Available online at www.sciencedirect.com

SCIENCE DIRECT

Genomics 84 (2004) 623–630

GENOMICS

www.elsevier.com/locate/ygeno

Estimation of genotype error rate using samples with pedigree information—an application on the GeneChip Mapping 10K array

Ke Hao,[a] Cheng Li,[a,b] Carsten Rosenow,[c] and Wing Hung Wong[a,d,*]

[a] *Department of Biostatistics, Harvard School of Public Health, 655 Huntington Avenue, Boston, MA 02115, USA*
[b] *Department of Biostatistics, Dana Farber Cancer Institute, Boston, MA, USA*
[c] *Genomics Collaboration, Affymetrix, Santa Clara, CA, USA*
[d] *Department of Statistics, Harvard University, Cambridge, MA, USA*

Received 16 January 2004; accepted 13 May 2004
Available online 13 July 2004

Abstract

Currently, most analytical methods assume all observed genotypes are correct; however, it is clear that errors may reduce statistical power or bias inference in genetic studies. We propose procedures for estimating error rate in genetic analysis and apply them to study the GeneChip Mapping 10K array, which is a technology that has recently become available and allows researchers to survey over 10,000 SNPs in a single assay. We employed a strategy to estimate the genotype error rate in pedigree data. First, the "dose–response" reference curve between error rate and the observable error number were derived by simulation, conditional on given pedigree structures and genotypes. Second, the error rate was estimated by calibrating the number of observed errors in real data to the reference curve. We evaluated the performance of this method by simulation study and applied it to a data set of 30 pedigrees genotyped using the GeneChip Mapping 10K array. This method performed favorably in all scenarios we surveyed. The dose–response reference curve was monotone and almost linear with a large slope. The method was able to estimate accurately the error rate under various pedigree structures and error models and under heterogeneous error rates. Using this method, we found that the average genotyping error rate of the GeneChip Mapping 10K array was about 0.1%. Our method provides a quick and unbiased solution to address the genotype error rate in pedigree data. It behaves well in a wide range of settings and can be easily applied in other genetic projects. The robust estimation of genotyping error rate allows us to estimate power and sample size and conduct unbiased genetic tests. The GeneChip Mapping 10K array has a low overall error rate, which is consistent with the results obtained from alternative genotyping assays.

Keywords: Genotyping error; Mendelian inheritance; GeneChip Mapping 10K array; Single nucleotide polymorphism

Genotyping error, defined as the proportion of mistypings in all called genotypes, may occur due to nonspecificity of experimental assay, inappropriate allele calling, or simply random assay instability. It can lead to (1) incorrect inference of allele frequency, map order, linkage disequilibrium, and marker distance [1,3,8] and (2) biased results or reduced statistical power of genetic linkage and association studies [1,3,4,8,9,11,13–16,21]. However, the genotype error rate is hard to estimate directly. Repeating the assays may confirm the reproducibility but is unable to rule out systematic errors. An alternative approach is to employ an alternative assay method, but the true genotype still remains unknown when inconsistency is observed, and the inconsistency rate cannot be treated directly as genotyping error rate.

Because the knowledge of genotyping error rate is valuable for sample size estimation [9] and unbiased genetic testing [21], statistical methods detecting errors in pedigree data have been proposed [6,10,22,23]. A natural approach is to check consistency with the Mendelian inheritance law. The detection rate of genotyping error using Mendelian law has been estimated to be 13–75% conditional on allele frequency and pedigree structure [5,7,11,12]. It has been suggested that true error rate can be estimated from the detection rate [7]. A likelihood approach was proposed by Gordon and Ott to estimate the average error rates [15].

* Corresponding author. Fax: +1-617-739-1781.
E-mail address: wwong@hsph.harvard.edu (W. Hung Wong).

0888-7543/$ - see front matter © 2004 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2004.05.003

The recently released GeneChip Mapping 10K array offers the ability to genotype over 10,000 SNPs on a single array [17,19]. This technology uses an innovative assay that eliminates the need for locus-specific PCR and requires only 250 of ng DNA for each sample [17,19]. The potential application includes linkage analysis, fine mapping, genomic control against population admixture, and loss of heterozygosity (LOH) study. To validate the accuracy of the technology, 538 SNPs across 40 individuals were randomly selected and genotyped by single-base extension and capillary sequencing. The genotypes were compared to the results generated by the GeneChip Mapping 10K array, and the inconsistency rate was found to be <0.4% (http://www.affymetrix.com/support/technical/datasheets/10k_datasheet.pdf). This result suggested a fairly high degree of accuracy; however, the number should not be considered as genotyping error rate since it did not take the inherent error rate of the other technologies into account.

In this report, we estimated genotyping errors by Mendelian inheritance law and likelihood of recombination events. We achieved this goal by systematically capturing the "dose–response" relationship between error rate and observable number of errors conditional on the given pedigree structures and genotypes. Once the reference dose–response curve was derived, the error rate could be estimated by calibrating observed error number to the reference curve. The performance of this method was first validated by simulation. The simulation results showed that the error rate estimate was accurate independent of the error model and the pedigree structure. The same accuracy is achieved whether we check Mendelian error only or both Mendelian error and unlikely genotypes. Finally, we applied this method to evaluate the average error rate of the GeneChip Mapping 10K array. The estimation was based on genotype data of 30 trio families. Our results suggested that the error rate of this technology was close to 0.1%, and this error rate was not affected by whether the SNP was frequent or rare or whether it was a transition or transversion polymorphism.

Results

Simulation results

Our simulation results showed that the relationship between $E(N)$ and ε was monotone and nearly linear, with a substantial slope in all three pedigree structures surveyed (Fig. 1). As expected, we observed more errors when we checked for unlikely genotypes in addition to Mendelian errors. Interestingly, the dose–response in the two cases (Mendelian error only versus both Mendelian error and unlikely genotypes) turned out to be nearly parallel. Thus it is sufficient to estimate the error rate by using only the Mendelian errors. We also found that the observed error numbers were almost the same under the random and the directed error models (Fig. 2), except in cases of extremely high error rate (i.e., >5%). Finally, we examined the ability of this method to estimate the average error rate in a population of SNPs with heterogeneous error rates and found it could still accurately estimate the average error rate in most cases (Fig. 3).

Error rate of GeneChip mapping 10K array

After validating the performance of the method by simulation, we applied it to evaluate the average error rate of the GeneChip Mapping 10K array. In total 30 trios were genotyped on 10,043 SNP markers, with an overall call rate of 96.4%. In this data set 8916 SNPs were polymorphic. Because a trio pedigree does not provide sufficient information to identify unlikely genotypes, we checked only for Mendelian inconsistency. $N = 210$ errors were observed in this data set. There was no SNP that had significantly more errors than the rest: 5 SNPs showed 2 errors, and the remaining SNPs showed 1 or no error. We surveyed several possible values of ε (0.1, 0.05, 0.01, 0.005, 0.001, 0.0005, and 0.0001) and calculated the $E(N)$ by simulation (Table 1). $E(N)$ was sensitive to ε in a wide range of parameter values. In the log–log plot, the dose–response curve was almost linear and had a slope that significantly deviated from 0. Calibrating $N = 210$ to this curve, we obtained an estimated genotyping error rate of 0.001. Moreover, we stratified the SNPs by their allele frequencies and mutation types and found no detectable difference in term of error rate between high- and low-frequency SNPs (Table 2) or between transitions and transversions (Table 3).

Discussion

In this report, we proposed a strategy utilizing pedigree information to estimate genotyping error rate through a simulation-based calibration. The method performed favorably in various scenarios of different pedigree structures, different error models, and heterogeneous error rates. We applied this method to a large data set of SNP genotypes obtained by using the GeneChip Mapping 10K array and estimated the error rate of this new technology.

Only a proportion of genotyping errors can lead to disagreement with the Mendelian inheritance law, especially in the case of SNP markers. But by conditioning on the given pedigree structure and genotypes, our method is sufficient to estimate validly the genotyping error rate. In our data set, there was a nearly linear relationship between error rate and Mendelian error number. The slope of the curve was fairly steep, which indicated that the variance in genotyping error rate estimation could be small. Detecting unlikely genotypes did not contribute greatly in estimation of error rate but required intensive computation and an arbitrary cutoff in claiming mistypings. Therefore, we suggest not utilizing unlikely genotypes in the current

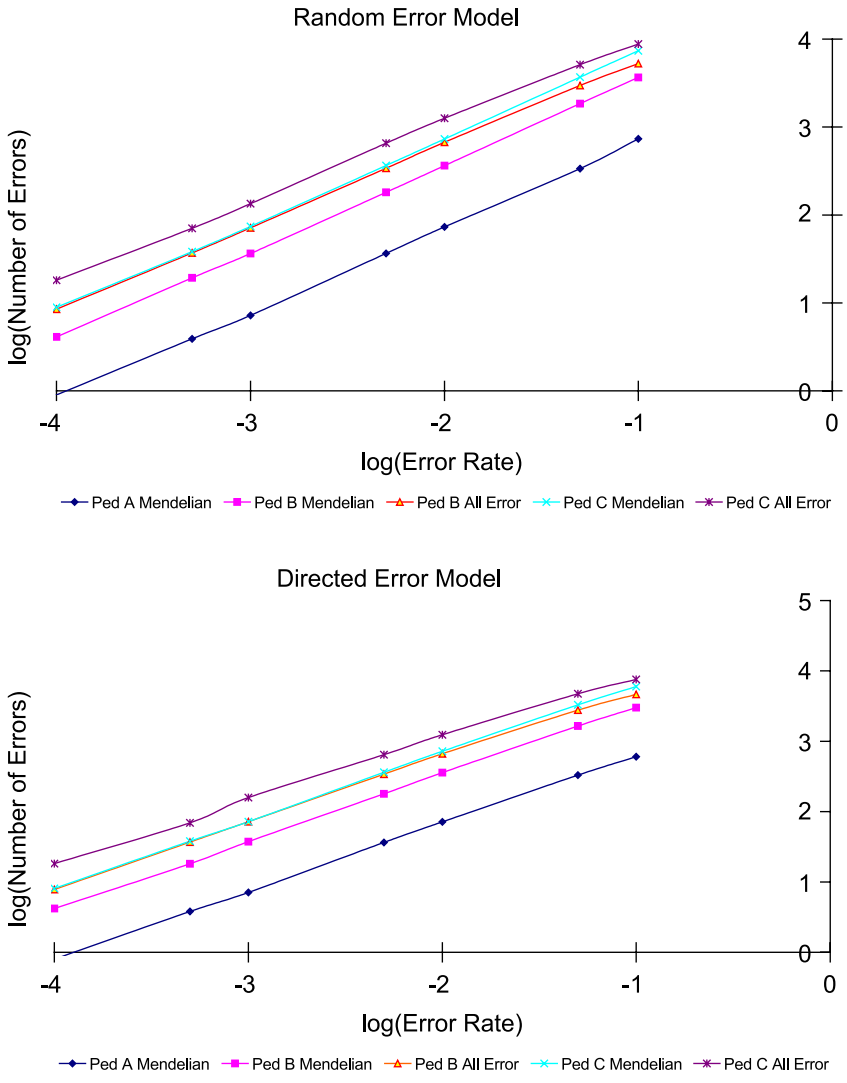

Fig. 1. The dose–response curves between error rate and number of observed errors. Base 10 was used in log transformation. Mendelian, detect only Mendelian inconsistency; All Error, both Mendelian inconsistency and unlikely genotype ($p_{\text{cutoff}} = 0.0001$).

setting with parental genotypes available. On the other hand, when parental genotypes are not available, SNPs become uninformative in Mendelian checking, and it is necessary to consider unlikely genotypes. If a SNP carries an extremely high error rate, it usually can be identified directly (i.e., extremely large number of Mendelian errors associated with this SNP). Hence, we studied only moderate error rate heterogeneity among multiple SNPs. In the current study, we estimated the average error rate of a group of SNPs, but this method can be applied directly to a single or a few markers if a sufficient number of pedigrees were genotyped.

An appealing feature of this method is that the estimation is achieved by conditioning on the structure of every pedigree in the given data set. In generating the dose–response reference curve, we corrected the observed errors and left most genotypes untouched. We noted there are multiple solutions to correct the Mendelian errors. If only one corrected genotype set was used for all simulation loops to derive the reference curve, the arbitrary choice could lead to unstable estimation, although it is still unbiased. In this report, we performed independent correction in each iteration. An alternative strategy is to remove the genotypes of a particular marker in a certain pedigree, if this marker shows Mendelian inconsistency in this family. However, this approach may reduce the number of total genotypes in the data set and lead to overestimation of the error rate. This bias could be more severe in the case of many Mendelian errors in the data set.

Two commonly used SNP genotyping error models were considered in this report. We found in the case of low-to-moderate error rate that the choice of error model did not affect the estimation very much. However, it ought to be noted that the error model depends on the chemistry of the genotyping assay and calling scheme and should be specified appropriately. For example, due to incomplete digestion in restriction fragment length polymorphism, the digested

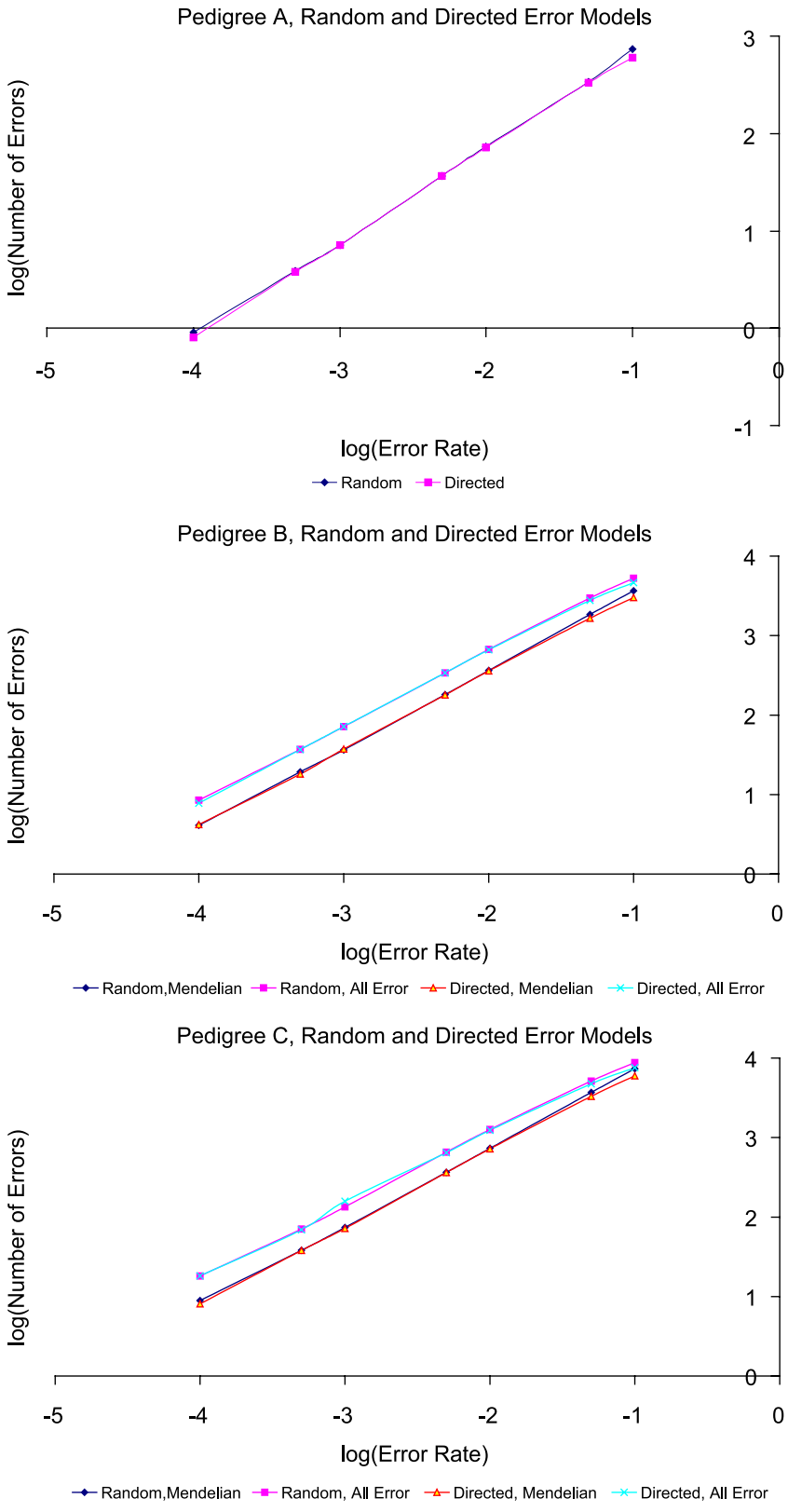

Fig. 2. Comparison of the dose–response curves between random and directed error models. Base 10 was used in log transformation. In the middle (pedigree B) and bottom (pedigree C) graphs, curves lie in two groups. The higher groups are the curves using All Errors, and the lower groups are the curves using Mendelian errors only.

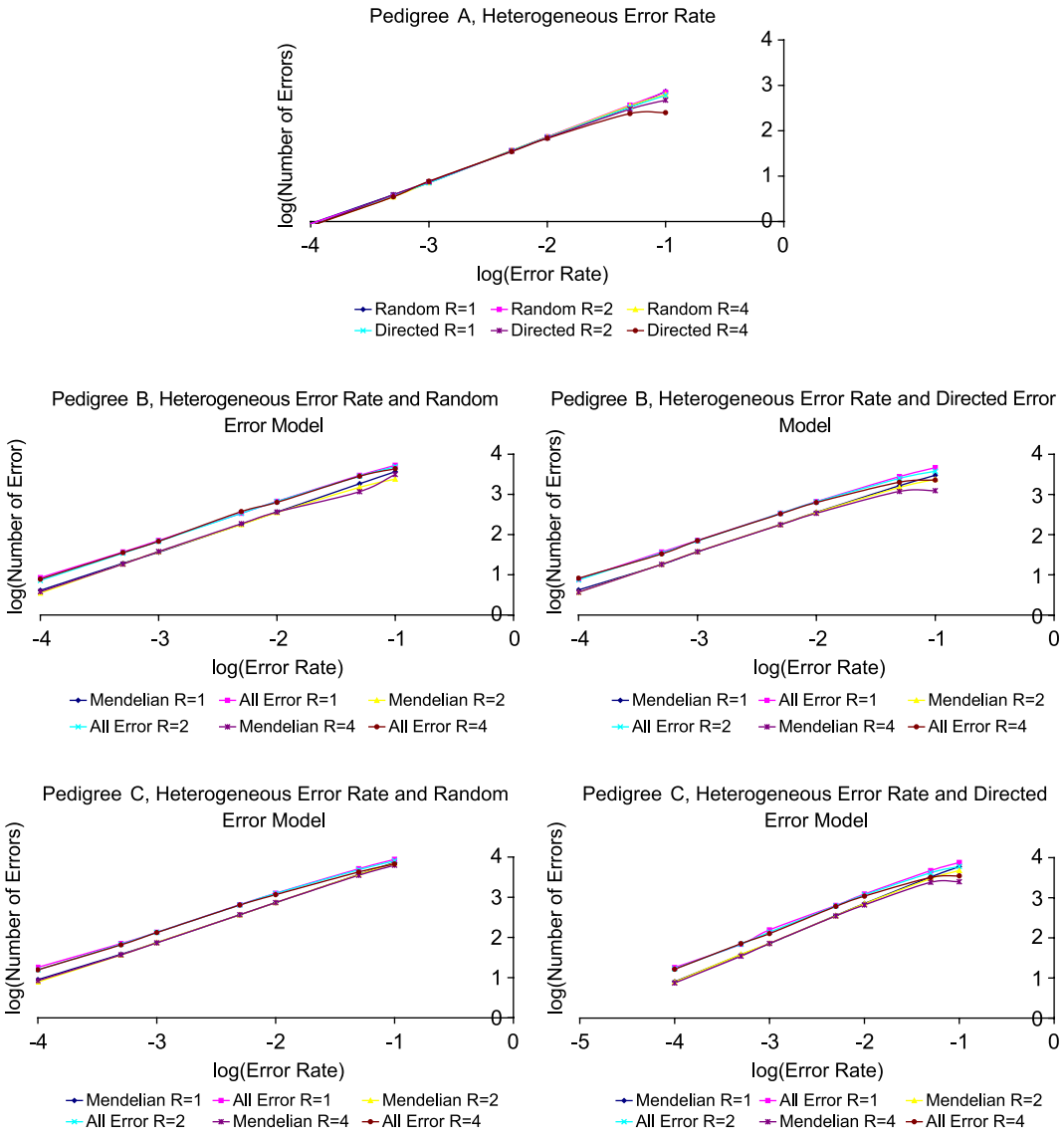

Fig. 3. Comparison of the dose–response curves between homogeneous and heterogeneous error rates. Base 10 was used in log transformation. $R = 1$ indicates homogeneous error rate and $R = 2$ or 4 indicates heterogeneous error rate. When $R = 2$, half of the SNPs do not carry genotyping errors, and the other half have errors with 2 times the specified error rate (x axis). When $R = 4$, three-quarters of the SNPs do not carry genotyping errors, and the other quarter have errors with 4 times the specified error rate (x axis).

allele is more likely to be misclassified as an undigested allele, for which a directed error model becomes more reasonable. Our method relies on the monotone relationship between the error rate and the number of observable errors. In the case of very high error rate (i.e., over 20%), we may no longer observe the monotonic pattern, and therefore the method becomes inapplicable. Moreover, at the current setting, the method cannot accommodate complicated error models containing more than one parameter [5,16,22].

In deriving the dose–response reference curves, we applied the observed allele frequency of the marker. When the genotyping error rate is small, as in the data set used in this paper, the estimation should be trustworthy. However, when error rate increases, the dose–response curve is biased due to incorrectly estimated allele frequency. One solution is to use known allele frequency of the marker in the given population. When such information is not available, we may partially overcome this problem by applying the estimation procedure twice. At the second time, we can recalculate the allele frequency according to the estimated error rate (from the first run) and the specified error model and simulate the genotypes using corrected allele frequency to draw the dose–response curve again. A simulation study was carried out to demonstrate that this recursive correction could somewhat alleviate the bias (Table 4). Under high error rate (i.e., >3%), the observed allele frequencies deviated from the true values and so did the initial estimations of the error rate ($\hat{e}^{(1)}$ in Table 4). Fortunately, the relative bias of $\hat{e}^{(1)}$ was only small to moderate in magnitude, even under 10% error rate. Furthermore, we corrected the allele frequency accord-

Table 1
Number of Mendelian errors on CEPH data set

	Number of Mendelian errors[a] (SNP = 8916, error = 210)	
Error rate	Random error model	Directed error model
0.1	21,628	17,633.3
0.05	10,866	9,774.0
0.01	2,162	2,129.7
0.005	1,069	1,074.3
0.001	216	214.4
0.0005	111	106.8
0.0001	26	22.0

[a] The average number of Mendelian errors was calculated over 1000 iterations.

ing to $\hat{e}^{(1)}$ and specified error model and obtained the second estimation $\hat{e}^{(2)}$. We found $\hat{e}^{(2)}$ was very close to the truth, thus in this case, a two-step recursive procedure was able to correct the estimation bias.

In addition to the point estimation, the confidence interval of the genotyping error rate is also of interest. Gordon and Ott observed a wide confidence interval when using maximum likelihood approaches [15]. Under the setting of this report, each CEPH (Centre d'Etude du Polymorphisme Humain) trio pedigree on one marker shows either 0 or 1 Mendelian error, which gives a Bernoulli distribution. Considering that all pedigrees on all markers are identical and independently distributed, the observed error number N follows a binomial distribution with $\mathrm{Var}(N) = p(1-p)N$. Under this assumption, we found that the 95% confidence interval of N and genotyping error rate were (181, 239) and (0.083%, 0.110%), respectively.

A caveat on our finding concerning the GeneChip Mapping 10K array is that the small error rate is demonstrated only on an application for which ample and high-quality DNA is available. This is true for many linkage and association studies but is unlikely to be true for some applications such as studies of genome amplification, deletion, or LOH in preserved tissue samples. Additional studies are necessary to assess the error rate in these situations. High-resolution SNP array has the potential to accelerate

Table 2
Estimation of error rate on CEPH data set stratified by allele frequency

	Number of Mendelian errors[a]			
	Low-frequency SNPs (SNP = 3134, error = 60)		High-frequency SNPs (SNP = 5782, error = 147)	
Error rate	Random	Directed	Random	Directed
0.1	6907.1	6001.1	14,361.9	11,870.0
0.05	3492.6	3114.1	7,107.2	6,538.5
0.01	706.2	675.0	1,427.5	1,390.5
0.005	348.7	342.0	706.8	714.3
0.001	69.1	72.8	146.6	146.7
0.0005	37.5	33.5	76.1	71.4
0.0001	7.9	8.4	14.0	17.8

[a] The average number of Mendelian errors was calculated over 1000 iterations. A low-frequency SNP was defined as a SNP with minor allele frequency less than or equal to 0.2, and a high-frequency SNP was defined as a SNP with minor allele frequency more than 0.2.

Table 3
Estimation of error rate on CEPH data set stratified by polymorphism type

	Number of Mendelian errors[a]			
	Transition (SNP = 6049, error = 149)		Transversion (SNP = 2867, error = 61)	
Error rate	Random	Directed	Random	Directed
0.1	14,724.8	11,977.0	6943.9	5638.4
0.05	7,416.4	6,624.4	3449.5	3175.2
0.01	1,458.6	1,450.9	692.5	685.9
0.005	724.8	725.2	344.7	341.3
0.001	145.0	148.8	66.8	70.6
0.0005	75.6	76.2	35.4	36.3
0.0001	17.5	16.8	7.7	8.5

[a] The average number of Mendelian errors was calculated over 1000 iterations.

greatly the genetic studies of complex traits. We found the accuracy of this technology to be rather high (around 99.9%). This feature is very important since genotyping errors could bias results or reduce statistical power of genetic studies [1,3,4,9]. It is especially critical in linkage studies, in which 1% error may double the required sample size to achieve a given power [20]. Finally, we found the technology to perform equally well on low- or high-fre-

Table 4
Bias in error rate estimation and recursive correction[a]

a	e	$\tilde{a}$	$\hat{e}^{(1)}$	$\hat{e}^{(2)}$
(A) Random error model				
0.2	1%	0.203	1.00%	1.00%
	3%	0.209	3.04%	3.01%
	5%	0.215	5.00%	5.00%
	10%	0.230	10.20%	10.05%
0.4	1%	0.401	1.00%	1.00%
	3%	0.403	3.01%	3.00%
	5%	0.404	4.99%	5.00%
	10%	0.410	10.04%	10.00%
(B) Directed error model				
0.2	1%	0.205	1.01%	1.00%
	3%	0.215	3.06%	3.00%
	5%	0.225	5.08%	5.01%
	10%	0.250	10.40%	9.95%
0.4	1%	0.405	1.00%	1.00%
	3%	0.415	3.03%	2.99%
	5%	0.425	5.02%	5.00%
	10%	0.450	10.45%	9.99%
0.6	1%	0.605	1.00%	1.00%
	3%	0.615	2.96%	3.00%
	5%	0.625	4.92%	4.98%
	10%	0.650	9.92%	10.00%
0.8	1%	0.805	1.00%	1.00%
	3%	0.815	3.00%	3.00%
	5%	0.825	5.01%	5.00%
	10%	0.850	10.75%	9.99%

[a] The simulation was carried out using the trio pedigree structure, and 1000 iterations were conducted for each set of parameters. Under high error rate, the estimation of allele frequency and error rate could be biased when the true allele frequency is unknown. A recursive procedure is applied to correct the estimation. a and $\tilde{a}$ denote the true and observed allele frequencies; e, $\hat{e}^{(1)}$, and $\hat{e}^{(2)}$ denote the true, first, and second estimations, respectively, of the error rate.

quency SNPs and on transitions or transversions. A computer software package, Genotype Error Rate Estimator on Pedigree Samples, has been developed for researchers to estimate the error rate on their data sets. This package is freely available via written request to the authors.

Methods

Mendelian inheritance error and unlikely genotypes

Checking for Mendelian transmission in a pedigree is a routine procedure in gene mapping studies [5]. However, only a fraction of the genotyping errors can lead to disagreement with the Mendelian inheritance law. Detection rate, among biallelic markers, is 13–75% in nuclear families [5]. Another strategy is to detect mistypings among tightly linked markers by considering the fact that recombination is a very rare event among these markers [23]. The methods can further identify errors that display Mendelian consistency, but an arbitrary cutoff is required to call an unlikely genotype as an error. In the current paper, we considered both Mendelian inconsistency and unlikely genotypes (using a p-value cutoff of 0.0001) as errors. The genetic distances among the SNPs were calculated based on the nearby short tandem repeat markers with accurate genetic map position [18] by assuming the genetic distance was proportional to physical distance in a small chromosomal region.

Error model

We used two error models to describe the relationship between allelotyping error rate (τ) and genotyping error rate (ε). These models were used in the simulation to propagate ε to τ, based on which allelotyping errors were randomly introduced.

For simplicity, we denote the average probability of misclassifying allele A to allele a as $P(\mathrm{A} \rightarrow \mathrm{a})$. In the random error model, it is assumed $P(\mathrm{A} \rightarrow \mathrm{a}) = P(\mathrm{a} \rightarrow \mathrm{A}) = \tau$. It follows that $P(\mathrm{AA} \rightarrow \mathrm{Aa}) = P(\mathrm{aa} \rightarrow \mathrm{Aa}) = 2\tau - 2\tau^2$, $P(\mathrm{Aa} \rightarrow \mathrm{AA}) = P(\mathrm{Aa} \rightarrow \mathrm{aa}) = \tau - \tau^2$, and $P(\mathrm{AA} \rightarrow \mathrm{aa}) = P(\mathrm{aa} \rightarrow \mathrm{AA}) = \tau^2$. Let p denote allele frequency of A, then the genotyping error rate ε satisfies the equation

$$\varepsilon = p^2(2\tau) + 2p(1-p)(2\tau) + (1-p)^2(2\tau) + O(\tau^2).$$

When τ is small, we can ignore the $O(\tau^2)$ term and obtain the relationship

$$\tau \approx \varepsilon/2. \quad (1)$$

In the directed error model, we assume $P(\mathrm{A} \rightarrow \mathrm{a}) = \tau$ and $P(\mathrm{a} \rightarrow \mathrm{A}) = 0$. Then $P(\mathrm{AA} \rightarrow \mathrm{Aa}) = 2\tau - 2\tau^2$, $P(\mathrm{Aa} \rightarrow \mathrm{aa}) = \tau$, $P(\mathrm{AA} \rightarrow \mathrm{aa}) = \tau^2$, and the probabilities of any other transmissions are 0, resulting in the equation

$$\varepsilon = p^2(2\tau) + 2p(1-p)\tau + O(\tau^2)$$

and

$$\tau \approx \varepsilon/2p. \quad (2)$$

Simulation studies

We studied three types of pedigrees: trio, nuclear family with multiple offspring, and extended pedigree. The structures of the three types of pedigrees are illustrated in Fig. 4. In each replication of the simulation, we performed the following computation: (1) simulating the genotypes of 8916 SNPs, (2) introducing genotyping errors on these SNPs, and (3) applying our method to estimate the average error rate. The allele frequencies of these 8916 SNPs were obtained from 30 CEPH trios genotyped by the GeneChip Mapping 10K array (see below). The allele frequency distribution was symmetric with respect to 0.5. We introduced genotyping errors under both random and directed error models and under a range of error rates (0.1, 0.05, 0.01, 0.005, 0.001, 0.0005, and 0.0001). In the analysis of pedigrees, the software package MERLIN was called to detect unlikely genotypes [2]. We performed 1000 replicates of the simulation for each set of parameters.

Evaluation of the impact of heterogeneous error rates

As mentioned earlier, this method estimates the average error rate. When the error rates are different among SNPs, we want to know whether this method can still accurately estimate the average error rate. We addressed this issue by randomly selecting 50 or 75% of the SNPs to be free of genotyping errors, and for the remaining 50 or 25% SNPs, we introduced errors using two or four times the specified error rate, respectively.

Estimation of the genotype error rate of GeneChip mapping 10K array

Thirty independent pedigrees from the CEPH research laboratory were used in the current study. In each pedigree

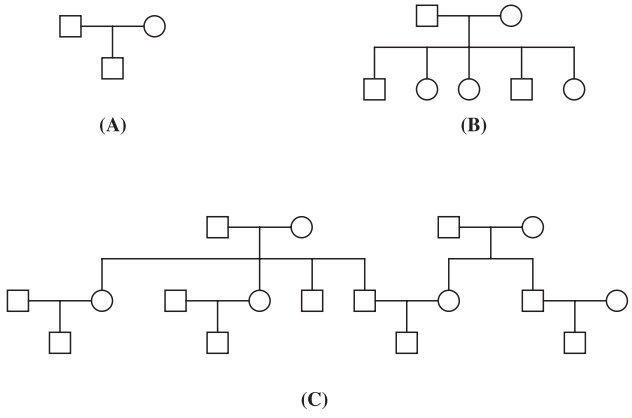

Fig. 4. The structures of pedigrees used in the simulation study. (A) Trio, (B) nuclear family with multiple offspring, and (C) extended pedigree.

we genotyped one trio (father, mother, and one offspring) using GeneChip Mapping 10K arrays, and only autosomal genotypes were examined for error rate.

Genomic DNA (250 ng) was digested with *Xba*I at 37°C for 4 h. The DNA fragments underwent ligation to a universal adaptor and then PCR amplification with a common primer. The amplicon was cleaved by partial DNase I digestion to shorter fragments and labeled with biotinylated ddATP using terminal deoxytransferase. The labeled DNA was injected into the microarray cartridge and incubated overnight. The hybridized microarray was washed and stained following a three-step protocol and scanned under the manufacturer's directions (Affymetrix, Santa Clara, CA, USA). Finally, the genotypes were determined using an automated calling software (Affymetrix). The detailed genotyping procedure used has been previously described [17].

Because only one offspring is typed in each pedigree, we cannot identify unlikely genotypes among tightly linked markers [23], and thus only Mendelian consistency checking was employed. We first counted the number of Mendelian inconsistencies in the data set and recorded it as N. Then we derived the reference dose–response curve between error rate and number of observed errors through simulation conditional on real data. In each iteration of the simulation, we (1) corrected the existing errors, (2) introduced genotyping errors according to ε and the error model, and (3) counted the number of observable errors and recorded it as n. Finally, we estimated the error rate by comparing N to the reference curve. Furthermore, we stratified the SNPs by allele frequency using a cutoff of 0.2 to study if the error rate was different in common and rare SNPs. We also stratified the SNPs by types of SNP (transition or transversion) and examined if the SNP type affected error rate.

Acknowledgments

We thank Dr. Rui Mei for providing the data set. We thank Dr. Xin Xu at Harvard School of Public Health for carefully reading the manuscript and providing insightful comments. This work is partially supported by NIH Grant 1R01HG02341.

References

[1] G.R. Abecasis, S.S. Cherny, L.R. Cardon, The impact of genotyping error on family-based analysis of quantitative traits, Eur. J. Hum. Genet. 9 (2001) 130–134.
[2] G.R. Abecasis, S.S. Cherny, W.O. Cookson, L.R. Cardon, Merlin—rapid analysis of dense genetic maps using sparse gene flow trees, Nat. Genet. 30 (2002) 97–101.
[3] J.M. Akey, K. Zhang, M. Xiong, P. Doris, L. Jin, The effect that genotyping errors have on the robustness of common linkage-disequilibrium measures, Am. J. Hum. Genet. 68 (2001) 1447–1456.
[4] K.H. Buetow, Influence of aberrant observations on high-resolution linkage analysis outcomes, Am. J. Hum. Genet. 49 (1991) 985–994.
[5] J.A. Douglas, A.D. Skol, M. Boehnke, Probability of detection of genotyping errors and mutations as inheritance inconsistencies in nuclear-family data, Am. J. Hum. Genet. 70 (2002) 487–495.
[6] M.G. Ehm, M. Kimmel, R.W. Cottingham Jr., Error detection for genetic data, using likelihood methods, Am. J. Hum. Genet. 58 (1996) 225–234.
[7] F. Geller, A. Ziegler, Detection rates for genotyping errors in SNPs using the trio design, Hum. Hered. 54 (2002) 111–117.
[8] D.R. Goldstein, H. Zhao, T.P. Speed, The effects of genotyping errors and interference on estimation of genetic distance, Hum. Hered. 47 (1997) 86–100.
[9] D. Gordon, S.J. Finch, M. Nothnagel, J. Ott, Power and sample size calculations for case–control genetic association tests when errors are present: application to single nucleotide polymorphisms, Hum. Hered. 54 (2002) 22–33.
[10] D. Gordon, S.C. Heath, X. Liu, J. Ott, A transmission/disequilibrium test that allows for genotyping errors in the analysis of single-nucleotide polymorphism data, Am. J. Hum. Genet. 69 (2001) 371–380.
[11] D. Gordon, S.C. Heath, J. Ott, True pedigree errors more frequent than apparent errors for single nucleotide polymorphisms, Hum. Hered. 49 (1999) 65–70.
[12] D. Gordon, S.M. Leal, S.C. Heath, J. Ott, An analytic solution to single nucleotide polymorphism error-detection rates in nuclear families: implications for study design, Pac. Symp. Biocomput. (2000) 663–674.
[13] D. Gordon, M.A. Levenstien, S.J. Finch, J. Ott, Errors and linkage disequilibrium interact multiplicatively when computing sample sizes for genetic case–control association studies, Pac. Symp. Biocomput. (2003) 490–501.
[14] D. Gordon, T.C. Matise, S.C. Heath, J. Ott, Power loss for multiallelic transmission/disequilibrium test when errors introduced: GAW11 simulated data, Genet. Epidemiol. 17 (Suppl 1) (1999) S587–592.
[15] D. Gordon, J. Ott, Assessment and management of single nucleotide polymorphism genotype errors in genetic association analysis, Pac. Symp. Biocomput. (2001) 18–29.
[16] S.J. Kang, D. Gordon, S.J. Finch, What SNP genotyping errors are most costly for genetic association studies? Genet. Epidemiol. 26 (2004) 132–141.
[17] G.C. Kennedy, H. Matsuzaki, S. Dong, W.M. Liu, J. Huang, G. Liu, X. Su, M. Cao, W. Chen, J. Zhang, W. Liu, G. Yang, X. Di, T. Ryder, Z. He, U. Surti, M.S. Phillips, M.T. Boyce-Jacino, S.P. Fodor, K.W. Jones, Large-scale genotyping of complex DNA, Nat. Biotechnol. (2003) 1233–1237.
[18] A. Kong, D.F. Gudbjartsson, J. Sainz, G.M. Jonsdottir, S.A. Gudjonsson, B. Richardsson, S. Sigurdardottir, J. Barnard, B. Hallbeck, G. Masson, A. Shlien, S.T. Palsson, M.L. Frigge, T.E. Thorgeirsson, J.R. Gulcher, K. Stefansson, A high-resolution recombination map of the human genome, Nat. Genet. 31 (2002) 241–247.
[19] H. Matsuzaki, H. Loi, S. Dong, Y.Y. Tsai, J. Fang, J. Law, X. Di, W.M. Liu, G. Yang, G. Liu, J. Huang, G.C. Kennedy, T.B. Ryder, G.A. Marcus, P.S. Walsh, M.D. Shriver, J.M. Puck, K.W. Jones, R. Mei, Parallel genotyping of over 10,000 SNPs using a one-primer assay on a high-density oligonucleotide array, Genome Res. 14 (2004) 414–425.
[20] A. Oliphant, D.L. Barker, J.R. Stuelpnagel, M.S. Chee, BeadArray technology: enabling an accurate, cost-effective approach to high-throughput genotyping. Biotechniques (Suppl 56–58) (2002) 60–51.
[21] K.M. Rice, P. Holmans, Allowing for genotyping error in analysis of unmatched case–control studies, Ann. Hum. Genet. 67 (2003) 165–174.
[22] E. Sobel, J.C. Papp, K. Lange, Detection and integration of genotyping errors in statistical genetics, Am. J. Hum. Genet. 70 (2002) 496–508.
[23] G. Zou, D. Pan, H. Zhao, Genotyping error detection through tightly linked markers, Genetics 164 (2003) 1161–1173.