A Cost-Effective Statistical Method to Correct for Differential Genotype Misclassification When Performing Case-Control Genetic Association

Douglas Londono^a Chad Haynes^b Francisco M. De La Vega^c Stephen J. Finch^d
Derek Gordon^a

^aDepartment of Genetics, Rutgers University, Piscataway, N.J., ^bLaboratory of Statistical Genetics, Rockefeller University, New York, N.Y., ^cApplied Biosystems, Foster City, Calif., and ^dDepartment of Applied Math and Statistics, Stony Brook University, Stony Brook, N.Y., USA

Key Words

Differential genotype misclassification • Genotyping errors • Method • Simulation • Power

Abstract

Background/Aims: There is a growing interest regarding the effect of differential misclassification on power and type I error rate in genome-wide association studies. We present an extension of a previously published test statistic: the likelihood ratio test allowing for errors (LRT_{AE}). This test uses double-sample information on a subset of individuals to increase power for genetic association in the presence of nondifferential misclassification. **Methods:** We extend the original LRT_{AE} by allowing for differential genotype misclassification between case and control populations. We label this new statistic as $LRT_{AE}^{D,M}$. We test the performance of this statistic with data simulated under differential misclassification specifications and two different types of genetic models: null and power. For simulations using the null model, we specify that there is no difference between case and control genotype frequencies before the introduction of errors. For simulations under power, we consider three modes of inheritance: dominant, multiplicative, and recessive. **Results:** We show that the $LRT_{AE}^{D,M}$, with p values computed using permutation,

maintains a correct type I error rate under the null model after the introduction of differential genotyping errors. Also, we find that as little as 10 to 15% of double-sampled genotype data is needed to achieve this effect. Aside from a few situations (particularly recessive mode of inheritance simulations) the $LRT_{AE}^{D,M}$ version that calculates p values through permutation requires 15 to 20% double sampling to maintain an 80% power for a 0.05 significance level and approximately 20% double sampling for a 0.01 significance level.

Copyright © 2010 S. Karger AG, Basel

Introduction

In whole genome-wide association studies, where hundreds of thousands to millions single nucleotide polymorphisms (SNPs) are used, differential and nondifferential genotype error rates among cases and controls is a major issue. It has been documented that although nondifferential error rates cause no increase in type I error rate they cause bias in the estimation of parameters and reduce the power to detect association [1–11]. In the context of differential genotype misclassification there have been a number of recent publications discussing its effect in association studies [12–18]. Clayton et al. [13] re-

ported inflation of the false-positive rate for a case-control study of type I diabetes in the United Kingdom. The authors proposed a weighting scheme for the test statistics to treat this inflation. As noted by these authors, however, the weighting can cause a decrease in power to detect association. Pearce et al. [15] showed that in a previously published association study of prostate cancer [19], the reported finding was a false positive. In the initial study cases were called much earlier than controls causing different group error rates. Moskvina et al. [14] performed a simulation study to document the effect of differential misclassification on the probability of false-positive association showing that unequal case-control error probabilities can cause substantial inflation in the type I error rate, as was indeed observed by Clayton et al. [13] and Pearce et al. [15]. Plagnol et al. [12] proposed a modified version of the SNP clustering algorithm of Moorhead et al. [20] in order to reduce the differential bias in genotype scoring between cases and controls. Since differential misclassification also affects genotype missing calls, the authors adapted association tests to deal with uncertain SNP calls (normally labeled as missing) in order to avoid the extra bias introduced by unequal missing rates. Marquard et al. [16] examined the effect of differential genotype misclassification in the type I error rate of three haplotype-based association methods and Cheng and Lin [17] evaluated its effect on studies of gene-environment interaction. Finally, Ahn et al. [18] showed that, for fixed differential error frequencies, when the allele frequencies are equal between cases and controls, under Hardy-Weinberg equilibrium the rejection rates of the χ^2 and linear trend tests increase as the minor allele frequency decreases and the sample size increases.

Differential error rates can occur due to differences in DNA quality or extraction protocols between cases and controls [13] or through the use of public controls. Different rates may also occur when cases and controls are genotyped at different times or when researches employ different genotype calling technologies. Furthermore, as the use of copy number polymorphisms in association studies increases, so will the impact of differential genotype misclassification [21].

In this article we propose an extension of a previously published test statistic: the likelihood ratio test allowing for errors (LRT_{AE}) [22]. This test uses double-sample information [23, 24] on a subset of individuals to increase power for genetic association in the presence of nondifferential misclassification. By double-sample, we mean that a subset of individuals are genotyped by both the standard genotyping mechanism, which is subject to mis-

classification, and a gold-standard genotyping mechanism that has much lower misclassification rates. We extend the original LRT_{AE} by allowing for differential genotype misclassification between case and control populations. This extension is accomplished by assuming that an individual's observed genotype is conditionally dependent upon both the individual's true genotype and true phenotype. In this work we do not consider phenotype misclassification. We perform simulations to evaluate the type I error rate and power of our method in the presence of differential genotype misclassification.

We use the notation LRT_{AE}^{DM} to denote the statistic that allows for differential misclassification.

Methods

Test Statistic

We derive our calculations under the specification that there is no phenotype misclassification, that is, the phenotype is known with certainty.

Under the alternative hypothesis, the log-likelihood of the double-sample data may be written as:

$$\ln(L_{AE,1}^{DM}) = \sum_{i'} \sum_{j'} \sum_j n_{i'j'} \ln[\Pr(Y_{i'}^t, X_{j'}^t, X_j)] + \sum_{i'} \sum_j n_{i'j} \ln[\Pr(Y_{i'}^t, X_j)], \quad (1)$$

where $n_{i'j}$ is the number of individuals who have phenotype classification i' ($i' = 0$ for cases; $i' = 1$ for controls), observed genotype classification j ($j = 0, 1$, or 2) and true (double-sampled) genotype classification j' ($j' = 0, 1$, or 2). These individuals have been double-sampled.

Analogously, n_{ij} is the number of individuals with only observed data who have phenotype classification i' ($i' = 0$ for cases; $i' = 1$ for controls), and observed genotype classification j ($j = 0, 1$, or 2).

Also, $Y_{i'}^t$ is the event that an individual has true phenotype classification i' ; $X_{j'}^t$, the event that the same individual has true genotype classification j' , and X_j the event that the same individual has observed genotype classification j .

Given that $\Pr(Y_{i'}^t, X_{j'}^t, X_j) = \Pr(X_j | X_{j'}^t, Y_{i'}^t) \Pr(X_{j'}^t | Y_{i'}^t) \Pr(Y_{i'}^t)$, to allow for differential genotype misclassification, we define $\theta_{i'jj} = \Pr(X_j | X_{j'}^t, Y_{i'}^t)$. Furthermore, if we define $p_{i'j'}^t = \Pr(X_{j'}^t | Y_{i'}^t)$ and $q_{i'}^t = \Pr(Y_{i'}^t)$, then we may rewrite the log-likelihood equation (1) above as:

$$\ln(L_{AE,1}^{DM}) = \sum_{i'} \sum_{j'} \sum_j n_{i'j'} \ln[\theta_{i'jj} p_{i'j'}^t q_{i'}^t] + \sum_{i'} \sum_j n_{i'j} \ln\left[\sum_{v'} \theta_{i'vj} p_{i'v'}^t q_{i'}^t\right]. \quad (2)$$

Under the null hypothesis that genotype frequencies do not differ between cases and controls, we may write the log-likelihood equation (2) as:

$$\ln(L_{AE,0}^{DM}) = \sum_{i'} \sum_{j'} \sum_j n_{i'j'} \ln[\theta_{i'jj} p_{i'j'}^t q_{i'}^t] + \sum_{i'} \sum_j n_{i'j} \ln\left[\sum_{v'} \theta_{i'vj} p_{i'v'}^t q_{i'}^t\right], \quad (3)$$

where $p_{i'j'}^t = \Pr(X_{j'}^t) = \Pr(X_{j'}^t | Y_0^t) = \Pr(X_{j'}^t | Y_1^t)$.

Because of misclassification, the true genotype frequencies $p_{i'j'}^t$ are not known. However, we showed previously [22] that they may be estimated using the expectation-maximization method [25].

The formulas for the $r + 1$ st step EM-algorithm estimates of the parameters $p_{ij}^{t(r+1)}$, $p_{*j}^{t(r+1)}$, $q_i^{t(r+1)}$ are given by:

$$\begin{aligned} p_{ij}^{t(r+1)} &= n_{ij}^{(r)} / n_{i+}, \\ p_{*j}^{t(r+1)} &= n_{*j}^{(r)} / n, \\ q_i^{t(r+1)} &= n_{i+} / n, \end{aligned} \quad (4)$$

where

$$n_{i+}^{(r)} = \left(\sum_j n_{ij}^{(r)} \right) + n_{2i+}^{(r)};$$

n_{i+} is the number of individuals with phenotype classification i ;

$$n_{+j}^{(r)} = \sum_i n_{ij}^{(r)}; n_{2i+}^{(r)} = \sum_j n_{ij}^{(r)} \left(\frac{\theta_{i+j} p_{ij}^{t(r)}}{\sum_{v'} \theta_{i+v'} p_{iv'}^{t(r)}} \right),$$

and n is the total number of individuals.

Maximum likelihood estimates of the misclassification parameters θ_{ij} are given by:

$$\theta_{ij} = n_{ij} / \sum_j n_{ij}.$$

The test statistic is $LRT_{AE}^{DM} = 2[\ln(L_{AE,1}^{DM}) - \ln(L_{AE,0}^{DM})]$, asymptotically distributed as χ^2 with 2 degrees of freedom (d.f.).

Simulations

Power and type I error rates are estimated for different parameter values and inheritance models as presented in table 1. From this point forward we use the following notation given in table 1: Na = number of cases, Nu = number of controls, err_{CA} = genotype error parameter for cases, err_{CO} = genotype error parameter for controls.

Without loss of generality, we consider a single susceptibility locus with only two alleles whose frequencies are identical to those of a diallelic marker locus. We also specify that the marker and disease susceptibility alleles are in complete linkage disequilibrium (LD).

Given that the wild-type and disease susceptibility (d) alleles are in Hardy-Weinberg proportions, the prevalence of the disease (K) in the general population can be expressed in terms of the genotype penetrances f_i and the disease susceptibility allele frequency P_d as:

$$K = (1 - P_d)^2 f_0 + 2P_d(1 - P_d)f_1 + P_d^2 f_2, \quad (5)$$

where $f_i = \Pr(\text{aff} | i \text{ copies of } d)$ for $i = 0, 1, 2$.

K may also be written in terms of the genotype relative risks $R_i (i = 1, 2)$ as

$$K = (1 - P_d)^2 f_0 + 2P_d(1 - P_d)R_1 f_0 + P_d^2 R_2 f_0, \quad (6)$$

where $R_i = f_i/f_0$ is obtained by dividing the genotype penetrances by the penetrance value in homozygotes for the wild type allele (the baseline disease penetrance).

Under the null hypothesis, both the heterozygote relative risk R_1 and the disease allele homozygote relative risk R_2 are equal to 1 (i.e. all penetrances are equal). Under the alternative hypothesis (with the exception of the recessive scenario) R_2 can be determined as a function of R_1 for each mode of inheritance:

Table 1. Simulation parameters

Parameter	Value
R_1 – under the null hypothesis	1
R_1 – under the alternative hypothesis*	1.5, 2.5
err_{CA}	0.1, 0.02
err_{CO}	1/4 err_{CA} , 1/10 err_{CA}
Na (Nu)	1,000, 2,000
P_d (P_m)	0.1, 0.25
K	0.1

Values of the different parameters used in the creation of the simulated data. A total of 16 simulations result in the combination of these values under the null and 32 under the alternative. In each simulation, 0, 5, 10, 15, 20, 50 and 100% double samples are taken.

R_1 = Heterozygote relative risk; err_{CA} = error rate in cases; err_{CO} = error rate in controls; Na = Nu = number of affected and unaffected; $P_m = P_d$ = marker (disease) allele frequency; K = population prevalence.

* When the mode of inheritance is recessive, $R_1 = 1$ and we let the disease allele homozygote relative risk (R_2) = 1.5 or 2.5.

- (1) for a dominant model, $R_2 = R_1$ and $R_1 > 1$;
- (2) for a recessive model, $R_2 > 1$ and $R_1 = 1$;
- (3) for a multiplicative model, $R_2 = R_1^2$ and $R_1 > 1$.

For a description of the steps for the simulated data generation and for the introduction of errors based on the Douglas Skol Boehnke (DSB) model [26] see our online supplementary material (www.karger.com/doi/10.1159/000314470).

Application of the LRT_{AE}^{DM} Statistic

Both asymptotic and permutation p values (for 1,000 permutations) are computed for 1,000 replicates in each of the simulations. These analyses provide a total of 2 p values per replicate. Confidence intervals for the permuted p values are created by using the LINKAGE utility program BINOM [27]. Power and type I error rates are calculated as the proportions of replicates out of 1,000 whose p values are less than a significance level of 0.05 or 0.01.

We note that the asymptotic p values for the LRT_{AE}^{DM} are calculated considering a central χ^2 distribution with 2 d.f. as the null distribution.

Results

Under the Null Hypothesis

The full set of results for all simulations under the null may be found in online supplementary tables 2a and b.

In results not shown, there is gross inflation of the type I error rate for the LRT and LRT_{AE} statistics with p values estimated both asymptotically and through permutation. This inflation in type I error rate increases as a function of increasing sample size and increasing genotype error

Table 2. Type I error rate of the LRT_{AE}^{DM} at the 0.05 level when errors are introduced according to the DSB model

DS%	LRT_{AE}^{DM}	LCI	UCI	LRT_{AE}^{DM} Perm	LCI	UCI
0	0.761	0.7333	0.7871	0.757	0.7292	0.7833
5	0.588	0.5568	0.6187	0.057	0.0435	0.0732
10	0.402	0.3714	0.4331	0.069	0.0541	0.0865
15	0.313	0.2843	0.3428	0.058	0.0443	0.0743
20	0.191	0.1671	0.2168	0.04	0.0287	0.0541
50	0.075	0.0594	0.0931	0.051	0.0382	0.0665
100	0.051	0.0382	0.0665	0.051	0.0382	0.0665

Heterozygote relative risk = 1; number of affected and unaffected = 1,000; marker (disease) allele frequency = 0.1; $err_{CA} = 0.1$; $err_{CO} = 0.025$. Perm = number of permutations performed = 1,000; DS% = double-sampling percentages. LRT_{AE}^{DM} = asymptotic p values computed assuming a χ^2 distribution with 2 d.f.; LRT_{AE}^{DM} Perm = p values computed via permutation; LCI (UCI) = 95% lower (upper) confidence interval value.

rates [14, 18]. The LRT does not allow for differential misclassification and the LRT_{AE} only allows for nondifferential misclassification. These results highlight the impact differential misclassification rates have on the performance of these tests, for example, in some instances the true null hypothesis was rejected 76% of the time.

On the other hand, the LRT_{AE}^{DM} (which allows for differential misclassification) with p values estimated using a χ^2 distribution with 2 d.f. presented a lesser extent of inflation. Nevertheless, for most simulations at least 50% double sampling is required to maintain correct type I error rate. Interestingly, when p values are computed using permutation this test requires 10% double sampling in almost all situations (in a few situations it requires 15% double sampling) to maintain the correct type I error rate. We provide an example of type I error rate results at the 0.05 level in table 2 (taken from sim02 in online suppl. table 2a).

In results not shown, we check the performance of the LRT_{AE}^{DM} method for null situations under a more general model than the DSB. The differential errors that most inflate the type I error rate are misclassifying the most common homozygote as the other homozygote or as the heterozygote [18]. We use the Mote and Anderson error model [3] which allows for both the miscoding of one homozygous genotype as the other and different probabilities of incorrectly coding the heterozygous genotype as either one of the homozygous genotypes (and vice versa). As expected, under this model, the LRT_{AE}^{DM} with p values computed asymptotically requires now more than 50% double sampling to maintain the correct 0.05 or 0.01 levels. However, when p values are calculated through per-

mutation, the LRT_{AE}^{DM} statistic still requires as low as 10% double sampling to maintain the correct type I error rates.

As indicated by the observed permutation results, even though the LRT_{AE}^{DM} model incorporates differential error rates, it does not follow a χ^2 distribution with 2 d.f. unless at least 50% double sampling is used to compute the statistic. The more double sampling data are used, the more data with no errors are used, and therefore the more correct information there is to estimate the statistic and to approximate the underlying null distribution. At 100% double sampling, only data with no errors are used. At this point, the LRT_{AE}^{DM} and LRT_{AE} statistics are equivalent (at 0% double sampling the LRT_{AE}^{DM} and LRT_{AE} statistics are equivalent to the standard LRT).

Under the Alternative Hypothesis (Power Simulations)

The full set of results for all simulations and all inheritance models are presented in online supplementary tables 3 through 5.

Dominant Model

The LRT_{AE}^{DM} statistic with p values computed asymptotically always rejects the null more often than with p values computed using permutation. To obtain a power of at least 80% using permutation, a 10 to 15% double sampling is almost always required at the 0.05 level (e.g. sim01, sim03, sim08 in online suppl. table 3a). In a few instances more than 20% double sampling is required. For example, with $R_1 = 1.5$, $N_a = N_u = 1,000$; $P_d = 0.1$, $err_{CA} = 0.1$ and $err_{CO} = 0.025$, the power is 66% for a 20% double sampling (table 3, taken from sim02 in online sup-

Table 3. Power of the LRT_{AE}^{DM} at the 0.05 level under a dominant model when errors are introduced according to the DBS model

DS%	LRT_{AE}^{DM}	LCI	UCI	$LRT_{AE}^{DM} Perm$	LCI	UCI
0	1	0.9963	1	1	0.9963	1
5	0.858	0.8348	0.8791	0.338	0.3087	0.3683
10	0.849	0.8253	0.8706	0.475	0.4437	0.5065
15	0.866	0.8433	0.8865	0.586	0.5548	0.6167
20	0.88	0.8582	0.8995	0.655	0.6246	0.6845
50	0.952	0.9369	0.9644	0.922	0.9036	0.9379
100	0.973	0.961	0.9821	0.974	0.9621	0.9829

Heterozygote relative risk = 1.5; number of affected and unaffected = 1,000; marker (disease) allele frequency = 0.1; $err_{CA} = 0.1$; $err_{CO} = 0.025$; Perm = number of permutations performed = 1,000; DS% = double-sampling percentages. LRT_{AE}^{DM} = asymptotic p values computed assuming a χ^2 distribution with 2 d.f.; $LRT_{AE}^{DM} Perm$ = p values computed via permutation; LCI (UCI) = 95% lower (upper) confidence interval value.

pl. table 3a). Similarly, at the 0.01 level, to obtain a power of at least 80% using permutation, a 20% double sampling is almost always required.

Recessive Model

As with the previous inheritance model, the LRT_{AE}^{DM} statistic with p values computed asymptotically always rejects the (false) null hypothesis more often than when permutation is used. Under this inheritance model, even calculating p values using permutation, when $R_2 = 1.5$ the LRT_{AE}^{DM} can have low power. In many instances it does not obtain a power of at least 80% even using 100% double sampling. When $R_2 = 2.5$, the power improves, although in some instances double sampling is still required to be up to 100%. This result is due to the mode of inheritance and not to the presence of errors.

Multiplicative Model

The LRT_{AE}^{DM} test with p values asymptotically always rejects the null more often than when permutation is used. To obtain a power of at least 80% when p values are calculated using permutation, a 10–15% double sampling is almost always required at the 0.05 level. In some instances, even as low as 5% double sampling is sufficient, particularly when $R_1 = 2.5$. At the 0.01 level, in a few situations more than 20% double sampling is required to attain 80% power. For example, with $R_1 = 1.5$, $N_a = N_u = 1,000$; $P_d = 0.1$, $err_{CA} = 0.1$ and $err_{CO} = 0.025$ (sim02 in online suppl. table 5b), at 20% double sampling, the power is 58%.

Not surprisingly, the best power results were obtained under the multiplicative model in which the homozygote

risk is increased by the square of the heterozygote relative risk. Interestingly, for the dominant model, even with error rates as high as 10% more than 80% simulations (26 out of 32) show power greater than 80% with as low as 10% double sampling. This proportion increases to 94% (30 out of 32 simulations) when we increase the double sampling to 20%. Under the recessive model, only 34% of the simulations with 10 to 20% double sampling have power greater than 80%. What is interesting here is that increasing the double sampling does not increase the power by much. In addition, we observe that changing the disease allele frequency from 0.1 to 0.25 appears to be one of the most important factors in increasing power. These results are consistent with data without error as checked with the web tool PAWE-PH [28].

In results not shown, for the simulations generated under the alternative hypothesis, the LRT and LRT_{AE} statistics using either asymptotic or permutation results always have power greater than or equal to the power of the LRT_{AE}^{DM} statistic with p values computed by permutation. However, this greater power comes at a price; the power is greater largely because there is a (sometimes gross) inflation in type I error rate.

Discussion

Arguably, a key advantage of our method is cost reduction. While our method requires that a proportion of individuals is double-sampled, researchers will routinely re-genotype an entire sample to come up with correct genotype calls. We can achieve accurate type I error rates

with as little as 10–15% double-sampling proportion for our most extreme situations. We can further control costs by restricting application of our method to those markers that show highly significant evidence for association. This can be performed as part of the replication of the significant associations found in the high-throughput setting, which typically is performed with a different genotyping platform for cost and logistic reasons. In addition, our method provides a direct way, for a given SNP, of testing whether there is differential genotype misclassification between case and control populations.

Note that in case-control genetic association studies, neither differential nor nondifferential misclassification may be detected without the use of reference genotypes [29]. In fact, Clayton et al. [13] determined differential misclassification by use of reference genotypes determined by TaqMan [30]. The underlying mechanism for the differential classification could be either platform-related or sample-related. Clayton et al. suggested that it is possible that this behavior could be replicated in different genotyping platforms. An explanation could be the presence of closely linked SNPs or insertion deletion polymorphisms that overlap the assay probes, present in the case but not in the control population [30]. While this could affect any genotyping chemistry, the probability that a fundamentally distinct secondary method would present the same artifact is low in practice. Due to the greater freedom to position primers in PCR based genotyping methods like TaqMan [30], these could be better suited for follow up of significant association for both testing for misclassification and replication. Ultimately, resequencing might be required to rule out misclassification of a hit on a putatively functional SNP.

One important question that arises from our work is: if a previously significant result becomes nonsignificant after applying our statistic, how are we to interpret the results? We provide two possible answers. First, the result truly is a false positive, or second, it is a true positive but the genotype error rates are so high and/or the sample size is so small for the genetic model parameters (e.g. R_1 , R_2 , P_d , K) that we lose a significant amount of power and therefore the signal goes away after using double-sampled data. However, under Hardy-Weinberg equilibrium, if we have a good estimate of the disease prevalence, we can estimate P_d from control genotype frequencies and further we can estimate genotype relative risks. This information can help us determine if our results are a false positive (genotype relative risks R_1 and R_2 close to 1) or a true positive (relative risks R_1 and R_2 substantially different than 1). For example, under the recessive model (power simula-

tions), if we consider the results for the parameter settings: significance level of 0.05, R_2 equal to 1.5, $N_a(N_u)$ equal to 1,000, $P_d(P_m)$ equal to 0.1, err_{CA} equal to 0.1, err_{CO} equal to 0.025 and 20% double sampling (online suppl. table 4a), for one replicate randomly selected we estimate the following genotype frequencies for cases and controls:

$$\begin{aligned}\hat{P}_{00}^t &= 0.02, \hat{P}_{01}^t = 0.185, \hat{P}_{02}^t = 0.794, \\ \hat{P}_{10}^t &= 0.012, \hat{P}_{11}^t = 0.168, \hat{P}_{12}^t = 0.820.\end{aligned}$$

Using Bayes' theorem and the fact that K is equal to 0.1, we can estimate the penetrances as $\hat{f}_0 = 0.097$, $\hat{f}_1 = 0.109$, $\hat{f}_2 = 0.160$, and the resultant genotype relative risks as $\hat{R}_1 = 1.12$ and $\hat{R}_2 = 1.65$. While these genotype relative risks values are incorrect, the fact that neither $\hat{R}_i$ is close to 1 suggests that the alternative hypothesis is true. We note also that the estimates $\hat{R}_i$ are fairly robust to misspecification in the prevalence K .

We propose that the follow-up of significant association hits by a second genotyping method should be considered the best practice when performing replication in an independent population cohort. Our proposed method can then be used to detect and correct any potential misclassification of the first high-throughput screen by including 10 to 15% of the samples typed in the first study. Most importantly, in this scenario our method would be more cost-effective, not incurring any power loss while maintaining correct type I error rates.

Future plans include the extension of the linear trend test allowing for errors [31] to one that is robust in the presence of differential misclassification.

Acknowledgements

We greatly acknowledge the editors of *Human Heredity* who were most patient in receiving our revised manuscript. We also extend our sincere thanks to two anonymous reviewers, whose comments significantly improved the manuscript.

References

- 1 Kang S, Finch S, Haynes C, Gordon D: Quantifying the percent increase in minimum sample size for SNP genotyping errors in genetic model-based association studies. *Hum Hered* 2004;58:139–144.
- 2 Kang S, Gordon D, Finch S: What SNP genotyping errors are most costly for genetic association studies? *Genet Epidemiol* 2004;26:132–141.
- 3 Mote V, Anderson R: An investigation of the effect of misclassification on the properties of chi-2-tests in the analysis of categorical data. *Biometrika* 1965;52:95–109.

- 4 Gordon D, Finch S, Nothnagel M, Ott J: Power and sample size calculations for case-control genetic association tests when errors are present: application to single nucleotide polymorphisms. *Hum Hered* 2002;54:22–33.
- 5 Gordon D, Ott J: Assessment and management of single nucleotide polymorphism genotype errors in genetic association analysis. *Pac Symp Biocomput* 2001:18–29.
- 6 Ahn K, Haynes C, Kim W, Fleur R, Gordon D, Finch S: The effects of SNP genotyping errors on the power of the Cochran-Armitage linear trend test for case/control association studies. *Ann Hum Genet* 2007;71:249–261.
- 7 Tintle N, Ahn K, Mendell N, Gordon D, Finch S: Characteristics of replicated single-nucleotide polymorphism genotypes from Coga: Affymetrix and Center for Inherited Disease Research. *BMC Genet* 2005;6:S154.
- 8 Tintle N, Gordon D, Van Bruggen D, Finch S: The cost effectiveness of duplicate genotyping for testing genetic association. *Ann Hum Genet* 2009;73:370–378.
- 9 Tintle N, Lantieri F, Lebecq J, Sohns M, Ballard D, Bickeböllner H: Inclusion of a priori information in genome-wide association analysis. *Genet Epidemiol* 2009;33:S74–S80.
- 10 Borchers B, Brown M, McLellan B, Bekmetjev A, Tintle N: Incorporating duplicate genotype data into linear trend tests of genetic association: methods and cost-effectiveness. *Stat Appl Genet Mol Biol* 2009;8:Article 24.
- 11 Tintle N, Gordon D, McMahon F, Finch S: Using duplicate genotyped data in genetic analyses: Testing association and estimating error rates. *Stat Appl Genet Mol Biol* 2007;6:Article 4.
- 12 Plagnol V, Cooper J, Todd J, Clayton D: A method to address differential bias in genotyping in large-scale association studies. *PLoS Genet* 2007;3:e74.
- 13 Clayton D, Walker N, Smyth D, Pask R, Cooper J, Maier L, Smink L, Lam A, Ovington N, Stevens H, Nutland S, Howson J, Faham M, Moorhead M, Jones H, Falkowski M, Hardenbol P, Willis T, Todd J: Population structure, differential bias and genomic control in a large-scale, case-control association study. *Nat Genet* 2005;37:1243–1246.
- 14 Moskvina V, Craddock N, Holmans P, Owen M, O'Donovan M: Effects of differential genotyping error rate on the type I error probability of case-control studies. *Hum Hered* 2006;61:55–64.
- 15 Pearce C, Van Den Berg D, Makridakis N, Reichardt J, Ross R, Pike M, Kolonel L, Henderson B: No association between the SRD5A2 gene A49T missense variant and prostate cancer risk: lessons learned. *Hum Mol Genet* 2008;17:2456–2461.
- 16 Marquard V, Beckmann L, Heid I, Lamina C, Chang-Claude J: Impact of genotyping errors on the type I error rate and the power of haplotype-based association methods. *BMC Genet* 2009;10:3.
- 17 Cheng K, Lin W: The effects of misclassification in studies of gene-environment interactions. *Hum Hered* 2009;67:77–87.
- 18 Ahn K, Gordon D, Finch S: Increase of rejection rate in case-control studies with the differential genotyping error rates. *Stat Appl Genet Mol Biol* 2009;8:Article 25.
- 19 Makridakis N, Ross R, Pike M, Crocitto L, Kolonel L, Pearce C, Henderson B, Reichardt J: Association of mis-sense substitution in SRD5A2 gene with prostate cancer in African-American and hispanic men in Los Angeles, USA. *Lancet* 1999;354:975–978.
- 20 Moorhead M, Hardenbol P, Siddiqui F, Falkowski M, Bruckner C, Ireland J, Jones H, Jain M, Willis T, Faham M: Optimal genotype determination in highly multiplexed SNP data. *Eur J Hum Genet* 2006;14:207–215.
- 21 Kim W, Gordon D, Sebat J, Ye K, Finch S: Computing power and sample size for case-control association studies with copy number polymorphism: application of mixture-based likelihood ratio test. *PLoS ONE* 2008;3:e3475.
- 22 Gordon D, Yang Y, Haynes C, Finch S, Mendell N, Brown A, Haroutunian V: Increasing power for tests of genetic association in the presence of phenotype and/or genotype error by use of double-sampling. *Stat Appl Genet Mol Biol* 2004;3:Article 26.
- 23 Tenenbein A: A double sampling scheme for estimating from binomial data with misclassifications. *J Am Stat Assoc* 1970;65:1350–1361.
- 24 Tenenbein A: A double sampling scheme for estimating from misclassified multinomial data with applications to sampling inspection. *Technometrics* 1972;14:187–202.
- 25 Dempster A, Laird N, Rubin D: Maximum likelihood from incomplete data via the EM algorithm. *J R Stat Soc B Met* 1977;39:1–38.
- 26 Douglas J, Skol A, Boehnke M: Probability of detection of genotyping errors and mutations as inheritance inconsistencies in nuclear-family data. *Am J Hum Genet* 2002;70:487–495.
- 27 Ott J: Analysis of Human Genetic Linkage. Baltimore, The Johns Hopkins University Press, 1999.
- 28 Edwards BJ, Haynes C, Levenstien MA, Finch SJ, Gordon D: Power and sample size calculations in the presence of phenotype errors for case/control genetic association studies. *BMC Genet* 2005;6:18.
- 29 Pompanon F, Bonin A, Bellemain E, Taberlet P: Genotyping errors: causes, consequences and solutions. *Nat Rev Genet* 2005;6:847–859.
- 30 Dela Vega F, Lazaruk K, Rhodes M, Wenz M: Assessment of two flexible and compatible SNP genotyping platforms: Taqman SNP Genotyping Assays and the SNplex Genotyping System. *Mutat Res* 2005;573:111–135.
- 31 Gordon D, Haynes C, Yang Y, Kramer P, Finch S: Linear trend tests for case-control genetic association that incorporate random phenotype and genotype misclassification error. *Genet Epidemiol* 2007;31:853–870.