Journal of Genetics (2019)98:99
https://doi.org/10.1007/s12041-019-1146-6

RESEARCH ARTICLE

X-chromosome genetic association test incorporating X-chromosome inactivation and imprinting effects

WEI LIU[1], BEI-QI WANG[1], GUOJUN LIU-FU[1], WING KAM FUNG[2]* and JI-YUAN ZHOU[1]*

[1]*State Key Laboratory of Organ Failure Research, Ministry of Education, and Guangdong Provincial Key Laboratory of Tropical Disease Research, Department of Biostatistics, School of Public Health, Southern Medical University, Guangzhou 510515, Guangdong, People's Republic of China*
[2]*Faculty of Science, Department of Statistics and Actuarial Science, The University of Hong Kong, Pok Fu Lam, Hong Kong, People's Republic of China*

*For correspondence. E-mail: Wing Kam Fung, wingfung@hku.hk; Ji-Yuan Zhou, zhoujiyuan5460@hotmail.com.

Received 10 January 2019; revised 29 July 2019; accepted 7 August 2019

Abstract. Studies have shown that many complex diseases are sex-determined. When conducting genetic association studies on X-chromosome, there are two important epigenetic factors which should be considered simultaneously: X-chromosome inactivation and genomic imprinting. Currently, there have been several association tests accounting for the information on X-chromosome inactivation. However, these tests do not take the imprinting effects into account. In this paper, we propose a novel association test simultaneously incorporating X-chromosome inactivation and imprinting effects based on case–parent trios and control–parent trios for female offspring and case–control data for male offspring, denoted by MLR_{XCII}. Extensive simulation studies are carried out to investigate the type I error rate and the test power of the proposed MLR_{XCII}. Simulation results demonstrate that the proposed test controls the type I error rate well and is more powerful than the existing method when imprinting effects exist. The proposed MLR_{XCII} test is valid and powerful in genetic association studies on X-chromosome for qualitative traits and thus is recommended in practice.

Keywords. association test; genomic imprinting; X-chromosome inactivation.

Introduction

Genomewide association study (GWAS) is a useful tool to identify candidate single-nucleotide polymorphism markers for complex diseases (Pearson and Manolio 2008; Hindorff *et al*. 2009). Many studies have shown that the mutants on X chromosome are related to X-linked diseases (Wise *et al*. 2013), such as autoimmune disorders (Voskuhl 2011), cardiovascular disease (Lerner and Kannel 1986; Mendelsohn and Karas 2005), cancer susceptibility (Cohn *et al*. 1996; Naugler *et al*. 2007) and psychiatric disorders (Breslau *et al*. 1997; Hankin and Abramson 2001). To date, many association tests have been developed for autosomes, of which only a few have been proposed for identifying X-chromosomal loci. However, most of these association tests have either ignored X chromosome or have limited power when testing for the sex-specific impact of genetic variation (Kukurba *et al*. 2016). There are several potential reasons to explain this phenomenon, such as the lack of coverage on GWAS chips, the different number of X chromosomes between two sexes and the difference in the minor-allele frequency of variants on X chromosome (Wise *et al*. 2013). Further, the expression of about 5–10% of genes in the mammalian genome is affected by X-chromosome inactivation (XCI) and genomic imprinting (Carrel and Willard 2005; Lee and Bartolomei 2013). Some disturbing diseases are caused by these classic epigenetic processes when not appropriately regulated in human (Lee and Bartolomei 2013). XCI is originally hypothesized by Lyon (1962). Females have two copies of X chromosomes while males have only one. Thus, XCI is an important process to equalize X-linked gene dosage between XX females and XY males. XCI takes place in early embryonic development in human, by which one of two copies of X chromosomes in females is inactivated and then stably inherits through subsequent somatic cell divisions, leading to each female diploid somatic cell having paternal or maternal X chromosome active (Lyon 1962; Lyon 1972; Migeon *et al*. 2001). In

Published online: 04 November 2019

human, the initiation of XCI contains three steps: counting, choosing and silencing (Avner and Heard 2001). The recognition step counts the number of X chromosomes and then chooses which X chromosome to be inactive. Finally, the X chromosome chosen to be inactive will be silenced by gene methylation. XCI is generally recognized as a random process (XCI-R) with an equal probability that either X chromosome will be inactivated in a given cell. However, some studies claimed that skewed or nonrandom XCI is a biological plausibility (Amos-Landgraf *et al*. 2006; Wong *et al*. 2011). The skewed XCI (XCI-S) is defined as the observation of inactivation of the same allele in 75% or 80% of cells (Penny *et al*. 1996; Kay 1998; Minks *et al*. 2008). Some studies even suggested extreme or severe skewed inactivation patterns with more than 90% of cells having one of the alleles inactive. Most X-linked genes are subject to XCI in human, while some X-linked genes escape from silencing and thus have higher expression levels in females than in males (Deng *et al*. 2014). It has been estimated that about three quarters of genes silence one copy of the two X chromosomes in females, while the remaining may undergo the escape from XCI (XCI-E). As such, under XCI-E, females carry two copies of active X chromosomes.

Note that XCI only occurs on X chromosome. However, genomic imprinting is a well-known epigenetic phenomenon, not only for autosomal genes but also for X chromosomal genes (Constância *et al*. 1998; Loesch *et al*. 2005). Genomic imprinting as an important mechanism of gene regulation causes genes to be expressed in a parent-of-origin-specific manner (Reik and Walter 2001; Bartolomei 2009; Ferguson-Smith 2011; Patten *et al*. 2014). Males have only one maternally inherited X chromosome, which is always expressed. As for females, if the allele or region on maternal X chromosome is functionally active and the paternal allele is silenced, then we refer to this phenomenon as paternal imprinting. On the contrary, maternal imprinting is the phenomenon with the allele or region on maternal X chromosome is not expressed. Imprinting happens in the parental germ cells, which influences embryonic growth in the womb and behaviour after birth (Zhou *et al*. 2018). Up to 1% of genes undergo imprinting effects (Tycko and Morison 2002). Then, the imprinted genes on X chromosome may also play an important role in studying the complex diseases and traits. Some studies claimed that XCI is initially affected by imprinting during early embryonic development of some mammals, which means the paternally inherited X chromosome is preferentially inactivated in the first cells to differentiate, which later give rise to extraembryonic tissues (Avner and Heard 2001). Over the past 50 years, many studies have found that there are some common features between XCI and imprinting (Lee 2000; Lee and Bartolomei 2013). Both XCI and genomic imprinting are regulated by the *cis*-acting master control regions. However, XCI-R is associated with *CpG* hypermethylation of genes on the inactive X, which has not been found for the imprinted mechanism. Moreover, XCI-R is subject to a zygotic X chromosome counting mechanism, while imprinted X inactivation is not subject to either mathematical or stochastic constraints (Lee 2000).

Currently, some statistical methods have been proposed to test for association between traits and genes on X chromosome. Zheng *et al*. (2007) proposed several association tests on X chromosome, which compared the frequency of each allele or genotype in cases and that in controls by assuming XCI-E. Two of them were based on the allele counting under the assumption of Hardy–Weinberg equilibrium, while other two were based on the genotypes under the assumption of Hardy–Weinberg disequilibrium. With different combinations of tests for male and female samples, Zheng *et al*. (2007) showed that the optimal choice of X-specific test statistic depends on whether Hardy–Weinberg equilibrium holds and whether males and females have the same risk allele at the locus. But recently published GWAS literature revealed that the methods of Zheng *et al*. (2007) may lose the power when XCI occurs (Hickey and Bahlo 2011; Wang *et al*. 2018). Clayton (2008) was the first one to take account of XCI-R in association study on X chromosome. By treating males as homozygous females, Clayton (2008) proposed two chi-square tests. One was analogous to a Cochran–Armitage trend test, which is applied under additive models, and the other was analogous to a Pearson's chi-square test which is adapted for other situations. By accounting for XCI-R, these methods greatly improve the test power, while also have a few limitations. Clayton's methods are sensitive to Hardy–Weinberg disequilibrium for requiring that the allele frequencies between two sexes are equal (Wang *et al*. 2018). On the other hand, these methods do not consider other XCI patterns, such as XCI-E and XCI-S. Since both Zheng *et al*.'s (2007) and Clayton's (2008) methods have seemingly not gained widespread use in GWAS, Gao *et al*. (2015) developed a chromosome X-wide analysis toolset to analyse genetic variation on X chromosome in association studies. To handle different XCI patterns on X chromosome, Wang *et al*. (2014) developed a unified test by maximizing likelihood ratio (LR) over four biological models: XCI-R, XCI-S in the direction of the deleterious allele, XCI-S in the direction of the normal allele and XCI-E. Because the true underlying XCI pattern is generally unknown in practice, the method of Wang *et al*. (2014) coded the three genotypes in females as 0, γ and 2, where γ varies between 0 and 2. By comparing with Clayton's 1-degree-of-freedom test, this method improves the test power in the scenarios where XCI is skewed, while loses the power when XCI-R or XCI-E happens. However, all of the methods mentioned above either assume XCI-R or account for both XCI-S and XCI-R, but they all ignore the information on genomic imprinting.

Therefore, in this paper, we propose a comprehensive model by simultaneously incorporating XCI and imprinting effects into association analysis on X chromosome. Note that the imprinting is based on the parent-of-origin of an allele on X chromosome, so we add two variables to our

model and divide the imprinting effects into the maternal and paternal effects based on case–parent trios and control–parent trios for female offspring and case–control data for male offspring by following Yang and Lin (2012) and Wang *et al.* (2014), respectively. By using the maximum likelihood estimation, we first construct the likelihood functions under the null hypothesis and the alternative hypothesis, respectively. Then, we obtain the LR. Note that the true underlying XCI pattern is usually unknown, therefore, we perform a grid search in which the γ value ranges from 0 to 2 and choose the maximum of LR's. Since the distribution of the proposed method is unknown under the null hypothesis of no association, we use a permutation procedure to obtain the empirical P value of the proposed method. An extensive simulation study is conducted to investigate the type I error rate and the power of the proposed test. The simulation results show that our proposed method is generally more powerful than the existing method when imprinting effects exist.

Materials and methods

Consider a candidate single-nucleotide polymorphism locus on X chromosome, having the deleterious allele D and the normal allele d. Males have only one copy of X chromosome, which is inherited from mother. Thus, males are hemizygote, denoted by D and d. As for female offspring, there are four possible genotypes D/D, D/d, d/D and d/d, where the left allele of the slash is paternal and the right one is maternal. However, the parent-of-origin of allele D in heterozygous female offspring requires the information on her parental genotypes. Here, we assume that the risk alleles are the same for two sexes. Let Y denote the disease status of an individual, with $Y = 1$ and 0 representing that the individual is affected and unaffected, respectively. In this paper, an affected daughter together with her parents is regarded as a case-parent trio and an unaffected daughter together with her parents is termed as a control–parent trio (Deng and Chen 2001; Li *et al.* 2016). Suppose that we collect N_f daughter–parent trios consisting of N_{f1} case–parent trios and N_{f0} control–parent trios, and N_m single males including N_{m1} male cases and N_{m0} male controls, with the sample size of N subjects where $N = N_f + N_m$. Let $N_1 = N_{f1} + N_{m1}$ and $N_0 = N_{f0} + N_{m0}$. As such, we have $N_f = N_{f1} + N_{f0}$, $N_m = N_{m1} + N_{m0}$ and $N = N_1 + N_0 = N_f + N_m = N_{f1} + N_{f0} + N_{m1} + N_{m0}$.

Suppose that (u, v, w) denotes the proportions of genotypes DD, Dd and dd for females and (s, t) is the proportions of genotypes D and d for males in the parental generation. Then, there are six possible mating types, which are listed in table 1, together with the corresponding random mating frequencies and the resulting female offspring compositions. From table 1, the frequencies of genotypes D/D, D/d, d/D and d/d for female offspring are $us + 1/2vs = s(u + 1/2v)$, $ws + 1/2vs = s(w + 1/2v)$, $ut + 1/2vt = t(u + 1/2v)$ and $wt + 1/2vt = t(w + 1/2v)$, respectively. If we define $p_F = u + 1/2v$ and $p_M = s$ as the frequencies of deleterious allele D for females and males in the parental generation, respectively, then when random mating holds in the parental generation, the probabilities of genotypes D/D, D/d, d/D and d/d for female offspring are $p_{f2} = p_M p_F$, $p_{f10} = p_M(1 - p_F)$, $p_{f01} = (1 - p_M)p_F$ and $p_{f0} = (1 - p_F)(1 - p_M)$. In the same way, the probabilities of genotypes D and d for male offspring are $p_{m1} = p_F$ and $p_{m0} = 1 - p_F$ (Hickey and Bahlo 2011).

Table 1. Mating types together with the corresponding random mating frequencies and the resulting female offspring composition.

Mating type	Random mating frequency	Female offspring composition
$D \times DD$	us	D/D
$d \times DD$	ut	d/D
$D \times Dd$	vs	$1/2D/D + 1/2D/d$
$d \times Dd$	vt	$1/2d/D + 1/2d/d$
$D \times dd$	ws	D/d
$d \times dd$	wt	d/d

To incorporate XCI-R, XCI-S and XCI-E, and imprinting effects into association analysis, we use the random variables X_1 and X_2 to respectively denote the paternal allele score and the maternal allele score of the daughter in a daughter-parent trio, and the codes of X_1 and X_2 of a single male are used to distinguish XCI from XCI-E, which are listed in table 2, with $\gamma \in (0, 2)$ being the degree of skewness of XCI. Under XCI-E, the males with D are regarded as the heterozygous female offspring with genotype d/D and $\gamma = 1$. Further, let X_3 denotes the sex of the individual in the offspring generation with 0 and 1 being male and female, respectively. Then, the association between the disease status Y and the allele scores X_1 and X_2 by adjusting the sex and other covariate variables $\mathbf{Z}$ is given by

$$\begin{aligned}&\text{Logit}(P(Y = 1|X_1, X_2, X_3, \mathbf{Z}))\\ &\quad = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \mathbf{b}^{\mathbf{T}}\mathbf{Z},\end{aligned} \tag{1}$$

Table 2. Coding of X_1 and X_2 for females and males under XCI/XCI-E and imprinting.

		XCI		XCI-E	
Sex	Genotype	X_1	X_2	X_1	X_2
Female	D/D	γ	$2 - \gamma$	1	1
	D/d	γ	0	1	0
	d/D	0	$2 - \gamma$	0	1
	d/d	0	0	0	0
Male	D	0	2	0	1
	d	0	0	0	0

where β_0 is the intercept, and β_1, β_2 and β_3 are the regression coefficients for X_1, X_2 and X_3, respectively. Note that $\mathbf{Z}$ denotes other covariates, such as environmental factors and $\mathbf{b}$ is the corresponding regression coefficients. For d/d female offspring and d males, suppose that $X_1 = 0$ and $X_2 = 0$ (i.e. the reference category). Then, $\beta_1 = \beta_2 = 0$ means that there is no association (i.e. the null hypothesis).

To explain equation (1) more clearly under the alternative hypothesis (i.e. association being present), suppose that the genotype distribution of female offspring in the case–parent trios and that in the control–parent trios follow multinomial distributions with probabilities $(h_{f2}, h_{f10}, h_{f01}, h_{f0})$ and $(g_{f2}, g_{f10}, g_{f01}, g_{f0})$, respectively, where $h_{f2}(g_{f2})$, $h_{f10}(g_{f10})$, $h_{f01}(g_{f01})$ and $h_{f0}(g_{f0})$ are the frequencies of genotypes D/D, D/d, d/D and d/d of female offspring in the case (control)–parent trios, respectively. Let $h_{f0}/g_{f0} = e_f$, $h_{f2}/g_{f2} = \lambda_{f2}e_f$, $h_{f10}/g_{f10} = \lambda_{f10}e_f$ and $h_{f01}/g_{f01} = \lambda_{f01}e_f$, where λ_{f2}, λ_{f10} and λ_{f01} are the odds ratios of genotypes D/D, D/d and d/D compared to d/d in female offspring, respectively. As such, $\lambda_{f2} = \exp(\beta_1\gamma + \beta_2(2-\gamma))$, $\lambda_{f10} = \exp(\beta_1\gamma)$ and $\lambda_{f01} = \exp(\beta_2(2-\gamma))$. Further, assume that the genotype distribution of male cases and that of male controls follow binomial distributions with probabilities (h_{m1}, h_{m0}) and (g_{m1}, g_{m0}), respectively. Let $h_{m0}/g_{m0} = e_m$. Then, $h_{m1}/g_{m1} = \lambda_m e_m$, where $\lambda_m = \exp(2\beta_2)$ is the odds ratio of D compared to d in male subjects. Then, $\lambda_{f2} = \lambda_m$ is indicative of dosage compensation. By $h_{f2} + h_{f10} + h_{f01} + h_{f0} = 1$ and $h_{m1} + h_{m0} = 1$, it is easy to show that

$$e_f = \frac{1}{g_{f2}\exp(\beta_1\gamma + \beta_2(2-\gamma)) + g_{f10}\exp(\beta_1\gamma) + g_{f01}\exp(\beta_2(2-\gamma)) + g_{f0}},$$

and

$$e_m = \frac{1}{g_{m1}\exp(2\beta_2) + g_{m0}}.$$

Taking γ= 0, 0.5, 1, 1.5 and 2 as examples, if $X_1 = 0$ and $X_2 = 2$ for D males (i.e. XCI is present) and $\lambda_m = \exp(2\beta_2)$, then:

(a) When $\gamma = 0$ and $\beta_2 \neq 0$, we have $\lambda_{f2} = \lambda_{f01} = \exp(2\beta_2)$ and $\lambda_{f10} = 1$, which means that complete paternal imprinting effect, all the cells in female offspring with genotype d/D have D active and all the cells in female offspring with genotype D/d have d active.

(b) When $\gamma = 0.5$, $\beta_1 \neq 0$ and $\beta_2 \neq 0$, we have $\lambda_{f2} = \exp(0.5\beta_1 + 1.5\beta_2)$, $\lambda_{f10} = \exp(0.5\beta_1)$ and $\lambda_{f01} = \exp(1.5\beta_2)$. If $\lambda_{f10} = \lambda_{f01}$, then there is no imprinting effects. If $\beta_1 = \beta_2 = \beta \neq 0$, then 75% of the cells in female offspring with genotype d/D have D active and the other 25% of the cells have d active; 25% of the cells in female offspring with genotype D/d keep D active and the other 75% of the cells keep d active.

(c) When $\gamma = 1$ and $\beta_1 = \beta_2 = \beta \neq 0$, we have $\lambda_{f2} = \exp(2\beta)$ and $\lambda_{f10} = \lambda_{f01} = \exp(\beta)$, which means that there is no imprinting effects and XCI is random.

(d) When $\gamma = 1.5$, $\beta_1 \neq 0$ and $\beta_2 \neq 0$, we have $\lambda_{f2} = \exp(1.5\beta_1 + 0.5\beta_2)$, $\lambda_{f10} = \exp(1.5\beta_1)$ and $\lambda_{f01} = \exp(0.5\beta_2)$. If $\lambda_{f10} = \lambda_{f01}$, then there is no imprinting effects. If $\beta_1 = \beta_2 = \beta \neq 0$, then 25% of the cells in female offspring with genotype d/D have D active and the other 75% of the cells have d active; 75% of the cells in female offspring with genotype D/d keep D active and the other 25% of the cells keep d active.

(e) When $\gamma = 2$ and $\beta_1 \neq 0$, we have $\lambda_{f2} = \lambda_{f10} = \exp(2\beta_1)$ and $\lambda_{f01} = 1$, which means that complete maternal imprinting effect, all the cells in female offspring with genotype d/D keep d active and all the cells in female offspring with genotype D/d keep D active.

(f) When $\lambda_{f10} \neq \lambda_{f01}$, $\lambda_{f10} \neq 1$ and $\lambda_{f01} \neq 1$, which means the incomplete imprinting effect.

On the other hand, $\beta_1 = \beta_2 = \beta \neq 0$, $\gamma = 1$ for female offspring, $X_1 = 0$ and $X_2 = 1$ for D males and $\lambda_m = \exp(\beta_2)$ indicate the scenario of XCI-E and no imprinting effects. Thus, equation (1) incorporates the information on XCI-R, XCI-S, XCI-E and imprinting effects.

From equation (1), the prevalence of the ith subject given his/her genotype is

$$P(y_i = 1 | x_{1i}, x_{2i}, x_{3i}, \mathbf{z}_i) = \frac{\exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \mathbf{b}^{\mathbf{T}}\mathbf{z}_i)}{1 + \exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \mathbf{b}^{\mathbf{T}}\mathbf{z}_i)}, \quad i = 1, 2, \ldots, N, \tag{2}$$

where y_i is the disease status of the ith subject, and x_{1i}, x_{2i}, x_{3i} and $\mathbf{z}_i$ are the values of X_1, X_2, X_3 and $\mathbf{Z}$ for the ith subject, respectively. Then, the likelihood function of the N subjects is as follows,

$$L(\mathbf{Y}|\mathbf{X_1}, \mathbf{X_2}, \mathbf{X_3}, \mathbf{Z}; \beta_0, \beta_1, \beta_2, \beta_3, \mathbf{b}) = \prod_{i=1}^{N} \left(\frac{\exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \mathbf{b}^{\mathbf{T}}\mathbf{z}_i)}{1 + \exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \mathbf{b}^{\mathbf{T}}\mathbf{z}_i)} \right)^{y_i} \times \left(\frac{1}{1 + \exp(\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \mathbf{b}^{\mathbf{T}}\mathbf{z}_i)} \right)^{1-y_i}, \tag{3}$$

where $\mathbf{Y} = (y_1, y_2, \ldots, y_N)^T$, $\mathbf{X_1} = (x_{11}, x_{12}, \ldots, x_{1N})^T$, $\mathbf{X_2} = (x_{21}, x_{22}, \ldots, x_{2N})^T$, $\mathbf{X_3} = (x_{31}, x_{32}, \ldots, x_{3N})^T$ and $\mathbf{Z} = (\mathbf{z_1}, \mathbf{z_2}, \ldots, \mathbf{z_N})^T$. The maximum likelihood estimates of β_0, β_1, β_2, β_3 and $\mathbf{b}$ obtained from equation (3) are denoted by $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$ and $\hat{\mathbf{b}}$, respectively. Similarly, under the null hypothesis of no association($H_0 : \beta_1 = \beta_2 = 0$), we have the following likelihood function

$$L(\mathbf{Y}|\mathbf{X_1}, \mathbf{X_2}, \mathbf{X_3}, \mathbf{Z}; \beta_0, \beta_3, \mathbf{b}) = \prod_{i=1}^{N} \left(\frac{\exp(\beta_0 + \beta_3 x_{3i} + \mathbf{b^T z}_i)}{1 + \exp(\beta_0 + \beta_3 x_{3i} + \mathbf{b^T z}_i)} \right)^{y_i} \times \left(\frac{1}{1 + \exp(\beta_0 + \beta_3 x_{3i} + \mathbf{b^T z}_i)} \right)^{1-y_i}, \quad (4)$$

and the corresponding maximum likelihood estimates of β_0, β_3 and $\mathbf{b}$ are denoted by $\tilde{\beta}_0$, $\tilde{\beta}_3$ and $\tilde{\mathbf{b}}$, respectively. Therefore, we get the LR as follows

$$LR_\gamma(\mathbf{X_1}, \mathbf{X_2}, \mathbf{X_3}, \mathbf{Z}) = \frac{L\left(\mathbf{Y}|\mathbf{X_1}, \mathbf{X_2}, \mathbf{X_3}, \mathbf{Z}; \hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\mathbf{b}}\right)}{L\left(\mathbf{Y}|\mathbf{X_1}, \mathbf{X_2}, \mathbf{X_3}, \mathbf{Z}; \tilde{\beta}_0, \tilde{\beta}_3, \tilde{\mathbf{b}}\right)}. \quad (5)$$

Note that the underlying biological XCI pattern and the imprinting effect models are generally unknown and a grid search for γ can be used. However, following Wang *et al.* (2014), a grid search using a small step function has very little effect on the *LR* values. Thus, we set γ=0, 1 and 2, and XCI-E is also included, just like Wang *et al.* (2014). Then, four LRs LR_0, LR_1, LR_2 and LR_E are obtained, where LR_E is the LR under XCI-E. Then, we regard the maximum of these four LRs as the test statistic for association, which simultaneously incorporates XCI and genomic imprinting and is then denoted by $MLR_{XCII} = \max(LR_0, LR_1, LR_2, LR_E)$ in this paper.

Permutation-based procedure for empirical P value of the proposed method

Note that the underlying distribution of MLR_{XCII} under the null hypothesis of no association is unknown. Thus, we use a permutation-based procedure which is stratified by sex to calculate the *P* value of MLR_{XCII}. Specifically, the steps are given as follows.

(a) For a collected sample of size N consisting of N_{f1} case–parent trios, N_{f0} control–parent trios, N_{m1} male cases and N_{m0} male controls, compute the value of MLR_{XCII}.

(b) Randomly permute the values of the disease statuses of daughters among N_f daughter–parent trios and those of N_m single males, respectively, and keep the genotypes of all the individuals in the sample unchanged.

(c) For each permuted sample, we calculate the value of MLR_{XCII}, denoted by MLR^*_{XCII}.

(d) Repeat steps (b) and (c) B times, which results in B test statistics $MLR^*_{XCII,1}$, $MLR^*_{XCII,2}$,…, and $MLR^*_{XCII,B}$.

(e) The empirical *P* value of the original MLR_{XCII} can be estimated as

$$P = \frac{1}{B} \sum_{j=1}^{B} I_{\{MLR^*_{XCII,j} > MLR_{XCII}\}},$$

where $I_{\{\cdot\}}$ is an indicator function.

Results

Simulation settings

To assess the performance of the proposed MLR_{XCII} method and compare MLR_{XCII} with the existing method of Wang *et al.* (2014), we conduct the following simulation study under various simulation settings. For simplicity, we did not take any other covariates into account except for sex. The sex of the subject is simulated based on the sex ratio. Let $r_f : r_m$ denote the ratio of the number of affected daughters among N_{f1} case–parent trios to the number of affected males (i.e. N_{m1}), which is fixed at 2:3, 1:1 and 3:2, while the sex ratio of the number of unaffected daughters among N_{f0} control–parent trios to the number of unaffected males (i.e. N_{m0}) is set at 1:1. The total sample size N is selected to be 1000 with N_1 = 500 and N_0 = 500. As such, according to the sex ratios, we respectively obtain N_{f1} = 200, 250 and 300 case–parent trios and N_{m1} = 300, 250 and 200 male cases, and N_{f0} = 250 control–parent trios and N_{m0} = 250 male controls.

Let (p_F, p_M), the frequencies of the deleterious allele D for females and males in the parental population, take the values of (0.2, 0.3), (0.3, 0.3) and (0.3, 0.2). Then, the genotype distributions of females and males in the offspring population are

$$(G_{f2}, G_{f10}, G_{f01}, G_{f0}) = (p_M p_F, p_M(1-p_F), (1-p_M)p_F, (1-p_M)(1-p_F)),$$

and $(G_{m1}, G_{m0}) = (p_F, 1-p_F)$, respectively. Since γ takes a value between 0 and 2, hence we set γ to be 0, 0.5, 1, 1.5 and 2. According to Wang *et al.* (2014), β_0 and β_3 are fixed at $\beta_0 = -2.5500$ and $\beta_3 = 0.4055$, respectively. To simulate the type I error rate of MLR_{XCII}, the regression coefficients β_1 and β_2 are fixed at $\beta_1 = \beta_2 = 0$. To investigate the empirical power of MLR_{XCII}, we consider the following different XCI patterns and/or imprinting scenarios. Under XCI, (β_1, β_2) is taken to be (0.2624, 0), (0, 0.2624) and (0.2624, 0.2624) with $\exp(0.2624) = 1.3$. There are some special cases under XCI. $\lambda_{f2} = \lambda_{f01} = \exp(2\beta_2)$ and $\lambda_{f10} = 1$ with $\beta_2 = 0.2624$ and γ=0 mean complete paternal imprinting

Table 3. Empirical type I error rates of MLR_{XCII} and the method of Wang *et al*. (2014) at significance level of $\alpha = 5\%$ based on 1000 replications with $\beta_0 = -2.5500$ and $\beta_3 = 0.4055$ under XCI.

			r_f:r_m=2:3		r_f:r_m = 1:1		r_f:r_m = 3:2	
p_F	p_M	γ	MLR_{XCII}	Wang	MLR_{XCII}	Wang	MLR_{XCII}	Wang
0.2	0.3	0	0.053	0.050	0.057	0.055	0.053	0.053
		0.5	0.050	0.057	0.057	0.048	0.051	0.062
		1	0.065	0.062	0.062	0.060	0.054	0.054
		1.5	0.052	0.049	0.052	0.046	0.054	0.048
		2	0.041	0.047	0.044	0.052	0.056	0.048
0.3	0.3	0	0.053	0.052	0.045	0.048	0.052	0.048
		0.5	0.061	0.062	0.059	0.067	0.056	0.058
		1	0.049	0.043	0.040	0.037	0.057	0.058
		1.5	0.052	0.046	0.054	0.056	0.043	0.047
		2	0.048	0.046	0.063	0.048	0.053	0.044
0.3	0.2	0	0.053	0.051	0.054	0.064	0.051	0.045
		0.5	0.051	0.055	0.057	0.062	0.059	0.056
		1	0.045	0.048	0.049	0.043	0.052	0.053
		1.5	0.045	0.048	0.048	0.044	0.051	0.059
		2	0.058	0.049	0.046	0.050	0.057	0.054

Table 4. Empirical type I error rates of MLR_{XCII} and the method of Wang *et al*. (2014) at significance level of $\alpha = 5\%$ based on 1000 replications with $\beta_0 = -2.5500$ and $\beta_3 = 0.4055$ under XCI-E.

		r_f:r_m = 2:3		r_f:r_m = 1:1		r_f:r_m = 3:2	
p_F	p_M	MLR_{XCII}	Wang	MLR_{XCII}	Wang	MLR_{XCII}	Wang
0.2	0.3	0.060	0.060	0.053	0.052	0.056	0.044
0.3	0.3	0.055	0.060	0.055	0.054	0.049	0.056
0.3	0.2	0.049	0.047	0.053	0.055	0.048	0.048

effect. Similarly, $\lambda_{f2} = \lambda_{f10} = \exp(2\beta_1)$ and $\lambda_{f01} = 1$ with $\beta_1 = 0.2624$ and $\gamma = 2$ are indicative of complete maternal imprinting effect. $\exp(\beta_1\gamma) = \exp(\beta_2(2-\gamma))$ suggests no imprinting effects. Further, $\exp(\beta_1\gamma) = \exp(\beta_2(2-\gamma))$ with $\beta_1 = \beta_2 = 0.2624$ and $\gamma = 1$ indicate that there is no imprinting effects and XCI is random. From equation (2), under XCI, the penetrances of the female offspring given genotypes *D*/*D*, *D*/*d*, *d*/*D* and *d*/*d* are respectively

$$\phi_{f2} = \frac{\exp(\beta_0 + \beta_1\gamma + \beta_2(2-\gamma) + \beta_3)}{1 + \exp(\beta_0 + \beta_1\gamma + \beta_2(2-\gamma) + \beta_3)},$$

$$\phi_{f10} = \frac{\exp(\beta_0 + \beta_1\gamma + \beta_3)}{1 + \exp(\beta_0 + \beta_1\gamma + \beta_3)},$$

$$\phi_{f01} = \frac{\exp(\beta_0 + \beta_2(2-\gamma) + \beta_3)}{1 + \exp(\beta_0 + \beta_2(2-\gamma) + \beta_3)},$$

$$\text{and } \phi_{f0} = \frac{\exp(\beta_0 + \beta_3)}{1 + \exp(\beta_0 + \beta_3)}.$$

Similarly, the penetrances of the male subjects given genotypes *D* and *d* are respectively $\phi_{m1} = \exp(\beta_0 + 2\beta_2)/1 + \exp(\beta_0 + 2\beta_2)$ and $\phi_{m0} = \exp(\beta_0)/1 + \exp(\beta_0)$. Thus, the genotype distribution of the female subjects in the case–parent trios and that in the control–parent trios follow multinomial distributions with probabilities:

$$(h_{f2}, h_{f10}, h_{f01}, h_{f0}) = \left(\frac{G_{f2}\phi_{f2}}{A_f}, \frac{G_{f10}\phi_{f10}}{A_f}, \frac{G_{f01}\phi_{f01}}{A_f}, \frac{G_{f0}\phi_{f0}}{A_f}\right)$$

and

$$(g_{f2}, g_{f10}, g_{f01}, g_{f0}) = \left(\frac{G_{f2}(1-\phi_{f2})}{1-A_f}, \frac{G_{f10}(1-\phi_{f10})}{1-A_f}, \frac{G_{f01}(1-\phi_{f01})}{1-A_f}, \frac{G_{f0}(1-\phi_{f0})}{1-A_f}\right),$$

respectively, where $A_f = G_{f2}\phi_{f2} + G_{f10}\phi_{f10} + G_{f01}\phi_{f01} + G_{f0}\phi_{f0}$ denotes the disease prevalence of the female subject. In the same way, the genotype distribution of the male cases and that in the male controls follow binomial distributions with probabilities

$$(h_{m1}, h_{m0}) = \left(\frac{G_{m1}\phi_{m1}}{A_m}, \frac{G_{m0}\phi_{m0}}{A_m}\right) \text{ and}$$

$$(g_{m1}, g_{m0}) = \left(\frac{G_{m1}(1-\phi_{m1})}{1-A_m}, \frac{G_{m0}(1-\phi_{m0})}{1-A_m}\right),$$

respectively, where $A_m = G_{m1}\phi_{m1} + G_{m0}\phi_{m0}$ is the disease prevalence of the male subject. Therefore, given $r_f : r_m$, (p_F, p_M), γ, β_0, β_1, β_2 and β_3, case–parent trios, control–parent trios, male cases and male controls are generated based on $(h_{f2}, h_{f10}, h_{f01}, h_{f0})$, $(g_{f2}, g_{f10}, g_{f01}, g_{f0})$, (h_{m1}, h_{m0}) and (g_{m1}, g_{m0}), respectively. Under XCI-E, we set (β_1, β_2) to be (0.4700, 0), (0.4055, 0.0953), (0.3365, 0.1823), (0.2624, 0.2624), (0.1823, 0.3365), (0.0953, 0.4055) and (0, 0.4700), and the corresponding $(\exp(\beta_1), \exp(\beta_2))$ value is (1.6, 1.0), (1.5, 1.1), (1.4, 1.2),

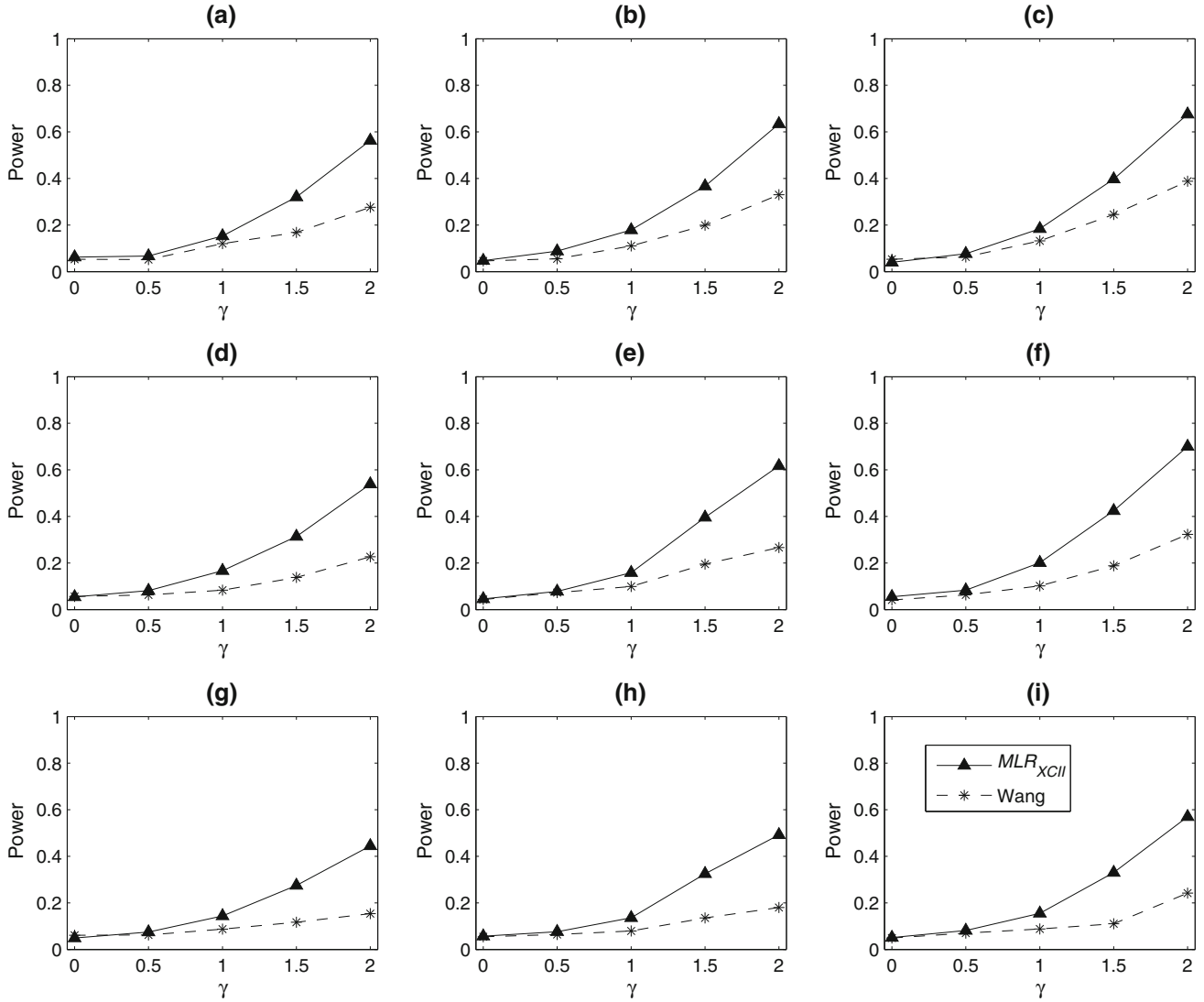

Figure 1. Estimated powers of MLR_{XCII} and the method of Wang *et al.* (2014) against γ when $\beta_1 = 0.2624$ and $\beta_2 = 0$ with different sex ratios and different (p_F, p_M) values based on 1000 replications for $N = 1000$ subjects at 5% significance level. (a) $(p_F, p_M) = (0.2, 0.3)$ and $r_f{:}r_m = 2{:}3$; (b) $(p_F, p_M) = (0.2, 0.3)$ and $r_f{:}r_m = 1{:}1$; (c) $(p_F, p_M) = (0.2, 0.3)$ and $r_f{:}r_m = 3{:}2$; (d) $(p_F, p_M) = (0.3, 0.3)$ and $r_f{:}r_m = 2{:}3$; (e) $(p_F, p_M) = (0.3, 0.3)$ and $r_f{:}r_m = 1{:}1$; (f) $(p_F, p_M) = (0.3, 0.3)$ and $r_f{:}r_m = 3{:}2$; (g) $(p_F, p_M) = (0.3, 0.2)$ and $r_f{:}r_m = 2{:}3$; (h) $(p_F, p_M) = (0.3, 0.2)$ and $r_f{:}r_m = 1{:}1$; (i) $(p_F, p_M) = (0.3, 0.2)$ and $r_f{:}r_m = 3{:}2$.

(1.3, 1.3), (1.2, 1.4), (1.1, 1.5) and (1.0, 1.6), respectively. Case–parent trios, control–parent trios, male cases and male controls under XCI-E are respectively generated from $(h_{f2}, h_{f10}, h_{f01}, h_{f0})$, $(g_{f2}, g_{f10}, g_{f01}, g_{f0})$, (h_{m1}, h_{m0}) and (g_{m1}, g_{m0}) based on the last two columns in table 2, where the detailed formulas are similar to those under XCI and are not shown here for brevity. For all the simulation settings, the nominal significance level is set to be $\alpha = 5\%$ and the number of replications are taken to be 1000. For each replication, 1000 permuted samples are generated.

Type I error rates

Tables 3 and 4 give the empirical size of the proposed MLR_{XCII} and the method of Wang *et al.* (2014) at the significance level of α=5% based on 1000 replications with $\beta_0 = -2.5500$ and $\beta_3 = 0.4055$ under XCI and XCI-E, respectively. From tables 3 and 4, we can see that the type I error rates of both methods can be well controlled under the null hypothesis of no association(β_1=β_2=0), regardless of the values of sex ratio, p_F, p_M, γ and XCI patterns.

Power comparisons

Figures 1–3 show the estimated powers of MLR_{XCII} and the method of Wang *et al.* (2014) against γ based on 1000 replications for $N = 1000$ at the 5% significance level under XCI when sex ratio $r_f : r_m$ =2:3, 1:1 and 3:2, (p_F, p_M) =(0.2, 0.3), (0.3, 0.3) and (0.3, 0.2), with (β_1, β_2) =(0.2624, 0), (0, 0.2624) and (0.2624, 0.2624),

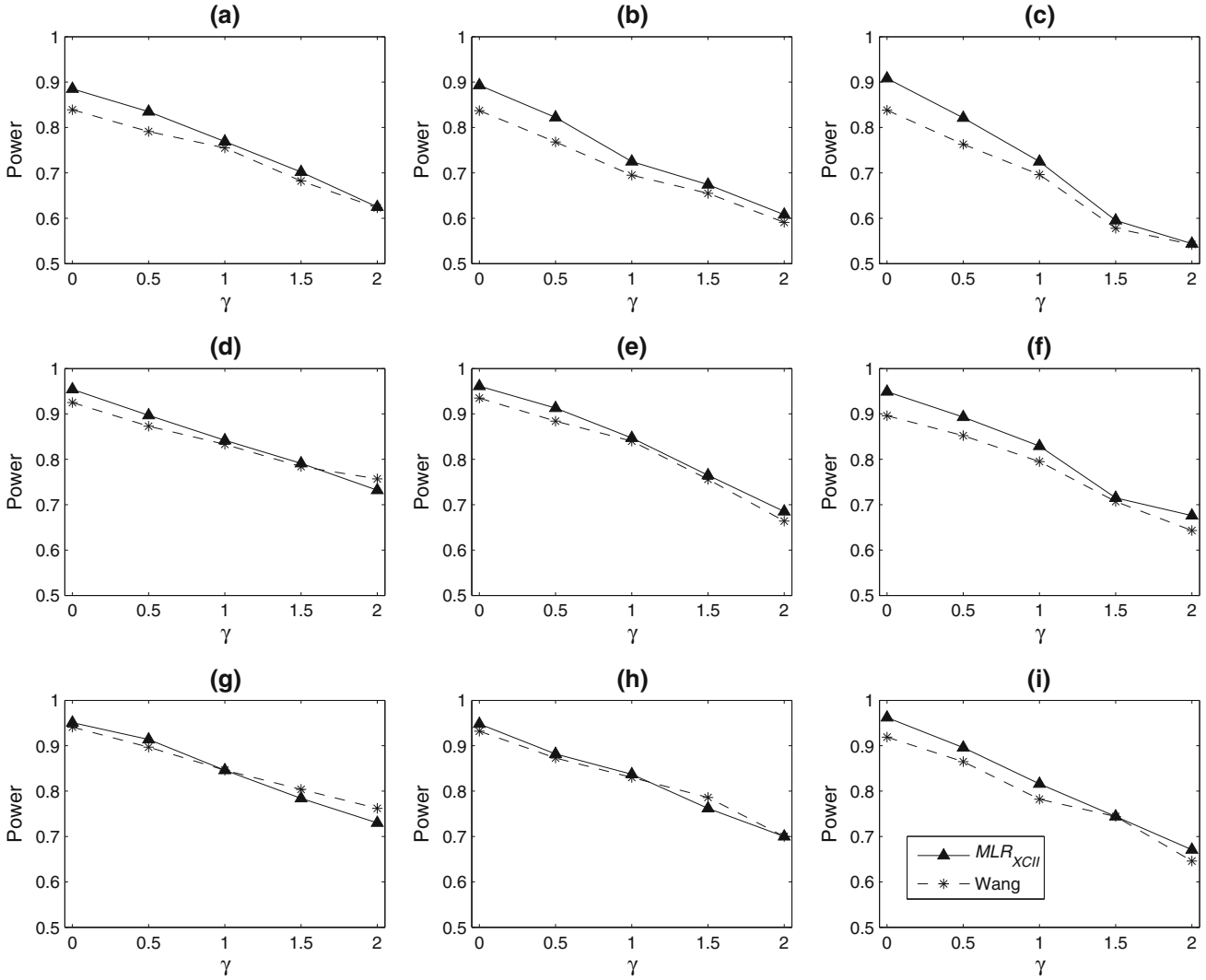

Figure 2. Estimated powers of MLR_{XCII} and the method of Wang *et al.* (2014) against γ when $\beta_1 = 0$ and $\beta_2 = 0.2624$ with different sex ratios and different (p_F, p_M) values based on 1000 replications for $N = 1000$ subjects at 5% significance level. (a) $(p_F, p_M) = (0.2, 0.3)$ and $r_f{:}r_m = 2{:}3$; (b) $(p_F, p_M) = (0.2, 0.3)$ and $r_f{:}r_m = 1{:}1$; (c) $(p_F, p_M) = (0.2, 0.3)$ and $r_f{:}r_m = 3{:}2$; (d) $(p_F, p_M) = (0.3, 0.3)$ and $r_f{:}r_m = 2{:}3$; (e) $(p_F, p_M) = (0.3, 0.3)$ and $r_f{:}r_m = 1{:}1$; (f) $(p_F, p_M) = (0.3, 0.3)$ and $r_f{:}r_m = 3{:}2$; (g) $(p_F, p_M) = (0.3, 0.2)$ and $r_f{:}r_m = 2{:}3$; (h) $(p_F, p_M) = (0.3, 0.2)$ and $r_f{:}r_m = 1{:}1$; (i) $(p_F, p_M) = (0.3, 0.2)$ and $r_f{:}r_m = 3{:}2$.

respectively. It is shown from figure 1 that MLR_{XCII} is more powerful than the method of Wang *et al.* (2014) under all the simulated situations, except for $\gamma = 0$. In fact, when $\gamma = 0$, $\beta_1 = 0.2624$ and $\beta_2 = 0$, we have $\beta_1\gamma = \beta_2(2-\gamma) = 0$ in female subjects and $\beta_1 X_1 = \beta_2 X_2 = 0$ in male subjects. As such, the powers of both methods are reduced to be the empirical size. On the other hand, $\beta_1\gamma$ increases as γ increases, while $\beta_2(2-\gamma)$ always stays at zero. This means that the degree of maternal imprinting effect increases and the powers of both methods become higher and higher. Further, the difference in power between these two methods is larger and larger. Meanwhile, figure 1 also demonstrates that the power of MLR_{XCII} with (p_F, p_M) =(0.2, 0.3) is the largest, the power with (p_F, p_M) =(0.3, 0.2) is the smallest and that with (p_F, p_M) =(0.3, 0.3) is between them when other parameters are unchanged, by comparing subplots 1(a–c) with subplots 1(d–f) and subplots 1(g–i), respectively. In addition, the bigger ratio of female offspring also improves the test power by comparing subplots 1(a), 1(d) and 1(g) with subplots 1(b), 1(e) and 1(h) (or comparing subplots 1(b), 1(e) and 1(h) with subplots 1(c), 1(f) and 1(i)), respectively.

From figure 2, we can see that the power of MLR_{XCII} is generally a little larger than the method of Wang *et al.* (2014). When $\gamma = 0$, $\beta_1 = 0$ and $\beta_2 = 0.2624$, we get $\beta_1\gamma = 0$ and $\beta_2(2-\gamma) = 2\beta_2$ in female subjects, which indicates the complete paternal imprinting effect. As such, the powers of both methods attain their respective maximum. With the increase of γ, $\beta_2(2-\gamma)$ decreases and $\beta_1\gamma$ is always fixed at zero. To this end, the degree of paternal imprinting effect is smaller and smaller, and then the powers of both methods decrease as γ increases. When $\gamma = 2$, $\beta_1\gamma = \beta_2(2-\gamma) = 0$

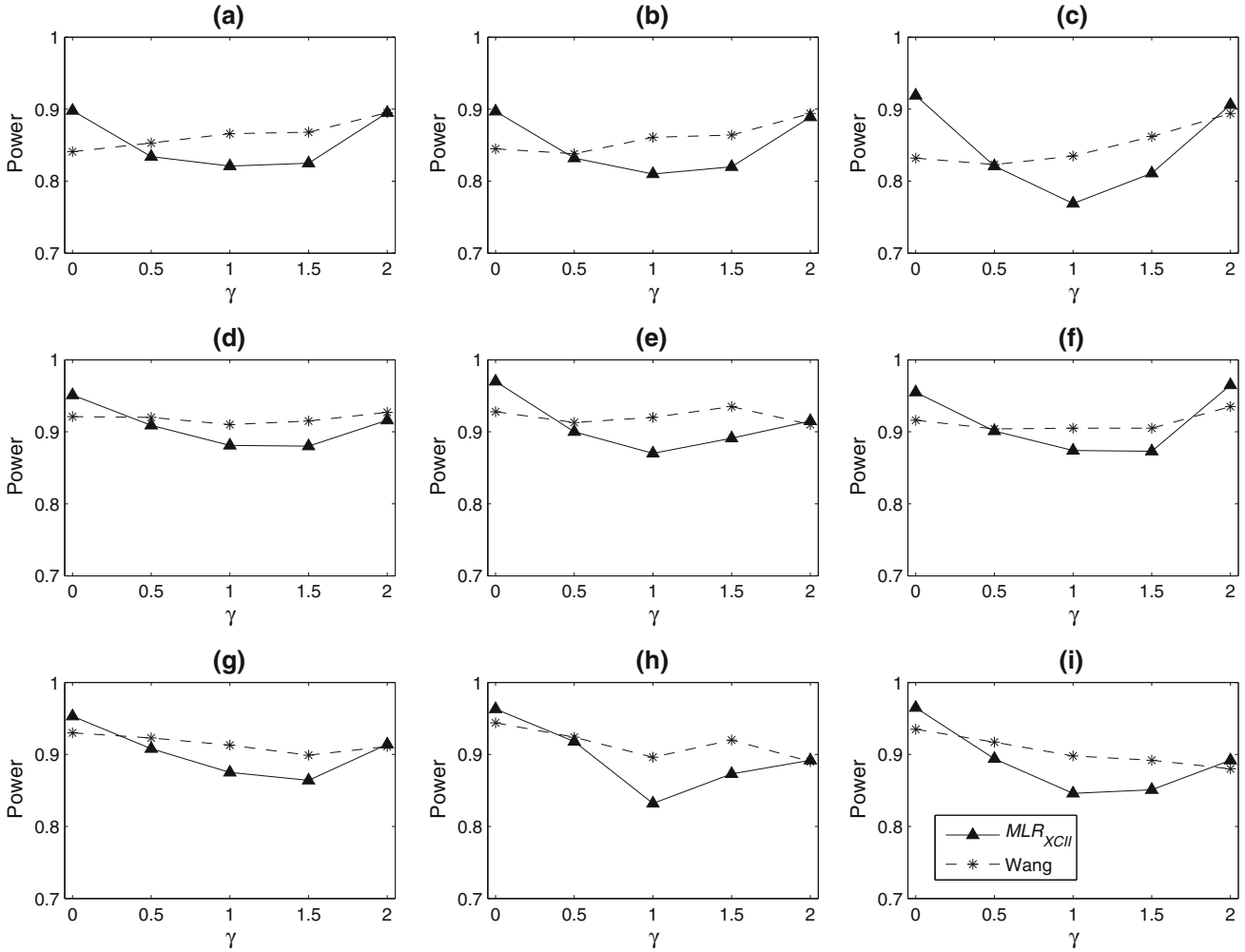

Figure 3. Estimated powers of MLR_{XCII} and the method of Wang *et al*. (2014) against γ when $\beta_1 = 0.2624$ and $\beta_2 = 0.2624$ with different sex ratios and different (p_F, p_M) values based on 1000 replications for $N = 1000$ subjects at 5% significance level. (a) $(p_F, p_M) = (0.2, 0.3)$ and $r_f{:}r_m = 2{:}3$; (b) $(p_F, p_M) = (0.2, 0.3)$ and $r_f{:}r_m = 1{:}1$; (c) $(p_F, p_M) = (0.2, 0.3)$ and $r_f{:}r_m = 3{:}2$; (d) $(p_F, p_M) = (0.3, 0.3)$ and $r_f{:}r_m = 2{:}3$; (e) $(p_F, p_M) = (0.3, 0.3)$ and $r_f{:}r_m = 1{:}1$; (f) $(p_F, p_M) = (0.3, 0.3)$ and $r_f{:}r_m = 3{:}2$; (g) $(p_F, p_M) = (0.3, 0.2)$ and $r_f{:}r_m = 2{:}3$; (h) $(p_F, p_M) = (0.3, 0.2)$ and $r_f{:}r_m = 1{:}1$; (i) $(p_F, p_M) = (0.3, 0.2)$ and $r_f{:}r_m = 3{:}2$.

in female subjects, while $\beta_1 X_1 = 0$ and $\beta_2 X_2 = 2\beta_2 = 0.5248 \neq 0$ in male subjects with allele D. Thus, the corresponding powers of both methods are not reduced to be the empirical size. In addition, figure 2 shows that the power of MLR_{XCII} with $(p_F, p_M) = (0.3, 0.2)$ is the largest and the power with $(p_F, p_M) = (0.2, 0.3)$ is the smallest when other parameters are unchanged, by comparing subplots 2(a–c) with subplots 2(g–i), respectively. The bigger ratio of female subjects reduces the test power by comparing subplots 2(a), 2(d) and 2(g) with subplots 2(b), 2(e) and 2(h) (or comparing subplots 2(b), 2(e) and 2(h) with subplots 2(c), 2(f) and 2(i)), respectively. These results about the effect of $r_f : r_m$ and (p_F, p_M) on the power of MLR_{XCII} are contrary to those in figure 1. In figure 3, note that (β_1, β_2) =(0.2624, 0.2624). When γ is close to 1, $\beta_1 X_1 = 0.2624 \times \gamma$ for D/d female subjects and $\beta_2 X_2 = 0.2624 \times (2 - \gamma)$ for d/D female subjects are very close to each other, which means that there is no imprinting effects or the degree of imprinting effect is very small. So, the power of MLR_{XCII} is less than that of the method of Wang *et al.* (2014). However, when $\gamma = 0$ (complete paternal imprinting) and $\gamma = 2$ (complete maternal imprinting), MLR_{XCII} can be more powerful than the method of Wang *et al*. (2014).

Figure 4 plots the estimated powers of MLR_{XCII} and the method of Wang *et al*. (2014) against seven simulation settings for different $(\exp(\beta_1), \exp(\beta_2))$ values based on 1000 replications for 1000 families at the 5% significance level under XCI-E when the sex ratio $r_f{:}r_m$ =2:3, 1:1 and 3:2, and (p_F, p_M) =(0.2, 0.3), (0.3, 0.3) and (0.3, 0.2), respectively. Note that larger difference between $\exp(\beta_1)$ and $\exp(\beta_2)$ in figure 4 is indicative of stronger imprinting effect and $\exp(\beta_1) = \exp(\beta_2)$ means no imprinting effects. So, for the simulation settings 1, 2 and 7 with strong maternal or paternal imprinting effect, MLR_{XCII} is generally more powerful than the method of Wang *et al*. (2014). For simulation setting 6, both methods have almost the same performance in

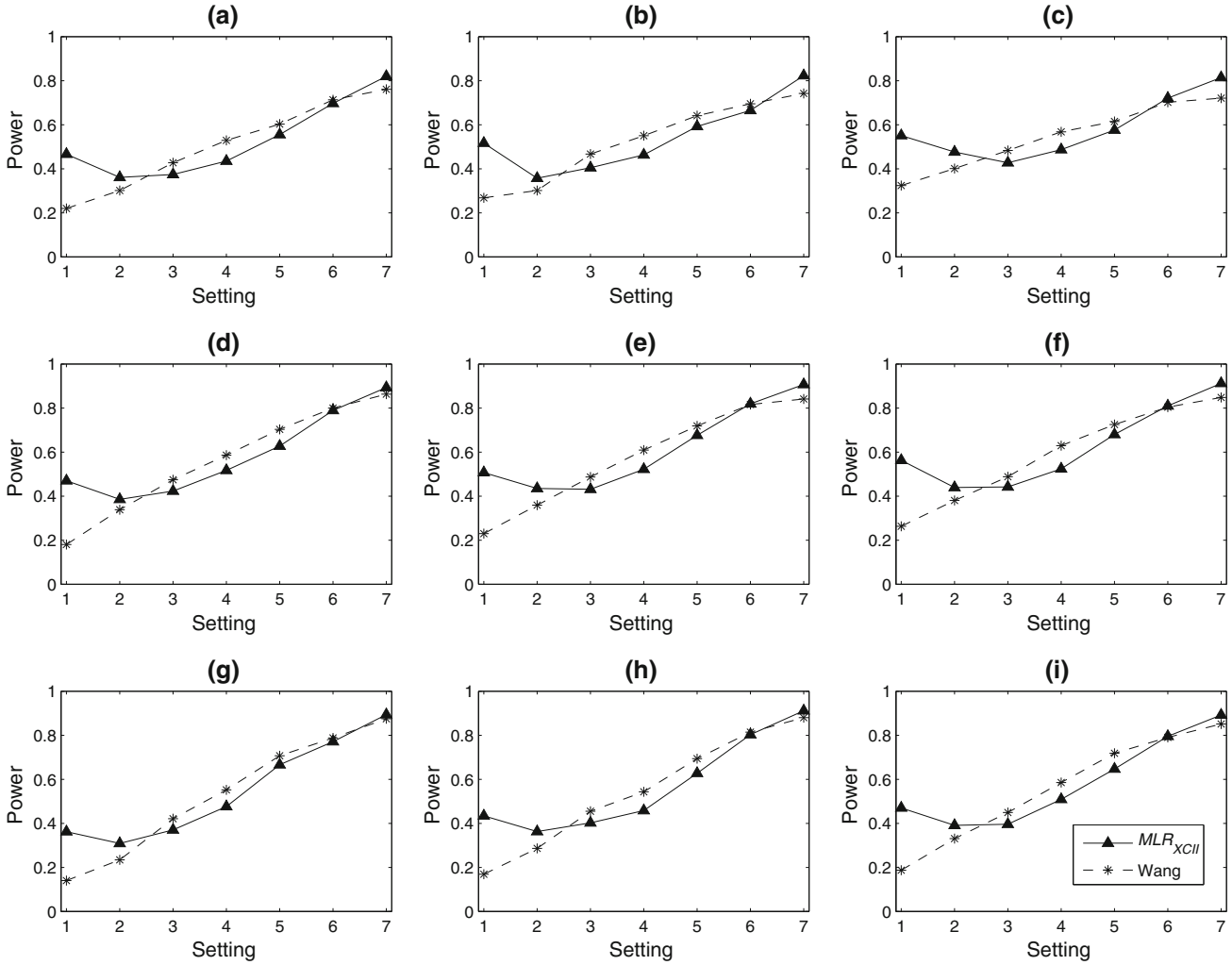

Figure 4. Estimated powers of MLR_{XCII} and the method of Wang *et al.* (2014) under XCI-E with different sex ratios and different (p_F, p_M) values based on 1000 replications for $N = 1000$ subjects at 5% significance level. Settings 1-7 indicate $(\exp(\beta_1), \exp(\beta_2))$=(1.6, 1.0), (1.5, 1.1), (1.4, 1.2), (1.3, 1.3), (1.2, 1.4), (1.1, 1.5) and (1.0, 1.6), respectively. (a) $(p_F, p_M) = (0.2, 0.3)$ and r_f:r_m = 2:3; (b) $(p_F, p_M) = (0.2, 0.3)$ and r_f:r_m = 1:1; (c) $(p_F, p_M) = (0.2, 0.3)$ and r_f:r_m = 3:2; (d) $(p_F, p_M) = (0.3, 0.3)$ and r_f:r_m = 2:3; (e) $(p_F, p_M) = (0.3, 0.3)$ and r_f:r_m = 1:1; (f) $(p_F, p_M) = (0.3, 0.3)$ and r_f:r_m = 3:2; (g) $(p_F, p_M) = (0.3, 0.2)$ and r_f:r_m = 2:3; (h) $(p_F, p_M) = (0.3, 0.2)$ and r_f:r_m = 1:1; (i) $(p_F, p_M) = (0.3, 0.2)$ and r_f:r_m = 3:2.

power. However, for simulation settings 3–5 with no or weak imprinting effect, the power of MLR_{XCII} is a little lower than that of the method of Wang *et al.* (2014).

Discussion

Different tests for association on X chromosome have been proposed in literature. But these studies only considered XCI-R or XCI-E. Wang *et al.* (2014) have put forward a unified method by combining XCI-R, XCI-S and XCI-E. However, it did not incorporate imprinting effects. Note that XCI and imprinting effects are two important genetic processes on X chromosome. Meanwhile, there has been no method available to simultaneously incorporate XCI and imprinting effects into analysis. Therefore, in this paper, we propose the novel and powerful method MLR_{XCII} for testing association incorporating XCI and genomic imprinting on X chromosome for qualitative traits. Simulation studies are conducted to compare the MLR_{XCII} with the existing method of Wang *et al.* (2014) under various simulation settings, including three groups of allele frequencies, three different sex ratios, different imprinting effect models, and different patterns of XCI. The simulation results show that both MLR_{XCII} and the method of Wang *et al.* (2014) control the type I error rates well. MLR_{XCII} is more powerful than the method of Wang *et al.* (2014) under the complete paternal imprinting model, the complete maternal imprinting model and the incomplete imprinting models. However, for the case of XCI-E and no imprinting effects $(\beta_1 X_1 = \beta_2 X_2)$, MLR_{XCII} has smaller powers than the method of Wang *et al.* (2014).

Notice that XPAT was proposed to test for imprinting effects on X chromosome based on affected daughter–parents trios (Zhou *et al*. 2018). To improve the power of MLR_{XCII} in the case of no imprinting effects, we may construct a two-stage test for association. Specifically, in the first stage, we detect imprinting effects by XPAT. In the second stage, if the P value of XPAT is less than its significance level, then we use our proposed method MLR_{XCII}, otherwise the method of Wang *et al*. (2014) is used. However, the distributions of both the proposed method MLR_{XCII} and the method of Wang *et al*. (2014) are unknown. Thus, how to control the overall significance level of the two-stage test for association may be challenging and then we will work on this in future.

Note that MLR_{XCII} simultaneously incorporates XCI-R, XCI-S and XCI-E, which are generally unknown. Thus, the distribution of MLR_{XCII} is unknown and we use a permutation-based procedure stratified by sex to calculate the P value of MLR_{XCII}, which is time-consuming. On the other hand, there have been methods proposed to estimate γ for case–control data (Wang *et al*. 2019) and family data (Xu *et al*. 2018), respectively. So, to reduce the computing time of the proposed method, one may first use the existing methods to estimate γ and then construct a LR test statistic to test for association. However, we are not sure if the LR test statistic still follows a chi-square distribution. So, we plan to investigate this as future work. In addition, the traits we consider in this paper are restricted to be qualitative and the proposed MLR_{XCII} method is applicable to family trios. As such, we will extend our proposed MLR_{XCII} method to pedigree data or quantitative traits in future. Finally, we will conduct a real data analysis in future to highlight the applicability of the proposed method.

Acknowledgements

The authors thank the two reviewers for helpful comments that greatly improve the presentation of the paper. This work was supported by the National Natural Science Foundation of China (grant 81773544), the National and Guangzhou University Students' Innovation and Enterprise Training Project of China (grant 201712121105) and the Hong Kong RGC GRF grant (17302919).

References

Amos-Landgraf J. M., Cottle A., Plenge R. M., Friez M., Schwartz C. E., Longshore J. *et al*. 2006 X chromosome-inactivation patterns of 1005 phenotypically unaffected females. *Am. J. Hum. Genet.* **79**, 493–499.

Avner P. and Heard E. 2001 X-chromosome inactivation: counting, choice and initiation. *Nat. Rev. Genet.* **2**, 59–67.

Bartolomei M. S. 2009 Genomic imprinting: employing and avoiding epigenetic processes. *Gene Dev.* **23**, 2124–2133.

Breslau N., Davis G. C., Andreski P., Peterson E. L. and Schultz L. R. 1997 Sex differences in posttraumatic stress disorder. *Arch. Gen. Psychiatry* **54**, 1044–1048.

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

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

Cohn B. A., Wingard D. L., Cirillo P. M., Cohen R. D., Reynolds P. and Kaplan G. A. 1996 Re: Differences in lung cancer risk between men and women: examination of the evidence. *J. Natl. Cancer Inst*. **88**, 1867.

Constância M., Pickard B., Kelsey G. and Reik W. 1998 Imprinting mechanisms. *Genome Res.* **8**, 881–900.

Deng H. W. and Chen W. M. 2001 The power of the transmission disequilibrium test (TDT) with both case-parent and control-parent trios. *Genet. Res.* **78**, 289–302.

Deng X., Berletch J. B., Nguyen D. K. and Disteche C. M. 2014 X chromosome regulation: diverse patterns in development, tissues and disease. *Nat. Rev. Genet*. **15**, 367–378.

Ferguson-Smith A. C. 2011 Genomic imprinting: the emergence of an epigenetic paradigm. *Nat. Rev. Genet.* **12**, 565–575.

Gao F., Chang D., Biddanda A., Ma L., Guo Y., Zhou Z. *et al*. 2015 XWAS: a software toolset for genetic data analysis and association studies of the X chromosome. *J. Hered.* **106**, 666–671.

Hankin B. L. and Abramson L. Y. 2001 Development of gender differences in depression: an elaborated cognitive vulnerability–transactional stress theory. *Psychol. Bull.* **127**, 773–796.

Hickey P. F. and Bahlo M. 2011 X chromosome association testing in genome wide association studies. *Genet. Epidemiol.* **35**, 664–670.

Hindorff L. A., Sethupathy P., Junkins H. A., Ramos E. M., Mehta J. P., Collins F. S. *et al*. 2009 Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci. USA* **106**, 9362–9367.

Kay G. F. 1998 *Xist* and X chromosome inactivation. *Mol. Cell. Endocrinol.* **140**, 71–76.

Kukurba K. R., Parsana P., Balliu B., Smith K. S., Zappala Z., Knowles D. A. *et al*. 2016 Impact of the X chromosome and sex on regulatory variation. *Genome Res.* **26**, 768–777.

Lee J. T. 2000 Disruption of imprinted X inactivation by parent-of-origin effects at *Tsix*. *Cell* **103**, 17–27.

Lee J. T. and Bartolomei M. S. 2013 X-inactivation, imprinting, and long noncoding RNAs in health and disease. *Cell* **152**, 1308–1323.

Lerner D. J. and Kannel W. B. 1986 Patterns of coronary heart disease morbidity and mortality in the sexes: a 26-year follow-up of the Framingham population. *Am. Heart J.* **111**, 383–390.

Li M., Li J., He Z., Lu Q., Witte J. S., Macleod S. L. *et al*. 2016 Testing allele transmission of an SNP set using a family based generalized genetic random field method. *Genet. Epidemiol.* **40**, 341–351.

Loesch D. Z., Bui Q. M., Kelso W., Huggins R. M., Slater H., Warne G. *et al*. 2005 Effect of Turner's syndrome and X-linked imprinting on cognitive status: analysis based on pedigree data. *Brain Dev.* **27**, 494–503.

Lyon M. F. 1962 Sex chromatin and gene action in the mammalian X-chromosome. *Am. J. Hum. Genet*. **14**, 135–148.

Lyon M. F. 1972 X-chromosome inactivation and developmental patterns in mammals. *Biol. Rev.* **47**, 1–35.

Mendelsohn M. E. and Karas R. H. 2005 Molecular and cellular basis of cardiovascular gender differences. *Science* **308**, 1583–1587.

Migeon B. R., Chowdhury A. K., Dunston J. A. and Mcintosh I. 2001 Identification of *TSIX*, encoding an RNA antisense to human *XIST*, reveals differences from its murine counterpart: implications for X inactivation. *Am. J. Hum. Genet.* **69**, 951–960.

Minks J., Robinson W. P. and Brown C. J. 2008 A skewed view of X chromosome inactivation. *J. Clin. Invest*. **118**, 20–23.

Naugler W. E., Sakurai T., Kim S., Maeda S., Kim K., Elsharkawy A. M. *et al*. 2007 Gender disparity in liver cancer due to sex

differences in *MyD88*-dependent *IL-6* production. *Science* **317**, 121–124.
Patten M. M., Ross L., Curley J. P., Queller D. C., Bonduriansky R. and Wolf J. B. 2014 The evolution of genomic imprinting: theories, predictions and empirical tests. *Heredity* **113**, 119–128.
Pearson T. A. and Manolio T. A. 2008 How to interpret a genome-wide association study. *J. Am. Med. Assoc.* **299**, 1335–1344.
Penny G. D., Kay G. F., Sheardown S. A., Rastan S. and Brockdorff N. 1996 Requirement for *Xist* in X chromosome inactivation. *Nature* **379**, 131–137.
Reik W. and Walter J. 2001 Genomic imprinting: parental influence on the genome. *Nat. Rev. Genet.* **2**, 21–32.
Tycko B. and Morison I. M. 2002 Physiological functions of imprinted genes. *J. Cell. Physiol.* **192**, 245–258.
Voskuhl R. 2011 Sex differences in autoimmune diseases. *Biol. Sex Differ.* **2**, 1.
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.
Wang P., Xu S. Q., Wang B. Q., Fung W. K. and Zhou J. Y. 2019 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 B. Q., Li J. L., Wang Y. X., Pan D. *et al*. 2019 A statistical measure for the skewness of X chromosome inactivation based on case-control design. *BMC Bioinformatics* **20**, 11.
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.
Wong C. C. Y., Caspi A., Williams B., Houts R., Craig I. W. and Mill J. 2011 A longitudinal twin study of skewed X chromosome-inactivation. *PLoS One* **6**, e17873.
Xu S. Q., Zhang Y., Wang P., Liu W., Wu X. B. and Zhou J. Y. 2018 A statistical measure for the skewness of X chromosome inactivation based on family trios. *BMC Genet*. **19**, 109.
Yang J. and Lin S. 2012 Likelihood approaches for detecting imprinting and maternal effects in family based association studies. *Biometrics* **68**, 477–485.
Zheng G., Joo J., Zhang C. and Geller N. L. 2007 Testing association for markers on the X chromosome. *Genet. Epidemiol.* **31**, 834–843.
Zhou J. Y., You X. P., Yang R. and Fung W. K. 2018 Detection of imprinting effects for qualitative traits on X chromosome based on nuclear families. *Stat. Methods Med. Res.* **27**, 2329–2343.

Corresponding editor: H. A. Ranganath