www.nature.com/hdy

the geneticssociety

ARTICLE

Robust association tests for quantitative traits on the X chromosome

Zi-Ying Yang[1,2,4], Wei Liu[1,2,4], Yu-Xin Yuan[1,2,4], Yi-Fan Kong[1], Pei-Zhen Zhao[1], Wing Kam Fung[3]✉ and Ji-Yuan Zhou[1,2]✉

The genome-wide association study is an elementary tool to assess the genetic contribution to complex human traits. However, such association tests are mainly proposed for autosomes, and less attention has been given to methods for identifying loci on the X chromosome due to their distinct biological features. In addition, the existing association tests for quantitative traits on the X chromosome either fail to incorporate the information of males or only detect variance heterogeneity. Therefore, we propose four novel methods, which are denoted as QXcat, QZ_{max}, $QMVX_{cat}$ and $QMVZ_{max}$. When using these methods, it is assumed that the risk alleles for females and males are the same and that the locus being studied satisfies the generalized genetic model for females. The first two methods are based on comparing the means of the trait value across different genotypes, while the latter two methods test for the difference of both means and variances. All four methods effectively incorporate the information of X chromosome inactivation. Simulation studies demonstrate that the proposed methods control the type I error rates well. Under the simulated scenarios, the proposed methods are generally more powerful than the existing methods. We also apply our proposed methods to data from the Minnesota Center for Twin and Family Research and find 10 single nucleotide polymorphisms that are statistically significantly associated with at least two traits at the significance level of 1×10^{-3}.

Heredity (2022) 129:244–256; https://doi.org/10.1038/s41437-022-00560-y

INTRODUCTION

The genome-wide association study is an elementary tool to assess the genetic contribution to complex human traits (Kang et al. 2010). Thousands of single nucleotide polymorphisms (SNPs) have been found to be associated with hundreds of complex traits by association tests (Chen et al. 2017; Ma et al. 2015a; Zheng et al. 2007). However, only a few association tests have focused on the X chromosome (Chang et al. 2014; Wang et al. 2019a), which consists of 1669 (almost 5%) known genes and affects ~7% of complex traits (Wise et al. 2013; Xu and Hao 2018). Unlike autosomes, the X chromosome has several distinct biological features. For instance, the number of copies of the X chromosome is different between sexes. In addition, gene expression in females is affected by X chromosome inactivation (XCI), where one copy of the two X chromosomes in females is silenced to compensate for the X chromosome dosage difference between sexes, i.e., complete dosage compensation is achieved (Hickey and Bahlo 2011; Wang et al. 2014). However, Carrel and Willard (2005) claimed that weak expression of the silenced X chromosome occurs in ~10% of genes, which is referred to as incomplete dosage compensation. XCI was discovered over fifty years ago (Lyon 1961). In XCI, which is usually regarded as a random process referred to as random XCI, ~50% of cells have the risk allele active, while the other ~50% of cells have the normal allele active (Jin et al. 2017; Wang et al. 2014). However, in recent studies, it has been reported that some X-linked genes in females may also undergo skewed XCI and escape from XCI (XCI-E) (Amos-Landgraf et al. 2006; Carrel and Willard 2005). The former is defined that one allele is inactivated in more than 50% of cells, such as 75% or even 90% of cells in some extreme cases (Minks et al. 2008; Wong et al. 2011). The latter implies that both alleles in female cells remain active, which is also referred to as no dosage compensation (Brown et al. 1997; Carrel et al. 2006). XCI is a complex biological mechanism that is not yet fully understood (Wu et al. 2014). Therefore, robust and powerful association tests on the X chromosome are needed to account for these characteristics.

Some methods for testing association have been developed to accommodate the X chromosome (Chung et al. 2007; Ding et al. 2006; Horvath et al. 2000; Zhang et al. 2008). Zheng et al. (2007) proposed several allele-based and genotype-based tests on the X chromosome, and compared their performance under Hardy-Weinberg equilibrium (HWE) and departure from HWE. However, these methods may lose power when XCI exists (Chen et al. 2017; Loley et al. 2011). To address this issue, Clayton (2008) suggested a 1 degree of freedom chi-square test and a 2 degrees of freedom chi-square test by treating males as homozygous females, without the assumption of HWE. In this case, three female genotypes were coded as 0, 1 and 2, and two male genotypes were coded as 0 and

[1]Department of Biostatistics, State Key Laboratory of Organ Failure Research, Ministry of Education, and Guangdong Provincial Key Laboratory of Tropical Disease Research, School of Public Health, Southern Medical University, Guangzhou, China. [2]Guangdong-Hong Kong-Macao Joint Laboratory for Contaminants Exposure and Health, Guangzhou, China. [3]Department of Statistics and Actuarial Science, The University of Hong Kong, Hong Kong, China. [4]These authors contributed equally: Zi-Ying Yang, Wei Liu, Yu-Xin Yuan. Associate editor: Sara Knott. ✉email: wingfung@hku.hk; zhoujiyuan5460@hotmail.com

Received: 15 July 2021 Revised: 24 August 2022 Accepted: 24 August 2022
Published online: 10 September 2022

2 (Hickey and Bahlo 2011). Nevertheless, Clayton's methods require the allele frequencies of the same allele to be equal between sexes, and only random XCI is considered (Clayton 2008). Using this coding strategy may lead to power loss when the XCI pattern is misspecified. As such, Wang et al. (2014) proposed a unified coding strategy, in which female genotypes were coded as 0, γ and 2, where γ ranges from 0 to 2. Here, $\gamma < 1$ represents XCI towards the risk allele, $\gamma > 1$ represents XCI towards the normal allele, and $\gamma = 1$ denotes random XCI. In the method proposed by Wang et al. (2014), the test power under skewed XCI is improved by maximizing the likelihood ratio over different biological models (random XCI, skewed XCI and XCI-E). However, the strategy is time-consuming because a permutation procedure is required to obtain the p value (Jin et al. 2017). Chen et al. (2017) proposed a test statistic that does not need to specify the underlying XCI pattern and HWE. It constructs the models for females and males separately and then combines them using Fisher's method (Fisher et al. 1967). The method proposed by Chen et al. (2017) effectively utilizes the information of both females and males. To further improve the test power, Wang et al. (2019a) provided an allelic test that considers different deviations from HWE. Instead of combining the test statistics of females and males by Fisher's method, Wang et al. (2019a) used the effective sample sizes of females and males to combine the information of both sexes. Different dosage compensation patterns can be incorporated in this method by selecting different weights.

All of the methods mentioned above were developed primarily for case–control studies. Some studies have shown that genetic loci on the X chromosome also affect quantitative traits (Al-Ayadhi et al. 2020; Auer et al. 2014; Gaukrodger et al. 2005; Konzman et al. 2020). Factors such as mutation, genetic interactions and parent-of-origin effects may influence the expression level of genes, thus changing the phenotypic means or variances across different genotypes (Brown et al. 2014; Cao et al. 2014; Ma et al. 2015b; Morley et al. 2004; Soave et al. 2015; Yang et al. 2012). As such, Ma et al. (2015b) assumed that XCI would cause extra phenotypic variance for heterozygous females and proposed three X-linked association tests, denoted as T_{Var}, T_W and T_S. T_{Var}, which can be regarded as a modification of the Brown-Forsythe test, directly tests for the inflated variance of the trait value for heterozygous females (Brown and Forsythe 1974). T_W uses a weighted linear regression to examine the means of the trait value and allows for variance heterogeneity in females. Finally, T_S first transforms the p values of T_{Var} and T_W to Z scores and then combines them using Stouffer's method (Stouffer et al. 1949). Since the methods proposed by Ma et al. (2015b) ignore the information of males, these methods should each lose test power. To effectively account for XCI, Chen et al. (2020) used a Bayesian model to average over different XCI patterns. However, the Bayesian model is known to be time-consuming because multiple Markov chains of parameters are generated. Deng et al. (2019) proposed a series of methods that simultaneously incorporate the information of females and males to investigate the variances among genotypes. One of the methods proposed by Deng et al. (2019) computes the p values of Levene's test for females and males separately (Levene 1961) and then combines them using Fisher's method (denoted as Fisher in this article). Deng et al. (2019) assumed that the association between the SNP and the quantitative trait being studied could be biased by sex-specific means or variances because of the different numbers of copies of the X chromosome between females and males. In this regard, two two-stage methods, wM3V3.2 and wM3VNA3.3, were proposed. For brevity, we refer to these methods as wM3V and wM3VNA, respectively, in this article. In the first stage, these methods regress the value of the quantitative trait on the genotype, sex and their interaction via a regression framework. In the second stage, the wM3V method tests for genotypic variances of the residuals obtained from the first stage via the generalized Levene's test under the additive genetic model, while the wM3VNA method does the same under the generalized genetic model (Chen and Ng 2012). Although the methods proposed by Deng et al. (2019) incorporate males' information and efficiently test for variance heterogeneity, the mean differences are only adjusted when conducting the generalized Levene's test. These methods are not designed to test for the mean differences, which may cause loss of power. In addition, Özbek et al. (2018) proposed an X chromosome association test statistic that considers the sex × SNP interaction term and is applicable to both quantitative and qualitative traits. This method can be directly implemented in PLINK, and in this article, we denote it for quantitative traits as T_{plink}. Song et al. (2021) further conducted extensive simulations to compare the performance of the model including the interaction term with that not including the interaction term and found that fitting the model with the interaction term can make the estimates of the effect sizes more robust to different XCI patterns. However, T_{plink} assumes the homogeneity of variances across different genotypes and only takes into account random XCI and XCI-E patterns. Chen et al. (2021) added a variable indicative of heterozygous females in T_{plink} and suggested an X chromosomal association approach that considers all three XCI patterns and is suitable for both quantitative and qualitative traits. We denote it for quantitative traits as T_{chen} in this article. However, T_{chen} only compares the difference in the means of the trait value across different genotypes under the assumption of variance homogeneity.

Therefore, in this article, we propose four novel statistical methods, denoted as QXcat, QZ_{max}, $QMVX_{cat}$ and $QMVZ_{max}$, to test for association between an SNP on the X chromosome and a quantitative trait. QXcat and QZ_{max} are designed for testing the mean differences of the trait value. In QXcat, we obtain the p values for females and males by testing the mean differences of the trait value via weighted linear regression models. Then, we combine these two p values using Fisher's method. In QZ_{max}, we use different sample sizes as weights, which represent different dosage compensation patterns according to Wang et al. (2019a), to combine the test statistics for females and males. In addition, we develop $QMVX_{cat}$ ($QMVZ_{max}$) by combining the p value of QXcat (QZ_{max}) with that of wM3VNA, to test for the difference in both means and variances. We perform extensive simulation studies to investigate the type I error rates and the test powers of the proposed methods. We also apply our proposed methods to data from the Minnesota Center for Twin and Family Research (MCTFR) for their practice.

MATERIALS AND METHODS

Notations

Consider an SNP on the X chromosome with alleles a and A. Let q_f and q_m be the frequencies of A in females and males, respectively, and let ρ be the inbreeding coefficient in the female population. Then, females have three genotypes, aa, Aa and AA, and males, who are hemizygous, only have two different genotypes, a and A. The frequencies of genotypes aa, Aa and AA for females are denoted as q_{aa}, q_{Aa} and q_{AA}, respectively. Thus, $q_{aa} = (1-q_f)^2 + \rho(1-q_f)q_f$, $q_{Aa} = 2(1-\rho)(1-q_f)q_f$ and $q_{AA} = q_f^2 + \rho(1-q_f)q_f$. Suppose that we collect a sample of N independent individuals consisting of n_f females and n_m males. Let n_{f0}, n_{f1} and n_{f2} be the number of females with genotypes aa, Aa and AA ($n_{f0} + n_{f1} + n_{f2} = n_f$), respectively. There are n_{m0} males with genotype a and n_{m1} males with genotype A ($n_{m0} + n_{m1} = n_m$). Let $\mathbf{Y}_f = (y_{f1}, y_{f2}, \ldots, y_{fnf})^T$ and $\mathbf{Y}_m = (y_{m1}, y_{m2}, \ldots, y_{mnm})^T$ denote the values of the quantitative trait for females and males, respectively. Here, we assume that $\mathbf{Y}_f$ and $\mathbf{Y}_m$ are normally distributed or approximately follow normal distributions after the rank-based inverse normal transformation (McCaw et al. 2019). For females, let G_{fi} denote the number of alleles A in female i ($i = 1, 2, \ldots, n_f$), i.e., G_{fi} takes the value of 0, 1 and 2 for aa, Aa and AA, respectively; for males, let G_{mi} denote the number of alleles A in male i ($i = 1, 2, \ldots, n_m$), i.e., G_{mi} takes the value of 0 and 1 for a and A,

respectively. In females, the means of the quantitative trait for *aa*, *Aa* and *AA* are denoted as μ_{f0}, μ_{f1} and μ_{f2}, respectively, while the variances of the quantitative trait for *aa*, *Aa* and *AA* are represented by σ^2_{f0}, σ^2_{f1} and σ^2_{f2}, respectively. Let $\mathbf{V}_f$ denote the variance-covariance matrix of $\mathbf{Y}_f$, a diagonal matrix with elements σ^2_{f0}, σ^2_{f1} and σ^2_{f2} for *aa*, *Aa* and *AA*, respectively. In males, the means of the quantitative trait for *a* and *A* are denoted as μ_{m0} and μ_{m1}, respectively, while the variances of the quantitative trait for *a* and *A* are represented by σ^2_{m0} and σ^2_{m1}, respectively. Let $\mathbf{V}_m$ be the variance-covariance matrix of $\mathbf{Y}_m$, a diagonal matrix with elements σ^2_{m0} and σ^2_{m1} for *a* and *A*, respectively. Here, we consider three types of null hypotheses of no association between the SNP and the quantitative trait. H_0^{MV}: both the means and the variances of the quantitative trait across genotypes are equal (i.e., $\mu_{f0}=\mu_{f1}=\mu_{f2}$, $\mu_{m0}=\mu_{m1}$, $\sigma^2_{f0}=\sigma^2_{f1}=\sigma^2_{f2}$ and $\sigma^2_{m0}=\sigma^2_{m1}$), H_0^{M}: only the means of the quantitative trait across genotypes are equal (i.e., $\mu_{f0}=\mu_{f1}=\mu_{f2}$, $\mu_{m0}=\mu_{m1}$ and no restrictions on the variances) and H_0^{V}: only the variances of the quantitative trait across genotypes are equal (i.e., $\sigma^2_{f0}=\sigma^2_{f1}=\sigma^2_{f2}, \sigma^2_{m0}=\sigma^2_{m1}$ and no restrictions on the means).

Sex-stratified X chromosome mean-based association test for quantitative traits considering various XCI patterns

Note that SNPs on the X chromosome of females may undergo different XCI patterns. To make our method robust to various XCI patterns, we first propose a general X chromosome association test for quantitative traits named QXcat, which aims to identify the mean differences of the trait value across genotypes. We construct the models for females and males separately because the numbers of X chromosomes are different between sexes and then combine their *p* values in an efficient way. Specifically, we first assume that *A* is the risk allele, and the risk allele in females is the same as that in males. In addition, similar to the work in Chen et al. (2017), the generalized genetic model is assumed for the SNP being studied for females, i.e., the genetic effect of carrying two risk alleles is not less than that of carrying one risk allele, and the genetic effect of carrying one risk allele is not less than that of carrying no risk allele ($\mu_{f2}\geq\mu_{f1}\geq\mu_{f0}$). Then, we consider two variables $X_{fi}^{(1)}=I_{\{G_{fi}\geq 1\}}$ and $X_{fi}^{(2)}=I_{\{G_{fi}=2\}}$ for female *i*, where $I_{\{\cdot\}}$ is the indicator function, $X_{fi}^{(1)}$ indicates that female *i* carries at least one risk allele and $X_{fi}^{(2)}$ means that the genotype of female *i* is *AA*. Based on the study by Wang et al. (2019b), $X_{fi}^{(1)}$ and $X_{fi}^{(2)}$ can be used to test for association between the SNP and the trait under different XCI patterns. Hence, the association between the quantitative trait and the SNP in females can be modeled as

$$y_{fi}=\beta_{f0}+\beta_{f1}X_{fi}^{(1)}+\beta_{f2}X_{fi}^{(2)}+\mathbf{b}_f^T\mathbf{Z}_{fi}+\varepsilon_{fi},\ i=1,2,\ldots,n_f \quad (1)$$

where β_{f0} is the intercept, and β_{f1} and β_{f2} are the regression coefficients of $X_{fi}^{(1)}$ and $X_{fi}^{(2)}$, respectively. $\mathbf{Z}_{fi}$ denotes a vector of covariates for female *i*, $\mathbf{b}_f$ is the vector of the regression coefficients of $\mathbf{Z}_{fi}$, and ε_{fi} is a random error that follows $N(0,\sigma^2_{f0})$, $N(0,\sigma^2_{f1})$ and $N(0,\sigma^2_{f2})$ for genotypes *aa*, *Aa* and *AA*, respectively. According to Wang et al. (2019b), under random XCI or XCI-E, $\beta_{f1}=\beta_{f2}\neq 0$ means that the SNP is associated with the quantitative trait. For the skewed XCI, $\beta_{f1}=0$ and $\beta_{f2}\neq 0$ when the risk allele is inactivated in 100% of the heterozygous female cells, while $\beta_{f1}\neq 0$ and $\beta_{f2}=0$ when all the cells in females with genotype *Aa* are normal allele inactive. In addition, $\beta_{f1}\neq 0$, $\beta_{f2}\neq 0$ and $\beta_{f1}\neq\beta_{f2}$ mean that *A* is associated with the quantitative trait for other skewed XCI patterns. Hence, Model (1) effectively incorporates all the XCI patterns when testing for association. Since some factors (such as mutation and XCI) may lead to unequal trait value variances across different genotypes, we use the weighted least square method to estimate $\boldsymbol{\beta}_f=(\beta_{f0},\beta_{f1},\beta_{f2},\mathbf{b}_f^T)^T$. Let $\mathbf{W}_f$ be a weight matrix for females. Here, we set $\mathbf{W}_f=\mathbf{V}_f^{-1}$ with elements $\frac{1}{\sigma^2_{f0}}$, $\frac{1}{\sigma^2_{f1}}$ and $\frac{1}{\sigma^2_{f2}}$ for genotypes *aa*, *Aa* and *AA*, respectively. We first fit Model (1) by the ordinary least square method and obtain the corresponding residuals. Then, $\frac{1}{\sigma^2_{f0}}$, $\frac{1}{\sigma^2_{f1}}$ and $\frac{1}{\sigma^2_{f2}}$ are estimated by the inverse of the residual variances for genotypes *aa*, *Aa* and *AA*, denoted as $\frac{1}{\hat\sigma^2_{f0}}$, $\frac{1}{\hat\sigma^2_{f1}}$ and $\frac{1}{\hat\sigma^2_{f2}}$, respectively. As a result, $\hat{\mathbf{W}}_f=\hat{\mathbf{V}}_f^{-1}$. To estimate $\boldsymbol{\beta}_f$, we minimize the following weighted residual sum of squares $\arg\min_{\boldsymbol{\beta}_f}\|\hat{\mathbf{W}}_f^{1/2}(\mathbf{Y}_f-\mathbf{X}_f\boldsymbol{\beta}_f)\|^2$ where $\mathbf{X}_f=(\mathbf{X}_f^{(0)},\mathbf{X}_f^{(1)},\mathbf{X}_f^{(2)},\mathbf{Z}_f)$ is a design matrix, and $\mathbf{X}_f^{(0)}=(1,1,\ldots,1)^T$, $\mathbf{X}_f^{(1)}=(X_{f1}^{(1)},X_{f2}^{(1)},\ldots,X_{fn_f}^{(1)})^T$, $\mathbf{X}_f^{(2)}=(X_{f1}^{(2)},X_{f2}^{(2)},\ldots,X_{fn_f}^{(2)})^T$, and $\mathbf{Z}_f=(\mathbf{Z}_{f1},\mathbf{Z}_{f2},\ldots,\mathbf{Z}_{fn_f})^T$. Specifically, $\mathbf{Z}_{fi}$ denotes a vector of covariates for female *i* in Model (1). Let $\hat{\boldsymbol{\beta}}_f=(\hat\beta_{f0},\hat\beta_{f1},\hat\beta_{f2},\hat{\mathbf{b}}_f^T)^T$ be the estimate of $\boldsymbol{\beta}_f$, and it can be expressed as

$$\hat{\boldsymbol{\beta}}_f=\left(\mathbf{X}_f^T\hat{\mathbf{W}}_f\mathbf{X}_f\right)^{-1}\mathbf{X}_f^T\hat{\mathbf{W}}_f\mathbf{Y}_f$$

The variance-covariance matrix of $\hat{\boldsymbol{\beta}}_f$ is estimated by

$$\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}}_f)=(\mathbf{X}_f^T\hat{\mathbf{W}}_f\mathbf{X}_f)^{-1}\mathbf{X}_f^T\hat{\mathbf{W}}_f\hat{\mathbf{V}}_f\hat{\mathbf{W}}_f^T\mathbf{X}_f(\mathbf{X}_f^T\hat{\mathbf{W}}_f^T\mathbf{X}_f)^{-1}$$

Since $\hat{\mathbf{W}}_f=\hat{\mathbf{V}}_f^{-1}$, $\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}}_f)=\left(\mathbf{X}_f^T\hat{\mathbf{W}}_f\mathbf{X}_f\right)^{-1}$, and the estimate of the variance-covariance matrix for $\hat\beta_{f1}$ and $\hat\beta_{f2}$ is $\hat\Sigma$, which is constructed by the four elements in Rows 2-3 and Columns 2-3 of $\widehat{\mathrm{Var}}(\hat{\boldsymbol{\beta}}_f)$, we define the following test statistics:

$$\begin{pmatrix}T_{f1}^A\\T_{f2}^A\end{pmatrix}=\hat\Sigma^{-1/2}\begin{pmatrix}\hat\beta_{f1}\\\hat\beta_{f2}\end{pmatrix}$$

Under the null hypothesis of H_0^{MV} or H_0^{M}, T_{f1}^A and T_{f2}^A are independent of each other and asymptotically follow the standard normal distribution. The corresponding proof of this independence is given in Appendix A. The one-sided *p* values of T_{f1}^A and T_{f2}^A are denoted as $p_{f1}^A=1-\Phi(T_{f1}^A)$ and $p_{f2}^A=1-\Phi(T_{f2}^A)$, respectively, where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. We combine p_{f1}^A with p_{f2}^A using Fisher's method and obtain the test statistic

$$Q_f^A=-2\ln(p_{f1}^Ap_{f2}^A)$$

Under H_0^{MV} or H_0^{M}, $Q_f^A\sim\chi^2_4$ (Chen et al. 2017). We denote the *p* value of Q_f^A as p_f^A.

For males, we use the following model to test for the association between the SNP and the trait

$$y_{mi}=\beta_{m0}+\beta_{m1}G_{mi}+\mathbf{b}_m^T\mathbf{Z}_{mi}+\varepsilon_{mi},\ i=1,2,\ldots,n_m \quad (2)$$

where β_{m0} is the intercept and β_{m1} is the regression coefficient of G_{mi}. $\mathbf{Z}_{mi}$ is a vector of covariates for male *i*, and $\mathbf{b}_m$ is the vector of the regression coefficients of $\mathbf{Z}_{mi}$. ε_{mi} is a random error that follows $N(0,\sigma^2_{m0})$ and $N(0,\sigma^2_{m1})$ for genotypes *a* and *A*, respectively. Similar to the case for females, we use the weighted least square method to estimate $\boldsymbol{\beta}_m=(\beta_{m0},\beta_{m1},\mathbf{b}_m^T)^T$. Here, we set the weight matrix $\mathbf{W}_m$ for males as $\mathbf{V}_m^{-1}$ with elements $\frac{1}{\sigma^2_{m0}}$ and $\frac{1}{\sigma^2_{m1}}$ for genotypes *a* and *A*, respectively. We denote the estimate of β_{m1} and its variance as $\hat\beta_{m1}$ and $\widehat{\mathrm{Var}}(\hat\beta_{m1})$, respectively, and then construct the test statistic as $T_m^A=\frac{\hat\beta_{m1}}{\sqrt{\widehat{\mathrm{Var}}(\hat\beta_{m1})}}$. If n_m is large enough, $T_m^A\sim N(0,1)$. We denote the one-sided *p* value of T_m^A as $p_m^A=1-\Phi(T_m^A)$. Then, we combine p_f^A with p_m^A and obtain the test statistic

$$Q^A=-2\ln(p_f^Ap_m^A)$$

Under H_0^{MV} or H_0^{M}, $Q^A\sim\chi^2_4$.

Note that the risk allele is generally unknown. Here, we also consider the case where the risk allele is *a*. We can obtain the test statistics Q_f^a for females and T_m^a for males, and the corresponding one-sided *p* values p_f^a and p_m^a, respectively, in the same way. Then, the test statistic can be derived as

$$Q^a=-2\ln(p_f^ap_m^a)$$

Similarly, $Q^a\sim\chi^2_4$ under H_0^{MV} or H_0^{M}. We define the final mean-based test statistic as

$$\mathrm{QXcat}=\max(Q^A,Q^a)$$

Based on the theorem proposed by Mosteller and Fisher (1948), the *p* value of QXcat can be approximated as follows:

$$2\xi-\xi^2\leq\Pr(\mathrm{QXcat}>\eta)\leq 2\xi$$

where $\xi=1-\chi^2_4(\eta)$. Here, we choose 2ξ to approximate the *p* value of QXcat, which is denoted as p_{QXcat}.

X chromosome mean-based association test for quantitative traits considering different dosage compensation patterns

Note that QXcat takes all the XCI patterns into account by introducing two indicator variables for females. In addition to this way of considering XCI, Wang et al. (2019a) combined the test statistics for females and males by

Table 1. Means and variances of the trait values across different genotypes under XCI and XCI-E.

Sex	Genotype	XCI			XCI-E		
		g_i	$E(y_i)$	$Var(y_i)$	g_i	$E(y_i)$	$Var(y_i)$
Female	*aa*	0	$\mu_{f0}=\beta_c+\beta_z$	$\sigma^2_{f0}=\sigma^2$	0	$\mu_{f0}=\beta_c+\beta_z$	$\sigma^2_{f0}=\sigma^2$
	Aa	γ	$\mu_{f1}=\beta_c+\gamma\beta_g+\beta_z$	$\sigma^2_{f1}=\sigma^2+\theta+\frac{\gamma}{2}\left(1-\frac{\gamma}{2}\right)b^2$	1	$\mu_{f1}=\beta_c+\beta_g+\beta_z$	$\sigma^2_{f1}=\sigma^2+\theta$
	AA	2	$\mu_{f2}=\beta_c+2\beta_g+\beta_z$	$\sigma^2_{f2}=\sigma^2+\tau$	2	$\mu_{f2}=\beta_c+2\beta_g+\beta_z$	$\sigma^2_{f2}=\sigma^2+\tau$
Male	*a*	0	$\mu_{m0}=\beta_c$	$\sigma^2_{m0}=\sigma^2$	0	$\mu_{m0}=\beta_c$	$\sigma^2_{m0}=\sigma^2$
	A	2	$\mu_{m1}=\beta_c+2\beta_g$	$\sigma^2_{m1}=\sigma^2+\tau$	1	$\mu_{m1}=\beta_c+\beta_g$	$\sigma^2_{m1}=\sigma^2+\tau$

different weights to account for different dosage compensation patterns in their method Z_{max} for case–control design. Adopting a similar idea, we put forward another mean-based association test, which also incorporates the information of dosage compensation by combining the test statistics for females and males based on different weights. Therefore, we propose our QZ_{max} test statistic as follows. Here, we assume that *A* is the risk allele, and the risk allele in females is the same as that in males. Furthermore, for females, the generalized genetic model is assumed at the SNP (Chen et al. 2017). For females, let $T_f^A=\frac{1}{\sqrt{2}}\left(T_{f1}^A+T_{f2}^A\right)$. Since T_{f1}^A and T_{f2}^A are independent of each other, $T_f^A\sim N(0,1)$ under H_0^{MV} or H_0^M. For males, we still use T_m^A, which is independent of T_f^A. Based on the work of Wang et al. (2019a), we combine T_f^A and T_m^A in the following way

$$T_{\lambda_k}=\sqrt{\lambda_k}T_f^A+\sqrt{1-\lambda_k}T_m^A$$

where $\lambda_k=2n_f/(kn_m+2n_f)$ $(1\le k\le 2)$. $k=1$ denotes no dosage compensation. $1<k<2$ indicates incomplete dosage compensation, and $k=2$ means complete dosage compensation. Note that the values of T_{λ_k} when *A* is the risk allele and when *a* is the risk allele have different signs, while their absolute values are still the same. Therefore, we only consider the corresponding test statistics when *A* is assumed to be the risk allele. Wang et al. (2019a) demonstrated that incomplete dosage compensation $(1<k<2)$ is much less common than no dosage compensation and complete dosage compensation, so we choose $k=1$ and $k=2$. Since the risk allele is generally unknown in practice, i.e., the signs of T_{λ_1} and T_{λ_2} are unknown, we propose the final mean-based test statistic as follows:

$$QZ_{max}=\max(|T_{\lambda_1}|,|T_{\lambda_2}|)$$

Here, T_{λ_1} and T_{λ_2} jointly follow a bivariate normal distribution. The correlation coefficient of T_{λ_1} and T_{λ_2} can be estimated by

$$\begin{aligned}r_{(T_{\lambda_1},T_{\lambda_2})}&=\frac{\mathrm{Cov}(T_{\lambda_1},T_{\lambda_2})}{\sqrt{\mathrm{Var}(T_{\lambda_1})\mathrm{Var}(T_{\lambda_2})}}\\&=\frac{\sqrt{\lambda_1\lambda_2}\mathrm{Var}(T_f^A)+\sqrt{(1-\lambda_1)(1-\lambda_2)}\mathrm{Var}(T_m^A)}{\sqrt{\left[\lambda_1\mathrm{Var}(T_f^A)+(1-\lambda_1)\mathrm{Var}(T_m^A)\right]\left[\lambda_2\mathrm{Var}(T_f^A)+(1-\lambda_2)\mathrm{Var}(T_m^A)\right]}}\\&=\sqrt{\lambda_1\lambda_2}+\sqrt{(1-\lambda_1)(1-\lambda_2)}\end{aligned}$$

The *p* value of QZ_{max} (denoted by $p_{QZ_{max}}$) can be obtained directly by the *mvtnorm* package (https://cran.r-project.org/web/packages/mvtnorm/index.html) in the R statistical software (R Core Team 2020) as follows:

$$\begin{aligned}p_{QZ_{max}}=1-pmvnorm(&lower=-rep(QZ_{max},2),\\&upper=rep(QZ_{max},2),corr=\boldsymbol{R}_{(T_{\lambda_1},T_{\lambda_2})}\Big)\end{aligned}$$

where $\boldsymbol{R}_{(T_{\lambda_1},T_{\lambda_2})}$ is a 2×2 correlation matrix, and element $r_{(T_{\lambda_1},T_{\lambda_2})}$ is the correlation coefficient of T_{λ_1} and T_{λ_2}.

Two X chromosome mean-variance-based association tests for quantitative traits

Note that QXcat and QZ_{max} can only test for the mean differences across different genotypes. However, the variances of the trait value across genotypes may also be affected by the mutation at the given SNP. To improve the test power, we propose the other two tests by combining the variance-based test wM3VNA proposed by Deng et al. (2019) with QXcat and QZ_{max} to test for both the mean differences and the variance heterogeneity. Here, we denote the *p* value of wM3VNA as p_{wM3VNA}. Referring to the proof by Soave et al. (2015), the mean-based association tests and the variance-based association tests for autosomal SNPs and normally distributed traits are independent, and we prove the independence of our proposed mean-based tests (i.e., QXcat and QZ_{max}) and the variance-based test wM3VNA for X chromosomal SNPs and show the proof in Appendix B. Based on this, we construct two mean-variance-based tests $QMVX_{cat}$, by combining p_{wM3VNA} with p_{QXcat}, and $QMVZ_{max}$, by combining p_{wM3VNA} with $p_{QZ_{max}}$, based on Fisher's method (Fisher et al. 1967), i.e.,

$$QMVX_{cat}=-2\ln(p_{QXcat}p_{wM3VNA})$$

and

$$QMVZ_{max}=-2\ln(p_{QZ_{max}}p_{wM3VNA})$$

Under H_0^{MV}, both $QMVX_{cat}$ and $QMVZ_{max}$ asymptotically follow a chi-square distribution with 4 degrees of freedom (Chen et al. 2017).

RESULTS

Simulation settings

We evaluate the type I error rates (sizes) and the powers of our proposed methods $QMVX_{cat}$, $QMVZ_{max}$, QXcat and QZ_{max} by extensive simulation studies. Furthermore, we include wM3VNA, wM3V, Fisher, T_{chen} and T_{plink} for the comparison. Note that T_{chen} and T_{plink} do not consider the unequal variances of the trait value across different genotypes, which leads to false-positive results in the presence of variance heterogeneity. Therefore, we also include T_{chenw} and T_{plinkw}, which use the weighted least square method to estimate the regression coefficients. To clearly differentiate these 11 tests, we categorize them into three groups: methods testing for means (i.e., QXcat, QZ_{max}, T_{chenw}, T_{plinkw}, T_{chen} and T_{plink}), methods testing for variances (i.e., wM3VNA, wM3V and Fisher) and methods simultaneously testing for means and variances (i.e., $QMVX_{cat}$ and $QMVZ_{max}$). (q_f, q_m) is set as (0.2, 0.2), (0.2, 0.3) and (0.3, 0.2). ρ is taken as 0 and 0.05, where $\rho=0$ means HWE and $\rho\neq 0$ indicates the departure from HWE. We set the sample size *N* at 6000, and the sex ratio $n_f{:}n_m$ is fixed at 2:1, 1:1 and 1:2, which corresponds to $(n_f, n_m)=(4000, 2000)$, (3000, 3000) and (2000, 4000), respectively. The genotypes of females are generated from a trinomial distribution with probabilities (q_{aa}, q_{Aa}, q_{AA}), while the genotypes of males are simulated from a binomial distribution with probabilities $(1-q_m, q_m)$. Let *z* and *g* denote the sex and the genotype score, respectively. *z* is set to 1 for females and 0 for males. Under XCI, *g* takes the possible values of 0, γ and 2 for genotypes *aa*, *Aa* and *AA* in females, respectively, and values of 0 and 2 are taken for genotypes *a* and *A* in males, respectively. Different γ values represent different XCI patterns when XCI exists. Here, γ is fixed as 0, 0.5, 1, 1.5 and 2. Under XCI-E, *g* is set to 0, 1 and 2 for *aa*, *Aa* and *AA* in females, respectively, and 0 and 1 for *a* and *A* in males, respectively.

The trait value y_i for individual *i* can be generated by the following linear regression model:

$$y_i=\beta_c+\beta_g g_i+\beta_z z_i+\varepsilon_i, i=1,2,\ldots,N$$

where g_i and z_i denote the values of *g* and *z* of individual *i*, respectively, β_c is the intercept, β_g and β_z are the corresponding regression coefficients of g_i and z_i, respectively, and ε_i is the random error. Assume that y_i follows a normal distribution. The

Table 2. Values of ψ, β_g, γ, b, θ and τ in five simulated scenarios.

Scenario	pattern	Effect of SNP		ψ	β_g	γ	b	θ	τ
		Mean	Variance						
1	–	–	–	0	0	–	0	0	0
2	–	–	√	0	0	–	0	0.2	0.2
3	XCI-E	√	–	{0.3%,0.4%}	{0.085,0.098}	–	0	0	0
4	XCI	√	√	{0.3%,0.4%}	{0.085,0.098}	{0, 0.5, 1, 1.5, 2}	{0.085,0.098}	0.2	0.2
5	XCI-E	√	√	{0.3%,0.4%}	{0.085,0.098}	–	0	0.2	0.2

corresponding mean and variance of y_i with different coding schemes of g are shown in Table 1. We fix $\beta_c = \beta_z = 0.133$. $\beta_g = \sqrt{\frac{\psi\sigma^2}{2q_g(1-q_g)}}$, where σ^2 is the variance of the trait value for genotype aa in females (σ^2_{f0}) and that for genotype a in males (σ^2_{m0}), ψ denotes the proportion of the phenotypic variation due to the SNP effect on the means of the trait value and q_g is the allele frequency (Struchalin et al. 2010). In our simulations, we set $\sigma^2 = 1$, and $q_g = 0.3$, which is the maximum of q_f and q_m, respectively. To simulate the type I error rates of the methods testing for means, ψ is set to 0, which indicates that $\beta_g = 0$. To simulate the test powers of the mean-based tests, we fix ψ at 0.3% and 0.4%, and the corresponding values of β_g are 0.085 and 0.098, respectively. According to Ma et al. (2015b), $\frac{\gamma}{2}\left(1-\frac{\gamma}{2}\right)b^2$ in σ^2_{f1} under XCI denotes the increased variance caused by XCI for heterozygous females when the SNP has an effect on the means of the trait value, where b is the additive effect of the SNP on the trait value. Hence, when $\beta_g \neq 0$ and XCI exists, b takes the same value as β_g (i.e., $b = 0.085$ (0.098) if $\beta_g = 0.085$ (0.098)), while it is fixed to 0 when $\beta_g = 0$ or under XCI-E. θ in σ^2_{f1} represents the increased variance caused by factors other than XCI for heterozygous females (Ma et al. 2015b). If σ^2_{f1} is affected by factors other than XCI, θ is set to 0.2; otherwise, $\theta = 0$. τ in σ^2_{f2} and σ^2_{m1} is the additional variance of the trait value introduced by genotype AA in females or A in males. When the SNP influences σ^2_{f2} and σ^2_{m1}, τ is 0.2, while it is set to 0 for variance homogeneity. Finally, we use Models (1) and (2) to fit these simulated data.

Since QXcat, QZ_{max}, T_{chenw}, T_{plinkw}, T_{chen} and T_{plink} only test for the mean difference of the trait value, wM3VNA, wM3V and Fisher only test for the variance heterogeneity, and $QMVX_{cat}$ and $QMVZ_{max}$ test for the differences of both means and variances, we consider the following five scenarios: (1) the means and the variances of the trait value are not influenced by the SNP, (2) the variances of the trait value are affected by the SNP due to factors other than XCI for Aa females and AA females or A males, while the SNP has no effect on the means, (3) under XCI-E, the SNP affects the means while it has no influence on the variances, (4) under XCI, the SNP affects the means and the variances of the trait value because of XCI, specific genotypes (i.e., Aa and AA females or A males) and other factors, and (5) under XCI-E, the SNP affects the means and the variances of the trait value owing to the factors other than XCI for Aa females and AA females or A males. Note that for the case of XCI, if the SNP has an effect on the means, then this SNP will also have an effect on the variances. Therefore, we do not simulate the scenario under XCI in which the SNP affects the means but not the variances. The corresponding values of ψ, β_g, γ, b, θ and τ under the five simulated scenarios are displayed in Table 2. In scenario (1) (i.e., no SNP effect), we evaluate the sizes of all the considered methods. In scenario (2) (i.e., SNP effect on variances only), the sizes of the six mean-based tests (QXcat, QZ_{max}, T_{chenw}, T_{plinkw}, T_{chen} and T_{plink}), the test powers of the two mean-variance-based tests ($QMVX_{cat}$ and $QMVZ_{max}$) and the three variance-based tests (wM3VNA, wM3V and Fisher) are assessed. In scenario (3) (i.e., SNP effect on means only under XCI-E), the sizes of the three methods testing for variances are presented, and the test powers of the two mean-variance-based tests and the six mean-based tests are compared. In scenarios (4) and (5) (i.e., SNP effect on both means and variances), we compare the test powers of all the methods. The number of replications is fixed at 10^5, and the significance level is $\alpha = 10^{-4}$. To further assess the robustness of our proposed methods, we consider the situations where the trait value follows a log-normal distribution with the parameters being the natural logarithm of the means and the variances listed in Table 1. In this case, the trait value will be transformed by the inverse normal transformation method in advance, as recommended by Deng et al. (2019).

Empirical type I error rates

Scenario (1): no SNP effect. Table 3 provides a summary of the sizes of our proposed methods (i.e., $QMVX_{cat}$, $QMVZ_{max}$, QXcat and QZ_{max}) and the seven existing methods (i.e., T_{chenw}, T_{plinkw}, T_{chen}, T_{plink}, wM3VNA, wM3V and Fisher) in scenario (1) under HWE (i.e., $\rho = 0$) when the trait value follows a normal distribution. In Table 3, we find that all of these methods control the sizes well regardless of allele frequencies and sex ratios. Supplementary Table S1 shows the empirical sizes of all these methods when $\rho = 0.05$. It can be seen that the sizes of all the methods still maintain levels close to the nominal level 10^{-4}, and the values of ρ have little effect on the empirical sizes.

Scenario (2): SNP effect on variances only. Table 4 shows the estimated sizes of the six mean-based tests (QXcat, QZ_{max}, T_{chenw}, T_{plinkw}, T_{chen} and T_{plink}) in scenario (2) when $\rho = 0$ and 0.05, and the trait value follows a normal distribution. It should be noted that only the sizes of QXcat, QZ_{max}, T_{chenw} and T_{plinkw} are controlled well when the variances of the trait value are unequal, while the type I error rates of T_{chen} and T_{plink} are higher.

Power comparison

Scenario (2): SNP effect on variances only. The simulated powers of the two mean-variance-based tests ($QMVX_{cat}$ and $QMVZ_{max}$) and the three variance-based tests (wM3VNA, wM3V and Fisher) against n_f:n_m in scenario (2) under HWE when the trait value is normally distributed are displayed in Supplementary Fig. S1. It is shown in Supplementary Fig. S1 that wM3VNA has better performance in terms of power than the other methods. Because the mean-based tests QXcat and QZ_{max} give the type I error rates under scenario (2), the powers of $QMVX_{cat}$ and $QMVZ_{max}$ are close to each other and are less than those of the three methods for testing variances. Generally, when (q_f, q_m) remains unchanged, the powers of the five methods gradually become less when n_f:n_m changes from 2:1, 1:1 to 1:2 (i.e., more male individuals). The powers of these methods for (q_f, q_m) = (0.2, 0.3) and (0.3, 0.2) are higher than those for (q_f, q_m) = (0.2, 0.2) when n_f:n_m is fixed (Supplementary Fig. S1b vs. Supplementary Fig. S1a and Supplementary Fig. S1c vs. Supplementary Fig. S1a). The corresponding test powers when $\rho = 0.05$ are presented in

Table 3. Empirical sizes ($\times 10^{-4}$) of the mean-variance-based tests ($QMVX_{cat}$ and $QMVZ_{max}$), mean-based tests (QXcat, QZ_{max}, T_{chenw}, T_{plinkw}, T_{chen} and T_{plink}) and variance-based tests (wM3VNA, wM3V and Fisher) at the significance level of $\alpha = 10^{-4}$ based on 10^5 replications in scenario (1) (i.e., no SNP effect) when $\rho = 0$ and the trait value follows a normal distribution.

q_f	q_m	n_f	n_m	$QMVX_{cat}$	$QMVZ_{max}$	QX_{cat}	QZ_{max}	T_{chenw}	T_{plinkw}	T_{chen}	T_{plink}	wM3VNA	wM3V	Fisher
0.2	0.2	4000	2000	1.0	0.7	0.7	0.9	0.5	0.8	0.3	0.5	1.4	1.0	1.6
		3000	3000	1.1	1.2	1.0	0.6	1.3	0.7	1.4	0.9	1.0	1.4	1.4
		2000	4000	0.6	0.5	1.0	0.6	1.1	0.7	0.9	0.8	0.4	0.5	0.5
0.2	0.3	4000	2000	0.8	0.9	1.1	0.7	1.3	0.9	1.6	0.8	1.0	1.3	1.5
		3000	3000	0.5	0.5	0.5	0.7	0.7	0.7	0.7	0.5	0.8	0.6	1.3
		2000	4000	0.7	0.7	1.1	0.7	1.2	1.2	1.1	1.2	0.9	0.9	1.2
0.3	0.2	4000	2000	1.4	1.2	1.4	1.3	1.2	1.1	1.2	1.3	1.0	0.5	0.7
		3000	3000	0.9	1.0	0.5	0.9	0.5	0.5	0.3	0.5	1.6	1.6	1.3
		2000	4000	1.1	1.0	0.8	0.8	1.2	0.6	1.2	0.6	1.2	1.4	1.0

Supplementary Fig. S2. We find that the performances of the tests in Supplementary Fig. S2 are similar to those in Supplementary Fig. S1.

Scenario (3): SNP effect on means only under XCI-E. Under scenario (3), the methods for testing variances (wM3VNA, wM3V and Fisher) present the type I error rates instead of the test powers (data not shown for brevity). In addition, Supplementary Table S2 shows that when $\beta_g = 0.085$ and $\rho = 0$ for a normally distributed trait value in scenario (3), the powers of the existing mean-based tests T_{chen} and T_{plink} are very close to those of T_{chenw} and T_{plinkw}, respectively. Hence, we remove the simulation results of the three variance-based tests, T_{chen} and T_{plink} from all the figures under this scenario for simplicity. The estimated powers of the two methods for simultaneously testing means and variances ($QMVX_{cat}$ and $QMVZ_{max}$) and the four methods for testing means (QXcat, QZ_{max}, T_{chenw} and T_{plinkw}) against n_f:n_m in scenario (3) when $\beta_g = 0.085$, $\rho = 0$ and the trait value follows a normal distribution are plotted in Fig. 1. From Fig. 1, we find that the mean-based test QZ_{max} performs the best and the performance of the mean-variance-based test $QMVX_{cat}$ is the worst. Testing means using QXcat is more powerful than testing means using the mean-variance-based test $QMVZ_{max}$ or the two existing mean-based tests (i.e., T_{chenw} and T_{plinkw}). $QMVZ_{max}$ and T_{chenw} have similar performance in terms of power, and the power of T_{plinkw} is larger. All the methods in Fig. 1 become less powerful as n_f:n_m decreases (i.e., more male individuals). When n_f:n_m is unchanged, the powers of these methods when $(q_f, q_m) = (0.2, 0.3)$ and $(0.3, 0.2)$ are higher than those when $(q_f, q_m) = (0.2, 0.2)$ (Fig. 1b vs. Fig. 1a and Fig. 1c vs. Fig. 1a). The powers of these methods in scenario (3) (i.e., SNP effect on means only under XCI-E) when $\beta_g = 0.098$ and $\rho = 0$ are given in Supplementary Fig. S3, and the corresponding results for $\rho = 0.05$ when $\beta_g = 0.085$ and 0.098 are shown in Supplementary Figs. S4 and S5, respectively. From these figures, we can see that the power when $\beta_g = 0.098$ is higher than those when $\beta_g = 0.085$ (Supplementary Fig. S3 vs. Fig. 1 and Supplementary Fig. S5 vs. Supplementary Fig. S4). Different values of ρ have minimal effect on the power.

Scenarios (4) and (5): SNP effect on both means and variances. Since T_{chen} and T_{plink} for testing means have increased empirical sizes when the variances of the trait value across genotypes are unequal, we remove them from all the figures in scenarios (4) and (5). Figure 2 gives the estimated power of the two mean-variance-based tests ($QMVX_{cat}$ and $QMVZ_{max}$), the four mean-based tests (QXcat, QZ_{max}, T_{chenw} and T_{plinkw}) and the three variance-based tests (wM3VNA, wM3V and Fisher) against different γ values in scenario (4) (i.e., SNP effect on both means and variances under XCI) when $\beta_g = b = 0.085$, $\rho = 0$ and the trait value follows a normal distribution. We can see from Fig. 2 that the two mean-variance-based tests have almost the same performance in terms of power and are more powerful than the other tests. For the four methods testing for means, when $\gamma = 2$ and n_f:n_m = 2:1 or 1:1 (subplots 2a-2f of Fig. 2), the powers of QXcat, T_{chenw} and T_{plinkw} are close to each other and are slightly larger than that of QZ_{max}. However, when $\gamma = 2$ and n_f:n_m = 1:2 (subplots 2g-2i of Fig. 2), the four mean-based tests perform similarly. For the cases when $\gamma = 0$, the proposed QXcat test generally performs the best, and the other three mean-based methods have similar powers, except for the situations where $(q_f, q_m) = (0.3, 0.2)$. For the cases when $\gamma = 0$ and $(q_f, q_m) = (0.3, 0.2)$, the existing T_{plinkw} test has the least power when n_f:n_m = 2:1 or 1:1 (subplots 2c and 2f of Fig. 2), while the two existing tests (T_{chenw} and T_{plinkw}) have similar powers and perform worse than the two proposed tests (QXcat and QZ_{max}) when n_f:n_m = 1:2 (subplot 2i of Fig. 2). When $\gamma = 0.5$, 1 and 1.5, the powers of the four mean-based tests are not much different when n_f:n_m = 2:1 and 1:1 (subplots 2a-2f of Fig. 2), while the existing T_{chenw} test has the smallest power when n_f:n_m = 1:2 (subplots 2g-

Table 4. Empirical sizes ($\times 10^{-4}$) of the mean-based tests (QXcat, QZ_{max}, T_{chenw}, T_{plinkw}, T_{chen} and T_{plink}) at the significance level of $\alpha = 10^{-4}$ based on 10^5 replications in scenario (2) (i.e., SNP effect on variances only) when the trait value follows a normal distribution.

ρ	q_f	q_m	n_f	n_m	QX_{cat}	QZ_{max}	T_{chenw}	T_{plinkw}	T_{chen}	T_{plink}
0	0.2	0.2	4000	2000	0.8	1.2	0.9	1.0	2.6	2.2
			3000	3000	1.3	1.1	1.3	1.0	1.9	1.9
			2000	4000	1.1	0.8	1.3	1.3	2.7	2.1
	0.2	0.3	4000	2000	0.9	1.3	0.5	1.0	1.5	1.2
			3000	3000	0.9	0.9	1.1	0.9	1.9	1.8
			2000	4000	1.0	0.6	1.3	0.9	2.8	2.2
	0.3	0.2	4000	2000	0.7	0.5	1.1	0.8	1.6	1.0
			3000	3000	0.9	0.7	1.2	1.1	2.0	1.5
			2000	4000	1.0	0.9	1.3	0.8	2.5	1.6
0.05	0.2	0.2	4000	2000	1.2	0.7	1.1	1.2	2.8	2.3
			3000	3000	1.0	1.0	1.1	0.8	3.0	2.0
			2000	4000	0.9	1.2	0.4	1.4	2.6	2.6
	0.2	0.3	4000	2000	0.8	1.3	0.8	0.8	2.0	1.5
			3000	3000	0.9	1.4	0.8	0.5	1.6	1.4
			2000	4000	0.6	0.4	0.4	0.5	1.5	1.0
	0.3	0.2	4000	2000	0.8	1.0	1.4	0.8	1.8	1.3
			3000	3000	1.0	0.9	0.6	1.1	1.5	1.7
			2000	4000	1.0	1.6	0.9	1.2	2.1	2.6

2i of Fig. 2). In addition, the powers of the two mean-variance-based tests and four mean-based tests increase as γ increases, while the powers of the methods testing for variances under different values of γ are not different because the extra variance for heterozygous females caused by XCI (i.e., $\frac{\gamma}{2}\left(1-\frac{\gamma}{2}\right)b^2$) attains the maximum value of 0.0018 when $\gamma = 1$, which is very small. For each fixed (n_f, n_m), all the methods when (q_f, q_m) = (0.2, 0.3) and (0.3, 0.2) perform better than those when (q_f, q_m) = (0.2, 0.2) (e.g., Fig. 2b vs. Fig. 2a and Fig. 2c vs. Fig. 2a). For each value of (q_f, q_m), the two methods for simultaneously testing means and variances and the four methods for testing means become more powerful when n_f: n_m changes from 2:1, 1:1 to 1:2 (e.g., Fig. 2a vs. Fig. 2d, Fig. 2a vs. Fig. 2g and Fig. 2d vs. Fig. 2g), while the powers of the methods for testing variances generally appear less. These results indicate that larger values of q_f and q_m may improve the powers of all the methods and that the three variance-based tests can be more efficient with higher n_f: n_m (i.e., larger female individuals). However, a lower n_f: n_m (i.e., more male individuals) may cause the two methods simultaneously testing for means and variances and the four mean-based tests to be more powerful.

We plot the powers of all these methods in scenario (4) (i.e., SNP effect on both means and variances under XCI) when $\beta_g = b = 0.098$ and $\rho = 0$, and the corresponding results for $\rho = 0.05$ when $\beta_g = b = 0.085$ and $\beta_g = b = 0.098$ in Supplementary Figs. S6–S8, respectively. By comparing Fig. 2 with Supplementary Fig. S6 or comparing Supplementary Fig. S7 with Supplementary Fig. S8, we find that for the methods testing for variances, the powers when $\beta_g = b = 0.085$ are similar to those when $\beta_g = b = 0.098$ because for different values of γ, the additional variances caused by XCI (i.e., $\frac{\gamma}{2}\left(1-\frac{\gamma}{2}\right)b^2$) for $b = 0.085$ are close to those for $b = 0.098$; for the two mean-variance-based tests and the four mean-based tests, the powers when $\beta_g = b = 0.098$ are higher than those when $\beta_g = b = 0.085$.

The estimated powers of the two methods for simultaneously testing means and variances ($QMVX_{cat}$ and $QMVZ_{max}$), four methods for testing means (QXcat, QZ_{max}, T_{chenw} and T_{plinkw}) and three methods for testing variances (wM3VNA, wM3V and Fisher) against n_f:n_m in scenario (5) (i.e., SNP effect on both means and variances under XCI-E) when $\beta_g = 0.085$ and $\rho = 0$ are presented in Fig. 3. The corresponding results when $\beta_g = 0.098$ and $\rho = 0$ and those with $\rho = 0.05$ when $\beta_g = 0.085$ and 0.098 are given in Supplementary Figs. S9–S11. It can be seen from these figures that under scenario (5), $QMVZ_{max}$ for simultaneously testing means and variances is the most powerful, the two mean-variance-based tests are more powerful than the other seven methods, and the power of T_{chenw} for testing means is the worst. Among the four mean-based tests (QXcat, QZ_{max}, T_{chenw} and T_{plinkw}), the order of the performance in terms of power is $QZ_{max} > QXcat > T_{plinkw} > T_{chenw}$. In addition, the power performances of the three variance-based tests in Fig. 3 and Supplementary Figs. S9–S11 are similar to those in Supplementary Figs. S1 and S2.

Other simulation results

We also simulate the type I error rates and powers for all the considered test statistics for all the abovementioned situations when the trait value follows a log-normal distribution. The simulation results are shown in Supplementary Tables S3–S5 and Supplementary Figs. S12–S25. From Supplementary Tables S3–S5, all the sizes stay close to the nominal level, except for the mean-based tests T_{chen} and T_{plink} under scenario (2), where the variances across genotypes can be unequal. From Supplementary Figs. S12–S25, we find that the power performances of all the methods and the impact of (q_f, q_m), n_f: n_m, γ, ρ, β_g and b on the powers of all the methods in scenarios (2)–(5) are similar to those when the trait value is normally distributed.

APPLICATION TO THE MCTFR DATA

The Minnesota Center for Twin and Family Research Genome-Wide Association Study of Behavioral Disinhibition is a family-based study that includes age (covariate) and five quantitative traits: the nicotine composite score (NIC), the alcohol consumption composite score (CON), the alcohol dependence composite score (DEP), the behavioral disinhibition composite score (BD) and the illicit drug composite score (DRG). This dataset is available from the database of Genotypes and Phenotypes (https://www.ncbi.nlm.nih.gov/gap/) with the accession number phs000620.v1.p1. This dataset includes 2183 families and 7377 individuals, including 3546 males and 3831 females. There are four

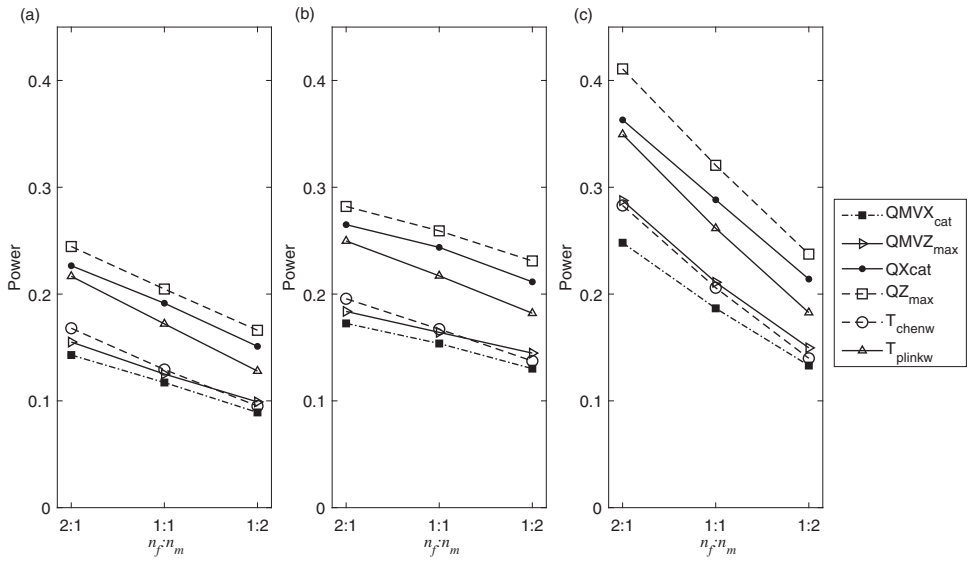

Fig. 1 Powers of the two mean-variance-based tests and the four mean-based tests against n_f:n_m. The two mean-variance-based tests are $QMVX_{cat}$ and $QMVZ_{max}$. The four mean-based tests are QXcat, QZ_{max}, T_{chenw} and T_{plinkw}. These results are based on 10^5 replications in scenario (3) (i.e., SNP effect on means only under XCI-E), where $N = 6000$, $\beta_g = 0.085$ and $\rho = 0$ at the significance level of $\alpha = 10^{-4}$ when the trait value follows a normal distribution. **a** $(q_f, q_m) = (0.2, 0.2)$. **b** $(q_f, q_m) = (0.2, 0.3)$. **c** $(q_f, q_m) = (0.3, 0.2)$.

types of offspring in this dataset, which are monozygotic twins, full biological nontwin siblings, adopted siblings and mixed siblings, which include one biological offspring and one adopted offspring. More details of the family structure in this dataset can be found in Fig. 7 of Li et al. (2021) and Supplementary Fig. S26 in this article for easy reference. In this dataset, 12,354 SNPs on the X chromosome are genotyped.

To ensure that the included individuals are independent, we only use the data of parents in the dataset. Then, the quality control procedures are conducted, in which we first exclude the individuals with a missing genotype rate greater than 10% and select the SNPs for which the minor allele frequencies are greater than 5%, the missing rates are less than 10%, the minimum genotype counts are larger than 20 and the p values of the HWE test are larger than 1×10^{-6} (Ma et al. 2015b; Soave et al. 2015; Marees et al. 2018). As a result, a total of 3649 independent individuals (1949 females and 1700 males) and 9963 SNPs are included in this application. We apply our proposed methods (i.e., $QMVX_{cat}$, $QMVZ_{max}$, QXcat and QZ_{max}) and the existing methods (i.e., T_{chenw}, T_{plinkw}, T_{chen}, T_{plink}, wM3VNA, wM3V and Fisher) to this subset of the MCTFR data.

Note that sex dimorphism of the five quantitative traits generally exists, and the histograms of the five traits for all the individuals, females only and males only are different in the MCTFR data, which are shown in Supplementary Fig. S27. Furthermore, all the residuals estimated from Models (1) and (2) fail to pass the normality tests. According to McCaw et al. (2019), we use the I-INT method to transform the five quantitative traits in females and males and then apply the 11 methods mentioned above to conduct the corresponding association analysis. Here, we include age as the covariate.

Since the five traits in this dataset share many similarities, similar to Schifano et al. (2013), we set the significance level to 1×10^{-3} to find the SNPs that are simultaneously associated with multiple traits. As a result, SNP rs808144 is identified to be simultaneously associated with four traits (BD, DEP, DRG and NIC). Table 5 shows the p values of all the methods for SNP rs808144, from which we discover that SNP rs808144 only influences the mean values of these four traits while having no effect on their variances (all the p values of the variance-based tests wM3VNA, wM3V and Fisher are larger than 1×10^{-3}). The p values of the proposed mean-based tests QXcat and QZ_{max} are close to those of the existing mean-based tests (T_{chenw}, T_{plinkw}, T_{chen} and T_{plink}). In addition, nine SNPs (rs808141, rs5934722, rs5926861, rs7064741, rs5942608, rs17261621, rs204332, rs5977759 and rs5925540) are found to be simultaneously associated with two traits. The p values of all the methods for these nine SNPs are given in Supplementary Table S6. Specifically, SNPs rs808141, rs5926861, rs7064741, rs204332, rs5977759 and rs5925540 only have effects on the mean values of the traits. Among these six SNPs, BD is statistically significantly associated with SNPs rs5926861, rs204332 and rs5925540; CON is only associated with SNP rs5977759; DEP is associated with SNPs rs5926861, rs7064741 and rs5977759; DRG is associated with SNPs rs808141, rs7064741 and rs204332; and NIC is associated with SNPs rs808141 and rs5925540. For SNPs rs808141, rs5926861, rs7064741 and rs5977759, the p values of six mean-based tests (QXcat, QZ_{max}, T_{chenw}, T_{plinkw}, T_{chen} and T_{plink}) are close to each other, while for SNP rs204332, the p values of QZ_{max} are much larger than those of the other five mean-based tests, and for SNP rs5925540, the p values of QZ_{max}, T_{plinkw} and T_{plink} are not much different and are much larger than those of QXcat, T_{chenw} and T_{chen}. In addition, we find that SNP rs5934722 only affects the variances of BD ($p_{wM3VNA} = 4.18 \times 10^{-4}$, $p_{wM3V} = 5.01 \times 10^{-4}$ and $p_{Fisher} = 5.04 \times 10^{-4}$) and DRG ($p_{wM3VNA} = 8.81 \times 10^{-4}$ and $p_{wM3V} = 8.70 \times 10^{-4}$). SNP rs17261621 is statistically significantly associated with the variance differences of BD ($p_{wM3V} = 7.66 \times 10^{-4}$). From the p values of all the methods for SNP rs17261621 and DRG, only the mean-variance-based test $QMVZ_{max}$ gives the statistically significant result ($p_{QMVZ_{max}} = 3.98 \times 10^{-4}$). Additionally, from the p values of all the methods for SNP rs5942608, only the p values of $QMVX_{cat}$ for simultaneously testing means and variances are lower than the significance level 1×10^{-3}, where the p values of $QMVX_{cat}$ for DEP and NIC are 4.53×10^{-4} and 8.02×10^{-4}, respectively. This indicates that either the means or the variances of the trait values across different genotypes are different, which needs to be further investigated.

We summarize the positions, minor alleles, major alleles, minor allele frequencies, p values of the HWE test and the genes consisting of the abovementioned 10 SNPs in Supplementary Table S7. We find that SNP rs5934722 is within the SHROOM2 gene, which is reported to be associated with autistic disorder and neurodevelopmental

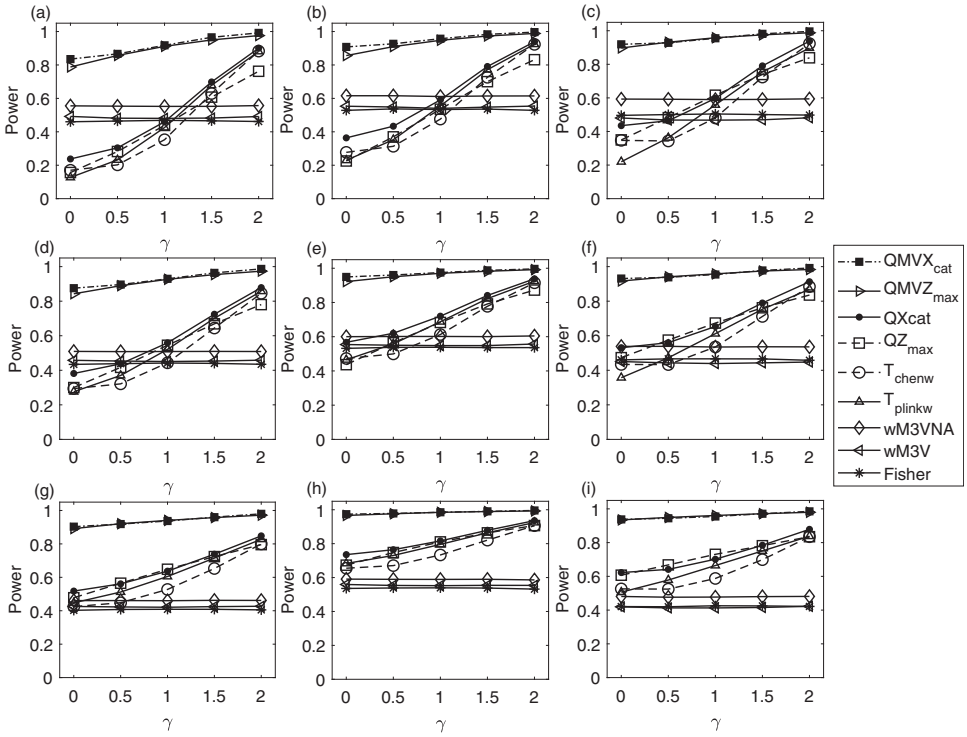

Fig. 2 Powers of the two mean-variance-based tests, four mean-based tests and three variance-based tests against γ. The two mean-variance-based tests are $QMVX_{cat}$ and $QMVZ_{max}$. The four mean-based tests are QXcat, QZ_{max}, T_{chenw} and T_{plinkw}. The three variance-based tests are wM3VNA, wM3V and Fisher. These results are based on 10^5 replications in scenario (4) (i.e., SNP effect on both means and variances under XCI), where $\beta_g = b = 0.085$ and $\rho = 0$ at the significance level of $\alpha = 10^{-4}$ when the trait value follows a normal distribution. **a** (n_f, n_m) = (4000, 2000) and (q_f, q_m) = (0.2, 0.2). **b** (n_f, n_m) = (4000, 2000) and (q_f, q_m) = (0.2, 0.3). **c** (n_f, n_m) = (4000, 2000) and (q_f, q_m) = (0.3, 0.2). **d** (n_f, n_m) = (3000, 3000) and (q_f, q_m) = (0.2, 0.2). **e** (n_f, n_m) = (3000, 3000) and (q_f, q_m) = (0.2, 0.3). **f** (n_f, n_m) = (3000, 3000) and (q_f, q_m) = (0.3, 0.2). **g** (n_f, n_m) = (2000, 4000) and (q_f, q_m) = (0.2, 0.2). **h** (n_f, n_m) = (2000, 4000) and (q_f, q_m) = (0.2, 0.3). **i** (n_f, n_m) = (2000, 4000) and (q_f, q_m) = (0.3, 0.2).

disorders (Kearney et al. 2011; Richards et al. 2015). SNP rs5926861 is included in the DCAF8L2 gene, which has been reported to be associated with autistic disorder, neurodevelopmental disorders and syndromic X-linked intellectual disability Lubs type (Kushima et al. 2018). SNP rs7064741 is located in the GLRA4 gene, which is related to intellectual disability, behavioral problems and craniofacial anomalies (Labonne et al. 2016). SNP rs5977759 is in the HS6ST2 gene, which is associated with the development of myopia and cognitive impairment (Paganini et al. 2019).

DISCUSSION

In this article, we propose four association tests ($QMVX_{cat}$, $QMVZ_{max}$, QXcat and QZ_{max}) for X-linked quantitative traits under the assumptions that the risk alleles for females and males are the same and the SNP being studied satisfies the generalized genetic model in females. Among these tests, QXcat and QZ_{max} focus on testing for the mean differences of quantitative traits, while $QMVX_{cat}$ and $QMVZ_{max}$ simultaneously test for both the mean and variance differences of quantitative traits. In addition, we choose two ways to incorporate the XCI information. In $QMVX_{cat}$ and QXcat, we introduce two indicator variables for females, which can be used in testing for the association under all the XCI patterns, and then directly combine the p values of the test statistics based on females and males. In $QMVZ_{max}$ and QZ_{max}, we combine the test statistics for females and males by different weights to consider different dosage compensation patterns and then obtain the test statistic by maximizing these combined test statistics. Extensive simulations are conducted to evaluate the type I error rates and the test powers of these proposed methods and the existing methods T_{chenw}, T_{plinkw}, T_{chen}, T_{plink}, wM3VNA, wM3V and Fisher. The simulation results show that our proposed methods control the type I error rates in various scenarios well. In the simulated scenarios where the mean values of the trait value are affected by the SNP, two proposed mean-based tests QXcat and QZ_{max} have better performance in terms of power than the existing methods for testing means under XCI-E and in some cases of XCI. In the simulated scenarios where both the means and the variances of the trait value are affected by the SNP, the two proposed mean-variance-based tests $QMVX_{cat}$ and $QMVZ_{max}$ outperform the others, as expected.

For the combination of p values, we use Fisher's method (Fisher et al. 1967), Stouffer's method (Stouffer et al. 1949) and Cauchy's method (Liu and Xie 2020) to combine the p value of wM3VNA for testing variances with those of QXcat and QZ_{max} for testing means to obtain the p values of $QMVX_{cat}$ and $QMVZ_{max}$ for simultaneously testing means and variances. In Stouffer's method, two p values are transformed to the p upper quantiles of the standard normal distribution, and then $\frac{1}{\sqrt{2}}$ times the sum of these two quantiles is used as the final test statistic, which follows the standard normal distribution under H_0^{MV}. In Cauchy's method, we first transform two p values to the corresponding quantiles of the standard Cauchy distribution and then calculate the average of these two quantiles as the final test statistic, which follows the standard Cauchy distribution under H_0^{MV}. We compare the test powers of the two mean-variance-based tests ($QMVX_{cat}$ and $QMVZ_{max}$) using the three combination methods under HWE for scenario (3) (i.e., SNP effect on means only under XCI-E) with $\beta_g = 0.085$ and scenario (4) (i.e., SNP effect on both means and variances under XCI) with $\beta_g = b = 0.085$ when the trait value follows a normal distribution. The estimated powers of both methods under scenario (3) are listed in Supplementary Table S8, and the corresponding results of $QMVX_{cat}$ and $QMVZ_{max}$ under scenario (4) are given in Supplementary Tables S9 and S10, respectively. From Supplementary Table S8, both $QMVX_{cat}$ and $QMVZ_{max}$ achieve the

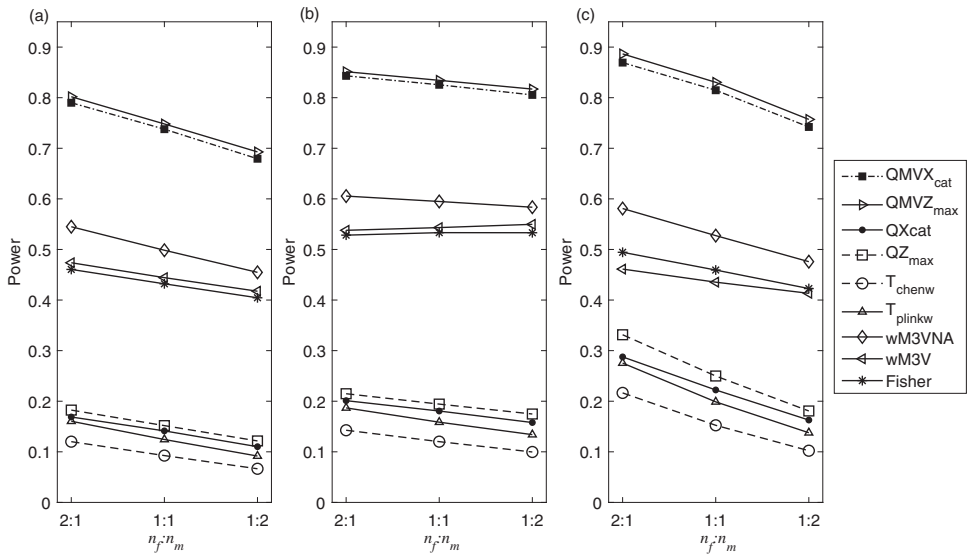

Fig. 3 Powers of the two mean-variance-based tests, four mean-based tests and three variance-based tests against n_f:n_m. The two mean-variance-based tests are $QMVX_{cat}$ and $QMVZ_{max}$. The four mean-based tests are QXcat, QZ_{max}, T_{chenw} and T_{plinkw}. The three variance-based tests are wM3VNA, wM3V and Fisher. These results are based on 10^5 replications in scenario (5) (i.e., SNP effect on both means and variances under XCI-E), where $N = 6000$, $\beta_g = 0.085$ and $\rho = 0$ at the significance level of $\alpha = 10^{-4}$ when the trait value follows a normal distribution. **a** (q_f, q_m) = (0.2, 0.2). **b** (q_f, q_m) = (0.2, 0.3). **c** (q_f, q_m) = (0.3, 0.2).

highest powers when using Cauchy's method in scenario (3), which are slightly larger than those with Fisher's method. The power using Stouffer's method are much less than those using the other two combination methods. In Supplementary Tables S9 and S10, we find that the test powers utilizing Fisher's method and Stouffer's method are close to each other, and both are much larger than that of Cauchy's method. Therefore, we finally choose the robust Fisher's method to construct the mean-variance-based tests $QMVX_{cat}$ and $QMVZ_{max}$. Additionally, Chen (2022a) recently proposed a method based on the constrained likelihood ratio test for combining independent p values and showed that this combination method is robust and powerful under many conditions. Moreover, two novel robust tests for combining dependent p values (i.e., MCM and CMC) were suggested by Chen (2022b). Both the simulation results and the real data application demonstrated that the MCM and CMC methods are robust and powerful under many situations and can be considered alternatives to Cauchy's method. We use the combination methods proposed in the work by Chen (2022a) and Chen (2022b) to calculate the p values of $QMVX_{cat}$ and $QMVZ_{max}$ for simultaneously testing the means and variances in the future and compare the powers of $QMVX_{cat}$ and $QMVZ_{max}$ utilizing these three methods with those using Fisher's method.

For the mean-based test QXcat, we consider three combination methods to construct the test statistic. The first way is to directly combine two p values for females (i.e., p_{f1}^A and p_{f2}^A if the risk allele is A) with the p value for males (i.e., p_m^A if the risk allele is A) based on Fisher's method and obtain the corresponding test statistic. The second way is to first combine two test statistics for females (i.e., T_{f1}^A and T_{f2}^A) to T_f^A and compute the corresponding p value, then combine it with the p value for males based on Fisher's method. The third way is the one we choose for QXcat in this article, which has been introduced in the Materials and methods section. The power performances of QXcat under three combinations are also compared in different scenarios, and we find that QXcat under the third combination achieves the highest power in general (data not shown for brevity).

For the mean-based test QZ_{max}, two test statistics that incorporate more dosage compensation patterns, i.e., $QZ_{max3} = \max(|T_{\lambda_1}|, |T_{\lambda_{1.5}}|, |T_{\lambda_2}|)$ and $QZ_{max5} = \max(|T_{\lambda_1}|, |T_{\lambda_{1.25}}|, |T_{\lambda_{1.5}}|, |T_{\lambda_{1.75}}|, |T_{\lambda_2}|)$, are also considered. We compare their power performance with $QZ_{max} = \max(|T_{\lambda_1}|, |T_{\lambda_2}|)$ under HWE and (q_f, q_m) = (0.2, 0.2) for scenario (3) (i.e., SNP effect on means only under XCI-E) with $\beta_g = 0.085$ and scenario (4) (i.e., SNP effect on both means and variances under XCI) with $\beta_g = b = 0.085$ when the trait value follows a normal distribution. The corresponding results are given in Supplementary Table S11, which shows that the powers of QZ_{max}, QZ_{max3} and QZ_{max5} are close to each other. Note that QZ_{max3} and QZ_{max5} are much more computationally intense than QZ_{max}. Therefore, we recommend choosing the test statistic $QZ_{max} = \max(|T_{\lambda_1}|, |T_{\lambda_2}|)$ in practice.

The proposed mean-based tests QXcat and QZ_{max} assume that the risk alleles for females and males are the same, and the SNP being studied satisfies the generalized genetic model in females (i.e., $\mu_{f2} \geq \mu_{f1} \geq \mu_{f0}$). When these two assumptions are satisfied in practice, the methods of constructing the test statistics QXcat and QZ_{max} can effectively incorporate the information from these two assumptions and hence can improve the test powers. For instance, if the risk alleles in females and males are both A and $\mu_{f2} > \mu_{f1} > \mu_{f0}$, the signs of β_{f1} and β_{f2} in Model (1) and that of β_{m1} in Model (2) are the same, and all of them are positive. For QXcat, the one-sided p values $p_{f1}^A = 1 - \Phi(T_{f1}^A)$ and $p_{f2}^A = 1 - \Phi(T_{f2}^A)$ are smaller than the one-sided p values $p_{f1}^a = 1 - \Phi(T_{f1}^a)$ and $p_{f2}^a = 1 - \Phi(T_{f2}^a)$, respectively, in females. Thus, $Q_f^A = -2\ln(p_{f1}^A p_{f2}^A)$ is larger than $Q_f^a = -2\ln(p_{f1}^a p_{f2}^a)$, and the corresponding p values satisfy $p_f^A < p_f^a$. Similarly, p_m^A is smaller than p_m^a in males. Then, $Q^A = -2\ln(p_f^A p_m^A)$ is larger than $Q^a = -2\ln(p_f^a p_m^a)$. By utilizing the information that p_{f1}^A, p_{f2}^A and p_m^A are smaller than p_{f1}^a, p_{f2}^a. and p_m^a, respectively, a final test statistic with a relatively large absolute value is obtained by maximizing Q^A and Q^a, so the test power of QXcat = max(Q^A, Q^a) will increase. For QZ_{max}, because both T_{f1}^A and T_{f2}^A should be positive with a high probability, $T_f^A = \frac{1}{\sqrt{2}}(T_{f1}^A + T_{f2}^A)$ is also larger than zero. In addition, note that both T_f^A and T_m^A should be greater than zero, so the signs of $T_{\lambda_1} = \sqrt{\lambda_1}T_f^A + \sqrt{1-\lambda_1}T_m^A$ and $T_{\lambda_2} = \sqrt{\lambda_2}T_f^A + \sqrt{1-\lambda_2}T_m^A$ are the same as those of T_f^A and T_m^A. Therefore, the test power of $QZ_{max} = \max(|T_{\lambda_1}|, |T_{\lambda_2}|)$ is improved by the weighted average of the test statistics T_{f1}^A, T_{f2}^A and T_m^A having the same signs.

However, if either of these two assumptions is violated, both QXcat and QZ_{max} may lose the test power. For example, considering

Table 5. *p* values of the mean-variance-based tests ($QMVX_{cat}$ and $QMVZ_{max}$), mean-based tests (QXcat, QZ_{max}, T_{chenw}, T_{plinkw}, T_{chen} and T_{plink}) and variance-based tests (wM3VNA, wM3V and Fisher) for SNP rs808144, which is statistically significantly associated with four traits at the significance level of 1×10^{-3}.

Trait	$QMVX_{cat}$	$QMVZ_{max}$	QXcat	QZ_{max}	T_{chenw}	T_{plinkw}	T_{chen}	T_{plink}	wM3VNA	wM3V	Fisher
BD[a]	2.70×10^{-3}	3.85×10^{-3}	3.95×10^{-4}	5.90×10^{-4}	9.05×10^{-4}	6.24×10^{-4}	8.23×10^{-4}	5.66×10^{-4}	7.48×10^{-1}	6.36×10^{-1}	6.12×10^{-1}
DEP[b]	2.25×10^{-4}	3.85×10^{-4}	2.01×10^{-5}	3.63×10^{-5}	5.78×10^{-5}	3.09×10^{-5}	6.55×10^{-5}	3.54×10^{-5}	9.41×10^{-1}	7.63×10^{-1}	8.27×10^{-1}
DRG[c]	2.33×10^{-4}	5.90×10^{-4}	4.83×10^{-5}	1.34×10^{-4}	1.03×10^{-4}	1.12×10^{-4}	1.08×10^{-4}	1.18×10^{-4}	4.07×10^{-1}	5.65×10^{-1}	6.88×10^{-1}
NIC[d]	1.15×10^{-4}	1.09×10^{-4}	6.71×10^{-4}	6.32×10^{-4}	1.79×10^{-3}	8.65×10^{-4}	2.27×10^{-3}	1.09×10^{-3}	1.36×10^{-2}	3.28×10^{-2}	1.50×10^{-2}

[a]BD: behavioral disinhibition composite score.
[b]DEP: alcohol dependence composite score.
[c]DRG: illicit drug composite score.
[d]NIC: nicotine composite score.

the situation where the risk allele in females is *A* while that in males is *a*, the signs of β_{f1} and β_{f2} are different from that of β_{m1} (i.e., $p_f^A < p_f^a$ and $p_m^A > p_m^a$). Then, both the test statistics $Q^A = -2\ln\left(p_f^A p_m^A\right)$ and $Q^a = -2\ln\left(p_f^a p_m^a\right)$ are less than $-2\ln\left(p_f^A p_m^a\right)$ (assuming that the risk alleles for females and males are known), and QXcat = max(Q^A, Q^a) is not the best combination of p_{f1}^A, p_{f2}^A, p_{f1}^a, p_{f2}^a, p_m^A and p_m^a, which may reduce the test power. For QZ_{max}, the signs of T_f^A and T_m^A should be different with a high probability; then, T_f^A and T_m^A may be canceled out in the calculation of $T_{\lambda_1} = \sqrt{\lambda_1}T_f^A + \sqrt{1-\lambda_1}T_m^A$ and $T_{\lambda_2} = \sqrt{\lambda_2}T_f^A + \sqrt{1-\lambda_2}T_m^A$. In this case, a smaller value of the final test statistic $QZ_{max} = \max(|T_{\lambda_1}|, |T_{\lambda_2}|)$ will be obtained, and hence, the power of QZ_{max} will be reduced. However, if the SNP being studied does not satisfy the generalized genetic model in females (e.g., $\mu_{f1} > \mu_{f2} > \mu_{f0}$), $\beta_{f1} > 0$ and $\beta_{f2} < 0$ (i.e., $p_{f1}^A < p_{f1}^a$ and $p_{f2}^A > p_{f2}^a$) when *A* is assumed to be the risk allele. As such, $Q_f^A = -2\ln\left(p_{f1}^A p_{f2}^A\right)$ and $Q_f^a = -2\ln\left(p_{f1}^a p_{f2}^a\right)$ are smaller than $-2\ln\left(p_{f1}^A p_{f2}^a\right)$. Hence, the final test statistic QXcat = max(Q^A, Q^a) may be very small, and the test power may be low. For QZ_{max}, T_{f1}^A (larger than zero) and T_{f2}^A (less than zero) can be canceled out in calculating $T_f^A = \frac{1}{\sqrt{2}}\left(T_{f1}^A + T_{f2}^A\right)$. Therefore, the final test statistic $QZ_{max} = \max(|T_{\lambda_1}|, |T_{\lambda_2}|)$ is reduced, and the corresponding test power is lower. Note that it is generally reasonable to assume that the risk alleles for females and males are the same and that the SNP being studied satisfies the generalized genetic model for females (Chen et al. 2017). Furthermore, the ideas of constructing the test statistics QXcat and QZ_{max} are similar to those in Chen et al. (2017) and Wang et al. (2019a), respectively, and both the simulation results of Chen et al. (2017) and Wang et al. (2019a) showed that the powers of their proposed methods are generally higher than those of other existing association tests. Additionally, under the simulated scenarios, both the proposed mean-based tests QXcat and QZ_{max} have better performance in power than the existing mean-based tests under XCI-E and in some cases of XCI. We also apply all the considered methods to the MCTFR data, and some further discussions on the violation of the assumptions can be found in Appendix C.

In this article, we consider the departure from HWE by fixing the inbreeding coefficient ρ at 0.05. To further assess the validity of our proposed methods without the HWE assumption, we simulate the following population stratification model by referring to the simulation settings of Haldar and Ghosh (2012) and Xia et al. (2013). Suppose that the whole population consists of two subpopulations, each of which is HWE. The sample of size $N = 6000$ is composed of N_1 and N_2 individuals from the first and second subpopulations, respectively. The ratio N_1:N_2 is set to be 2:3 and 1:1, and the sex ratio in each subpopulation is fixed at 2:1, 1:1 and 1:2. Let q_{f1} and q_{m1} (q_{f2} and q_{m2}) denote the frequencies of allele *A* for females and males in the first (second) subpopulation, respectively, and (q_{f1}, q_{m1}, q_{f2}, q_{m2}) are assumed to be (0.1, 0.1, 0.9, 0.9) and (0.2, 0.2, 0.5, 0.5), respectively. The simulated type I error rates of four proposed tests ($QMVX_{cat}$, $QMVZ_{max}$, QXcat and QZ_{max}) under scenario (1) (i.e., no SNP effect) when $\rho = 0$ and the trait value follows a normal distribution are shown in Supplementary Table S12, while the empirical sizes of two mean-based tests QXcat and QZ_{max} under scenario (2) (i.e., SNP effect on variances only) when $\rho = 0$ and the trait value follows a normal distribution are presented in Supplementary Table S13. It can be seen from these two tables that our proposed methods can control the sizes well, which verifies their validity under population stratification.

Our proposed methods have several advantages. First, the proposed mean-variance-based tests have higher powers than the existing methods in the simulated scenarios where both the means and the variances of the trait value across different genotypes are different. Second, our methods incorporate XCI information in two different ways that are necessarily considered when conducting X chromosome association tests. Third, we use the information of the two sexes, which improves the test power. Nonetheless, there are some limitations in our methods. When two assumptions (i.e., the risk alleles in females and males are the

same and the genetic effect of heterozygous females is between those of two homozygous females) are not satisfied in practice, the powers of the proposed association tests may decrease. In addition, these methods cannot test for the association between SNP sets and a trait. These methods cannot incorporate the information of family structure, which results in a loss of power and needs to be improved in the future. In summary, our proposed methods not only effectively consider the XCI but are also powerful under XCI-E and in some cases of XCI.

SOFTWARE

The R package QMVtest is publicly available at https://github.com/yuxinyuanqt/QMVtest, which is implemented by R software (version 4.1.2).

DATA AVAILABILITY

The MCTFR data used for the analyses described in this article can be found on the database of Genotypes and Phenotypes with accession number phs000620.v1.p1, and dbGaP request numbers 86747-6 and 95621-5 (https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000620.v1.p1).

REFERENCES

Al-Ayadhi LY, Qasem HY, Alghamdi HAM, Elamin NE (2020) Elevated plasma X-linked neuroligin 4 expression is associated with autism spectrum disorder. Med Princ Pr 29:480–485

Amos-Landgraf JM, Cottle A, Plenge RM, Friez M, Schwartz CE, Longshore J et al. (2006) X chromosome-inactivation patterns of 1,005 phenotypically unaffected females. Am J Hum Genet 79:493–499

Auer PL, Teumer A, Schick U, O'Shaughnessy A, Lo KS, Chami N et al. (2014) Rare and low-frequency coding variants in CXCR2 and other genes are associated with hematological traits. Nat Genet 46:629–634

Brown AA, Buil A, Viñuela A, Lappalainen T, Zheng HF, Richards JB et al. (2014) Genetic interactions affecting human gene expression identifed by variance association mapping. Elife 3:e01381

Brown CJ, Carrel L, Willard HF (1997) Expression of genes from the human active and inactive X chromosomes. Am J Hum Genet 60:1333–1343

Brown MB, Forsythe AB (1974) Robust tests for the equality of variances. J Am Stat Assoc 69:364–367

Cao Y, Wei P, Bailey M, Kauwe JSK, Maxwell TJ (2014) A versatile omnibus test for detecting mean and variance heterogeneity. Genet Epidemiol 38:51–59

Carrel L, Park C, Tyekucheva S, Dunn J, Chiaromonte F, Makova KD (2006) Genomic environment predicts expression patterns on the human inactive X chromosome. PLoS Genet 2:e151

Carrel L, Willard HF (2005) X-inactivation profle reveals extensive variability in X-linked gene expression in females. Nature 434:400–404

Chang D, Gao F, Slavney A, Ma L, Waldman YY, Sams AJ et al. (2014) Accounting for eXentricities: analysis of the X chromosome in GWAS reveals X-linked genes implicated in autoimmune diseases. PLoS One 9:e113684

Chen B, Craiu RV, Strug LJ, Sun L (2021) The X factor: A robust and powerful approach to X-chromosome-inclusive whole-genome association studies. Genet Epidemiol 45:694–709

Chen B, Craiu RV, Sun L (2020) Bayesian model averaging for the X-chromosome inactivation dilemma in genetic association study. Biostatistics 21:319–335

Chen ZX (2022a) Optimal tests for combining p-values. Appl Sci 12:322

Chen ZX (2022b) Robust tests for combining p-values under arbitrary dependency structures. Sci Rep. 12:3158

Chen ZX, Ng HKT (2012) A robust method for testing association in genome-wide association studies. Hum Hered 73:26–34

Chen ZX, Ng HKT, Li J, Liu Q, Huang H (2017) Detecting associated single-nucleotide polymorphisms on the X chromosome in case control genome-wide association studies. Stat Methods Med Res 26:567–582

Chung RH, Morris RW, Zhang L, Li YJ, Martin ER (2007) X-APL: an improved family-based test of association in the presence of linkage for the X chromosome. Am J Hum Genet 80:59–68

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

Deng WQ, Mao S, Kalnapenkis A, Esko T, Mägi R, Paré G et al. (2019) Analytical strategies to include the X-chromosome in variance heterogeneity analyses: evidence for trait-specifc polygenic variance structure. Genet Epidemiol 43:815–830

Ding J, Lin S, Liu Y (2006) Monte Carlo pedigree disequilibrium test for markers on the X chromosome. Am J Hum Genet 79:567–573

Fisher B, Costich ER, Ganz M, Stanford JW (1967) Questions & answers. J Am Dent Assoc 75:799

Gaukrodger N, Mayosi BM, Imrie H, Avery P, Baker M, Connell JMC et al. (2005) A rare variant of the leptin gene has large effects on blood pressure and carotid intima-medial thickness: a study of 1428 individuals in 248 families. J Med Genet 42:474–478

Haldar T, Ghosh S (2012) Effect of population stratifcation on false positive rates of population-based association analyses of quantitative traits. Ann Hum Genet 76:237–245

Hickey PF, Bahlo M (2011) X chromosome association testing in genome wide association studies. Genet Epidemiol 35:664–670

Horvath S, Laird NM, Knapp M (2000) The transmission/disequilibrium test and parental-genotype reconstruction for X-chromosomal markers. Am J Hum Genet 66:1161–1167

Jin H, Park T, Won S (2017) Efficient statistical method for association analysis of X-linked variants. Hum Hered 82:50–63

Kang HM, Sul JH, Service SK, Zaitlen NA, Kong SY, Freimer NB et al. (2010) Variance component model to account for sample structure in genome-wide association studies. Nat Genet 42:348–354

Kearney HM, Thorland EC, Brown KK, Quintero-Rivera F, South ST (2011) American college of medical genetics standards and guidelines for interpretation and reporting of postnatal constitutional copy number variants. Genet Med 13:680–685

Konzman D, Abramowitz LK, Steenackers A, Mukherjee MM, Na HJ, Hanover JA (2020) O-GlcNAc: regulator of signaling and epigenetics linked to X-linked intellectual disability. Front Genet 11:605263

Kushima I, Aleksic B, Nakatochi M, Shimamura T, Okada T, Uno Y et al. (2018) Comparative analyses of copy-number variation in autism spectrum disorder and schizophrenia reveal etiological overlap and biological insights. Cell Rep. 24:2838–2856

Labonne JDJ, Graves TD, Shen YP, Jones JR, Kong IK, Layman LC et al. (2016) A microdeletion at Xq22. 2 implicates a glycine receptor GLRA4 involved in intellectual disability, behavioral problems and craniofacial anomalies. BMC Neurol 16:132

Levene H (1961) Robust tests for equality of variances. Contributions to Probability and Statistics: 279–292.

Li BH, Yu WY, Zhou JY (2021) A statistical measure for the skewness of X chromosome inactivation for quantitative traits and its application to the MCTFR data. BMC Genom Data 22:24

Liu Y, Xie J (2020) Cauchy combination test: a powerful test with analytic p-value calculation under arbitrary dependency structures. J Am Stat Assoc 115:393–402

Loley C, Ziegler A, König IR (2011) Association tests for X-chromosomal markers–a comparison of different test statistics. Hum Hered 71:23–36

Lyon MF (1961) Gene action in the X-chromosome of the mouse (Mus musculus L.). Nature 190:372–373

Ma C, Boehnke M, Lee S, GoT2D Investigators (2015a) Evaluating the calibration and power of three gene-based association tests of rare variants for the X chromosome. Genet Epidemiol 39:499–508

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

Marees AT, Kluiver HD, Stringer S, Vorspan F, Curis E, Marie-Claire C et al. (2018) A tutorial on conducting genome-wide association studies: quality control and statistical analysis. Int J Methods Psychiatr Res 27:e1608

McCaw ZR, Lane JM, Saxena R, Redline S, Lin X (2019) Operating characteristics of the rank-based inverse normal transformation for quantitative trait analysis in genome-wide association studies. Biometrics 76:1262–1272

Minks J, Robinson WP, Brown CJ (2008) A skewed view of X chromosome inactivation. J Clin Invest 118:20–23

Morley M, Molony CM, Weber TM, Devlin JL, Ewens KG, Spielman RS et al. (2004) Genetic analysis of genome-wide variation in human gene expression. Nature 430:743–747

Mosteller F, Fisher RA (1948) Questions and answers. Am Stat 2:30–31

Özbek U, Lin HM, Lin Y, Weeks DE, Chen W, Shaffer JR et al. (2018) Statistics for X-chromosome associations. Genet Epidemiol 42:539–550

Paganini L, Hadi LA, Chetta M, Rovina D, Fontana L, Colapietro P et al. (2019) A HS6ST2 gene variant associated with X-linked intellectual disability and severe myopia in two male twins. Clin Genet 95:368–374

R Core Team (2020) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. Austria, http://www.R-project.org/

Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J et al. (2015) Standards and guidelines for the interpretation of sequence variants: a joint consensus

recommendation of the american college of medical genetics and genomics and the association for molecular pathology. Genet Med 17:405–423
Schifano ED, Li L, Christiani DC, Lin X (2013) Genome-wide association analysis for multiple continuous secondary phenotypes. Am J Hum Genet 92:744–759
Soave D, Corvol H, Panjwani N, Gong J, Li W, Boëlle PY 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
Song YL, Biernacka JM, Winham SJ (2021) Testing and estimation of X-chromosome SNP effects: Impact of model assumptions. Genet Epidemiol 45:577–592
Stouffer SA, Suchman EA, DeVinney LC, Star SA, Williams Jr RM (1949) *The american soldier: adjustment during army life. (studies in social psychology in World War II)*. Princeton Univ. Press.
Struchalin MV, Dehghan A, Witteman JCM, Duijn CV, Aulchenko YS (2010) Variance heterogeneity analysis for detection of potentially interacting genetic loci: method and its limitations. BMC Genet 11:92
Wang J, Yu R, 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
Wang P, Xu SQ, Wang BQ, Fung WK, Zhou JY (2019a) A robust and powerful test for case-control genetic association study on X chromosome. Stat Methods Med Res 28:3260–3272
Wang P, Zhang Y, Wang BQ, Li JL, Wang YX, Pan D et al. (2019b) A statistical measure for the skewness of X chromosome inactivation based on case-control design. BMC Bioinforma 20:11
Wise AL, Gyi L, Manolio TA (2013) eXclusion: toward integrating the X chromosome in genome-wide association analyses. Am J Hum Genet 92:643–647
Wong CCY, Caspi A, Williams B, Houts R, Craig IW, Mill J (2011) A longitudinal twin study of skewed X chromosome-inactivation. PLoS One 6:e17873
Wu H, Luo J, Yu H, Rattner A, Mo A, Wang Y et al. (2014) Cellular resolution maps of X-chromosome inactivation: implications for neural development, function, and disease. Neuron 81:103–119
Xia F, Zhou JY, Fung WK (2013) Powerful tests for association on quantitative trait loci incorporating imprinting effects. J Hum Genet 58:384–390
Xu W, Hao M (2018) A unifed partial likelihood approach for X-chromosome association on time-to-event outcomes. Genet Epidemiol 42:80–94
Yang J, Loos RJF, Powell JE, Medland SE, Speliotes EK, Chasman DI et al. (2012) FTO genotype is associated with phenotypic variability of body mass index. Nature 490:267–272
Zhang L, Martin ER, Chung RH, Li YJ, Morris RW (2008) X-LRT: a likelihood approach to estimate genetic risks and test association with X-linked markers using a case-parents design. Genet Epidemiol 32:370–380
Zheng G, Joo J, Zhang C, Geller NL (2007) Testing association for markers on the X chromosome. Genet Epidemiol 31:834–843

ACKNOWLEDGEMENTS

The authors are grateful to the editor, the associate editor and three anonymous reviewers for insightful comments which greatly improved the presentation of the materials.

FUNDING

This work was supported by the National Natural Science Foundation of China (grant numbers 82173619 and 81773544), the Science and Technology Planning Project of Guangdong Province (grant number 2020B1212030008), the Hong Kong RGC GRF (grant number 17302919) and the National and Guangzhou University Students' Innovation and Enterprise Training Project of China (grant number 201912121019). The Minnesota Center for Twin and Family Research was supported by the National Institute on Drug Abuse (grant number U01 DA024417). The sample ascertainment and data collection in MCTFR data were supported by the National Institute on Drug Abuse (grant numbers R37 DA05147 and R01 DA13240), the National Institute on Alcohol Abuse and Alcoholism (grant numbers R01 AA09367 and R01 AA11886), and the National Institute of Mental Health (grant number R01 MH66140). The funding bodies played no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41437-022-00560-y.

Correspondence and requests for materials should be addressed to Wing Kam Fung or Ji-Yuan Zhou.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.