Association Tests for X-Chromosomal Markers – A Comparison of Different Test Statistics

Christina Loley[a, b]   Andreas Ziegler[a]   Inke R. König[a]

[a]Institut für Medizinische Biometrie und Statistik, Universität zu Lübeck, und [b]Medizinische Klinik II, Universitätsklinikum Schleswig-Holstein, Campus Lübeck, Lübeck, Deutschland

Abstract

Objective: Genome-wide association studies have successfully elucidated the genetic background of complex diseases, but X chromosomal data have usually not been analyzed. A reason for this is that there is no consensus approach for the analysis taking into account the specific features of X chromosomal data. This contribution evaluates test statistics proposed for X chromosomal markers regarding type I error frequencies and power. **Methods:** We performed extensive simulation studies covering a wide range of different settings. Besides characteristics of the general population, we investigated sex-balanced or unbalanced sampling procedures as well as sex-specific effect sizes, allele frequencies and prevalence. Finally, we applied the test statistics to an association data set on Crohn's disease. **Results:** Simulation results imply that in addition to standard quality control, sex-specific allele frequencies should be checked to control for type I errors. Furthermore, we observed distinct differences in power between test statistics which are determined by sampling design and sex specificity of effect sizes. Analysis of the Crohn's disease data detects two previously un-known genetic regions on the X chromosome. **Conclusion:** Although no test is uniformly most powerful under all settings, recommendations are offered as to which test performs best under certain conditions.

Copyright © 2011 S. Karger AG, Basel

Introduction

In the past, genome-wide association (GWA) studies have been successful in elucidating the genetic background of complex diseases. In these studies, single nucleotide polymorphisms (SNPs) are tested for association with a disease or phenotype. The focus of GWA studies and subsequent meta-analyses, however, has been on the autosomes, whereas X chromosomal data have usually been collected but not analyzed [1–4].

The probable reason for this neglect is that handling of X chromosomal genotype data is not yet standardized: there are no standard statistics established to test for association, and although special criteria for quality control have been defined, no standard thresholds for these have been approved [5]. Also, different genotype-calling algorithms are needed for the X chromosome which are not implemented in all programs. As a consequence, only few associations have been reported for the X chromosome,

in contrast to numerous established associations on the autosomes [6]. Thus, the genetic information located on the X chromosome is to a great extent lost. This is even more dramatic since the inheritance patterns of many complex diseases are known to be sex-determined, which should ascribe vital importance to the analysis of X chromosomal data.

To acknowledge the specific characteristics of X chromosomal data, we need to bear in mind that for the autosomes, males and females carry two copies of each chromosome, resulting in three possible genotypes per SNP. For females, this also holds for the X chromosome. Males, however, have one X chromosome and one Y chromosome. For some loci on the X chromosome, the so-called 'pseudo-autosomal' loci, homologous loci on the Y chromosome exist, so that both males and females, again, carry two alleles per SNP. For the remaining loci, males carry only one allele at each SNP, resulting in only two possible genotypes per SNP. While SNPs in pseudo-autosomal regions can be analyzed by known association tests for autosomes [7], SNPs on other X chromosomal loci need special treatment as soon as the sample analyzed contains both male and female subjects.

A further special feature of the X chromosome is the process of inactivation. Early in embryonic development, great parts of one of the two female X chromosomes are silenced by the XIST RNA [8]. It has been suggested that this is a mechanism of dosage compensation, resulting in equal effects for one copy of the X chromosome in males and two copies in females. Considering disease predisposing loci, this means that males with one risk allele have a comparable risk to females homozygous for the risk allele. It is estimated that about three quarters of X chromosomal genes are silenced on one of the female X chromosomes, while the remaining loci may escape inactivation in some females [9].

So far, there has been little work regarding association tests for markers on the X chromosome. Zheng et al. [10] proposed different tests for a model without taking the possible inactivation of the female X chromosome into account. They also provide results from a simulation study demonstrating power and error levels of the presented tests in some basic situations. A number of interesting situations that include differences between male and female sub-samples were not considered, such as different numbers of males and females in the sample and different proportions of males and females in cases and controls, or different effect sizes for males and females.

Another approach was taken by Clayton [11], who derived tests modeling one of the female X chromosomes as

inactivated. Clayton [11] does not present a simulation study concerning power and error levels of the suggested tests, but they were applied in a GWA setting [12]. A systematic comparison of all test statistics proposed by Zheng et al. [10] and Clayton [11] is still missing. The purpose of this article is therefore to compare these tests with regard to power and error levels in a wide range of settings. In order to do so, we simulated SNP association data for a sample of males and females. We considered models with and without inactivation of one female X chromosome. Besides characteristics of the general population, we investigated a variety of sex-specific parameters, including differences in effect sizes, allele frequencies and prevalence between males and females. Furthermore, we considered unbalanced sampling procedures resulting in different numbers of females and males in the sample or in a sex-specific composition of case and control samples. Additionally, we applied the test statistics to the analysis of a sample of 300 Crohn's disease cases and 432 unrelated controls.

Materials and Methods

Existing tests for associations on the X chromosome make different assumptions on the influence of a single allele in females compared to males. The intuitive approach is to simply count alleles in males and females, suggesting that one allele in a male subject has the same influence as one allele in a female subject. This corresponds to a model with both female X chromosomes active. Under this assumption, males with one risk allele are treated like heterozygous females. Another strategy takes the idea of X chromosome inactivation into account. Here, males are treated like homozygous females. Zheng et al. [10] proposed different tests for a model without X inactivation, whereas Clayton [11] derived tests modeling one of the female X chromosomes as inactivated.

Allele and genotype counts for SNPs on the X chromosome are presented in table 1 (adapted from Zheng et al. [10]), and details of all test statistics are summarized in table 2.

Test Statistics Assuming No X Inactivation
The first test proposed by Zheng et al. [10] is the allele-based test for the entire sample of males and females, Z_A^2. Here, differences in allele counts between cases and controls are compared jointly for males and females (table 1d). This is the most intuitive approach, but under departure from Hardy-Weinberg equilibrium (HWE), allele-based test statistics like Z_A^2 are known to deviate from the proposed distribution, which reduces the reliability of the test results [13]. Additionally, allele frequencies are estimated over the entire sample, which, again, may lead to a violation of distribution assumptions for sex-specific allele frequencies.

The further tests are based on separate test statistics for male and female sub-samples. For males, the allele-based test statistic Z_m^2 is calculated (allele counts according to table 1a). For females, two test statistics are considered, the genotype-based trend test

Z_{fG}^2 and the allele-based test Z_{fA}^2 (table 1b, c). Tests Z_C^2, Z_{mfA}^2, and Z_{mfG}^2 are now different combinations of Z_m^2 and Z_{fA}^2 or Z_{fG}^2 (table 2). The test Z_C^2 is genotype based, and allele frequencies are estimated separately for males and females. Hence, this test does not require HWE or equal allele frequencies. However, it has 2 degrees of freedom (d.f.) and thus potentially less power than a 1-d.f. test. Therefore, Zheng et al. [10] construct tests Z_{mfA}^2 and Z_{mfG}^2. As these are weighted sums of tests Z_m^2 and Z_{fA}^2 or Z_{fG}^2, these tests have only 1 d.f. Z_{mfA}^2, again, is based on allele counts for females and requires HWE in that cohort. As differences in allele frequencies between sexes also imply departure from HWE [10], this test may not be valid for sex-specific allele frequencies either.

The test statistics Z_m^2, Z_{fA}^2 and, for a purely female sample, Z_A^2 correspond to the typically used allele-based χ^2 test with 1 d.f. for association on the autosomes. Likewise, Z_{fG}^2 is equivalent to the ordinary Cochran-Armitage trend test [7].

Zheng et al. [10] also proposed two tests for situations where the effect alleles for males and females are different. These are not considered here, since we do not regard this as biologically plausible.

Test Statistics Assuming X Inactivation

We first consider two score test statistics proposed by Clayton [11], one testing for an additive genetic model and the other testing for both additive and dominant models of inheritance. Both tests are derived for a general phenotype and a biallelic locus.

Clayton's [11] additive test, T_A, corresponds to the usual Cochran-Armitage trend test for a completely female sample and differs from this as soon as the sample contains males. The genotype counts in females, A_i, take values 0, 1 or 2 corresponding to 0, 1 or 2 risk alleles, respectively. For males, values 0 or 2 are possible corresponding to 0 or 1 risk allele. This genotype coding is based on X chromosomal inactivation and the assumption that two risk alleles in females have the same effect as one risk allele in males. The score of the additive model, U_A, is calculated over the entire sample. This is based on the assumption of equal allele frequencies for males and females. If these differ, the distribution assumptions for T_A do not hold. Since variances in female and male samples differ, these must be calculated separately for the two subgroups even for equal allele frequencies in males and females.

To model a dominant effect, Clayton [11] introduces a heterozygosity indicator, D, taking the value 1 for heterozygotes and 0 for homozygotes. As there are no heterozygous males, D is set to 0 for all men. Combining the additive score U_A and the dominant score U_D, Clayton [11] derives a 2-d.f. score test, which should be able to detect both additive and dominant genetic effects. Again, the covariances of U_A and U_D have to be calculated separately for male and female subsamples. The test T_{AD} also requires equal allele frequencies in males and females. But since both T_A and T_{AD} are based on genotype counts rather than allele counts, they remain valid under departure from HWE in females.

If allele frequencies differ between sexes, stratified tests need to be calculated. Clayton [11] proposed to calculate the above-mentioned test T_{AD} in exactly the same way separately for males and females and to then add both test statistics. Since for males no dominance term is calculated, this test has 3 d.f. Equivalently, a stratified version of the test T_A can be calculated. This would lead to a 2-d.f. test statistic, T_A^s, asymptotically equivalent to the test Z_C^2 proposed by Zheng et al. [10], and will therefore not be considered. Another possibility to calculate a stratified test for the

Table 1. Genotype and allele counts for X chromosomal SNPs

a Male allele counts

	A	a	Total
Cases	r_{m0}	r_{m1}	r_m
Controls	s_{m0}	s_{m1}	s_m
Total	n_{m0}	n_{m1}	n_m

b Female genotype counts

	AA	Aa	aa	Total
Cases	r_{f0}	r_{f1}	r_{f2}	r_f
Controls	s_{f0}	s_{f1}	s_{f2}	s_f
Total	n_{f0}	n_{f1}	n_{f2}	n_f

c Female allele counts

	A	a	Total
Cases	$2r_{f0} + r_{f1}$	$2r_{f2} + r_{f1}$	$2r_f$
Controls	$2s_{f0} + s_{f1}$	$2s_{f0} + s_{f2}$	$2s_f$
Total	$2n_{f0} + n_{f1}$	$2n_{f0} + n_{f2}$	$2n_f$

d Male and female allele counts

	AA	Aa	Total
Cases	$2r_{f0} + r_{f1} + r_{m0}$	$2r_{f2} + r_{f1} + r_{m1}$	$2r_f + r_m$
Controls	$2s_{f0} + s_{f1} + s_{m0}$	$2s_{f2} + s_{f1} + s_{m1}$	$2s_f + s_m$
Total	$2n_{f0} + n_{f1} + n_{m0}$	$2n_{f2} + n_{f1} + n_{m1}$	$2n_f + n_m$

additive model is to weigh the (additive) scores for males and females with their inverse variances. This yields the test S_A proposed by Ziegler and König [7], which is χ^2 distributed with 1 d.f.

Simulation Study

We simulated a sample of 400 subjects. Estimates for type I error frequencies and power of the different test statistics are based on 10,000 replications for every scenario, which results in a precision of at least 0.99 at a confidence of at least 0.95 for any proportion. The alternative hypothesis of association between phenotype and genotype was specified by the genotypic relative risks. Let a and A be the risk and the other allele, respectively. Case-control genotype data for females were generated according to Wittke-Thompson et al. (their Appendix A and B) [14] by using the heterozygous relative risk

$$\gamma_1 = \frac{P(case|Aa)}{P(case|AA)}$$

and the homozygous relative risk

$$\gamma_2 = \frac{P(case|aa)}{P(case|AA)}.$$

Table 2. Mathematical details of proposed test statistics

Test statistic	Formula[a]	d.f.
Z_A^2	$\left(\dfrac{(n_m + n_f)^{\frac{1}{2}} \left[(n_m + 2n_f)(2r_{f0} + r_{f1} + r_{m0}) - (2r_f + r_m)(2n_{f0} + n_{f1} + n_{m0}) \right]}{\left[(2r_f + r_m) - (2s_f + s_m)(2n_{f0} + n_{f1} + n_{m0})(2n_{f2} + n_{f1} + n_{m1}) \right]^{\frac{1}{2}}} \right)^2$	1
Z_m^2	$\left(\dfrac{n_m^{\frac{1}{2}} (r_m s_{m0} - s_m r_{m0})}{(n_{m0} n_{m1} r_m s_m)^{\frac{1}{2}}} \right)^2$	1
Z_{fA}^2	$\left(\dfrac{n_f^{\frac{1}{2}} \left[2r_f (2s_{f0} + s_{f1}) - 2s_f (2r_{f0} + r_{f1}) \right]}{\left[2r_f s_f (2n_{f0} + n_{f1})(n_{f1} + 2n_{f2}) \right]^{\frac{1}{2}}} \right)^2$	1
Z_{fG}^2	$\left(\dfrac{n_f^{\frac{1}{2}} \left[s_f \left(\frac{1}{2} r_{f1} + r_{f2} \right) - r_f \left(\frac{1}{2} s_{f1} + s_{f2} \right) \right]}{\left[r_f s_f \left(n_f \left(\frac{1}{4} n_{f1} + n_{f2} \right) - \left(\frac{1}{2} n_{f1} + n_{f2} \right)^2 \right) \right]^{\frac{1}{2}}} \right)^2$	1
Z_C^2	$Z_{fG}^2 + Z_m^2$	2
Z_{mfA}^2	$\left(\sqrt{\dfrac{2n_f}{n_m + 2n_f}}\, Z_{fA} + \sqrt{\dfrac{n_m}{n_m + 2n_f}}\, Z_m \right)^2$	1
Z_{mfG}^2	$\left(\sqrt{\dfrac{n_f}{n_m + n_f}}\, Z_{fG} + \sqrt{\dfrac{n_m}{n_m + n_f}}\, Z_m \right)^2$	1
T_A	$\dfrac{U_A^2}{\hat{V}(U_A)}$	1
T_{AD}	$(U_A, U_D) \hat{V}^{-1} \begin{pmatrix} U_A \\ U_D \end{pmatrix}$	2
T_A^s	$\dfrac{U_{fA}^2}{\hat{V}(U_{fA})} + \dfrac{U_{mA}^2}{\hat{V}(U_{mA})}$	2
T_{AD}^s	$(U_{fA}, U_D) \hat{V}_f^{-1} \begin{pmatrix} U_{fA} \\ U_D \end{pmatrix} + \dfrac{U_{mA}^2}{\hat{V}(U_{mA})}$	3
S_A	$\dfrac{\hat{V}(U_{fA}) \cdot \hat{V}(U_{mA})}{\hat{V}(U_{fA}) + \hat{V}(U_{mA})} \left(\dfrac{U_{fA}}{\hat{V}(U_{fA})} + \dfrac{U_{mA}}{\hat{V}(U_{mA})} \right)^2$	1

d.f. = Degrees-of-freedom of the χ^2 distribution. $U_A = \Sigma_{i=1}^{n} (Y_i - \bar{Y}) A_i$ with Y_i phenotype of subject i, A_i genotype counts for additive model and $\bar{Y}$ mean of Y_i; $U_D = \Sigma_{i=1}^{n_f} (Y_i - \bar{Y}_f) D_i$ with D_i genotype counts in dominant model and $\bar{Y}_f$ mean of Y_i calculated over female sample;

$\hat{V} = \begin{pmatrix} \hat{V}(U_A) & \widehat{Cov}(U_A, U_D) \\ \widehat{Cov}(U_A, U_D) & \hat{V}(U_D) \end{pmatrix}$ with $\hat{V}_m(U_A) = 4\hat{p}(1 - \hat{p}) \Sigma_{i=1}^{n_m} (Y_i - \bar{Y})^2$, $\hat{V}_f(U_A) = \dfrac{1}{n_f - 1} \sum_{i=1}^{n_f} (A_i - \bar{A})^2 \sum_{i=1}^{n_f} (Y_i - \bar{Y})^2$ where $\bar{A}$ mean of A_i

and $\hat{p} = \dfrac{\bar{A}}{2}$, $\hat{V}(U_A) = \hat{V}_f(U_A) + \hat{V}_m(U_A)$, $\hat{V}(U_D) = \dfrac{1}{n_f - 1} \sum_{i=1}^{n_f} (D_i - \bar{D}_f)^2 \sum_{i=1}^{n_f} (Y_i - \bar{Y})^2$ and $\widehat{Cov}(U_A, U_D) = \dfrac{1}{n_f - 1} \sum_{i=1}^{n_f} (A_i - \bar{A})(D_i - \bar{D}_f) \sum_{i=1}^{n_f} (Y_i - \bar{Y})^2$;

U_{fA} (U_{mA}): U_A calculated over female (male) sample only. $\hat{V}(U_{fA})$ ($\hat{V}(U_{mA})$): $\hat{V}(U_A)$ calculated over female (male) sample only. $\hat{V}_f$: $\hat{V}$ calculated over female sample only. [a] Notations according to table 1.

For males, case-control data were generated using the relative risk

$$\gamma = \frac{P\left(case|a\right)}{P\left(case|A\right)}$$

accordingly (online suppl. A; for all online supplementary material, see www.karger.com/doi/10.1159/000323768). The different genetic models can now be modeled as shown in table 3. Each genetic model was simulated for a range of different minor allele frequencies (MAFs, between 0.05 and 0.5) and at a disease prevalence of 0.1 in the total population. We fixed the significance level α at 0.05.

Sample Designs

We considered balanced and unbalanced sample designs. For the balanced design, we generated a sample of 100 female cases, 100 female controls, 100 male cases and 100 male controls. For the unbalanced design, we first simulated a sample of 150 females and 250 males and a sample of 250 females and 150 males, keeping the distribution of females and males to cases and controls balanced. Second, we simulated samples of 200 females and 200 males, now changing the ratios of males and females in cases and controls. Specifically, we simulated one sample with 67% females in cases and 33% females in controls and a second sample with 33% females in cases and 67% females in controls.

Departure from HWE

To examine the influence of departure from HWE, we simulated a balanced sample design where males and females do not differ in any parameter. For X chromosomal SNPs, HWE can be violated in the female sample only. When HWE holds, genotype frequencies of females are given by $P(AA) = (1 - p_f)^2$, $P(Aa) = 2p_f(1 - p_f)$ and $P(aa) = p_f^2$, where p_f is the risk allele frequency for females. We now measured departure from HWE by the difference $\Delta = P(AA) - (1 - p_f)^2$. The extent of departure from HWE is here modeled by the proportion of excess or deficiency of heterozygosity, $\Delta = \varepsilon * p_f(1 - p_f)$, where a positive value of ε indicates a deficiency of heterozygotes and a negative one an excess of heterozygotes.

Sex-Specific Allele Frequencies

Allele frequency differences between males and females were considered in balanced as well as unbalanced sample designs because unbalanced sample designs can have an additional influence on power and type I error frequencies.

Simulation of Type I Error

To assess the validity of the test statistics, we investigated type I error frequencies under the null hypothesis of no association. Specifically, we simulated departures from HWE with values for ε between -0.4 and 0.4 and differences in allele frequencies between males and females of 0.02–0.1.

Simulation of Power

To determine the power of the different tests, we simulated case-control data under each genetic model listed in table 3. Relative risks were varied between 1.0 and 3.0. We simulated data in HWE and under departure from HWE in females ($\varepsilon = \pm 0.2$). Besides characteristics of the general population, we investigated a variety of sex-specific parameters. We simulated differences in

Table 3. Values of relative risks for different genetic models

	Relative risks
Null model	$\gamma_1 = \gamma_2 = \gamma = 1$
Model with X inactivation	
Recessive	$\gamma_1 = 1,\ \gamma_2 = \gamma > 1$
Additive	$\gamma_1 = \dfrac{(1 + \gamma_2)}{2},\ \gamma_2 = \gamma > 1$
Dominant	$\gamma_1 = \gamma_2 = \gamma > 1$
Model without X inactivation	
Recessive	$\gamma_1 = \gamma = 1,\ \gamma_2 > 1$
Additive	$\gamma_1 = \gamma = \dfrac{(1 + \gamma_2)}{2},\ \gamma_2 > 1$
Dominant	$\gamma_1 = \gamma_2 = \gamma > 1$

γ_1 = Heterozygous relative risk for females; γ_2 = homozygous relative risk for females; γ = relative risk for males.

effect sizes by reducing the homozygous relative risk γ_2 in females or the relative risk γ in males by 50%. To consider differences in prevalence between males and females, we simulated a prevalence of 0.05 in males and 0.15 in females and vice versa. These parameters were simulated under a balanced sample design and changed separately, keeping all other parameters identical for both sexes. Additionally, we simulated differences in allele frequencies between males and females of 0.05 and 0.1 for balanced and unbalanced sample designs.

Data – Crohn's Disease

We applied the test statistics to real data from a Crohn's disease GWA study previously described by Duerr et al. [15] and Rioux et al. [16]. The study was funded by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) IBD Genetics Consortium (IBDGC). The data set contained genetic information of a sample of 300 Crohn's disease cases with ileal involvement and 432 unrelated controls matched to cases based on sex and year of birth. Both groups had Jewish ancestry. The entire sample consists of 336 females and 396 males. Individuals with a call rate <90% or genotypes incompatible with the recorded sex were excluded from analysis. Genotyping was done using the Illumina HumanHap300 Genotyping BeadChip, which contains 8,706 X chromosomal SNPs outside of the pseudo-autosomal regions. Male heterozygous calls were SNP-wise excluded. Likewise, SNPs with ≥2% heterozygous calls were excluded from analysis. According to the Travemunde criteria [1, 2, 5], we excluded SNPs with a MAF <0.01, SNPs with a genotypic call fraction <0.98 in either cases or controls, and SNPs with p < 10^{-4} in the test for departure from HWE in the female control sample.

Table 4. Type I error frequencies for differences in allele frequencies between sexes under a balanced sample design

$q_f - q_m$	q	Z_A^2	Z_C^2	Z_{mfA}^2	Z_{mfG}^2	T_A	T_{AD}	T_{AD}^s	S_A
−0.1	0.1	0.0496	0.0497	0.0505	0.0520	0.0879	0.0805	0.0170	0.0506
	0.3	0.0482	0.0489	0.0508	0.0484	0.0525	0.0523	0.0484	0.0493
	0.5	0.0522	0.0544	0.0552	0.0497	0.0477	0.0501	0.0553	0.0520
−0.05	0.1	0.0516	0.0546	0.0513	0.0522	0.0696	0.0682	0.0279	0.0514
	0.3	0.0487	0.0515	0.0490	0.0509	0.0547	0.0506	0.0485	0.0498
	0.5	0.0516	0.0494	0.0541	0.0501	0.0485	0.0480	0.0498	0.0502
0.05	0.1	0.0457	0.0469	0.0464	0.0460	0.0295	0.0326	0.0364	0.0458
	0.3	0.0473	0.0510	0.0484	0.0484	0.0432	0.0433	0.0513	0.0486
	0.5	0.0478	0.0500	0.0500	0.0477	0.0478	0.0523	0.0528	0.0476
0.1	0.1	0.0488	0.0460	0.0503	0.0503	0.0202	0.0259	0.0383	0.0507
	0.3	0.0511	0.0515	0.0521	0.0512	0.0436	0.0446	0.0473	0.0520
	0.5	0.0521	0.0489	0.0553	0.0487	0.0465	0.0509	0.0522	0.0502

Nominal error level $\alpha = 0.05$. q_f = MAF in females; q_m = MAF in males; q = mean MAF in entire sample.

Results

Type I Error – Departure from HWE

Results of the type I error analysis under departure from HWE are shown in online supplementary table 1. We only present results for a MAF of 0.3 in the overall sample; results for other allele frequencies do not differ from these essentially, although T_{AD}^s tends to be conservative with smaller MAFs independently of departures from HWE. If there is no departure from HWE, all tests are close to the nominal type I error level of 0.05. As expected, the distribution of the allele-based tests Z_A^2 and Z_{mfA}^2 diverges from a χ^2 distribution with an increasing departure from HWE in the (female) population. They are conservative if a deficiency of heterozygotes is observed and liberal if an excess of heterozygotes occurs.

Type I Error – Sex-Specific Allele Frequencies

Under a balanced sample design, differences in allele frequencies between sexes (table 4) seem to be relevant primarily for the tests T_A and T_{AD} proposed by Clayton [11]. These show increased error frequencies when the allele frequencies of males are higher than those of females, and vice versa they show decreased error frequencies. This is especially pronounced with mean MAFs ≤ 0.3 in the overall sample, where the proportional difference relative to the mean MAF is higher.

Z_A^2 and Z_{mfA}^2 show a slight inflation of type I error for MAFs near 0.5 when numbers of males and females in the total sample differ (online suppl. table 2). Such differences show no additional effect on the error frequencies of

T_A and T_{AD} as long as the ratios of males and females in cases and controls are equal, but error frequencies increase with different ratios (online suppl. table 3: unbalanced design of 67% female cases and 33% female controls; results for 33% female cases and 67% female controls are comparable and therefore omitted). Differences in allele frequencies for this sample design inflate error frequencies of the tests Z_A^2, T_A and T_{AD}, while all other tests stay close to the nominal error level. At a mean MAF of 0.1, error frequencies are increased even for very small deviations in allele frequencies of ± 0.02. Again, it can be recognized that T_{AD}^s is conservative for MAFs <0.3.

Power – Balanced Sample Designs

Figure 1 shows the empirical power at a MAF of 0.3 of the eight tests under the different genetic models summarized in table 3. To facilitate better discrimination of test statistics in power curves, a colored line and a specific symbol is drawn for each test. Note that for a dominant model of inheritance, X chromosomal inactivation is not relevant since both inactivation and no inactivation lead to the same model. The presented results were simulated under the assumption of HWE. Under departure from HWE in females, the allele-based tests are not valid and were therefore not evaluated. Power of all other test statistics was not essentially affected by departures from HWE; therefore these results are not shown.

For a dominant model of inheritance, there are only small differences in power between the different tests. Under all but the recessive models, the 3-d.f. test T_{AD}^s has considerably less power than all other tests.
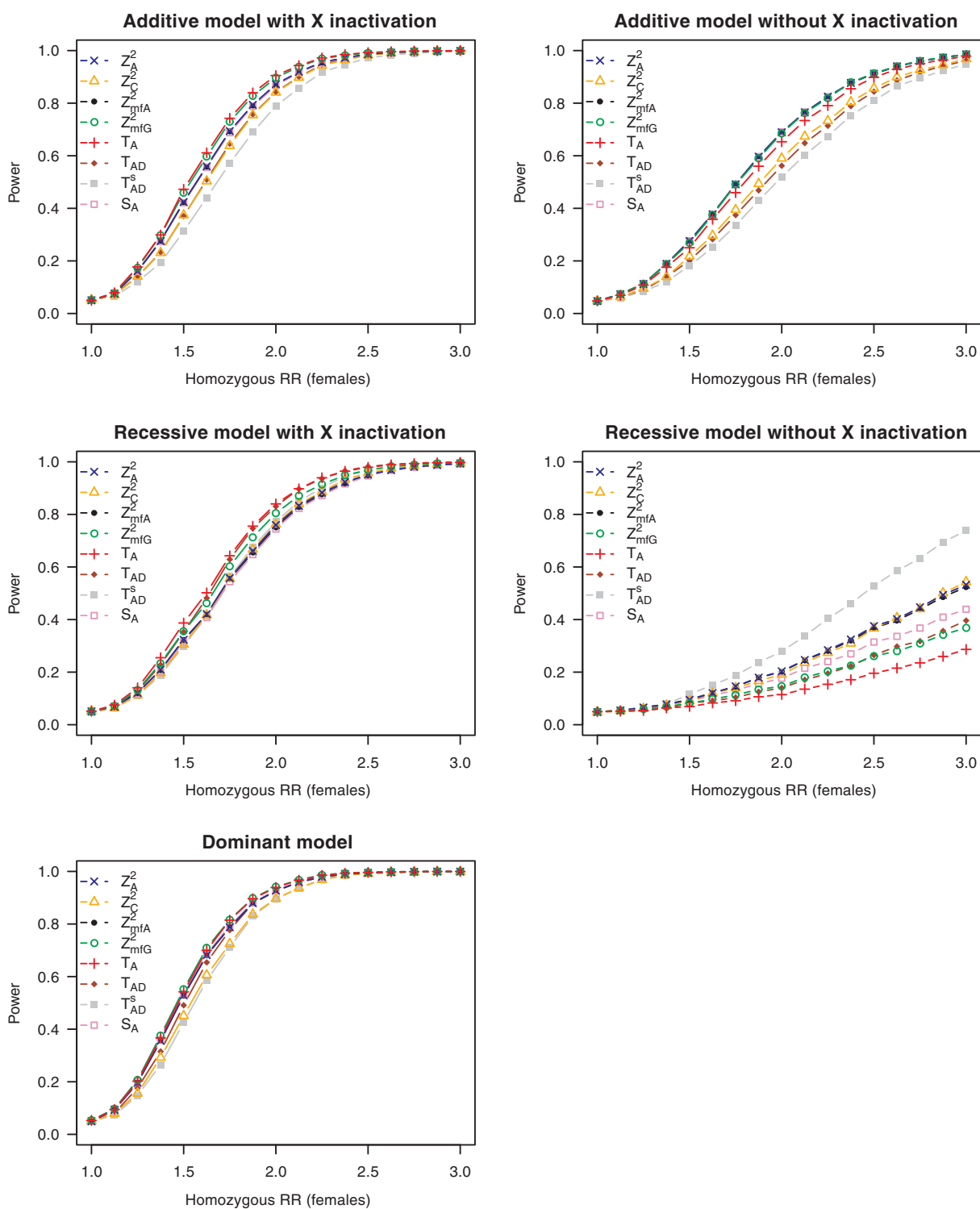
Fig. 1. Power of test statistics under different genetic models under a balanced sample design. All parameters are equal in males and females. MAF = 0.3, prevalence = 0.1. RR = Relative risk.

In situations with X inactivation, the test T_A proposed by Clayton [11] shows generally good power properties, closely followed by Z^2_{mfG} proposed by Zheng et al. [10]. Remarkably, the test T_{AD} shows no obvious advantage compared to T_A in the recessive and dominant models, although it has a special term measuring heterozygosity in females.

In models without X chromosomal inactivation power is generally lower. The reason for this is that under this assumption the effect of males is equal to the heterozygous relative risk instead of the homozygous relative risk so that, in the additive model, males show only half the effect compared to a model with inactivation, and in the recessive model, there is no effect in the male sample. Generally, the allele-based tests Z^2_A and Z^2_{mfA} and the genotype-based tests Z^2_{mfG} and S_A perform well under an additive model without X inactivation. In the recessive model, test T^s_{AD} shows considerably better power than all other tests, whereas the additive test T_A performs considerably poorer.

In general, different MAFs lead to comparable power results (additional figures are shown in online suppl. fig. 1–4). However, differences between tests become more prominent with smaller MAFs in the recessive model with X chromosomal inactivation (online suppl. fig. 1, 2). Here, tests T_A and T_{AD} show considerable superiority, and test S_A seems to have little power to detect an effect. In a recessive model without inactivation, there is almost no power to detect an effect for MAFs around 0.1 and lower (online suppl. fig. 1, 2).

Power – Unbalanced Sample Designs

Different numbers of males and females in the sample have nearly no effect on power as long as the ratio of males to females is the same for cases and controls (online suppl. fig. 5, 6). If, however, there are different ratios of males to females in cases and controls, we do observe distinct differences in power. Results for the recessive model are shown in figure 2. With 67% females in cases and 33% in controls, differences compared to the balanced design are small in the dominant model and both additive models but become more pronounced in the recessive models. In models without X inactivation and for MAFs ≤ 0.1, in the dominant model, Z^2_A now proves to be the most powerful test. For MAFs of 0.3 and especially smaller, T_A and T_{AD} lose considerable power in the recessive model with inactivation. With smaller MAFs, Z^2_C performs best and retains about the same power as it has in the balanced design (online suppl. fig. 7–14). With 33% female cases and 67% female controls, differences in power between tests

become even more pronounced. Nevertheless, the tests with highest and lowest power in most cases remain the same as in the balanced sample design. But again, test Z^2_A now shows superiority in the recessive model without inactivation and, for MAFs ≤ 0.1, in the dominant model.

Power – Sex-Specific Prevalences, Effect Sizes and Allele Frequencies

Differences in disease prevalence between sexes seem to have no influence on the power performance of the different test statistics (online suppl. fig. 15, 16). Differences in effect size between males and females, though, seem to be crucial. We did not consider the recessive model without X inactivation in these settings since there is no effect in the male sample. Results for the additive models with and without inactivation for a MAF of 0.3 are shown in figure 3.

As the effect is reduced in one of the two sexes, power is generally expected to be lower, which can clearly be observed for all genetic models (fig. 3, online suppl. fig. 17–24). When the effect in the male sample is reduced by 50%, the allele-based tests Z^2_A, Z^2_{mfA}, and S_A are most powerful for the dominant and both additive models. These differences between tests become more pronounced for smaller MAFs. Only for the recessive model with inactivation do tests T_A and T_{AD} stay most powerful. But for higher MAFs near 0.5, the allele-based test Z^2_A and Z^2_{mfA} and test S_A are superior again. When the effect is reduced in the female sample, on the other hand, T_A has the highest power over all genetic models and all MAFs. Especially for the additive model with X inactivation, this difference becomes more distinct in this situation.

Under sex-specific allele frequencies, tests T_A, T_{AD} and, for sex-specific compositions of cases and controls, Z^2_A show increased error frequencies and were therefore not evaluated. Among the remaining tests, Z^2_{mfG} performs best over almost all MAFs and all genetic models but the recessive model without X inactivation (fig. 4) where test T^s_{AD} has best power. For mean MAFs of around 0.1, tests Z^2_C and T^s_{AD} show better power in the recessive model with X inactivation for higher allele frequencies in males (online suppl. fig. 25–30). Note that unbalanced sample designs show no additional effect on the power of test statistics valid under sex-specific allele frequencies.

Analysis of Crohn's Disease Data

After quality control, 294 cases and 431 controls were left for analysis genotyped at 7,546 X chromosomal SNPs. After a conservative Bonferroni correction for multiple testing, we fixed the chromosome-wide significance
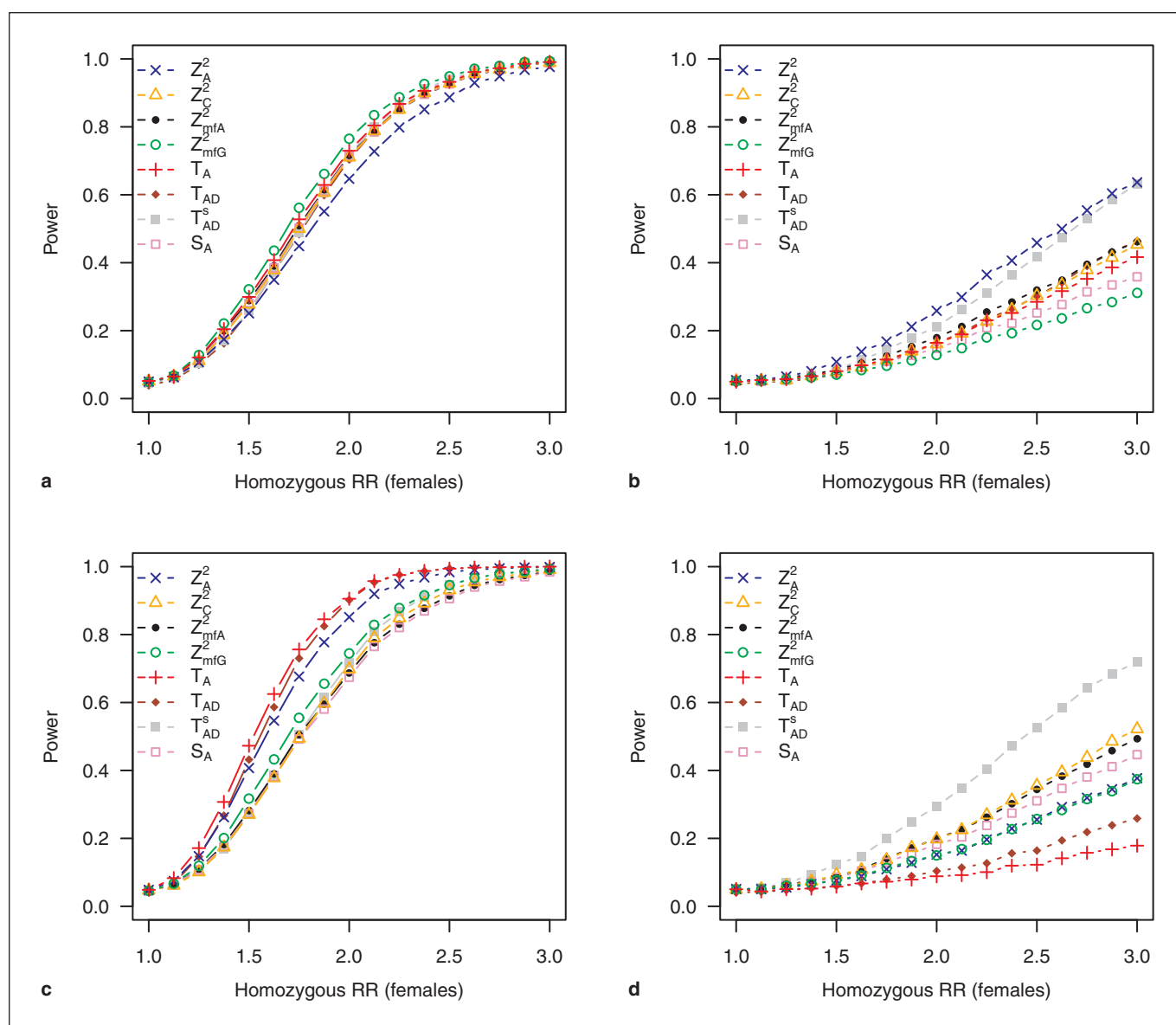
Fig. 2. Power of test statistics under an unbalanced sample design of different sex ratios in cases and controls. All parameters are equal in males and females. MAF = 0.3, prevalence = 0.1. RR = Relative risk. Results for recessive models: **a** 67% females in cases, 33% females in controls, model with X inactivation; **b** 67% females in cases, 33% females in controls, model without X inactivation; **c** 33% females in cases, 67% females in controls, model with X inactivation, and **d** 33% females in cases, 67% females in controls, model without X inactivation.

threshold at $5 \cdot 10^{-6}$. Although absolute numbers of males and females as well as absolute numbers of cases and controls differ, we have about the same proportion of females in cases (47%) and in controls (45%). For all SNPs reported below, MAFs were >0.2, and the difference in allele frequencies between males and females was <0.05. Therefore, we do not expect the tests Z_A^2, T_A, and T_{AD} to be in-

flated essentially. Since we controlled for departure from HWE in females, the distribution assumptions for the allele-based tests Z_A^2 and Z_{mfA}^2 should remain valid, too.

Online supplementary figure 31 shows the logarithmic p values of all X chromosomal SNPs for the eight test statistics. The top region can be identified as a peak near 141 mega base pairs (Mbp). Here, one SNP showed sig-
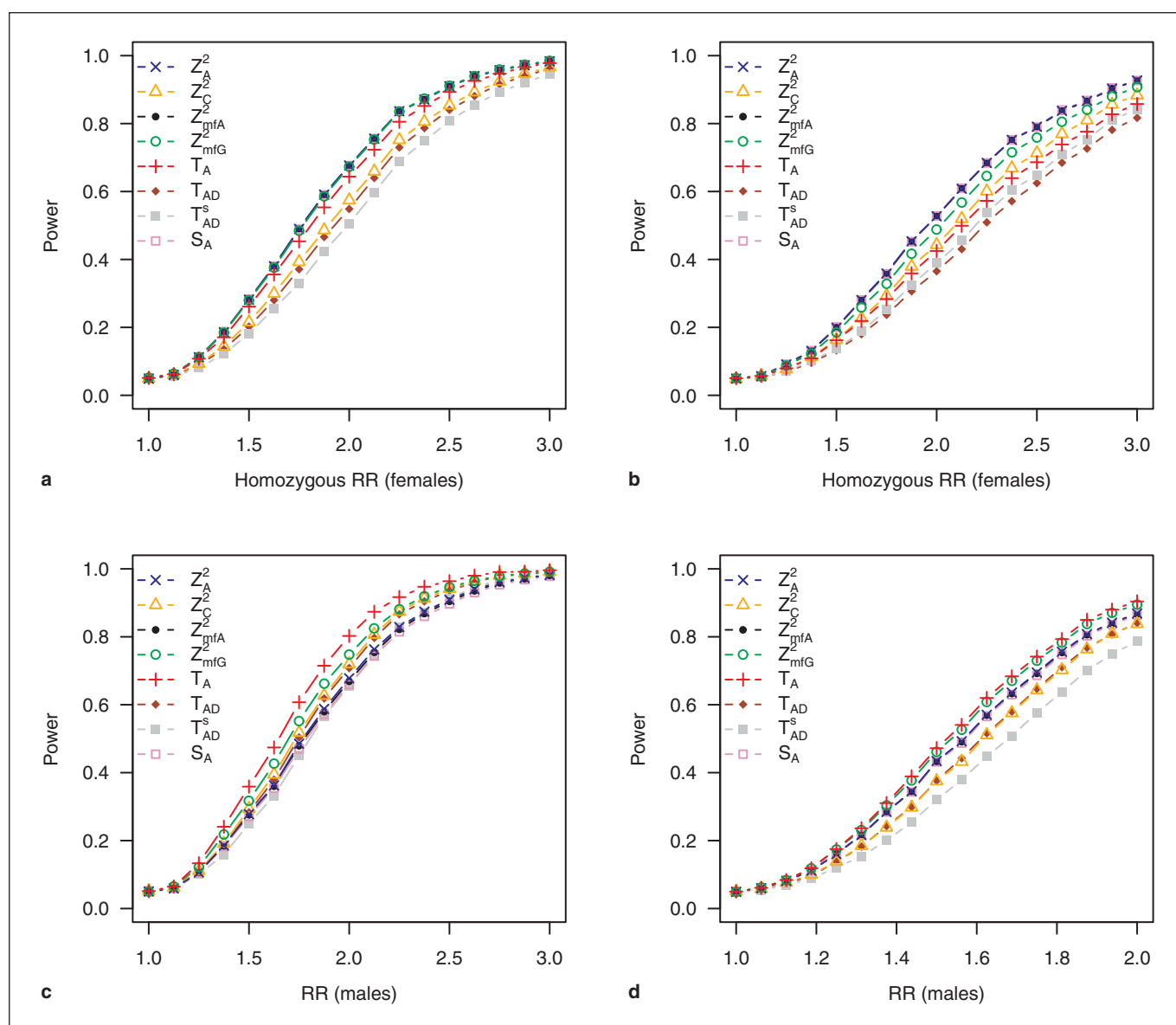
Fig. 3. Power of test statistics under sex-specific effect sizes. All other parameters are equal in males and females; balanced sample design. MAF = 0.3, prevalence = 0.1. RR = Relative risk. Results for additive models: **a** effect in males reduced by 50%, model with X inactivation; **b** effect in males reduced by 50%, model without X inactivation; **c** effect in females reduced by 50%, model with X inactivation, and **d** effect in females reduced by 50%, model without X inactivation.

nificant association to Crohn's disease in the overall sample regardless of the test (rs2038265, minimal p value $2.0 \cdot 10^{-7}$, test statistic T_{AD}; table 5, online suppl. table 5) and another one showed significant association for all but the 3-d.f. test T_{AD}^s (rs7889974, minimal p value $5.9 \cdot 10^{-7}$, test statistic Z_{mfG}^2). Both lie in a region near the genes *MAGEC1, MAGEC2* and *MAGEC3* (online suppl. fig. 32).

Looking at the genotype counts for the top SNP rs2038265, it may be suggested that the risk for Crohn's disease is increased especially for female homozygotes for the C allele and for males with a C allele, hinting at a recessive model of inheritance with X inactivation. For SNP rs7889974, the effect seems to be more pronounced in females since the p value in the male sample is considerably higher than
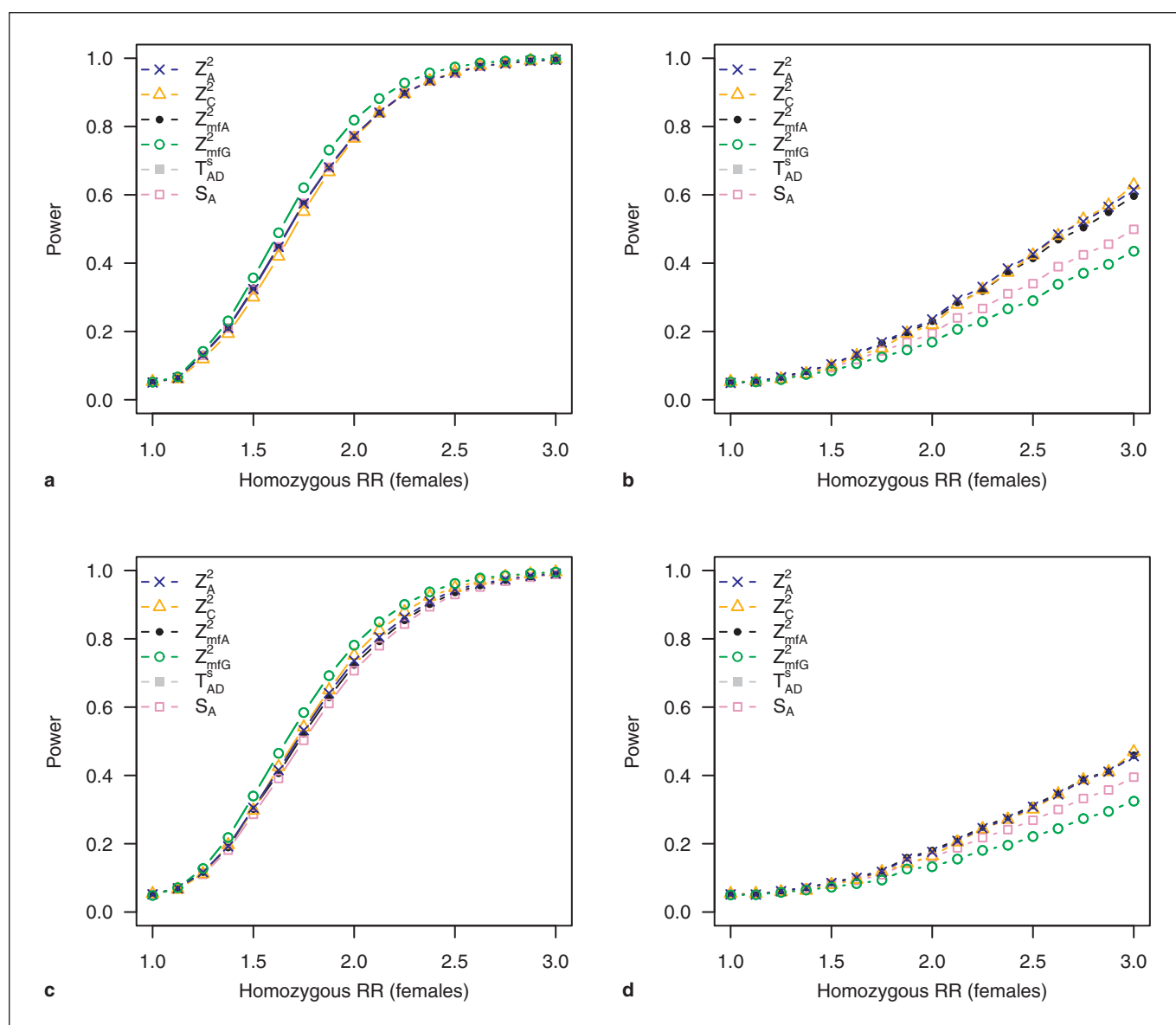
Fig. 4. Power of test statistics under sex-specific allele frequencies. All other parameters are equal in males and females; balanced sample design. Mean MAF = 0.3, prevalence = 0.1. RR = Relative risk. Results for recessive models: **a** MAF in males 0.325, MAF in females 0.275, model with X inactivation; **b** MAF in males 0.325, MAF in females 0.275, model without X inactivation; **c** MAF in males 0.275, MAF in females 0.325, model with X inactivation, and **d** MAF in males 0.275, MAF in females 0.325, model without X inactivation.

in the female sample. In both cohorts, though, the risk of Crohn's disease seems to increase with the number of G alleles. For two further SNPs in this region (rs2207272, rs2144096) T_{AD} yields the minimal p value. For SNPs rs7056485 and rs5908216, T_A yields the minimal p value; here, the effect appears to result primarily from males since p values for females are >0.05.

SNP rs4829424 near *NR0B1* and *MAGEB* (online suppl. fig. 33) displays a purely female effect on the susceptibility for Crohn's disease (Z_{fG}^2: p = 3.3 · 10^{-6}; Z_m^2: p = 0.22; table 5), which is only captured by test S_A. One further SNP in this region (rs4829169) shows a similar female-specific effect. Here, T_{AD}^s performs best.

Table 5. Association results of the Crohn's disease case-control data for the top SNPs in regions Xp21.2 and Xq27.2

SNP	Position	MAF		p values									
		f	m	Z_A^2	Z_m^2	Z_{fG}^2	Z_C^2	Z_{mfA}^2	Z_{mfG}^2	T_A	T_{AD}	T_{AD}^s	S_A
Xp21.2													
rs6526959	30132716	0.54	0.55	0.0018	0.0041	0.0841	0.0037	0.0017	0.0010	8.5E-04	1.7E-04	5.4E-04	0.0025
rs4829424	30231897	0.20	0.24	1.5E-05	0.2197	3.3E-06	9.5E-06	1.0E-05	5.0E-05	4.4E-04	5.2E-05	3.1E-05	4.2E-06
rs4829169	30238067	0.23	0.25	4.8E-05	0.1133	5.0E-05	7.6E-05	4.4E-05	9.3E-05	4.8E-04	5.2E-04	2.9E-04	1.9E-05
Xq27.2													
rs2038265	140951668	0.26	0.25	1.2E-06	6.5E-04	3.2E-04	4.7E-06	1.1E-06	7.6E-07	1.7E-06	2.0E-07	6.7E-07	1.1E-06
rs2207272	140955698	0.40	0.38	4.4E-05	0.0045	0.0022	1.6E-04	4.1E-05	3.1E-05	5.7E-05	2.0E-05	6.6E-05	3.9E-05
rs5908101	140959050	0.51	0.47	9.2E-05	0.0043	0.0078	4.9E-04	8.4E-05	9.4E-05	1.2E-04	1.7E-04	4.5E-04	1.4E-04
rs2144096	140981653	0.45	0.42	3.3E-04	0.0033	0.0247	0.0011	3.0E-04	2.3E-04	2.5E-04	6.1E-05	2.1E-04	4.6E-04
rs7889974	140992709	0.21	0.20	9.3E-07	0.0010	1.4E-04	3.3E-06	8.9E-07	5.9E-07	1.6E-06	4.1E-06	1.3E-05	6.7E-07
rs7056485	140994631	0.54	0.53	8.4E-04	0.0024	0.0672	0.0019	8.1E-04	5.1E-04	4.1E-04	0.0016	0.0048	0.0013
rs5908216	141098096	0.27	0.28	7.6E-04	0.0011	0.0767	0.0010	7.5E-04	3.2E-04	2.0E-04	0.0010	0.0031	0.0012

Position = Physical position in base pairs.

Although online supplementary figure 31 shows that no test is uniformly best over the entire X chromosome, for a variety of loci the tests T_{AD}, T_{AD}^s, and Z_C^2 show substantially lower p values than all other tests.

At about 84 Mbp, there are seven SNPs indicating only effects for Z_C^2, T_{AD}, or T_{AD}^s (online suppl. table 4). Four of these SNPs show gender-specific effects, two for males and two for females. For the SNPs indicating an effect for males only, Z_C^2 produces the smallest p values, while for the SNPs with effect for females only, T_{AD}^s and T_{AD} have the smallest p values. Note, however, that for the last SNP the MAF is at 0.07 for females and 0.10 for males, so that test T_{AD} may be inflated in this case. Z_C^2 also detects an effect for one SNP which has only slight effects for males and females (Z_m^2: p = 0.0777; Z_{fG}^2: p = 0.0526) which no other test is able to detect. The remaining two SNPs are only detected by T_{AD} and T_{AD}^2. They show no sex-specific effects either; specifically Z_{fG}^2 results in p values >0.7. For these SNPs, the effect seems to arise from a heterozygosity deficit in female cases (online suppl. table 5) which all other tests have less power to measure.

A similar situation can be observed for a region at about 146 Mbp (online suppl. table 4). Here, we observe a peak solely for T_{AD} and T_{AD}^s. Three SNPs in this region have p values <0.01 only for T_{AD} and T_{AD}^s and show no sex-specific effects (p values for both Z_m^2 and Z_{fG}^2 are >0.05). For these SNPs, the effect seems to arise from a heterozygosity excess in female cases (online suppl. table 5).

Discussion

Different tests for association on the X chromosome have been proposed in the literature [10, 11], but there has so far been no systematic comparison between them. To close this gap, we conducted a broadly conceived simulation study comparing tests in a wide range of settings. Additionally, we applied the tests to the analysis of a real data set. The aim was to facilitate researchers' choice of an appropriate test for detecting associations for X chromosomal SNPs.

Results from simulations under the null hypothesis of no association reveal that the tests Z_A^2, T_A, and T_{AD} are not valid under sex-specific allele frequencies, especially when ratios of males to females differ between cases and controls. For these unbalanced sample designs, even minor differences in allele frequencies between sexes can lead to a crucial inflation of type I error frequencies; here, tests Z_A^2, T_A, and T_{AD} should not be used. It is therefore recommended to check for differences in allele frequencies between males and females before applying tests. As expected, departure from HWE is problematic for the allele-based tests Z_A^2 and Z_{mfA}^2 which should only be used after compatibility with HWE is established.

Results from power simulations as well as from the real data analysis demonstrate that no test is uniformly best over all genetic models. As could be expected, Clayton's [11] additive test T_A is most powerful for most models assuming X inactivation. If the effect size in females is

reduced, test T_A performs best even in models without inactivation. In most models without X inactivation, T_A is less powerful. Here, test Z_A^2 shows good power under a variety of models. For balanced sample designs, tests S_A, Z_{mfG}^2, and Z_{mfA}^2 show comparable power results. If the effect size is reduced in males, these tests perform better than T_A under all genetic models apart from the recessive model assuming X inactivation.

Remarkably, simulations show no obvious advantage of T_{AD} compared to T_A in the recessive and dominant models, although it has a special term measuring heterozygosity in females. Apparently, the gain by the heterozygosity indicator does not compensate for the additional d.f. In contrast, in the recessive model assuming no X inactivation, test T_{AD}^s is most powerful, although it has 3 d.f. However, this model seems to be biologically speculative. Results from the Crohn's disease data offer some evidence for an advantage of T_{AD} and T_{AD}^s in a few specific situations. As described above, these two tests seem to perform best to detect an effect when this is primarily caused by a heterozygote excess or deficit in females which does not necessarily indicate a dominant or recessive model.

The Crohn's disease data also indicate some advantage of Z_C^2 and T_{AD}^s when effect sizes are sex specific which was not seen in the simulations. Z_C^2 and T_{AD}^s measure effects for males and females separately and should therefore have highest power to detect effects merely arising for one sex. But since Z_C^2 has 2 and T_{AD}^s 3 d.f., presumably more distinct differences in effects between males and females need to be simulated before this effect can be observed.

The analysis of the Crohn's disease data uncovered a SNP in region Xq27.2 showing chromosome-wide significance regardless of the test statistic used. This SNP lies in the vicinity of three genes of the melanoma antigen family *(MAGEC1*, *MAGEC2* and *MAGEC3)* which are known to be expressed specifically in tumors of various histological types. One additional SNP in region Xp21.2 showed a chromosome-wide significant p value only in females. This SNP lies near a nuclear receptor subfamily *(NR0B1)* which encodes a protein containing a DNA-binding domain. Although these results are based on only 294 cases and 431 controls of a very specific population of Jewish ancestry and need to be followed up in further studies, they might indicate novel regions, since they were not described in the literature on Crohn's disease so far [17].

To conclude, we recommend that for a specific study, the validity of all available tests firstly needs to be checked, i.e., whether they have the correct significance level. As shown, this depends mostly on whether genotype frequencies in females are in HWE, on allele frequency differences between males and females, and on the sex ratio in cases and controls. Since these factors are easily checked before analysis, they can be used to guide further analysis.

Secondly, there might be genetic models that are more plausible than others in a given situation, so that a test statistic can be chosen that is more powerful. As described above, T_A showed good power under models assuming X inactivation, while Z_A proved good power under most models assuming no X inactivation. However, in a typical GWA setting (including X chromosomal data), the model is usually unknown to the investigator before completing the analysis and it remains challenging to decide which test to use. Calculating all tests and then using the test with the smallest p value would require adjustment of the p value for the number of tests calculated, which would in turn reduce the power. A plausible compromise to calculating all six test statistics could be to calculate T_A and Z_A^2 only, if they are valid in a given situation. This reduces the number of tests to be corrected for to a minimum and leaves good power to detect effects under models with and without X inactivation.

When departure from HWE cannot be excluded but allele frequencies are equal in males and females, test S_A instead of Z_A^2 can be calculated which shows good power in most situations where T_A has little power, specifically under most models not assuming X inactivation. Under sex-specific allele frequencies, Z_{mfG}^2 is a robust alternative, being the most powerful test under almost all models. However, further research is still necessary to reveal whether a uniformly most powerful test across all genetic models is possible. Moreover, to be applicable under a wide range of different settings, this test would need to be robust to differences in allele frequencies between the sexes. However, we showed that even the available test statistics are able to detect previously unknown genetic regions in the case of Crohn's disease. Especially when hypotheses on the underlying genetic model exist, this contribution should offer good assistance for the analysis of X chromosomal data.

References

1 Erdmann J, Grosshennig A, Braund PS, König IR, Hengstenberg C, Hall AS, Linsel-Nitschke P, Kathiresan S, Wright B, Tregouet DA, Cambien F, Bruse P, Aherrahrou Z, Wagner AK, Stark K, Schwartz SM, Salomaa V, Elosua R, Melander O, Voight BF, O'Donnell CJ, Peltonen L, Siscovick DS, Altshuler D, Merlini PA, Peyvandi F, Bernardinelli L, Ardissino D, Schillert A, Blankenberg S, Zeller T, Wild P, Schwarz DF, Tiret L, Perret C, Schreiber S, El Mokhtari NE, Schafer A, Marz W, Renner W, Bugert P, Kluter H, Schrezenmeir J, Rubin D, Ball SG, Balmforth AJ, Wichmann HE, Meitinger T, Fischer M, Meisinger C, Baumert J, Peters A, Ouwehand WH, Deloukas P, Thompson JR, Ziegler A, Samani NJ, Schunkert H: New susceptibility locus for coronary artery disease on chromosome 3q22.3. Nat Genet 2009;41:280–282.

2 Samani NJ, Erdmann J, Hall AS, Hengstenberg C, Mangino M, Mayer B, Dixon RJ, Meitinger T, Braund P, Wichmann HE, Barrett JH, König IR, Stevens SE, Szymczak S, Tregouet DA, Iles MM, Pahlke F, Pollard H, Lieb W, Cambien F, Fischer M, Ouwehand W, Blankenberg S, Balmforth AJ, Baessler A, Ball SG, Strom TM, Braenne I, Gieger C, Deloukas P, Tobin MD, Ziegler A, Thompson JR, Schunkert H: Genomewide association analysis of coronary artery disease. N Engl J Med 2007;357:443–453.

3 Scherag A, Dina C, Hinney A, Vatin V, Scherag S, Vogel CI, Muller TD, Grallert H, Wichmann HE, Balkau B, Heude B, Jarvelin MR, Hartikainen AL, Levy-Marchal C, Weill J, Delplanque J, Korner A, Kiess W, Kovacs P, Rayner NW, Prokopenko I, McCarthy MI, Schafer H, Jarick I, Boeing H, Fisher E, Reinehr T, Heinrich J, Rzehak P, Berdel D, Borte M, Biebermann H, Krude H, Rosskopf D, Rimmbach C, Rief W, Fromme T, Klingenspor M, Schurmann A, Schulz N, Nothen

MM, Muhleisen TW, Erbel R, Jockel KH, Moebus S, Boes T, Illig T, Froguel P, Hebebrand J, Meyre D: Two new loci for bodyweight regulation identified in a joint analysis of genome-wide association studies for early-onset extreme obesity in French and German study groups. PLoS Genet 2010; 6:e1000916.

4 Kathiresan S, Willer CJ, Peloso GM, Demissie S, Musunuru K, Schadt EE, Kaplan L, Bennett D, Li Y, Tanaka T, Voight BF, Bonnycastle LL, Jackson AU, Crawford G, Surti A, Guiducci C, Burtt NP, Parish S, Clarke R, Zelenika D, Kubalanza KA, Morken MA, Scott LJ, Stringham HM, Galan P, Swift AJ, Kuusisto J, Bergman RN, Sundvall J, Laakso M, Ferrucci L, Scheet P, Sanna S, Uda M, Yang Q, Lunetta KL, Dupuis J, de Bakker PI, O'Donnell CJ, Chambers JC, Kooner JS, Hercberg S, Meneton P, Lakatta EG, Scuteri A, Schlessinger D, Tuomilehto J, Collins FS, Groop L, Altshuler D, Collins R, Lathrop GM, Melander O, Salomaa V, Peltonen L, Orho-Melander M, Ordovas JM, Boehnke M, Abecasis GR, Mohlke KL, Cupples LA: Common variants at 30 loci contribute to polygenic dyslipidemia. Nat Genet 2009;41: 56–65.

5 Ziegler A: Genome-wide association studies: quality control and population-based measures. Genet Epidemiol 2009; 33(suppl 1):S45–S50.

6 Manolio TA: Genomewide association studies and assessment of the risk of disease. N Engl J Med 2010;363:166–176.

7 Ziegler A, König IR: A statistical approach to genetic epidemiology, concepts and applications, ed. 2. Weinheim, Wiley-VCH, 2010.

8 Chow JC, Yen Z, Ziesche SM, Brown CJ: Silencing of the mammalian X chromosome. Annu Rev Genomics Hum Genet 2005;6:69–92.

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

10 Zheng G, Joo J, Zhang C, Geller NL: Testing association for markers on the X chromosome. Genet Epidemiol 2007;31:834–843.

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

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

13 Sasieni PD: From genotypes to genes: doubling the sample size. Biometrics 1997;53: 1253–1261.

14 Wittke-Thompson JK, Pluzhnikov A, Cox NJ: Rational inferences about departures from Hardy-Weinberg equilibrium. Am J Hum Genet 2005;76:967–986.

15 Duerr RH, Taylor KD, Brant SR, Rioux JD, Silverberg MS, Daly MJ, Steinhart AH, Abraham C, Regueiro M, Griffiths A, Dassopoulos T, Bitton A, Yang H, Targan S, Datta LW, Kistner EO, Schumm LP, Lee AT, Gregersen PK, Barmada MM, Rotter JI, Nicolae DL, Cho JH: A genome-wide association study identifies il23r as an inflammatory bowel disease gene. Science 2006;314:1461–1463.

16 Rioux JD, Xavier RJ, Taylor KD, Silverberg MS, Goyette P, Huett A, Green T, Kuballa P, Barmada MM, Datta LW, Shugart YY, Griffiths AM, Targan SR, Ippoliti AF, Bernard EJ, Mei L, Nicolae DL, Regueiro M, Schumm LP, Steinhart AH, Rotter JI, Duerr RH, Cho JH, Daly MJ, Brant SR: Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. Nat Genet 2007;39:596–604.

17 Invernizzi P, Gershwin ME: The genetics of human autoimmune disease. J Autoimmun 2009;33:290–299.