RESEARCH ARTICLE

The X factor: A robust and powerful approach to X-chromosome-inclusive whole-genome association studies

Bo Chen[1] | Radu V. Craiu[1] | Lisa J. Strug[1,2,3,4] | Lei Sun[1,2]

[1]Department of Statistical Sciences, University of Toronto, Toronto, Ontario, Canada

[2]Biostatistics Division, Dalla Lana School of Public Health, University of Toronto, Toronto, Ontario, Canada

[3]Department of Computer Science, University of Toronto, Toronto, Ontario, Canada

[4]Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, Ontario, Canada

Correspondence
Lei Sun, Department of Statistical Sciences, University of Toronto, Toronto, ON, Canada.
Email: sun@utstat.toronto.edu

Abstract

The X-chromosome is often excluded from genome-wide association studies because of analytical challenges. Some of the problems, such as the random, skewed, or no X-inactivation model uncertainty, have been investigated. Other considerations have received little to no attention, such as the value in considering nonadditive and gene–sex interaction effects, and the inferential consequence of choosing different baseline alleles (i.e., the reference vs. the alternative allele). Here we propose a unified and flexible regression-based association test for X-chromosomal variants. We provide theoretical justifications for its robustness in the presence of various model uncertainties, as well as for its improved power when compared with the existing approaches under certain scenarios. For completeness, we also revisit the autosomes and show that the proposed framework leads to a more robust approach than the standard method. Finally, we provide supporting evidence by revisiting several published association studies. Supporting Information for this article are available online.

KEYWORDS
confounding, dominance, interaction, model uncertainty, regression

1 | INTRODUCTION

Genome-wide association studies (GWAS) are ubiquitous, delivering significant insights into the genetic determinants of complex traits over the past decade (Visscher et al., 2017). For this reason, it is surprising that it is not a common practice to include the X-chromosome in GWAS (Konig et al., 2014; Wise et al., 2013). The

X-chromosome differs from the autosomes in that males have only one copy of the X-chromosome while females have two, and at any given genomic location one of the two copies in females may be silenced (Gendrel & Heard, 2011), referred to as X-chromosome inactivation (XCI). The choice of the silenced copy could be random or skewed toward a specific copy (Wang et al., 2014). These unique aspects lead to more complex analytic

considerations for genetic association analysis of X-chromosomal variants, such as bi-allelic single-nucleotide polymorphisms (SNPs).

A bi-allelic SNP has two alleles, r and R, of which one is the reference allele and the other is the alternative allele with allele frequency f. An autosomal SNP has three genotypes regardless of sex, namely $G = (rr, rR, RR)$. In association analysis of an autosomal SNP, the common practice is to simply model a binary or continuous phenotype Y as an additive function of the number of copies of the non-baseline allele present in G; that is, coding G additively as $G_A = (0, 1, 2)$. Here, without loss of generality, r is chosen to be the baseline allele in a statistical model and R the non-baseline allele. When Y is binary, this regression-based additive test is also equivalent to the Cochran–Armitage trend test (Wellek & Ziegler, 2012). Although both dominant and recessive genetic models of inheritance are possible, among these one degrees of freedom (1 df) models, a common practice for GWAS is to use the additive model, because it has reasonable power to detect both additive and dominance effects at a causal variant, and at variants in linkage disequilibrium (LD) with the causal variant (Bush & Moore, 2012; Hill et al., 2008). An alternative parameterization is the 2 df genotypic model that includes both the additive $G_A = (0, 1, 2)$ term and the dominance $G_D = (0, 1, 0)$ term. In the case of recessive genetic inheritance, Zhou et al. (2017) showed that the 2 df genotypic test outperforms the 1 df additive test for binary outcomes, and Dizier et al. (2017) reached the same conclusion for continuous traits. In the case of additive genetic inheritance being true, the genotypic test is known to be less powerful than the additive test due to the increased df, which is unnecessary. The preferred test for unknown genetic inheritance in terms of power and robustness to different genetic models is, however, not clear across different true genetic effect sizes, sample sizes, and significance levels.

For an X-chromosomal SNP, the most commonly used approach assumes additivity and XCI. However, recent work (Tukiainen et al., 2017) showed that up to one-third of X-chromosomal genes are expressed from both the active and inactive X-chromosomes in female cells, with varying degrees of "escape" from inactivation between genes and individuals. Several additional points also require attention. Table 1 describes eight analytical considerations and challenges (C1–C8) present in an X-chromosome-inclusive GWAS, including a method's suitability for analyzing both binary and continuous traits (C1), which is related to the type of method used, that is, genotype-based or allelic association tests (C2); the (under-appreciated) consequence of the choice of the baseline allele on association analysis of an X-chromosomal SNP (C3); the importance of including sex as a covariate (C4) and its analytical connection with C3; the value in considering gene-sex interaction effect (C5) and its connection with the assumption of XCI (C6); and the assumption of random versus skewed XCI (C7) and its connection with nonadditive effects (C8).

Several association methods have been developed for the X-chromosome, and they are computationally efficient for conducting X-chromosome-wide association analysis. However, each method solves only some of the C1–C8 challenges. For example, Zheng et al. (2007) considered only binary outcomes for which both genotype- and allele-based association tests are applicable. The classical allelic association test, comparing allele frequencies between case and control groups, is locally most powerful but sensitive to the Hardy–Weinberg equilibrium (HWE) assumption and not applicable to continuous traits (Sasieni, 1997; Zhang & Sun, 2021; Zheng, 2008). Clayton (2008, 2009) discussed analytical strategies assuming the X-chromosome is always inactivated. Hickey and Bahlo (2011) and Loley et al. (2011) performed simulation studies, each providing a thorough method comparison, for example, between tests of Zheng et al. (2007) and Clayton (2008). Konig et al. (2014) gave detailed guidelines for including the X-chromosome in GWAS, recommending different tests for different model assumptions (e.g., presence or absence of an interaction effect or XCI), but it is difficult to validate these assumptions in practice. Gao et al. (2015) developed a toolset for conducting X-chromosome association studies, implementing some of the existing methods. More recently Z. Chen et al. (2017) improved sex-stratified analysis by eliminating genetic model assumptions, but their method is limited to analyzing genetic main effects on binary traits. Focusing on XCI uncertainty, Wang et al. (2014) proposed a frequentist maximum likelihood solution to deal with no, random or skewed X-inactivation, and in their follow-up work Wang et al. (2017) provided a model selection method. In contrast, B. Chen et al. (2020) applied the Bayesian model averaging principle (Draper, 1995) to deal with the XCI uncertainty problem. However, these approaches assumed additive genetic effects. The value in considering dominance and gene-sex interaction effects, and the inferential consequence of defining different baseline allele (i.e., the reference or the alternative allele) when analyzing an X-chromosomal SNP, have received little to no attention.

Here we propose a theoretically justified and robust X-chromosome association method that can simultaneously deal with all eight challenges (C1–C8) outlined in Table 1. We emphasize the robustness of the proposed method to genetic assumptions as our understanding is

TABLE 1 Eight analytical considerations and challenges, C1–C8, present in X-chromosome-inclusive association studies

Problem	Solution	Relevant sections
C1: Quantitative traits vs. binary outcomes		
C2: Genotype-based vs. allele-based association methods Allele-based association tests, comparing allele frequency differences between cases and controls, are locally most powerful. However, they analyze binary outcomes only and are sensitive to the Hardy–Weinberg equilibrium (HWE) assumption (Sasieni, 1997).	Genotype-based regression models, Y-on-G, support various types of outcome data, account for covariate effects with ease, and are robust to the HWE assumption.	Sections 1 and 2
C3: The choice of the baseline allele for association analysis, r vs. R For the autosomes, switching the two alleles does not affect the association inference. Is this true for the X-chromosome?	It is *not* always true for the X-chromosome, unless S is included in the model.	Sections 2.1 and 2.2, and C4
C4: Sex as a covariate vs. no S main effect Unlike the autosomes, sex is a confounder when analyzing the X-chromosome for traits exhibiting sexual dimorphism (e.g., height and weight). Even for the autosomes, sex can be a confounder if allele frequencies differ significantly between males and females.×	To maintain the correct type I error rate control, the sex main effect must be considered particular when analyzing the X-chromosome. The resulting association test is also invariant to the choice of the baseline allele.	Section 2.2 and C3
C5: Gene–sex interaction vs. no $G \times S$ interaction effect Gene–sex interaction might exist, but there is a concern over loss of power due to increased degrees of freedom. In addition, what is the interpretation of gene–sex interaction effect in the presence of X-inactivation?	Under no interaction, power loss of modeling interaction is capped at 11.4%. Models including the $G \times S$ covariate also lead to tests invariant to the assumption of X-chromosome inactivation status.	Sections 2.3 and 3, and C6
C6: X-chromosome inactivation (XCI) vs. no XCI XCI occurs if one of the two alleles in a genotype of a female is silenced. Individual-level XCI status requires additional biological information that are not typically available to genetic association studies. Assuming XCI or no XCI at the sample level leads to different genotype coding strategies (Table 2), and it was thought that this will always lead to different association results.	XCI uncertainty implies sex-stratified genetic effect which can be analytically represented by the $G \times S$ interaction effect. Teasing apart these different biological phenomenon require other "omic" data and additional analyses.	Sections 2.3 and 5, and C5
C7: If XCI, random vs. skewed X-inactivation If the choice of the silenced allele in females is skewed toward a specific allele, the average effect of the rR genotype is no longer the average of those of r and R.	XCI skewness is statistically equivalent to a dominance genetic effect.	Section 2.4, and C8
C8: Dominance effect vs. no G_D dominance effect For both the autosomes and X-chromosome, the most common practice is to use the additive test which has better power than the genotypic test under (approximate) additivity, but it cannot capture dominance effects. The exact trade-off, however, is not clear.	We provide analytical and empirical evidence supporting the use of genotypic model when analyzing either the autosomes or X-chromosome. For an X-chromosomal variant, including the dominance effect term has the added benefit of resolving of the skewed X-inactivation uncertainty issue.	Sections 2.4, 3 and 4, and C7

TABLE 2 Covariate coding schemes for examining the additive, dominance, gene–sex interaction, and sex effects under different assumptions of the X-chromosome inactivation status and the choice of the baseline allele

Effect interpretation	Covariate notation	Non-baseline allele	X-chromosome inactivation (XCI) status	Coding schemes Females			Males	
				rr	rR	RR	r	R
	$G_{A,R,I}$	R	Yes	0	0.5	1	0	1
Additive G_A	$G_{A,r,I}$	r	Yes	1	0.5	0	1	0
G_A	$G_{A,R,N}$	R	No	0	1	2	0	1
	$G_{A,r,N}$	r	No	2	1	0	1	0
Dominance G_D	G_D	Either	Either	0	1	0	0	0
Gene–sex interaction	GS_R	R	Either	0	0	0	0	1
$GS = G_A \times S$	GS_r	r	Either	0	0	0	1	0
Sex S	S	Either	Either	0	0	0	1	1

Note: The subscripts A and D represent additive and dominance effects, R or r represents the non-baseline allele of which we count the number of copies present in a genotype, and I or N denotes X-chromosome inactivated or not inactivated.

evolving. For example, although most published X-chromosome-inclusive GWAS assumed XCI, recent work has shown that up to a third of genes "escape" XCI (Tukiainen et al., 2017).

The proposed method is regression- and genotype-based (robust to departure from HWE), analyzing either a continuous or binary trait while adjusting for covariate effects. The recommended test has three degrees of freedom, including both additive and dominance genetic effects, as well as a gene–sex interaction effect. We show analytically why the proposed method is robust to the various model uncertainties, including no, random or skewed XCI, as well as the choice of the baseline allele. Desirably, the power of the proposed test is robust to different alternative genetic models, despite its increased degrees of freedom over a simple additive test. We note that the work here focuses on efficient association testing, not parameter estimation or model selection which requires additional biological data (Busque et al., 1996).

We first present our main theory to address the eight challenges associated with X-chromosome-inclusive GWAS in Section 2. We then provide analytical results of power study across all possible genetic models, sample sizes and type I error rates, as well as empirical results from simulation studies in Section 3. For methodology completeness, this section also briefly discusses merit of the genotypic model in the familiar context of analyzing autosomal SNPs. We then provide corroborating evidence from several applications in favor of the proposed approach in Section 4. Finally we discuss the limitations of our approach and possible future work in Section 5.

2 | METHOD FOR X-CHROMOSOME-INCLUSIVE ASSOCIATION ANALYSIS

The proposed method relies on the generalized linear model (McCullagh & Nelder, 1989) as it is flexible, analyzing both binary and continuous traits (C1 of Table 1). As a result, the method is a genotype-based approach (C2) that is robust to the assumption of HWE by regressing the phenotype data (Y) on genetic data (G) while accounting for other covariate effects.

For robust and powerful association analysis of a bi-allelic X-chromosomal SNP, we recommend the following model:

$$g(E(Y)) = \beta_0 + \beta_S S + \beta_A G_A + \beta_D G_D + \beta_{GS} GS, \quad (1)$$

and the corresponding 3 df test, jointly testing

$$H_0: \beta_A = \beta_D = \beta_{GS} = 0, \quad (2)$$

where notations for the covariates are defined in Table 2. Other relevant covariates such as environmental factors (Es) should also be included in the model but omitted here for notation simplicity.

We show later (a) why the association result from the proposed approach is invariant to the different G_A (e.g., $G_{A,R,I}$ or $G_{A,r,I}$) and GS (e.g., GS_R or GS_r) coding schemes as defined in Table 2, and (b) why the proposed method also solves the C3–C8 issues simultaneously. But before we do so, we first provide more details about the notations presented in Table 2.

2.1 | X-chromosome specific genotype and covariate coding schemes

Table 2 summarizes the various covariate coding schemes for analyzing an X-chromosomal SNP, when considering all the analytical challenges outlined in Table 1. Note that when the choice of the baseline allele is varied (i.e., either r or R) and the XCI status is unknown, there are four ways to code the additive covariate G_A, and two ways to code the gene–sex interaction covariate GS. The specific coding for sex does not have an impact on our proposed method. In Table 2, a female is coded as 0 and a male as 1, and the interaction $G_D \times S$ term vanishes. If a female were coded as 1 and a male as 0, then $G_D \times S$ is the same as G_D. Thus, in either case it is redundant to include $G_D \times S$ in our proposed regression model.

Using the notations in Table 2, it is immediately clear why the choice of the baseline allele (C3) matters for association analysis of an X-chromosomal SNP. Under no XCI, if r were assumed to be the baseline allele there would be one copy of allele R in genotype rR of a female, and R of a male. Thus, genotypes rR and R would be grouped together for association analysis. However, if R were chosen to be the baseline allele, genotypes rR and r would be grouped together, resulting in different inference. In contrast, the choice of the baseline allele does not affect association evidence when analyzing an autosomal SNP. It is well-known that although the estimate of the effect size changes direction, the magnitude of the association remains the same when analyzing an autosomal SNP. But, this is not always true when analyzing an X-chromosomal SNP.

2.2 | Sex as a confounder (C4) and its connection with the choice of the baseline allele (C3)

Sex is a confounder for phenotype-genotype association analysis of an X-chromosomal SNP for traits displaying sexual dimorphism. When sex, but not the SNP, is associated with a trait of interest, omitting sex in the analysis leads to false positives. This is because sex is inherently associated with the genotypes of an X-chromosomal SNP (Table 2); see Ozbek et al. (2018) for empirical evidence from simulation studies. Thus, accuracy of a test provides the first argument for always including S as a covariate in association analysis of an X-chromosomal SNP.

The second advantage of modeling the S main effect is more subtle. As shown in Table 2, the coding of G_A

depends on the choice of the baseline allele (i.e., R or r) and the X-inactivation status (I for XCI and N for no XCI), resulting in a total of four different ways of coding the five genotype groups, namely $G_{A,R,I} = (0, 0.5, 1, 0, 1)'$, $G_{A,r,I} = (1, 0.5, 0, 1, 0)'$, $G_{A,R,N} = (0, 1, 2, 0, 1)'$, and $G_{A,r,N} = (2, 1, 0, 1, 0)'$. Furthermore, $G_{A,R,N}$ and $G_{A,r,N}$ yield different test statistics, because the two coding schemes lead to different groupings of the genotypes as discussed in 2.1. Note that, in contrast to $G_{A,R,I} = 1 - G_{A,r,I}$ under XCI, under no XCI there is no linear transformation that makes $G_{A,R,N}$ and $G_{A,r,N}$ equivalent. An inference that is invariant to the coding choices may seem difficult, but we show that this is achievable for models that include sex as a covariate.

Theorem 1. *Let $\mathcal{M}_1$ and $\mathcal{M}_2$ be two generalized linear models (McCullagh & Nelder, 1989) with the same link function g, $g(E(Y)) = X_1\boldsymbol{\beta_1}$ and $g(E(Y)) = X_2\boldsymbol{\beta_2}$, where Y is the response vector of length n, X_1 and X_2 are two $n \times p$ design matrices, and $\boldsymbol{\beta_1}$ and $\boldsymbol{\beta_2}$ are the corresponding parameter vectors of length p. Let $X_1 = (X_{11}, X_{12})$, where X_{11} and X_{12} are $n \times (p - q)$ and $n \times q$ matrices corresponding to, respectively, the $(p - q)$ secondary covariates not being tested and the q primary covariates of interest, and similarly for $X_2 = (X_{21}, X_{22})$, and partition the regression coefficients accordingly as $\boldsymbol{\beta_1} = (\beta', \beta')'$ and $\boldsymbol{\beta_2} = (\beta', \beta')'$. If there exists an invertible $p \times p$ matrix*

$$T = \begin{pmatrix} T_{11} & T_{12} \\ 0 & T_{22} \end{pmatrix}, \text{ such that } X_2 = X_1\,T,$$

where T_{11} and T_{22} are, respectively, invertible $(p - q) \times (p - q)$ and $q \times q$ matrices, then any of the Wald, Score or LRT tests for testing

$$H_0: \beta_{12} = 0 \quad \text{and} \quad H_0: \beta_{22} = 0$$

are identical under the two models $\mathcal{M}_1$ and $\mathcal{M}_2$, resulting in the same association inference for evaluating the q primary covariates of interest. Note that given the structure of matrix T, $X_2 = X_1\,T$ implies $X_{21} = X_{11}\,T_{11}$.

We provide the proof of Theorem 1 in Supporting Information – Appendix A. Here we emphasize that the two sets of q primary covariates being tested, X_{22} and X_{12}, are *not* required to be linear transformation of each other, for example, between $G_{A,R,N} = (0, 1, 2, 0, 1)'$ and $G_{A,r,N} = (2, 1, 0, 1, 0)'$. Instead, X_{21} and X_{11}, corresponding to the $p - q$ secondary covariates (including the unit vector if modeling the intercept), that are *not* being tested must be invertible linear transformations of each other, $X_{21} = X_{11}\,T_{11}$, in addition to $X_2 = X_1\,T$. This result may seem surprising, but the two requirements imply that the

two design matrices are equivalent to each other either in general or under the null, resulting in identical F-test statistics; see Supporting Information Appendix A for technical details.

In our setting when sex is included in the model, consider only the additive effect for the moment, $g(E(Y)) = \beta_0 + \beta_S S + \beta_A G_A$. Then the two design matrices, corresponding to r or R being the baseline allele and under no XCI, are

$$X_1 = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 2 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix} \quad \text{and} \quad X_2 = \begin{pmatrix} 1 & 0 & 2 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix}.$$

In this case, $T_{11} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $T_{12} = (2, -1)'$ and $T_{22} = -1$ satisfy the two requirements. Thus, even though $G_{A,R,N} = (0, 1, 2, 0, 1)'$ and $G_{A,r,N} = (2, 1, 0, 1, 0)'$ are not linked by a linear transformation, Theorem 1 allows us to conclude that a Wald, Score or LRT test of $H_0: \beta_A = 0$ is invariant to the two G_A coding schemes $G_{A,R,N}$ and $G_{A,r,N}$, if sex is included as a covariate.

Note that the known result that two tests are equivalent to each other if X_{12} and X_{22}, corresponding to q primary covariates, are linear transformation of each other is a special case of Theorem 1, where all elements except the first row of T_{12} are zero; the exception allows for a location shift. For example, under the XCI assumption, $X_{12} = G_{A,R,I} = (0, 0.5, 1, 0, 1)'$ and $X_{22} = G_{A,r,I} = (1, 0.5, 0, 1, 0)'$, and $X_{22} = 1 - X_{12}$. Thus, $T_{11} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $T_{12} = (1, 0)'$ and $T_{22} = -1$ satisfy the requirements.

At this point in the methodology development, the preferred model $g(E(Y)) = \beta_0 + \beta_S S + \beta_A G_A$ controls the type I error rate and is invariant to the choice of the baseline allele. However, in practice the XCI status is unknown and if we assume there is XCI, $G_{A,R,I} = (0, 0.5, 1, 0, 1)'$ and $G_{A,r,N} = (2, 1, 0, 1, 0)'$, and

$$X_1 = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0.5 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix} \quad \text{and} \quad X_2 = \begin{pmatrix} 1 & 0 & 2 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 0 \end{pmatrix}.$$

In this case, it is not difficult to show that a matrix T satisfying the requirements of Theorem 1 does not exist, because $Rank(X_1) < Rank((X_1, X_2))$ implies that the linear system $X_2 = X_1 T$ has no solution, and the XCI uncertainty remains a challenge.

2.3 | Gene–sex interaction effect (C5) and its connection with unknown XCI status (C6)

Throughout the paper, we define the GS interaction term as $G_A \times S$. Depending on the choice of the baseline allele, GS has two different codings, namely GS_R and GS_r (Table 2). In the previous section, we have shown that when S is included in the model, that is, $g(E(Y)) = \beta_0 + \beta_S S + \beta_A G_A$, the choice of the baseline allele is no longer of a concern if we test $H_0: \beta_A = 0$ within a particular XCI assumption. Interestingly, when both S and GS are included in the model, $g(E(Y)) = \beta_0 + \beta_S S + \beta_A G_A + \beta_{GS} GS$, by applying Theorem 1 again, testing $H_0: \beta_A = \beta_{GS} = 0$ is statistically equivalent between the different choices of the baseline allele *and* the assumption of the XCI status. For example, consider

$$X_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0.5 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix} \quad \text{and} \quad X_2 = \begin{pmatrix} 1 & 0 & 2 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \end{pmatrix},$$

respectively, for a model assuming XCI and choosing r as the baseline allele (i.e., tracking the number of copies of allele R), and for a model assuming no XCI and choosing R as the baseline allele, we can show that

$$T_{11} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad T_{12} = \begin{pmatrix} 2 & 0 \\ -1 & 1 \end{pmatrix} \quad \text{and} \quad T_{22} = \begin{pmatrix} -2 & 0 \\ 1 & -1 \end{pmatrix}$$

satisfy the linear transformation requirements of Theorem 1. That is, for association analysis of an X-chromosomal SNP, testing $H_0: \beta_A = \beta_{GS} = 0$ based on $g(E(Y)) = \beta_0 + \beta_S S + \beta_A G_A + \beta_{GS} GS$ is invariant to the choice of the baseline allele and the assumption of the X-inactivation status. Figure 1 summarizes the equivalency between the design matrices that correspond to the different coding schemes studied so far; all the theoretical results have been confirmed empirically via simulations. Our findings are supported by a recent simulation study (Song et al., 2021), where estimation biases in SNP coefficient estimate were observed "in several situations, particularly if the assumptions about XCI made by the coding scheme used and the assumptions made about sex differences in SNP effect of the fitted model were incorrect," while "Fitting models with SNP × Sex interaction terms can avoid reliance on assumptions."

(a) $g(E(Y)) = \beta_0 + \beta_A G_A$
$H_0 : \beta_A = 0$

(b) $g(E(Y)) = \beta_0 + \beta_A G_A + \beta_D G_D$
$H_0 : \beta_A = \beta_D = 0$

(a) $g(E(Y)) = \beta_0 + \beta_S S + \beta_A G_A$
$H_0 : \beta_A = 0$

(b) $g(E(Y)) = \beta_0 + \beta_S S + \beta_A G_A + \beta_D G_D$
$H_0 : \beta_A = \beta_D = 0$

(a) $g(E(Y)) = \beta_0 + \beta_S S + \beta_A G_A + \beta_{GS} GS$
$H_0 : \beta_A = \beta_{GS} = 0$

(b) $g(E(Y)) = \beta_0 + \beta_S S + \beta_A G_A + \beta_D G_D + \beta_{GS} GS$
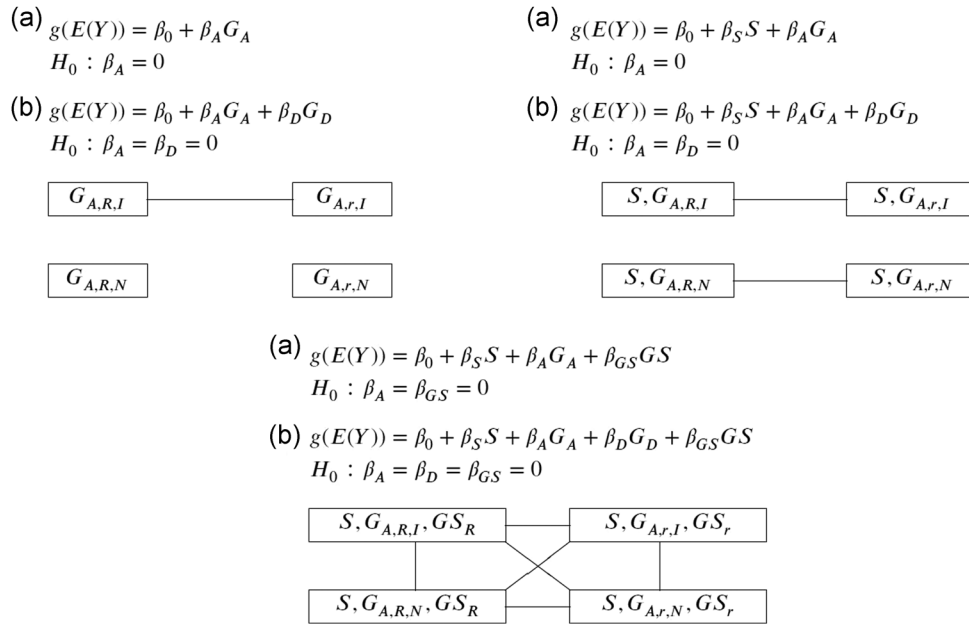$H_0 : \beta_A = \beta_D = \beta_{GS} = 0$

FIGURE 1 Equivalency between different regression models for association analysis of an X-chromosomal bi-allelic SNP. The subscript R or r represents the non-baseline allele of which we count the number of copies present in a genotype, and I or N denotes X-chromosome inactivated or not inactivated; see Table 2 for additional covariate coding details. Two groups of coding connected by a line if there is an invertible linear transformation between the design matrices as specified in Theorem 1, and the resulting test statistics for testing the specified H_0 will be identical to each other. Part (a) corresponds to models and tests without the dominance G_D covariate, and part (b) corresponds to models and tests with G_D included. Inclusion of G_D has no effect on the linear relationships established in part (a), because coding of G_D in Table 2 is invariant to the choice of the baseline allele or the XCI status. However, G_D effect is statistically equivalent to skewed XCI as shown in Section 2.4

2.4 | Random versus skewed X-inactivation (C7) and its connection with genetic dominance effect (C8)

Similar to analyzing an autosomal SNP, the first reason for modeling the dominance effect is to capture potential departure from additivity; see Section 3 for additional discussion. For an X-chromosomal SNP, another important reason is that the dominance effect can also *statistically* capture skewness of X-inactivation, if present.

Intuitively, if we assume the effects of rr and RR to be, respectively, 0 and 1, the effect of rR will be either 0 or 1 for each individual, depending on the inactivated allele of the sample collected. If the two alleles are equally likely to be inactivated (i.e., random XCI) across all individuals, the average statistical effect of rR is 1/2. If r is more (or less) likely to be inactivated (i.e., skewed XCI), the average effect of rR is greater (or less) than 1/2. However, this XCI skewness is *analytically* equivalent to a dominance effect (i.e., effect of rR deviating from 1/2), even though dominance effect is at the population level whereas skewed XCI is a sample-specific property. This analytical equivalency also shows that knowing the true

underlying biological model requires more than the standard GWAS data.

Table 3 summarizes the statistical behaviors of all the regression models and corresponding tests discussed in this section. Notably, jointly testing H_0: $\beta_A = \beta_D = \beta_{GS} = 0$, based on the $g(E(Y)) = \beta_0 + \beta_S S + \beta_A G_A + \beta_D G_D + \beta_{GS} GS$ model M_4, ensures that the inference is invariant to the assumptions of the XCI status and baseline allele, and accounts for dominance effect and XCI-skewness if present.

3 | ANALYTICAL AND SIMULATION-BASED METHOD EVALUATION

The proposed method is easy-to-implement and has good type I error control, because regression-based approach is known to be well behaved, as long as sample size is not too small and allele frequency is not too low, which are satisfied by most GWAS of common variants. Thus, we focus on evaluating power of the proposed method. We first provide a general analytical finding then present some simulation-based results.
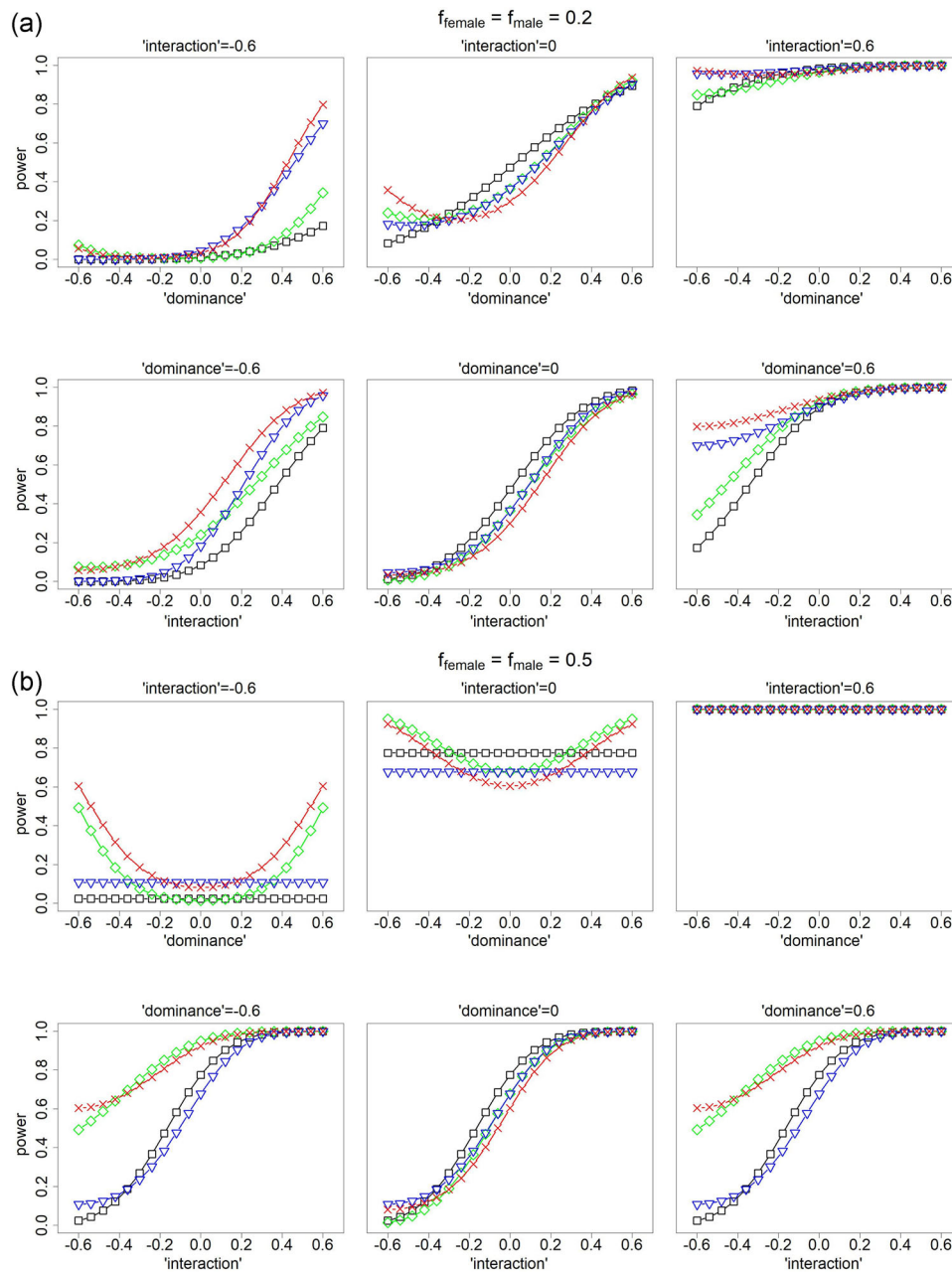
FIGURE 2 Power comparison for analyzing X-chromosomal single nucleotide polymorphisms (SNPs). Black $\square$ curves for testing $\beta_A = 0$ based on model M_1 as specified in Table 3, green $\diamond$ curves for testing $\beta_A = \beta_D = 0$ based on model M_2, blue $\triangledown$ curves for testing $\beta_A = \beta_{GS} = 0$ based on model M_3, and red $\times$ curves for testing $\beta_A = \beta_D = \beta_{GS} = 0$ based on the proposed model M_4. (a) $f_{female} = f_{male} = 0.2$ and (b) $f_{female} = f_{male} = 0.5$. Upper panels in (a) and (b) examine power as a function of the "dominance" effect. Lower panels in (a) and (b) examine power as a function of the gene–sex "interaction" effect. Note that biological dominance effect and skewed X-chromosome inactivation (XCI), and gene–sex interaction effect and the XCI status are statistically confounded with each other; see Section 3.2. Results for other parameter values including differential f between males and females are shown in Figures S5. The analyses related to M_1–M_3 assume that the true baseline allele is known and f being the allele frequency of the non-baseline allele, and the true XCI status is known at the population level. Unlike the other methods (M_1–M_3), the proposed method (M_4) is invariant to the assumptions of the baseline allele and the XCI status

3.1 │ Using the general theory of χ^2 distributions

One concern with the use of the proposed 3 df test is the potential loss of power due to the increased degrees of freedom. Indeed, if the true model for an X-chromosomal SNP is without a dominance effect and skewed inactivation, without gene-sex interaction, and the true inactivation status is known so that the additive genotype variable G_A can be correctly coded, then the corresponding 1 df test

Model, $g(E(Y))=$	Testing H_0:	C3/C4	C5/C6	C7/C8
M_1: $\beta_0 + \beta_S S + \beta_A G_A$	$\beta_A = 0$	√	×	×
M_2: $\beta_0 + \beta_S S + \beta_A G_A + \beta_D G_D$	$\beta_A = \beta_D = 0$	√	×	√
M_3: $\beta_0 + \beta_S S + \beta_A G_A + \beta_{GS} GS$	$\beta_A = \beta_{GS} = 0$	√	√	×
M_4: $\beta_0 + \beta_S S + \beta_A G_A + \beta_D G_D + \beta_{GS} GS$	$\beta_A = \beta_D = \beta_{GS} = 0$	√	√	√

TABLE 3 Properties of different regression models in the presence of the eight analytical challenges, as detailed in Table 1

Note: Whole-genome considerations such as C1 (continuous vs. binary traits) and C2 (Hardy–Weinberg equilibrium vs. disequilibrium) are naturally dealt with by the genotype-based regression approach. X-chromosome-specific considerations include C3 (choice of the baseline allele), C4 (sex as a confounder and type I error control), C5 (gene-sex interaction), C6 (X-chromosome inactivation (XCI) vs. no XCI), C7 (random vs. skewed XCI), and C8 (the dominance effect). In the table, × indicates a problem for the corresponding model and test, and √ means no problem. Relevant covariates E's should be included in the model but omitted here for notation simplicity. Joint testing of H_0: $\beta_A = \beta_D = \beta_{GS} = 0$ based on M_4 is the recommended, most robust approach; see Figures 2 and S5 for power comparisons among M_1–M_4.

will be more powerful than the proposed 3 *df* test. However, we show that even under the worst-case scenario and irrespective of sample size and the nominal type I error α level, the maximum power loss of the proposed 3 *df* is, surprisingly, capped at 18.8%, while the potential maximum power gain is $1 - \alpha$ (i.e., close to 100%).

Let $W_1 \sim \chi^2_{(1,ncp_1)}$, $W_2 \sim \chi^2_{(2,ncp_2)}$, and $W_3 \sim \chi^2_{(3,ncp_3)}$ be the 1, 2, and 3 *df* test statistics derived from the different regression models listed in Table 3. The power difference between the different W's depends on both the non-centrality parameters and α. When all the *ncp*'s are close to 0, all tests have no power. At the other extreme when all *ncp*'s are sufficiently large or α close to 1, all tests have power close to 1. Thus, we expect meaningful power comparison when *ncp*'s, and α have moderate values.

First, we assume that there are no dominance or interaction effects and the true XCI status is known to study the maximum power loss induced by unnecessarily including the G_D and GS terms. In that case, $ncp_1 = ncp_2 = ncp_3 = ncp$ and W_1, derived from $g(E(Y)) = \beta_0 + \beta_S S + \beta_A G_A$ with the correct genotype coding, is the optimal test. Varying *ncp* and α values, we numerically compute the power of the W's for $ncp \in [0, 100]$ and $-\log_{10}\alpha \in [0, 15]$. Figure S1 provides a heat plot for power as a function of *ncp* and α for the two tests. Results show that the maximum power loss of W_3 compared with W_1 is capped at 18.8%, regardless of the true additive effect size, sample size and the α level. The maximum occurs at $\alpha = 0.0008$ and $ncp = 13.4$ (Figure S1). At the genome-wide significance level $\alpha = 5 \times 10^{-8}$ (Dudgridge & Gusnanto, 2008), the maximum power loss is 17.7% occurring at $ncp = 32.6$.

Notably, the maximum of 18.8% power loss holds for comparing any 3 *df* χ^2 test with a 1 *df* χ^2 test that was

derived from the *correctly* specified 1 *df* model. This is because the derivation is based on *ncp* and α alone. Second, we emphasize that although a 18.8% loss of power is substantial, the fact that this is the maximum power loss for the 3 *df* test, under any true 1 *df* genetic model and regardless of the true genetic effect size, sample size, and significance level, is encouraging, as the potential power gain of the proposed 3 *df* test under other models can be much greater than 18.8% as we show next.

In the presence of dominance effect/skewed XCI or interaction effect/misspecified XCI, $ncp_3 = ncp_1 + \Delta_{13}$, where $\Delta_{13} > 0$. Compared with the maximum power loss of using the proposed 3 *df* for a 1 *df* (correctly specified) model, the maximum power gain under other genetic models can be theoretically as large as $1 - \alpha$. To provide specific numerical results, we consider $\alpha = 0.0008$ (the worse-case scenario derived above for the 3 *df* test), $ncp_1 = 5$, 10 or 15, and Δ_{13} ranging from 0 to 10. Results in Figure S2 show that once Δ_{13} is as large as half of ncp_1 (i.e., $ncp_3 \approx 1.5 \cdot ncp_1$), the 3 *df* test is more powerful than the 1 *df* test.

Together these two observations suggest that the proposed 3 *df* test is not only robust to the various model uncertainties associated with analyzing X-chromosomal variants, but it is reasonably powered as compared with the standard 1 *df* additive test. Compared with a 2 *df* test derived from correctly specified $g(E(Y)) = \beta_0 + \beta_S S + \beta_A G_A + \beta_D G_D$ or $g(E(Y)) = \beta_0 + \beta_S S + \beta_A G_A + \beta_{GS} G_{GS}$, the global maximum power loss of the proposed 3 *df* test is capped at 7.7%, occurring at $\alpha = 9.12 \times 10^{-5}$ and $ncp = 19$. At $\alpha = 5 \times 10^{-8}$, the maximum power loss is 7.5% occurring at $ncp = 34.2$ (Figure S1). In contrast, if the 2 *df* model is misspecified the potential power gain of the proposed 3 *df* test can be greater than 95%.

Power comparison between a 1 *df* test and a 2 *df* test is more relevant to the analysis of an autosomal SNP, but

the conclusion is similar to above. For example, under additivity, the maximum power loss of a 2 df genotypic test is capped at 11.4% across all parameter values and sample sizes. The maximum occurs at $\alpha = 0.0025$ and $ncp = 10.6$, and at the genome-wide significance level of $\alpha = 5 \times 10^{-8}$ (Dudgridge & Gusnanto, 2008), the maximum power loss is 10.3% when $ncp = 31.4$ (Figure S1). Supporting Information Appendix B also provides power comparison between the additive and genotypic tests for association analysis of an autosomal SNP across a range of dominance effects and allele frequencies (Figure S3). For each combination of parameter values considered, Figure S4 and Table S1 also provide the corresponding ncp_1 and ncp_2.

3.2 | Using different genetic models for the X-chromosome

Here we provide some empirical results based on different genetic models for an X-chromosomal variant and sample sizes. Note that tests derived from models that do not include sex as a covariate are susceptible to type I error rate inflation. Thus, power comparisons here focus on M_1–M_4 as specified in Table 3.

To compare the empirical power, we first derive the non-centrality parameters of the tests as functions of sample size, additive, dominance, and interaction effects, and under different assumptions of the baseline allele and X-inactivation status. We provide the technical details in Supporting Information Appendices C and D. We then considered $n = 1,000$, $\alpha = 0.0008$ (the worst case scenario for the 3 df test as shown in Section 3.1), and allele frequency $f_{male} = f_{female} = 0.2$ or 0.5. Results for other parameter values, including differential allele frequency values between males and females, are provided as Supporting Information; sex-specific allele frequencies may occur due to sex-specific selection.

Because of the various analytical equivalencies between GS interaction and XCI status, and between dominance effect and skewed XCI, the corresponding interaction, dominance, and skewed effect sizes are statistically confounded with each other. Thus, we specified the averaged statistical effect size for each of the five genotype groups, that is, $\mu_{rr}, \mu_{rR}, \mu_{RR}, \mu_r$, and μ_R. We fixed $\mu_{rr} = -0.3$, $\mu_{RR} = 0.3$ and $\mu_r = 0$, and varied μ_{rR} and μ_R from -0.6 to 0.6. Note that fixing μ_{rr} and μ_{RR} is equivalent to fixing the additive effect $\beta_A = 0.6$ under XCI or $\beta_A = 0.3$ under no XCI; varying μ_{rR} is equivalent to varying the dominance effect β_D from -0.6 to 0.6. The link with the interaction effect β_{GS} is less clear. Under the XCI assumption, $\beta_{GS} = (\mu_R - \mu_r) - (\mu_{RR} - \mu_{rr})$ $/2 = \mu_R - 0.3$, while under the no XCI assumption,

$\beta_{GS} = (\mu_R - \mu_r) - (\mu_{RR} - \mu_{rr})/4 = \mu_R - 0.15$. Thus, for the μ_R values considered here, β_{GS} ranged from -0.9 to 0.3 under XCI, and from -0.75 to 0.45 under no XCI. For ease of interpretation, Figure 2 uses the "dominance" and "interaction" terms to denote the varying degrees of μ_{rR} and μ_R.

Results in Figure 2 demonstrate the merits of the proposed method (testing $\beta_A = \beta_D = \beta_{GS} = 0$ jointly, the red × curves). While there could be some power loss in the worse case scenario (no G_D dominance or GS interaction effects), it is theoretically capped at 18.8% regardless of the parameter values. On the other hand, compared with the standard 1 df additive test (testing $\beta_A = 0$ and assuming the correct genotype coding, the black □ curves), power gain can be 70% for the cases considered here. When the allele frequency is 0.2 (Figure 2a), the performance of the 2 df additive and interaction test (testing $\beta_A = \beta_{GS} = 0$, the blue ∇ curves) is close to the proposed 3 df test. However, that is no longer the case when $f = 0.5$ (Figure 2b), where the 2 df additive and dominance test (testing $\beta_A = \beta_D = 0$, the green ◇ curves) is better and close to the proposed 3 df test. Figures S5 provides additional results for other parameter values, all showing the robustness of the proposed method, which is testing H_0: $\beta_A = \beta_D = \beta_{GS} = 0$ based on M_4, $g(E(Y)) = \beta_0 + \beta_S S + \beta_A G_A + \beta_D G_D + \beta_{GS} GS$. In practice the regression model should include relevant E's, which are omitted here for notation simplicity.

The proposed method not only resolves all C1–C8 analytical challenges simultaneously, but also has the best overall performance across the different underlying genetic models. However, we note that our robust method cannot identify the underlying true genetic model. This is true for any method that uses GWAS data alone, because we have shown, for example, XCI uncertainty is analytically equivalent to a gene–sex interaction effect, while XCI skewness is analytically equivalent to dominance effect. The lack of model identifiability of the proposed method, however, does not prevent a robust and powerful association analysis of X-chromosomal SNPs.

4 | APPLICATIONS TO THREE PREVIOUSLY PUBLISHED ASSOCIATION STUDIES

4.1 | Reanalyses of the X-chromosome-inclusive GWAS of Sun et al. (2012)

This data set consists of 3199 unrelated individuals with cystic fibrosis (CF) and 570,724 genome-wide bi-allelic SNPs after standard quality control (Sun et al., 2012).

Among the 570,724 SNPs, 14,279 are for the X-chromosome and 556,445 are from the autosomes. And among the 3199 CF subjects, 574 are cases with meconium ileus, an intestinal obstruction at birth seen in ≈15% of CF patients (Dupuis et al., 2016), and the remaining 2625 CF subjects are controls; 1722 are males and 1477 are females. The rates of meconium ileus are 17.7% and 18.3%, respectively, in the male and female groups, which is not statistically different.

A previous X-chromosome-inclusive GWAS of meconium ileus in CF has been conducted based on this data set (Sun et al., 2012), where the standard 1 df additive test was used for analyzing the autosomal SNPs, and X-chromosome being inactivated was further assumed for analyzing the X-chromosomal SNPs (i.e., using model M_1 in Table 3 with genotype coding under the assumption of XCI). Here we reanalyze both the autosomal and X-chromosomal SNPs to demonstrate the utility of the proposed approach.

For the X-chromosome, we compared the M_1–M_4 models and their corresponding tests as detailed in Table 3 in Section 2. For each SNP, we performed six different association tests, depending on which of the M_1–M_4 models was used and if the XCI status needed to be specified, because (a) sex must be included to ensure correct type I error rate control and models including S are invariant to the choice of the baseline allele (Section 2.2), and (b) models including the gene-sex interaction effect are invariant to the assumption of XCI (Section 2.3). Figure 3a shows the results for the top 15 ranked X-chromosomal SNPs, ordered by the minimal p value of all six tests; the lines connecting the SNPs are used only for visualization purposes to demonstrate the robustness of a method.

The application results here are consistent with our earlier analytical and simulation results in Section 3, showing that joint modeling and testing the additive, dominance, and gene–sex interaction effects is the most robust association approach for analyzing X-chromosomal variants. For example, the result of Sun et al. (2012) (M_1 and assuming XCI) marked by the black □ curve is clearly less "stable" than the red × curve (the proposed M_4) across different SNPs. In this particular application, we observed that the performance of the orange + curve (M_2 assuming no XCI) is similar to the proposed method. However, interestingly, the green ◇ curve (also M_2 but assuming XCI) is noticeably different from the orange + curve.

For the autosomal SNPs, we contrast the standard 1 df additive test with the proposed 2 df genotypic test as briefly discussed in Section 3.1 and detailed in Supporting Information Appendix B. Figure 3b shows the results for the top 15 ranked autosomal SNPs, ordered by the minimal p value of additive and genotypic tests; Figures S7 and S8 provide genome-wide results. It is clear that if the p values of the standard 1 df additive test (the black □ curve) are smaller, then those from the recommended 2 df genotypic test (the red × curve) are close in magnitude, while the reverse is not true. For example, p value of the recommended 2 df genotypic test for the 6th SNP (rs2657147) in the plot is more than four orders of magnitude smaller than that of the 1 df additive test; there is no evidence for genotyping error at this SNP as the p value of HWE test in the control group is 0.026. The genotype counts for rr, rR, and RR are (210, 312, 52) in the case group and (1012, 1192, 421) in the control group, which yields case/control ratios of (0.208, 0.262, 0.124), clearly suggesting a dominance pattern. Whether this is a true new finding, however, requires further investigation.

4.2 | Evidence from the first (autosome only) GWAS of WTCCC (2007)

We then examined the results of the first (autosome only) GWAS, conducted by the Wellcome Trust Case Control Consortium (WTCCC, 2007). Their tab. 3 lists regions of the genome showing the strongest association signals and provides results from both the 1 df trend test (statistically equivalent to the additive test considered here) and the 2 df genotypic tests.

Consistent with the autosomal results of the CF meconium ileus application above, the results in tab. 3 of WTCCC (2007) also show that if the 1 df additive test provides a smaller p value, the p value of the 2 df genotypic test is at most one order of magnitude larger. For example, the p values are 1.16×10^{-13} and 1.79×10^{-14}, respectively, for the 1 df additive and 2 df genotypic tests, testing association between coronary artery disease and rs1333049, the second SNP in Table 3 of WTCCC (2007). On the other hand, the p value of the 2 df genotypic test can be several orders of magnitude smaller that of the 1 df additive test. For example, the p values are 2.19×10^{-4} and 6.29×10^{-8}, respectively, for the 1 df additive and 2 df genotypic tests, testing association between bipolar disorder and rs420259, the first SNP in Table 3 of WTCCC (2007); the association between rs420259 and bipolar disorder has since been replicated by other studies (Gonzalez et al., 2016; Tesli et al., 2010). We can draw similar conclusions based on the Bayes factors provided in their tab. 3, obtained under the 1 df additive or 2 df genotypic models.
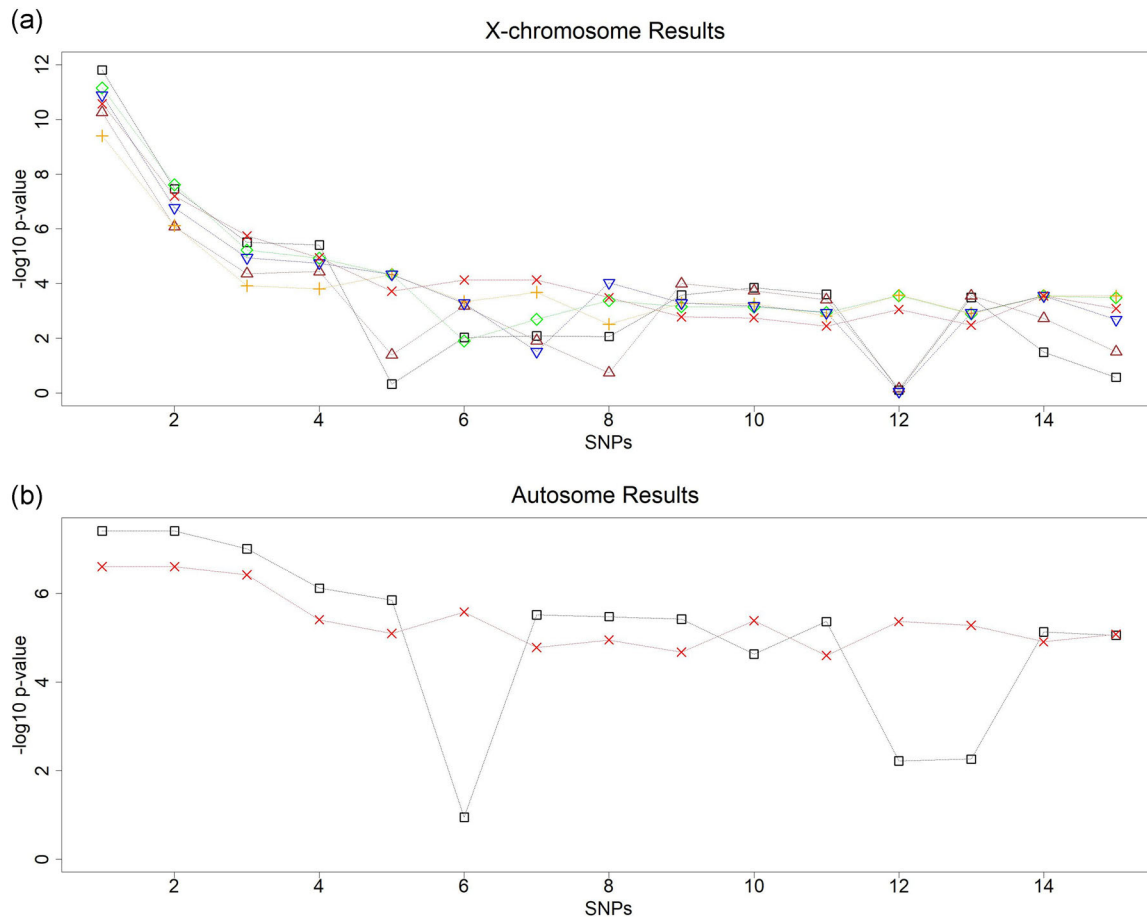
FIGURE 3 Results of a genome-wide association study of meconium ileus in cystic fibrosis subjects. In total, 3199 independent cystic fibrosis subjects, 14,279 X-chromosomal single nucleotide polymorphisms (SNPs), and 556,445 autosomal SNPs are analyzed. The SNPs are ordered by the minimal p value of the different tests considered, and the lines connecting the SNPs are used only for visualization purposes to demonstrate the robustness of a particular method. (a) X-chromosome results. These top 15 ranked X-chromosomal SNPs are selected based on any of the six tests based on M_1–M_4 models in Table 3: the Black □ curve for testing $\beta_A = 0$ based on M_1 assuming X-chromosome inactivation (XCI), the brown △ curve for testing $\beta_A = 0$ based on M_1 assuming no XCI, the green ◇ curve for testing $\beta_A = \beta_D = 0$ based on M_2 assuming XCI, the orange + curve for testing $\beta_A = \beta_D = 0$ based on M_2 assuming no XCI, the blue ▽ curve for testing $\beta_A = \beta_{GS} = 0$ based on M_3 (invariant to the XCI assumptions if GS is included in the model and tested), and the red × curve for testing $\beta_A = \beta_D = \beta_{GS} = 0$ based on the recommended model M_4 that is most robust for analyzing the X-chromosome. (b) Autosome results. These top 15 ranked autosomal SNPs are selected based on either the 1 df additive test or the 2 df genotypic test. The black □ curve for testing $\beta_A = 0$ using the standard additive model, and the red × curve for testing $\beta_A = \beta_D = 0$ using the recommend genotypic model that is most robust for analyzing the autosomes

4.3 | Re-analyses of the 60 autosomal SNPs potentially associated with various complex traits, selected by Wittke-Thompson et al. (2005)

Finally, we re-analyzed the 60 autosomal SNPs selected by Wittke-Thompson et al. (2005) from 41 case-control association studies of various complex traits, including Alzheimer's disease and breast cancer; the genotype count data are available from Table 1 of Wittke-Thompson et al. (2005). Although these SNPs were originally selected by Wittke-Thompson et al. (2005) for a

study of departure from HWE, genotype-based methods are robust to the HWE assumption (Sasieni, 1997; Zhang & Sun, 2021). Here we focused on comparing the standard 1 df additive test with the recommended 2 df genotypic test for analyzing these 60 autosomal SNPs, which are presumed to be associated with complex traits based on the earlier 41 studies.

We observed that the genotypic test leads to 31 SNPs with p values less than $\alpha = 0.05$, while the additive test results in 22 SNPs (Figure S6). Using the Bonferroni threshold of $\alpha = 0.05/60$ the numbers are 7 and 6, respectively, for the genotypic and additive tests. Although

these autosomal SNPs can only be presumed to be associated with the various complex traits, the empirical evidence here is consistent with the analytical and simulation results in Section 3.1 and Supporting Information Appendix B.

5 | DISCUSSION

We have shown that in association analysis of an X-chromosomal variant, the sex main effect must be included to achieve correct type I error rate control. The inclusion of sex also addresses the complication of baseline allele specification that otherwise affects association inference for an X-chromosomal SNP, in contrast to an autosomal SNP. Although the method developed here is motivated by genetic association studies of the X-chromosome, Theorem 1 is applicable to other settings where model uncertainty plays a role. For association studies of autosomal variants, sex is not routinely included. However, sex can be a confounder for an autosomal SNP as well, for example, when there is sex difference in allele frequency due to sex-specific selection. When the allele frequency difference is small, including sex does not substantively change the association result, because sex is not directly included in the genetic association test. Thus, we recommend to always include sex as a covariate in association analysis of either autosomal or X-chromosomal variants.

We have also shown that modeling the genetic dominance effect β_D is beneficial for analyzing both the autosomes and X-chromosome. The proposed model can significantly increase test power when β_D is large. When β_D is close to 0, the model is still robust and maintains 'comparable' power with that of the additive model; "comparable" is in the context of the trade-off between the maximum power loss and gain across different models. For an autosomal SNP, we have shown analytically that even under true additivity, compared with the classical 1 df additive test, the maximum power loss of the 2 df genotypic test is capped at 11.4%, regardless of the sample, genetic effect and test sizes, but the power gain can be as high as $1 - \alpha$. Similarly, for an X-chromosome SNP, with a 3 df test that includes β_A, β_D, and β_{GS} interaction effects, power loss is capped at 18.8%; this assumes that the standard 1 df additive test used the correct XCI model and there is no skewed XCI or dominance effect. If these assumptions do not hold, the potential power gain of the 3 df test can be as high as $1 - \alpha$. However, not all alternative genetic models are equally likely in practice. Consistent with the earlier work of Hill et al. (2008) and Bush and Moore (2012), two recent studies showed that "genetic variance for complex traits is predominantly additive" (Hivert et al., 2021; Pazokitoroudi et al., 2021). To this end, a Bayesian alternative that incorporates prior evidence for the different genetic models can be considered.

When the true genetic model is unknown, one alternative frequentist's approach is to consider all possible models and use the "best" or weighted average. But, such an approach is difficult to implement in practice; see Bagos (2013) for a review. For example, selection bias inherent in choosing the best-fitted model must be corrected for, often through computationally intensive simulation studies, and power of this bias-corrected inferential procedure is not clear. On the other hand, ways to obtain a weighted average of the test statistics or p values across all models can be quite ad hoc, and the optimal weighting factors are difficult to derive. The recent Cauchy method can be used to combine correlated p values derived from all possible genetic models (Liu & Xie, 2020). Finally, the method proposed here is tailored for analyzing one common SNP at a time, and joint analysis of multiple common or rare SNPs (Derkach et al., 2014) requires further consideration.

When the true genetic model is unknown, another alternative is to use sex-stratified analysis, followed by meta-analysis combining the female and male groups (Willer et al., 2010). This approach appears to be robust to the XCI assumption when analyzing an X-chromosomal variant, because association evidence in females are the same between the XCI and no-XCI assumptions. However, the two assumptions lead to different effect size estimates by a factor of two (the standard errors also differ by a factor of two), resulting in different results using the inverse-variance-based meta-analysis. The sample-size-based meta-analysis can overcome this limitation, but other issues remain including difficulty of modeling non-additive or gene–sex interaction effects.

Summary statistics from the proposed 3 df test for X-chromosome SNPs can be used to perform meta-analysis by using, for example, Fisher's combined p-value approach. The classical inverse-variance-based method, however, is not applicable for two reasons. Firstly, there are multiple genotype-related β estimates, β_A, β_D, and β_{GS}. Secondly, and more importantly, some of the β estimates are not meaningful on their own as we have shown that skewed XCI and dominance effects are statistically confounded with each other, so are the $G \times S$ interaction effect and the assumption of XCI. Even if we limit our attention to the genetic main additive effect, the effect size estimate changes by a factor of two depending on the XCI assumption (i.e., the genotype coding scheme). Thus, our work here also highlights new challenges associated with other analyses of X-chromosomal SNPs. For example, how

to aggregate association evidence across multiple SNPs or multiple traits (Zhao & Sun, 2021), and how to perform X-chromosome-inclusive polygenic risk score (PRS) analysis (Dudbridge et al., 2018), both of which we will address in future research.

The proposed full model for analyzing an X-chromosomal SNP, $g(E(Y)) = \beta_0 + \beta_S S + \beta_A G_A + \beta_D G_D + \beta_{GS} GS$, is robust to various model uncertainties, analytically. However, as noted earlier it is not capable of differentiating between the scenarios. Indeed, the recent work by Song et al. (2021), focusing on effect size estimation as opposed to association testing, showed that genetic effect estimates are sensitive to model assumptions. Using the available genetic association data, Ma et al. (2015) proposed a variance-based test for detecting X-inactivation by comparing phenotypic variance of the rR group with that of the rr and RR groups in females, but this method is limited to a continuous trait (Deng et al., 2019; Soave & Sun, 2017). Wang et al. (2014) explicitly introduced a parameter to represent the amount of skewness of X-inactivation. Our work here, however, shows that the interpretation of their parameter is statistically confounded with dominance genetic effect using GWAS data alone. How to incorporate additional 'omic' data (Carrel & Willard, 2005) to tease apart different biological phenomenon is an interesting problem that deserves further investigation.

ACKNOWLEDGMENTS

CONFLICT OF INTERESTS

The authors declare that there are no conflict of interests.

AUTHOR CONTRIBUTIONS

Bo Chen developed the method and designed the study with significant input from Lei Sun and Radu Craiu. Bo Chen performed the data analysis with significant input from Lei Sun and Lisa J. Strug. Bo Chen wrote the manuscript with significant contributions from all co-authors. All authors have reviewed and approved the final manuscript.

DATA AVAILABILITY STATEMENT

The meconium ileus application data is available by application to the Cystic Fibrosis Canada National data registry for researchers who meet the criteria for access to confidential clinical data for the purpose of CF research. The 60-SNP application data is publicly available from Table 1 of Wittke-Thompson et al. (2005). The WTCCC application used only the summary statistics reported in Table 3 of WTCCC (2007).

ORCID

Bo Chen http://orcid.org/0000-0002-5916-4443
Lei Sun http://orcid.org/0000-0002-5640-937X

REFERENCES

Bagos, P. G. (2013). Genetic model selection in genome-wide association studies: Robust methods and the use of meta-analysis. *Statistical Applications in Genetics and Molecular Biology*, *12*(3), 285–308.

Bush, W. S., & Moore, J. H. (2012). Chapter 11: Genome-wide association studies. *PLOS Computational Biology*, *8*(12), e1002822.

Busque, L., Mio, R., Mattioli, J., Brais, E., Blais, N., Lalonde, Y., Maragh, M., & Gilliland, D. G. (1996). Nonrandom X-inactivation patterns in normal females: lyonization ratios vary with age. *Blood*, *88*(1), 59–65.

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

Chen, B., Craiu, R. V., & Sun, L. (2020). Bayesian model averaging for the X-chromosome inactivation dilemma in genetic association study. *Biostatistics*, *21*(2), 319–335.

Chen, Z., Ng, H. K. T., Li, J., Liu, Q., & Huang, H. (2017). Detecting associated single-nucleotide polymorphisms on the X chromosome in case control genome-wide association studies. *Statistical Methods in Medical Research*, *26*(2), 567–582.

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

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

Deng, W., Mao, S., Kalnapenkis, A., Esko, T., Magi, R., Paré, G., & Sun, L. (2019). Analytical strategies to include the X-chromosome in variance heterogeneity analyses: Evidence for trai-specific polygenic variance structure. *Genetic Epidemiology.*, *43*(7), 815–830.

Derkach, A., Lawless, J. F., & Sun, L. (2014). Pooled association tests for rare genetic variants: A review and some new results. *Statistical Science*, 29(2), 302–321.

Dizier, M. H., Demenais, F., & Mathieu, F. (2017). Gain of power of the general regression model compared to Cochran-Armitage trend tests: simulation study and application to bipolar disorder. *BMC Genetics*, 181(24).

Draper, D. (1995). Assessment and propagation of model uncertainty. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1), 45–97.

Dudbridge, F., Pashayan, N., & Yang, J. (2018). Predictive accuracy of combined genetic and environmental risk scores. *Genetic Epidemiology*, 42(1), 4–19.

Dudbridge, F., & Gusnanto, A. (2008). Estimation of significance thresholds for genomewide association scans. *Genetic Epidemiology*, 32(3), 227–234.

Dupuis, A., Keenan, K., Ooi, C. Y., Dorfman, R., Sontag, M. K., Naehrlich, L., Castellani, C., Strug, L. J., Rommens, J. M., & Gonska, T. (2016). Prevalence of meconium ileus marks the severity of mutations of the Cystic Fibrosis Transmembrane Conductance Regulator (CFTR) gene. *Genetics in Medicine*, 18(4), 333–340.

Gao, F., Chang, D., Biddanda, A., Ma, L., Guo, Y., Zhou, Z., & Keinan, A. (2015). XWAS: A software toolset for genetic data analysis and association studies of the X chromosome. *Journal of Heredity*, 106(5), 666–671.

Gendrel, A. V., & Heard, E. (2011). Fifty years of X-inactivation research. *Development*, 138, 5049–5055.

Gonzalez, S., Gupta, J., Villa, E., Mallawaarachchi, I., Rodriguez, M., Ramirez, M., Zavala, J., Armas, R., Dassori, A., Contreras, J., Flores, D., Jerez, A., Ontiveros, A., Nicolini, H., & Escamilla, M. (2016). Replication of genome-wide association study (GWAS) susceptibility loci in a Latino bipolar disorder cohort. *Biopolar Disorders*, 18(6), 520–527.

Hickey, P. F., & Bahlo, M. (2011). X chromosome association testing in genome wide association studies. *Genetic Epidemiology*, 35, 664–670.

Hill, W. G., Goddard, M. E., & Visscher, P. M. (2008). Data and theory point to mainly additive genetic variance for complex traits. *PLOS Genetics*, 4(2), e1000008.

Hivert, V., Sidorenko, J., Rohart, F., Goddard, M. E., Yang, J., Wray, N. R., Yengo, L., & Visscher, P. M. (2021). Estimation of non-additive genetic variance in human complex traits from a large sample of unrelated individuals. *American Journal of Human Genetics*, 108(5), 786–798.

Konig, I. R., Loley, C., Erdmann, J., & Ziegler, A. (2014). How to include chromosome X in your genome-wide association study. *Genetic Epidemiology*, 38, 97–103.

Liu, Y., & Xie, J. (2020). Cauchy combination test: A powerful test with analytic p-value calculation under arbitrary dependency structures. *Journal of the American Statistical Association*, 115(529), 393–402.

Loley, C., Ziegler, A., & Konig, I. R. (2011). Association tests for X-chromosomal markers—A comparison of different test statistics. *Human Heredity*, 71, 23–36.

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

McCullagh, P., & Nelder, J. A. (1989). *Generalized linear models* (2nd ed.). Chapman & Hall.

Ozbek, U., Lin, H. M., Lin, Y., Weeks, D. E., Chen, W., Shaffer, J. R., Purcell, S. M., & Feingold, E. (2018). Statistics for X-chromosome associations. *Genetic Epidemiology*, 42(6), 539–550.

Pazokitoroudi, A., Chiu, A. M., Burch, K. S., Pasaniuc, B., & Sankararaman, S. (2021). Quantifying the contribution of dominance deviation effects to complex trait variation in biobank-scale data. *American Journal of Human Genetics*, 108(5), 799–808.

Sasieni, P. D. (1997). From genotypes to genes: Doubling the sample size. *Biometrics*, 53(4), 1253–1261.

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

Song, Y., Biernacka, J., & Winham, S. (2021). Testing and estimation of x-chromosome SNP effects: Impact of model assumptions. *Genetic Epidemiology*, 1–16.

Sun, L., Rommens, J. M., Corvol, H., Li, W., Li, X., Chiang, T. A., , Lin, F., Dorfman, R., Busson, P. F., Parekh, R. V., Zelenika, D., Blackman, S. M., Corey, M., Doshi, V. K., Henderson, L., Naughton, K. M., O'Neal, W. K., Pace, R. G., Stonebraker, J. R., ... Strug, L. J. (2012). Multiple apical plasma membrane constituents are associated with susceptibility to meconium ileus in individuals with cystic fibrosis. *Nature Genetics*, 44(5), 562–569.

Tesli, M., Athanasiu, L., Mattingsdal, M., Kahler, A. K., Gustafsson, O., Andreassen, B. K., Werge, T., Hansen, T., Mors, O., Mellerup, E., Koefoed, P., Jönsson, E. G., Agartz, I., Melle, I., Morken, G., Djurovic, S., & Andreassen, O. A. (2010). Association analysis of palb2 and brca2 in bipolar disorder and schizophrenia in a Scandinavian case-control sample. *American Journal of Medical Genetics Part B Neuropsychiatric Genetics*, 13B(7), 1276–1282.

Tukiainen, T., Villani, A.-C., Yen, A., Rivas, M. A., Marshall, J. L., Satija, R., Aguirre, M., Gauthier, L., Fleharty, M., Kirby, A., Cummings, B. B., Castel, S. E., Karczewski, K. J., Aguet, F., Byrnes, A., GTEx Consortium, Lappalainen, T., Regev, A., Ardlie, K. G., ... MacArthur, D. G. (2017). Landscape of X chromosome inactivation across human tissues. *Nature*, 550(7675), 244–248.

Visscher, P. M., Wray, N. R., Zhang, Q., Sklar, P., McCarthy, M. I., Brown, M. A., & Yang, J. (2017). 10 years of GWAS discovery: Biology, function, and translation. *American Journal of Human Genetics*, 101(1), 5–22.

Wang, J., Talluri, R., & Shete, S. (2017). Selection of X-chromosome inactivation model. *Cancer Informatics*, 16, 1–8.

Wang, J., Yu, R., & Shete, S. (2014). X-chromosome genetic association test accounting for X-inactivation, skewed X-inactivation, and escape from X-inactivation. *Genetic Epidemiology*, 38, 483–493.

Wellek, S., & Ziegler, A. (2012). Cochran-Armitage test versus logistic regression in the analysis of genetic association studies. *Human Heredity*, 73(1), 14–17.

Willer, C. J., Li, Y., & Abecasis, G. R. (2010). METAL: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*, 26(17), 2190–219.

Wise, A. L., Gyi, L., & Manolio, T. A. (2013). eXclusion: Toward integrating the X chromosome in genome-wide association analyses. *American Journal of Human Genetics*, *92*, 643–647.

Wittke-Thompson, J. K., Pluzhnikov, A., & Cox, N. J. (2005). Rational inferences about departures from Hardy-Weinberg equilibrium. *American Journal of Human Genetics*, *76*, 967–986.

WTCCC. (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, *447*, 661–678.

Zhang, L., & Sun, L. (2021). A generalized robust allele-based genetic association test. *Biometrics*.

Zhao, Y., & Sun, L. (2021). On set-based association tests: Insights from a regression using summary statistics. *The Canadian Journal of Statistics*.

Zheng, G. (2008). Can the allelic test be retired from analysis of case-control association studies? *Annals of Human Genetics*, *72*, 848–851.

Zheng, G., Joo, J., Zhang, C., & Geller, N. L. (2007). Testing association for markers on the X chromosome. *Genetic Epidemiology*, *31*, 834–843.

Zhou, Z., Ku, H. C., Huang, Z., Xing, G., & Xing, C. (2017). Differentiating the Cochran-Armitage trend test and Pearson's chi-squared test: Location and dispersion. *Annals of Human Genetics*, *81*(5), 184–189.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.