONDITION 4A		CONDITION 4B	
β_E	0	0.5	0	0.5	0	0	0.5	0.5	0	0	0.5	0.5
β_S	0	0	0.5	0.5	0	0	0	0	0.5	0.5	0.5	0.5
β_{SE}	0	0	0	0	0.25	-0.25	0.25	-0.25	0.25	-0.25	0.25	-0.25
LEV3	0.0544	0.0513	**0.0620**	0.0465	**0.0957**	**0.0754**	**0.0778**	**0.0589**	**0.0552**	**0.0683**	**0.1217**	**0.0794**
LEV5	0.0472	0.0504	0.0475	0.0482	**0.8584**	**0.8734**	**0.9983**	**0.1807**	**0.8932**	**0.9053**	**0.9989**	**0.1738**
FEMALE	0.0525	0.0525	0.0502	0.0476	0.0528	0.0444	0.0253	0.0514	0.0508	0.0434	0.0484	0.0478
MALE	0.0513	0.0485	0.0496	0.0509	0.0496	0.0494	0.0395	0.0518	0.0501	0.0505	0.0508	0.0535
FISHER	0.0517	0.0505	0.0502	0.0478	0.0506	0.0472	0.0286	0.0522	0.0487	0.0454	0.0472	0.0480
M1V1	0.0540	0.0460	0.0526	0.0456	**0.1130**	**0.0941**	**0.119**	**0.0614**	**0.0639**	**0.0884**	**0.1660**	**0.0709**
M1V2	0.0535	0.0461	**0.0575**	0.0424	0.0489	0.0548	0.0375	0.0518	**0.0662**	**0.0566**	0.0276	0.0501
M1V3	0.0492	0.0472	**0.1579**	**0.1234**	0.0459	0.0438	0.0280	0.0538	**0.1311**	**0.1292**	0.0292	**0.1695**
M1V3.2	0.0508	0.0490	**0.1474**	**0.0990**	0.0473	0.0501	0.0320	0.0536	**0.1465**	**0.1269**	0.0234	**0.1461**
M2V1	0.0536	0.0455	0.0497	0.0479	**0.1136**	**0.0948**	**0.1184**	**0.061**	**0.0998**	**0.1142**	**0.1914**	**0.0610**
M2V2	0.0539	0.0459	0.0510	0.0465	0.0484	0.0539	0.0378	0.0526	0.0510	0.0507	0.0318	0.0487
M2V3	0.0485	0.0474	0.0494	0.0499	0.0458	0.0440	0.0278	0.0548	0.0448	0.0433	0.0440	0.0537
M2V3.2	0.0514	0.0492	0.0504	0.0492	0.0479	0.0490	0.0315	0.0534	0.0486	0.0471	0.0351	0.0525
M3V1	0.0533	0.0456	0.0499	0.0482	**0.1109**	**0.092**	**0.1152**	**0.0619**	**0.0971**	**0.1106**	**0.1886**	**0.0617**
M3V2	0.0545	0.0458	0.0501	0.0458	0.0489	0.0544	0.0384	0.0528	0.0523	0.0505	0.0313	0.0485
M3V3	0.0483	0.0466	0.0494	0.0511	0.0461	0.0441	0.0267	0.0545	0.0444	0.0428	0.0455	0.0532
M3V3.2	0.0516	0.0492	0.0508	0.0493	0.0477	0.0487	0.0314	0.0522	0.0487	0.0477	0.0353	0.0528

Figure 1. Empirical examples of sexual dimorphism: quantitative trait distribution stratified by sex.

Phenotype data from UK Biobank (UKB, top row) and the Multi-Ethnic Study of Atherosclerosis (MESA, bottom row) were used to illustrate the possible types of sexual dimorphism as characterized by a location shift in mean of height, a scale difference in the variance of hip circumference, and changes in both mean and variance of waist circumference and BMI. Each trait (Y) was inversely normal transformed so the overall distribution (solid curve) is normal with mean 0 and variance 1. Areas under the sex-stratified distributions (dashed curves) are colored by blue for male and orange for female, respectively.

Figure 2. Defining null and alternative hypotheses for X-chromosome variance heterogeneity test allowing for sexual dimorphism.

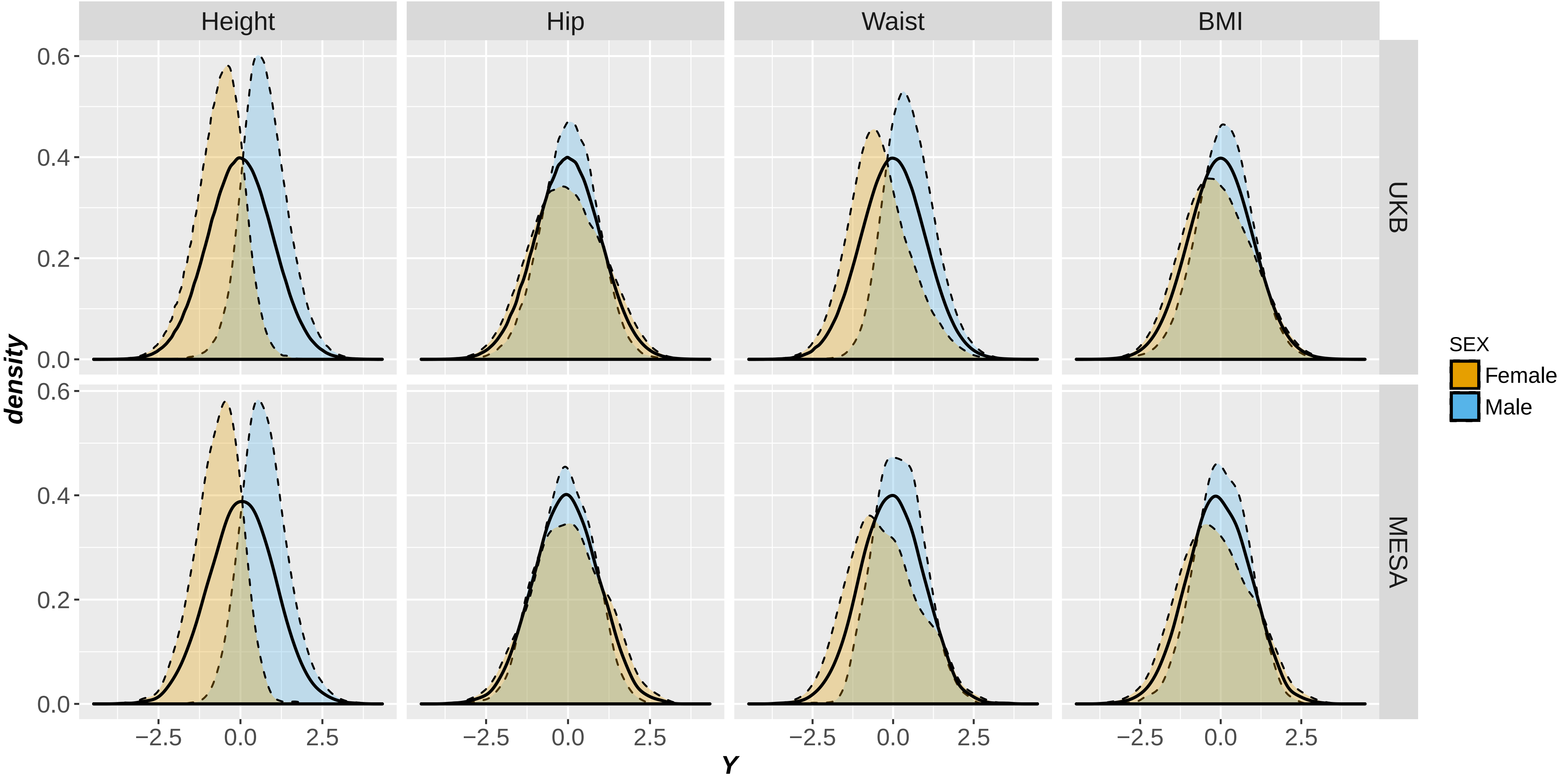
Upper panel: Figures A-D showcase the different types of conceptual *null* distributions, where the variance of a quantitative trait does not vary across the different genotype groups, but is subjected to a possible sex-specific difference in either mean (B), variance (C) or both (D). The black curve is for the overall distribution, and without loss of generality the orange curve is for female and the blue curve is for male as in Figure 1. Lower panel: Figures E-H represent the respective *alternative* distributions. The different genotype groups are marked by different line types and visible only under the alternative conditions when there is phenotypic variance heterogeneity among the genotype groups.

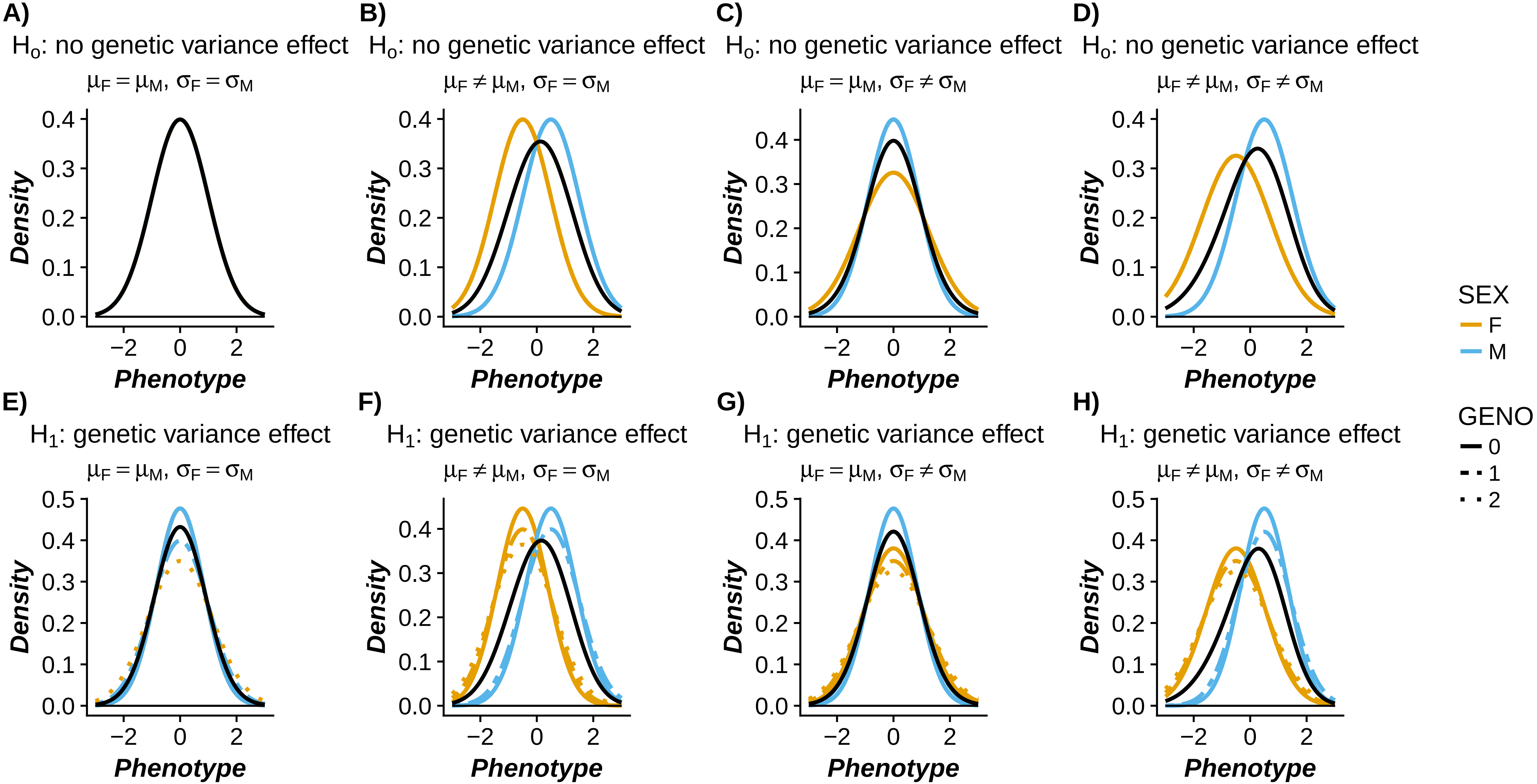
Figure 3. XCHR-wide variance heterogeneity test results for waist circumference using the UK Biobank (upper panel) and MESA (lower panel) data.

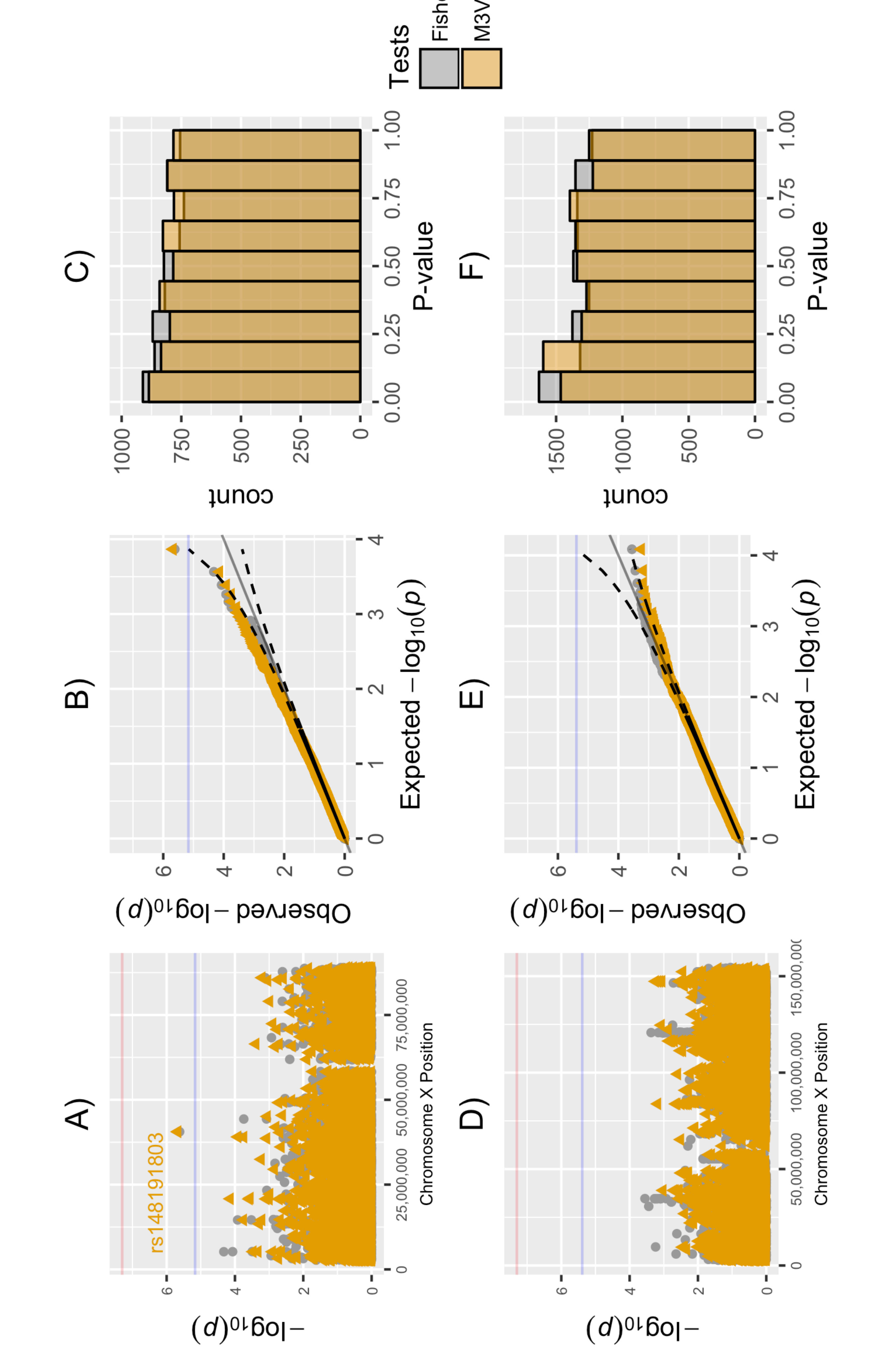
For each XCHR SNP, the variance heterogeneity p-value was calculated using a sex-stratified Fisher's method (grey color) or the model-based M3V3.2 approach (orange color). Manhattan plots (A and D), quantile-quantile plots (B and E), and histograms (C and F) of the p-values using data from the UK Biobank are shown on the top row, and on the bottom row for the Multi-Ethnic Study of Atherosclerosis (MESA). In Figure 3A, SNP rs148191803 (M3V3.2 test p = 2.08E-06 and Fisher's method p = 2.4E-06) in the *MED14* locus was annotated for passing the XCHR-wide significance at 6.8E-06 in the UK Biobank data; but with p>0.05 in the MESA data. Results for other traits are in Figures S2-S4 and Tables S5-6.

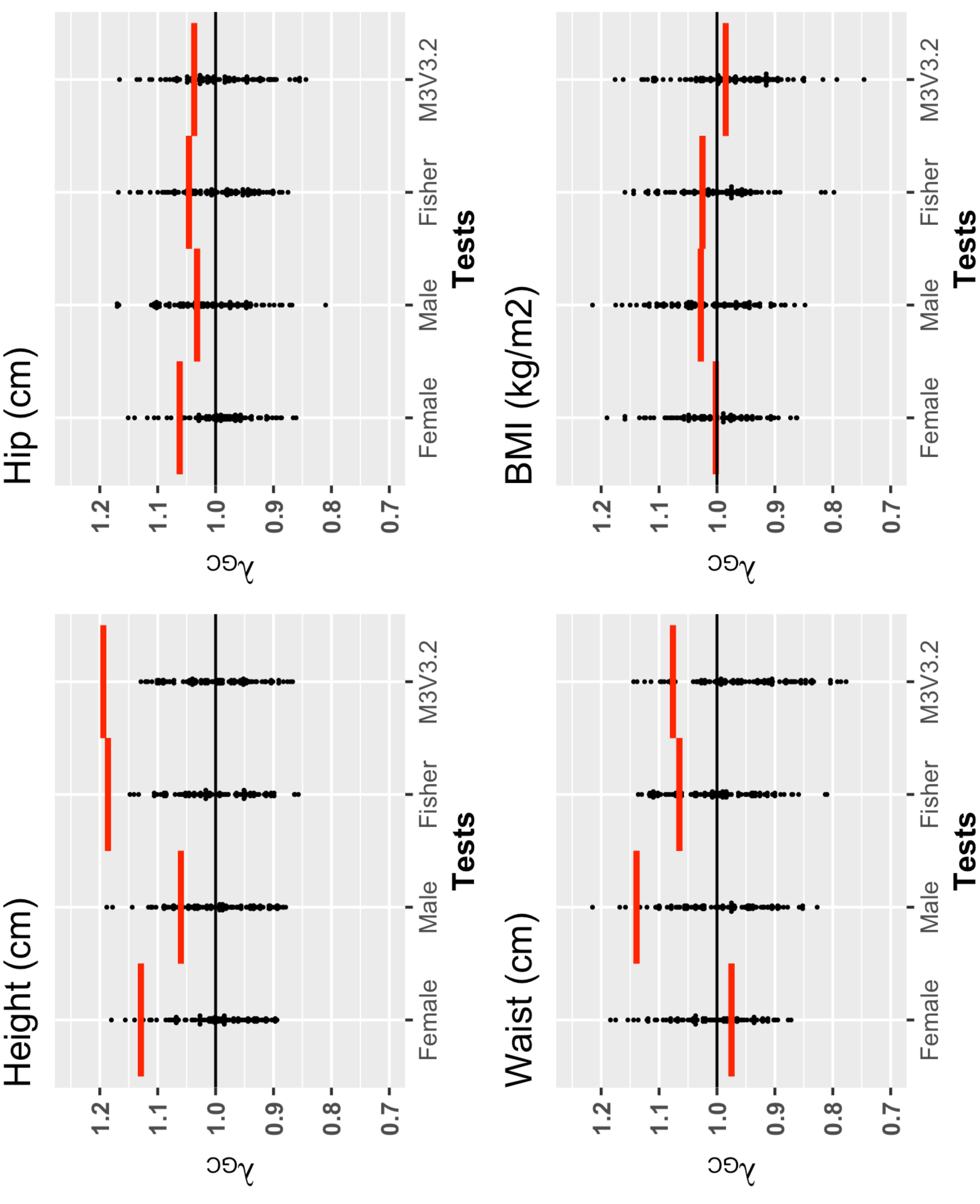
Figure 4. Estimated genomic control lambda values using the observed and permuted individual MESA data.

The permuted dataset was obtained by permuting each quantitative trait stratified by sex, independently, 100 times. The genomic lambda λ_{GC} was computed for each trait in a permuted dataset and is shown as a black dot under each test. The red line represents the genomic lambda estimated for the original observed data, and the black horizontal line represents the reference line at of 1. The corresponding estimates of the proportion of truly associated variants are shown in Figure S6.


Height
Hip
Waist
BMI
UKB
MESA
density
Y
0.0
0.2
0.4
0.6
−2.5
0.0
2.5
SEX
Female
Male


A)
H_o: no genetic variance effect
$\mu_F = \mu_M$, $\sigma_F = \sigma_M$
B)
H_o: no genetic variance effect
$\mu_F \neq \mu_M$, $\sigma_F = \sigma_M$
C)
H_o: no genetic variance effect
$\mu_F = \mu_M$, $\sigma_F \neq \sigma_M$
D)
H_o: no genetic variance effect
$\mu_F \neq \mu_M$, $\sigma_F \neq \sigma_M$
E)
H_1: genetic variance effect
$\mu_F = \mu_M$, $\sigma_F = \sigma_M$
F)
H_1: genetic variance effect
$\mu_F \neq \mu_M$, $\sigma_F = \sigma_M$
G)
H_1: genetic variance effect
$\mu_F = \mu_M$, $\sigma_F \neq \sigma_M$
H)
H_1: genetic variance effect
$\mu_F \neq \mu_M$, $\sigma_F \neq \sigma_M$
Density
0.0
0.1
0.2
0.3
0.4
0.5
Phenotype
−2
0
2
SEX
F
M
GENO
0
1
2


A)
rs148191803
−log10(p)
0
2
4
6
0
25,000,000
50,000,000
75,000,000
Chromosome X Position
B)
Observed −log10(p)
Expected −log10(p)
0
1
2
3
4
C)
count
0
250
500
750
1000
0.00
0.25
0.50
0.75
1.00
P-value
D)
−log10(p)
0
2
4
6
0
50,000,000
100,000,000
150,000,000
Chromosome X Position
E)
Observed −log10(p)
Expected −log10(p)
F)
count
0
500
1000
1500
P-value
Tests
Fisher
M3VS2


Height (cm)
Hip (cm)
Waist (cm)
BMI (kg/m2)
λGC
1.2
1.1
1.0
0.9
0.8
0.7
Tests
Female
Male
Fisher
M3V3.2