Original Paper

Human Heredity

Hum Hered 2009;67:116–127 DOI: 10.1159/000179559 Received: February 13, 2008 Accepted after revision: April 29, 2008 Published online: December 12, 2008

Detection of Parent-of-Origin Effects in Complete and Incomplete Nuclear Families with Multiple Affected Children Using Multiple Tightly Linked Markers

Ji-Yuan Zhou^{a, b} Shili Lin^c Wing K. Fung^a Yue-Qing Hu^a

^aDepartment of Statistics and Actuarial Science, The University of Hong Kong, Hong Kong, and ^bDepartment of Mathematics, Huaihua University, Huaihua, Hunan, China; ^cDepartment of Statistics, The Ohio State University, Columbus, Ohio, USA

Key Words

Parent-of-origin effects • Haplotype analysis • Single-marker analysis • Missing parent • Incomplete nuclear family • Complete nuclear family • Multiple testing • Population stratification demographic model • Assortative mating demographic model

Abstract

For a diallelic marker locus, the parental-asymmetry test (PAT) based on case-parents trios and its extensions to accommodate incomplete unclear families (1-PAT and C-PAT) are simple and powerful approaches to test for parent-oforigin effects. However, haplotype analysis is generally regarded as advantageous over single-marker analysis in genetic study of common complex diseases. This is mainly due to the fact that complex diseases are often associated with multiple markers. As such, HAP-PAT was constructed to test for parent-of-origin effects in the framework of haplotype analysis. However, its applicability is limited due to the need for complete parental information. In this paper, for nuclear families with only one parent and multiple affected children, we develop HAP-1-PAT to test for parent-of-origin effects using multiple tightly linked markers. We further propose HAP-C-PAT to combine data from families with both parents and those with only one parent. We carry out a simulation

KARGER

Fax +41 61 306 12 34 E-Mail karger@karger.ch www.karger.com © 2008 S. Karger AG, Basel 0001–5652/09/0672–0116\$26.00/0

Accessible online at: www.karger.com/hhe study to evaluate the validity and power of the test statistics in various settings, including incomplete family rates, marker/disease-locus linkage disequilibrium patterns, and population models. We perform analysis for all possible combinations of the markers being considered. A permutation-based Monte Carlo procedure is devised to determine the significance of the tests; the corrected global p values taking into account of multiple testing are used for inferences. The results show that HAP-1-PAT and HAP-C-PAT would work well even under the population stratification demographic model and assortative mating demographic model. Furthermore, for the disease models considered, there are significant gains in power from haplotype analysis compared to single-marker analysis, and from combined analysis using HAP-C-PAT compared to analysis using HAP-PAT for the complete family data only. Copyright © 2008 S. Karger AG, Basel

Introduction

Parent-of-origin effects, also known as 'genomic imprinting', play an important role in genetic traits. Recently, considerable research effort has been devoted to detection of parent-of-origin effects. Many single-marker analysis methods were proposed. Based on case-parents

Wing K. Fung Department of Statistics and Actuarial Science The University of Hong Kong, Pokfulam Road Hong Kong (China) Tel. +852 2859 1988, Fax +852 2858 9041, E-Mail wingfung@hku.hk trios, Weinberg et al. [1] introduced a versatile framework of the log-linear model to test for parent-of-origin effects, maternal effects, and linkage disequilibrium (LD). Weinberg [2] developed the parental-asymmetric test (PAT) in the case of no maternal genotype effects and found that PAT is very simple and powerful. The log-linear approach [1] can work well and is robust against population stratification, but it is difficult to extend the approach to general nuclear families or haplotypes [3]. On the other hand, because complex traits are often associated with multiple (interacting) markers [4-9], haplotype-based analysis has gained increasing attention as it could potentially be more efficient than a single-marker-based analysis. As such, Becker et al. [3] proposed an extension of PAT, the HAP-PAT, to test for parent-of-origin effects using multiple tightly linked markers in the case of nuclear families with two genotyped parents and multiple affected children. They found that the relative performance of haplotype analysis versus single-marker analysis depends strongly on the marker/disease-locus LD pattern; the gain in power for haplotype analysis can be significant in most cases.

In practical studies, there may be some families for which only the genotypes of one parent are available, and as such, HAP-PAT is not applicable to such data. Since incorporation of information from these incomplete nuclear families would likely lead to appreciable increase in the statistical power of a test, it would be desirable to extend HAP-PAT to accommodate such families with incomplete data. By incomplete families, we refer to nuclear families with missing data for one, but not both, parents, because nuclear families with data missing for both parents are not informative for testing for parent-of-origin effects [10]. In the literature, some statistical tests have been developed for testing parent-of-origin effects with incomplete family data when a single marker is considered. Rampersaud et al. [10] suggested the combined likelihood ratio test (Combined LRT), an extension of Weinberg et al. [1], for testing for parent-of-origin effects in the presence of missing parental genotypes by incorporating additional information from the genotypes of unaffected siblings to improve inference of missing parental data. Zhou et al. [11], on the other hand, proposed the 1-PAT for incomplete families, and C-PAT for a data set with both complete and incomplete families, and showed that C-PAT is more powerful than Combined LRT.

In this paper, we focus on analyzing multiple tightly linked markers jointly. When only one parent is available for each family with an arbitrary number of affected children, we develop a haplotype-based statistic, HAP-1-PAT, to test for parent-of-origin effects. We further propose HAP-C-PAT, which is amenable with a dataset comprising both complete and incomplete families. We carry out a simulation study to evaluate the validity and power of the test statistics in various settings of missing father rates, incomplete family rates, parent-of-origin effects, haplotype diversity, marker/disease-locus LD patterns and population models. We perform analyses for all possible combinations of the markers being considered. A permutation-based Monte Carlo procedure is devised to determine the significance of the tests; the corrected global p values taking into account of multiple testing are used for inferences. The results show that HAP-1-PAT and HAP-C-PAT would work well even under the population stratification demographic model and assortative mating demographic model. Furthermore, for the disease models considered, there are significant gains in power from haplotype analysis compared to single-marker analysis, and from combined analysis using HAP-C-PAT compared to analysis using HAP-PAT for the complete family data only.

Methods

Background and Notation

We consider *m* tightly linked markers so that the recombination frequency between any pair can be assumed to be zero. A child inherits one copy of each parent's two haplotypes with equal probability, and the maternal haplotype (the copy inherited from the mother) and the paternal haplotype (the copy inherited from the father) are assumed to be inherited independently. We term any nonempty subset *Q* of the marker set {1, 2, ..., *m*} as a 'marker combination' [12]. For a marker combination *Q*, let $\mathbf{H} = \{h_1, h_2, ..., h_n\}$ be the set of all haplotypes (compatible with the observed genotype data) and $\boldsymbol{\Theta} = \{\theta_1, \theta_2, ..., \theta_n\}$ be the corresponding population haplotype frequencies.

Suppose *D* and *d* are the disease and normal alleles at a disease susceptibility locus (DSL). Let *P*_D denote the frequency of disease allele *D*, and let $\phi_{D/D}$, $\phi_{D/d}$, $\phi_{d/D}$, and $\phi_{d/d}$ be the penetrances of genotypes *D*/*D*, *D*/*d*, *d*/*D* and *d*/*d* at the DSL, respectively, where the allele before / is paternal and the allele after / is maternal. The parent-of-origin effects are measured by the degree of imprinting $I = (\phi_{D/d} - \phi_{d/D})/2$, which is half the difference of the two heterozygote penetrances, with I = 0 indicating no parent-of-origin effects. Furthermore, *I* ranges from $(\phi_{d/d} - \phi_{D/D})/2$ (complete paternal parent-of-origin effect) to $(\phi_{D/D} - \phi_{d/d})/2$ (complete maternal parent-of-origin effect) [13]. For convenience, we also represent the genotype relative risks as $\gamma_2 = \phi_{D/D}/\phi_{d/d}$, $\gamma_{1p} = \phi_{D/d}/\phi_{d/d}$, and $\gamma_{1m} = \phi_{d/D}/\phi_{d/d}$ [14].

Like Weinberg [2], Becker et al. [3], and Hu et al. [15], we need to make some assumptions throughout the paper. We assume that there is no maternally-mediated genetic effects, i.e., there is no maternal genotype effects on the phenotype of her children. Mat-

Detecting Imprinting Using Haplotype

ing symmetry in the population is also presumed. When some parental genotypes are missing, we assume that the missingness of a parental genotype is independent of his/her underlying genotype [16]. For incomplete families, we further assume that the haplotype configurations of the underlying complete nuclear family conditional on the observed data have the same distributions irrespective of whether it is the father's, or the mother's genotype that is missing.

Existing Methods

We begin by describing two existing single-marker analysis methods for detecting parent-of-origin effects. Suppose the marker locus of interest has two alleles, M_1 and M_2 . For convenience, let 0, 1 and 2 represent the marker genotypes M_2M_2 , M_1M_2 and M_1M_1 , respectively. We also let *F*, *M* and *C* represent the genotypes of the father, mother and child, respectively, and therefore *F*, *M* and *C* take possible values of 0, 1 or 2. For n_C case-parents trios with known marker genotypes for the father, mother, and affected child, Weinberg [2] proposed the following PAT to test for parent-of-origin effects

PAT =
$$\frac{\left(N_{F>M,C=1} - N_{FM,C=1} + N_{F,$$

where $N_{F>M, C=1}$ ($N_{F<M, C=1}$) is the number of case-parents trios with heterozygous child in which the father carries more (fewer) copies of marker allele M_1 than the mother, i.e., the number of heterozygous children who inherit marker allele M_1 from the father (mother).

For n_I case-parent pairs with known marker genotypes for the available parent and affected child (comprising n_M case-mother pairs and n_F case-father pairs), the 1-PAT statistic was developed by Zhou et al. [11] as follows

$$1 - PAT = \frac{w \left(N_{M < C,C=1} - N_{M > C,C=1} \right) + (1 - w) \left(N_{F > C,C=1} - N_{F < C,C=1} \right)}{\sqrt{\frac{w^2 \left(N_{M < C,C=1} + N_{M > C,C=1} \right) + (1 - w)^2 \left(N_{F > C,C=1} + N_{F < C,C=1} \right)}{\sqrt{\frac{w^2 \left(N_{M < C,C=1} - N_{M > C,C=1} \right) + (1 - w)^2 \left(N_{F > C,C=1} - N_{F < C,C=1} \right)}}}}}$$

where the weight $w = n_F / (n_M + n_F)$; $N_{M < C, C = 1} (N_{M > C, C = 1})$ is the number of heterozygous children with homozygous mother M =0 (M = 2), i.e., the number of heterozygous children whose marker allele M_1 is inherited from the father (mother) in n_M case-mother pairs; $N_{F>C, C=1}$ and $N_{F<C, C=1}$ are respectively the numbers of heterozygous children whose marker allele M_1 is inherited from the father and mother in n_F case-father pairs. It is noted that under the null hypothesis of no parent-of-origin effects the joint genotype distribution of the case-mother pair and that of the case-father pair are the same [15]. Therefore, we have $E[wN_{M < C, C = 1}] = E[(1 - w)N_{F < C, C = 1}]$ and $E[wN_{M > C, C = 1}] =$ $E[(1 - w)N_{F>C, C=1}]$, and then the expectation of $w(N_{M < C, C=1} - w)$ $N_{M>C, C=1}$ + (1 - w)($N_{F>C, C=1}$ - $N_{F<C, C=1}$) is zero under the null hypothesis. Further, $w^2(N_M < C, C = 1 + N_M > C, C = 1) +$ $(1 - w)^2 (N_{F > C, C = 1} + N_{F < C, C = 1}) + (n_M + n_F)^{-1} (N_{M < C, C = 1} - N_{M < C, C = 1})$ $N_{M>C, C=1}$)($N_{F>C, C=1} - N_{F<C, C=1}$) is an unbiased estimator of the variance of $w(N_{M < C, C=1} - N_{M > C, C=1}) + (1 - w)(N_{F > C, C=1} - N_{M > C, C=1})$ $N_{F < C, C = 1}$) under the null hypothesis.

Becker et al. [3] extended PAT to a haplotype-based statistic, HAP-PAT, for data from complete nuclear families. Suppose we have n_C independent complete nuclear families, each with known marker genotypes for the father, mother, and affected offspring. For each $h_i \in \mathbf{H}$, let t_{Ci1} and t_{Ci2} denote the numbers of heterozygous children who inherit haplotype h_i from the father and mother, respectively. Then the HAP-PAT statistic for detecting parent-of-origin effects is

HAP-PAT =
$$\frac{n-1}{n} \sum_{i=1}^{n} \frac{(t_{Ci1} - t_{Ci2})^2}{t_{Ci1} + t_{Ci2}}$$
.

If we only consider one diallelic marker locus with alleles M_1 and M_2 and there is only one affected child in each family, then t_{C11}/t_{C22} and t_{C12}/t_{C21} are simply $N_{F>M, C=1}$ and $N_{F<M, C=1}$ in the expression of the PAT, respectively. So the HAP-PAT is reduced to the PAT. The significance of the observed HAP-PAT statistic is determined via a Monte-Carlo simulation scheme: in each replicate, one either exchange the parental genotypes or leave them unchanged with equal probability within each family.

Method for Incomplete Family Data – Only One Parent Is Available

We first consider the case with complete phase information. Suppose we have n_M incomplete nuclear families with the mothers available and n_F incomplete nuclear families with the fathers available. The total number of incomplete nuclear families is denoted by $n_I = n_M + n_F$. If there are n_j affected children in the *j*-th incomplete family, then we have $n_{CMP} = \sum_{j=1}^{n_M} n_j$ case-mother pairs and $n_{CFP} = \sum_{j=n_M+1}^{n_I} n_j$ case-father pairs. For the n_M incomplete nuclear families with the mothers available, we construct a parent-of-origin contingency table $(t_{Mik})_{n \times 2}$ with rows representing different haplotypes and columns representing parental origin (paternal or maternal). More precisely, for each $i (1 \le i \le n)$, t_{Mi1} is the total number of heterozygous children who inherit haplotype h_i from the father and t_{Mi2} is the total number of heterozygous children who inherit haplotype h_i from the mother. In other words, each heterozygous child contributes two counts to the table. For example, if the mother has phase-known genotype (h_1, h_2) and her heterozygous child has (h_1, h_3) for the marker combination Q, then this child contributes one unit to the ' h_1 row' and 'maternal column' (i.e., t_{M12}) and one unit to the ' h_3 row' and 'paternal column' (i.e., t_{M31}), respectively. For the n_F incomplete nuclear families with the fathers available, we construct another parent-of-origin contingency table (t_{Fik}) , $1 \le i \le n$, $1 \le k \le 2$, where t_{Fi1} and t_{Fi2} are similarly defined for each haplotype h_i .

Using both kinds of incomplete nuclear families (either with missing father or missing mother), we consider the weighted sum $w(t_{Mi1} - t_{Mi2}) + (1 - w)(t_{Fi1} - t_{Fi2})$ as contribution to evidence of parent-of-origin effects for haplotype h_i , where the weight $w = n_{CFP}/(n_{CMP} + n_{CFP})$. An omnibus test (combining evidence from all *n* haplotypes) is then devised as follows:

HAP-1-PAT =
$$\frac{n-1}{n} \sum_{i=1}^{n} \frac{\left[w(t_{Mi1}-t_{Mi2})+(1-w)(t_{Fi1}-t_{Fi2})\right]^2}{w^2(t_{Mi1}+t_{Mi2})+(1-w)^2(t_{Fi1}+t_{Fi2})}$$
. (1)

The asymptotic distribution of HAP-1-PAT is unclear, so we develop a permutation procedure to assess its significance. It is noted that the variance of the square root of the numerator has a cross-product term (see the expression of 1-PAT), but it is being omitted from the denominator of (1) for simplicity, which would not affect the validity of the test because the significance is being assessed by permutation method. Note that, under the null hy-

pothesis of no parent-of-origin effects, the joint genotype distribution of incomplete nuclear families with missing father and that of incomplete nuclear families with missing mother are the same, i.e., the joint genotype distribution of the available parent and his/her children is independent of this parent's sex, which leads to the following permutation-based Monte Carlo procedure to find the p value of HAP-1-PAT. Without loss of generality, we assume that, in a sample with n_I incomplete nuclear families, the first n_M are families with missing fathers while the remaining $n_F = n_I - n_M$ are with missing mothers. We permute the order of the families to obtain a new sample with the same size n_I . The first n_M families in the permuted sample are now relabeled as missing fathers, and the remaining n_F families are treated as having missing mothers. The resulted value from (1) based on the permuted sample is denoted as HAP-1-PAT^{*}₁. Repeating this permutation procedure B times leads to a collection, $\{HAP-1-PAT_1^*, ..., \}$ HAP-1-PAT^{*}_B, of the HAP-1-PAT statistic under the null hypothesis of no parent-of-origin effects for marker combination $Q(H_0^Q)$. From this estimated sampling distribution, we find the (uncorrected) p value of HAP-1-PAT for the observed data as follows

$$p \text{ value} = \frac{\left|\left\{b: 1 \le b \le B, \text{HAP} - 1 - \text{PAT}_b^* \ge \text{HAP} - 1 - \text{PAT}\right\}\right|}{B}$$

where $|\cdot|$ denotes the number of elements of a set.

We now turn our attention to the case in which phased haplotypes cannot always be inferred unambiguously. Let G_i be the observed genotypes of all members in the *j*-th family, $1 \le j \le n_I$, and $\mathbf{G} = \{G_i\}$ be the collection of genotype data from all n_I families. For haplotype $h_i \in \mathbf{H}$, since each *t* of four quantities t_{Mi1}, t_{Mi2} , t_{Fi1} , and t_{Fi2} in HAP-1-PAT cannot be directly calculated, we replace it by $E(t|\mathbf{G}) = \sum_{Z} [t(Z)P(Z|\mathbf{G})]$, where the summation is over all the possible haplotype explanations Z compatible with G. Note that this is one way, but definitely not the only way, for constructing a statistic with incomplete (unphased genotype) data, a frequently employed technique for non-parametric tests, such as the one proposed in (1). The probability $\pi_Z = P(Z|\mathbf{G})$ can be interpreted as a weight to signify the likelihood of haplotype Z given the observed data. It can be expressed in terms of population haplotype frequencies, which are estimated by the expectation-maximization (EM) algorithm [17] in our simulation study. The corresponding significance is also obtained via the permutation procedure as in the phase known case.

We now describe, in more details, how one may calculate the expected contribution of the data to each entry t in the contingency table. First note that since the families in a sample are regarded as independent, the expected contribution to the table from each family can be calculated separately. More specifically, for family j, let s_j be the number of possible haplotype explanations that are compatible with the observed genotypes G_j . For haplotype h_i and the l-th haplotype explanation in the j-th family, denote the total numbers of heterozygous children who inherit haplotype h_i from the father and mother for the l-th haplotype explanation as t_{Mijl1} and t_{Mijl2} , respectively (for $1 \le j \le n_M$), or as t_{Fijl1} and t_{Fijl2} , respectively (for $n_M < j \le n_l$). Then the following four quantities

$$\sum_{j=1}^{n_M} \sum_{l=1}^{s_j} \pi_{jl} t_{Mijl1}, \sum_{j=1}^{n_M} \sum_{l=1}^{s_j} \pi_{jl} t_{Mijl2}, \sum_{j=n_M+1}^{n_l} \sum_{l=1}^{s_j} \pi_{jl} t_{Fijl1}, \sum_{j=n_M+1}^{n_l} \sum_{l=1}^{s_j} \pi_{jl} t_{Fijl2}$$

are used to estimate t_{Mi1} , t_{Mi2} , t_{Fi1} , and t_{Fi2} , respectively. The weight $\pi_{jl} = P(Z_l|G_j)$ is as described above, but written more explicitly in the subscript here, and there is no need to distinguish between the two kinds of missing families. It can be found by computing the probabilities of the *l*-th haplotype configuration, $p_{jl}(\hat{\Theta})$ ($1 \le l \le s_j$) and then normalizing them, where $\hat{\Theta}$ is the estimated haplotype frequencies (from the EM algorithm [17]) of $\hat{\Theta} = \{\theta_1, ..., \theta_n\}$. As an illustration, we offer table 1 to show all the possible values of $p_{jl}(\hat{\Theta})$ (last column) for a family with two affected children.

Multiple Testing

In equation (1), we propose a statistic to test for parent-oforigin effects for a given marker combination Q. However, prior to any analysis it is usually unknown as to which marker combination provides the greatest power for testing for parent-of-origin effects. For each marker combination Q (out of a total of $2^m - 1$), let H^Q_0 be the null hypothesis, as defined before, and let HAP-1-PAT^Q be the corresponding test statistic associated with it. This amounts to a fairly complicated multiple testing problem, as all the test statistics are correlated. Therefore, as in Becker et al. and Becker and Knapp [3, 12], we consider $H_0 = \bigcap_Q H_0^Q$, which corresponds to the global null hypothesis that there is no parentof-origin effects for any marker combination. As argued in the above references, a small $p^{\min} = \min_Q p_0^Q$ would provide evidence against the null hypothesis, where p_0^Q is the (uncorrected) p value of the test statistic HAP-1-PAT^Q for the null hypothesis H_0^Q . To assess whether p^{\min} is sufficiently small to warrant the rejection of H_0 , one needs to compare p^{\min} to its sampling distribution under H_0 . This can be accomplished by recycling the permuted samples (data generated under H_0 to estimate p_0^Q) without much further computational burden. Furthermore, it has been shown that such a procedure for multiplicity adjustment can lead to considerable power gains [12].

Method for a Mixture of Complete and Incomplete Families

Now suppose we have a mixture of n_C complete families and n_I incomplete families (with n_M missing fathers and n_F missing mothers). We propose the following combined statistic to test for parent-of-origin effects:

HAP - C - PAT =

$$\frac{n-1}{n} \sum_{i=1}^{n} \frac{\left[(t_{Ci1} - t_{Ci2}) + w(t_{Mi1} - t_{Mi2}) + (1-w)(t_{Fi1} - t_{Fi2}) \right]^{2}}{(t_{Ci1} + t_{Ci2}) + w^{2}(t_{Mi1} + t_{Mi2}) + (1-w)^{2}(t_{Fi1} + t_{Fi2})}.$$
 (2)

It is noted from the combined statistic that the contribution of families with both parents and that of families with only one parent are the same [16]. However, different weights for the complete and incomplete families (say λ and $(1 - \lambda)$, respectively) can be entertained if so desire. For example, λ can be set to be a function of the relative sample sizes and the relative informativeness of these two types of families. Significance of HAP-C-PAT is determined via the Monte Carlo procedures as described above. Note that we jointly employ the estimation methods for complete nuclear families [18] and for incomplete nuclear families [17] to estimate haplotype frequencies for the combined data.

Detecting Imprinting Using Haplotype

	F/M ^a	C_1^{a}	C_2^{a}	Joint probability
1	$h_i h_i$	$h_i h_i$	$h_i h_i$	$0.5 \theta_i^3 (1+\theta_i)$
2	$h_i h_i$	$h_i h_i$	$h_i h_j \ (j \neq i)$	$0.5 \theta_i^3 \theta_j$
3	$h_i h_i$	$h_i h_i \ (j \neq i)$	$h_i h_j$	$0.5 \theta_i^2 \theta_j (1+\theta_j)$
4	$h_i h_i$	$h_i h_i \ (j \neq i)$	$h_i h_k \ (k \neq i, j)$	$0.5 \ \theta_i^2 \theta_j \theta_k$
5	$h_i h_i \ (j \neq i)$	$h_i h_i$	$h_i h_i$	$0.25 \theta_i^2 \theta_j (1+\theta_i)$
6	$h_i h_i \ (j \neq i)$	$h_j h_j$	$h_j h_j$	$0.25 \theta_i \theta_j^2 (1+\theta_j)$
7	$h_i h_i \ (j \neq i)$	$h_i h_i$	$h_j h_j$	$0.25 \ \theta_i^2 \theta_j^2$
8	$h_i h_i \ (j \neq i)$	$h_i h_i$	$h_i h_j$	$0.25 \theta_i^2 \theta_j (1 + \theta_i + \theta_j)$
9	$h_i h_j \ (j \neq i)$	$h_j h_j$	$h_i h_j$	$0.25 \theta_i \theta_j^2 (1 + \theta_i + \theta_j)$
10	$h_i h_i \ (j \neq i)$	$h_i h_i$	$h_i h_k \ (k \neq i, j)$	$0.25 \ \theta_i^2 \theta_j \theta_k$
11	$h_i h_j \ (j \neq i)$	$h_i h_i$	$h_i h_k \ (k \neq i, j)$	$0.25 \ \theta_i^2 \theta_j \theta_k$
12	$h_i h_i \ (j \neq i)$	$h_j h_j$	$h_i h_k \ (k \neq i, j)$	$0.25 \ \theta_i \theta_j^2 \theta_k$
13	$h_i h_j \ (j \neq i)$	$h_j h_j$	$h_i h_k (k \neq i, j)$	$0.25 \ \theta_i \theta_j^2 \theta_k$
14	$h_i h_j \ (j \neq i)$	$h_i h_j$	$h_i h_j$	$0.25 \theta_i \theta_j (\theta_i + \theta_j)(1 + \theta_i + \theta_j)$
15	$h_i h_j \ (j \neq i)$	$h_i h_j$	$h_i h_k \ (k \neq i, j)$	0.25 $\theta_i \theta_j \theta_k (\theta_i + \theta_j)$
16	$h_i h_j \ (j \neq i)$	$h_i h_j$	$h_i h_k \ (k \neq i, j)$	0.25 $\theta_i \theta_j \theta_k (\theta_i + \theta_j)$
17	$h_i h_j \ (j \neq i)$	$h_i h_k (k \neq i, j)$	$h_i h_k$	0.25 $\theta_i \theta_j \theta_k (1 + \theta_k)$
18	$h_i h_j \ (j \neq i)$	$h_i h_k \ (k \neq i, j)$	$h_j h_k$	$0.25 \theta_i \theta_j \theta_k (1 + \theta_k)$
19	$h_i h_j \ (j \neq i)$	$h_i h_k \ (k \neq i, j)$	$h_j h_k$	0.25 $\theta_i \theta_j \theta_k (1 + \theta_k)$
20	$h_i h_j \ (j \neq i)$	$h_i h_k \ (k \neq i, j)$	$h_i h_l \ (l \neq i, j, k)$	$0.25 \theta_i \theta_j \theta_k \theta_l$
21	$h_i h_j \ (j \neq i)$	$h_i h_k (k \neq i, j)$	$h_i h_l (l \neq i, j, k)$	$0.25 \theta_i \theta_j \theta_k \theta_l$
22	$h_i h_j \ (j \neq i)$	$h_i h_k \ (k \neq i, j)$	$h_j h_l (l \neq i, j, k)$	$0.25 \ \theta_i \theta_j \theta_k \theta_l$
	-			

Table 1. Classification of all 22 family types for incomplete nuclear families each with father/mother and two children, together with the joint probabilities of phase-known genotypes of each family type

^a F, M, C_1 , and C_2 denote the phase-known genotypes of the father, mother and two children, respectively.

Simulation Study

A simulation study is conducted to assess the performance of the proposed test under various settings of parameter values. In the following, we will discuss, in turn, the settings for parent-of-origin effect and genotype relative risks, missing father and incomplete family rates, marker/disease-locus LD patterns and haplotype diversity scenarios in the single nucleotide polymorphism (SNP) markers, and population models. In all the simulations, the total number of families (complete or incomplete) is fixed at 200, and each family is assumed to have two affected offspring. The nominal significance level is set at 5%.

Settings

PEM Models

The relative risk with two copies of the disease allele, γ_2 , is fixed at 2. Because of symmetry of paternal and maternal parent-of-origin effects under the assumption of

no maternal effects, it suffices to consider the following three models that represent various scenarios of imprinting degree: no parent-of-origin effects (PEM0: $\gamma_{1p} = \gamma_{1m} = 1.5$), incomplete maternal parent-of-origin effect (PEM1: $\gamma_{1p} = 2$, $\gamma_{1m} = 1.5$), and complete maternal parent-of-origin effect (PEM2: $\gamma_{1p} = 2$, $\gamma_{1m} = 1$). The disease allele frequency P_D is set to be either 0.25 or 0.5, depending on the population model considered.

Missing Rates

For convenience, a parameter $\beta = P$ (the missing parent is father one parent is missing in a family) is introduced, which means that for each one-parent family the probability that the father is missing is β and the probability that the mother is missing is $1-\beta$. It is termed missing father rate (among incomplete families) hereafter for ease of reference. The missing father rate β ranges from 0.2 to 0.8 in increments of 0.1 in the first simulation and is fixed at 0.5 in all the remaining simulations in this section. The incomplete family rate τ is defined as the ratio of the number of incomplete families to that of all families. The incomplete family rate ranges within the interval 0–1. Note that HAP-C-PAT is reduced to HAP-PAT when the incomplete family rate τ is 0 and to HAP-1-PAT when τ is 1.

LD Patterns and Haplotype Diversity (HD) Scenarios Our HD structure and LD patterns over a 5-SNPs system follow that in Becker et al. [3] to facilitate comparison with HAP-PAT. The middle three markers (SNPs 2-4) are regarded as the core SNPs while the two end markers (SNPs 1 and 5) are interpreted as markers on the border of the LD region. Three HD scenarios for the three core SNPs are given as A, B, and C in table 2. As explained [3], scenario A depicts a setting without evidence for historical recombinations, whereas scenario B cannot be explained without at least one historical recombination event. Under scenario C, all haplotypes at the core region exist. Four marker/disease-locus LD patterns for the core SNPs (2, 3 and 4) are considered: LD1: perfect LD with SNP 3; LD2: perfect LD with haplotype 111; LD3: incomplete LD with SNP 3; LD4: incomplete LD with haplotype 111. For the two border SNPs of the LD region, the allele at SNP 1 (or SNP 5) is allele 1 with probability 0.65 and 2 with 0.35 if the DSL carries the disease allele D on the haplotype. If, on the other hand, the DSL carries the normal allele d, then the allele at SNP 1 (SNP 5) being allele 1 is with probability 0.35 and 2 with 0.65.

Population Models

Three population models are considered: homogeneous population model (HPM), population stratification demographic model (PSM) and assortative mating demographic model (AMM) [11, 19]. In the homogeneous population model, we consider 12 combinations of three parameters, (P_D , LD, HD), with $P_D = 0.25$, LD = LD1 – LD4, and HD = A - C. The population stratification demographic model consists of two very different homogeneous subpopulations. It is assumed that the families are ascertained from the first subpopulation with probability 0.6 and from the second one with 0.4. In the first subpopulation, we set $(P_D, LD, HD) = (0.25, LD1, A)$, whereas in the second subpopulation, $(P_D, LD, HD) = (0.5, LD4,$ C). Finally, in the assortative mating demographic model, we set $(P_D, LD, HD) = (0.25, LD2, B)$. We assume that 70% of the families were generated through random mating and 30% of the families were generated through assortative mating where the father and mother have the same affection status.

Table 2. Haplotype diversity scenarios for SNPs 2, 3 and 4^a

A	В	С	
111	111	111	
112	112	112	
121	121	121	
		122	
211	211	211	
		212	
		221	
	222	222	
^a Ado	pted from table	2 of [3].	

To generate data under each of the three population models, we first simulate the father's and mother's haplotypes at the 5 SNP markers and a DSL based on the specified setting of (P_D , LD, HD). Then, the haplotypes of the children are generated from the parents' haplotypes assuming no recombination. The affection status of the parents and their children are assigned according to their genotypes at the DSL and the corresponding four risks $\phi_{D/D}$, $\phi_{D/d}$, $\phi_{d/D}$ and $\phi_{d/d}$. The simulated haplotype data are converted to the multiple-locus genotype data prior to analysis. From these genotype data, we begin by estimating the haplotype frequencies, inferring the phase and then calculating the statistic HAP-C-PAT.

We assess the actual type I error rate and power of HAP-C-PAT based on 1,000 replicates. For each replicate, the global *p* values are determined on the basis of 1,000 Monte Carlo samples as described in the Methods section. The actual type I error rates/powers are estimated as the proportions of simulated data sets of rejecting the null hypothesis of no parent-of-origin effects at significance level 5% when the null/alternative hypothesis holds.

Sizes and Powers of HAP-1-PAT

We begin by exploring the effect of various missing father rates (or β values) on the performance of HAP-1-PAT when the sum $n_M + n_F$ is fixed at 200 and the population model is assumed to be HPM. Simulations under the null hypothesis of no parent-of- origin effects (PEM0) yield the simulated type I error rates of HAP-1-PAT. Table 3 shows that the actual sizes of HAP-1-PAT based on both single-marker analysis and haplotype analysis are generally quite close to the nominal 5% level, with almost all within two standard errors (0.014 = $2\sqrt{0.05 \times 0.95/1000}$) from 0.05.

Detecting Imprinting Using Haplotype



Fig. 1. Powers of HAP-1-PAT for singlemarker analysis (SA) and haplotype analysis (HA) against missing father rate β under homogeneous population model (HPM), two parent-of-origin effects models, PEM1 and PEM2, marker/diseaselocus LD pattern 2 and three haplotype diversity (HD) scenarios, HDA-HDC. **a** HDA and SA. **b** HDA and HA. **c** HDB and SA. **d** HDB and HA. **e** HDC and SA. **f** HDC and HA.

Table 3. Type I error rates (in %) of HAP-1-PAT under the homogeneous population model HPM and no parent-of-origin effects mod-el PEM0

Haplotype	Analysis and missing father rate β													
diversity scenario	Single-marker analysis						Haple	Haplotype analysis						
	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.2	0.3	0.4	0.5	0.6	0.7	0.8
Marker/disease	e-locus LD	pattern	1											
А	5.0	3.6	6.2	4.5	5.7	5.3	5.5	3.9	3.8	5.9	5.1	4.6	5.5	5.5
В	4.1	5.4	5.6	4.6	4.3	5.2	5.1	3.8	5.1	4.4	4.9	4.4	4.9	4.1
С	5.2	5.0	4.1	3.7	5.1	5.7	5.0	3.7	4.7	4.3	4.1	5.3	4.2	4.7
Marker/disease	-locus LD	pattern	2											
А	4.8	4.5	4.1	5.4	5.6	4.6	4.8	5.7	4.0	4.5	5.4	5.0	4.5	4.5
В	5.6	5.7	4.5	5.2	6.1	4.9	5.3	4.5	4.8	5.0	4.9	4.5	4.4	4.5
С	5.1	4.7	3.7	6.6	4.2	4.2	4.9	4.2	4.7	4.6	3.8	4.4	5.3	3.5
Marker/disease	-locus LD	pattern	3											
А	5.4	6.1	5.2	4.5	4.4	5.3	5.1	4.7	4.5	3.5	4.7	3.8	5.3	3.4
В	4.2	4.5	5.1	5.7	4.3	5.7	4.0	4.5	4.0	4.3	4.3	3.5	4.7	3.6
С	3.6	5.5	5.1	4.6	5.2	4.0	5.3	3.2	5.0	4.7	3.3	3.5	4.4	4.8
Marker/disease	-locus LD	pattern	4											
А	5.9	4.9	5.0	4.8	6.5	4.6	5.4	4.6	4.6	5.4	4.5	4.6	4.1	4.6
В	4.8	5.7	6.6	5.6	5.4	5.2	5.8	4.6	4.2	4.7	4.8	5.1	3.6	5.7
С	4.3	5.9	5.7	5.9	4.2	4.4	3.5	4.1	5.2	5.2	4.5	4.2	5.0	4.2



Fig. 2. Powers of HAP-C-PAT and HAP-PAT for SA and HA against incomplete family rate τunder HPM, PEM1 and PEM2, LD2 and three HD scenarios, HDA-HDC. **a** HDA and SA. **b** HDA and HA. **c** HDB and SA. **d** HDB and HA. **e** HDC and SA. **f** HDC and HA. The hollow and filled squares represent the powers of HAP-C-PAT under PEM1 and PEM2, respectively. The hollow and filled triangles represent the powers of HAP-PAT excluding incomplete families under PEM1 and PEM2, respectively.

Simulations under the alternative hypothesis of parent-of-origin effects (PEM1 and PEM2) are performed to compare powers of haplotype-based analysis versus those of single-marker analysis. Shown in figure 1 are the results for three of the 12 (P_D , LD, HD) combinations, when P_D and LD are fixed to be 0.25 and LD2 (haplotype association), respectively, while HD varies from A to C. As can be seen from the figure, haplotype analysis generally has more power regardless of the PEM models, HD scenarios, or β values. However, the gain in power is more profound for PEM2 (complete imprinting), as expected, and for HD = A (no historical recombinations). The latter observation is not surprising either, because under HD scenario A, the LD between markers is the strongest, leading to a greater power gain for haplotype analysis, which is consistent with results in Becker et al. [3]. The observed general patterns hold for the other nine combinations of (P_D, LD, HD) , although the powers for haplotype and single-marker analyses are comparable for LD1 and LD3. We defer further discussion on this to the next segment. As expected, it is observed from the figure that the effect of β on the power of HAP-1-PAT is non-negligible, with the maximum power occurring when the missing rates are the same for father and mother (i.e., $\beta = 0.5$). For simplicity, we fix $\beta = 0.5$ for the remaining simulations.

Sizes and Powers of HAP-C-PAT

We consider three incomplete family rates τ : 100, 50, 0%, with the total number of families fixed at 200. Again, all 12 combinations of (P_D , LD, HD) are considered under population model HPM. Under PEM0, we assess the type I error rate of HAP-C-PAT. The results in table 4 show that almost all of the actual type I error rates are within two standard errors of the nominal 0.05, signifying the validity of the test. We further investigate the type I error rates under the two other population models, PSM and AMM, and once again the type I error rates are all reasonably close to the nominal (rows PEM0 of table 5).

For power comparison, PEM1 and PEM2 are used. We analyze the data using both HAP-C-PAT (taking the entire sample, complete and incomplete families, into consideration) and HAP-PAT (using only the complete families, i.e., excluding the incomplete families). Figures 2 and

Detecting Imprinting Using Haplotype

LD1, SA LD1, HA 100 100 Power (%) Power (%) 50 50 0 0 b а 0 25 50 75 100 0 25 50 75 100 LD2, SA LD2, HA 100 100 Power (%) Power (%) 50 50 0 0 d 75 С 25 50 100 0 25 50 75 100 0 LD3, SA LD3, HA 80 80 Power (%) Power (%) 40 40 0 0 f 100 100 25 50 50 75 75 25 e 0 0 LD4, SA LD4, HA 70 70 Power (%) Power (%) 35 35 0 0 h 100 25 g 0 25 50 75 0 50 75 100 τ

Fig. 3. Powers of HAP-C-PAT and HAP-PAT for SA and HA against τ under HPM, PEM1 and PEM2, HDA, and four LD patterns, LD1-LD4. **a** LD1 and SA. **b** LD1 and HA. **c** LD2 and SA. **d** LD2 and HA. **e** LD3 and SA. **f** LD3 and HA. **g** LD4 and SA. **h** LD4 and HA. The hollow and filled squares represent the powers of HAP-C-PAT under PEM1 and PEM2, respectively. The hollow and filled triangles represent the powers of HAP-PAT excluding incomplete families under PEM1 and PEM2, respectively.

Table 4. Type I error rates (in %) of HAP-C-PAT under the homogeneous population model HPM and no parent-of-origin effects model PEM0

Haplotype	Analysis and incomplete family rate $ au$								
diversity scenario	Single-	marker	analysis	Haplot	Haplotype analysis				
	100%	50%	0%	100%	50%	0%			
Marker/disea	se-locus I	D patter	rn 1						
А	4.5	4.8	4.0	5.1	4.5	4.9			
В	4.6	4.7	4.2	4.9	5.1	4.4			
С	3.7	3.9	4.4	4.1	3.9	4.6			
Marker/disea	se-locus I	D patter	rn 2						
А	5.4	5.8	5.8	5.4	4.8	5.9			
В	5.2	5.4	5.2	4.9	4.7	4.9			
С	6.6	4.5	4.6	3.8	4.8	4.4			
Marker/disea	se-locus I	D patter	rn 3						
А	4.5	4.4	4.4	4.7	4.4	5.0			
В	5.7	5.5	4.9	4.3	3.8	4.9			
С	4.6	6.4	5.4	3.3	6.2	5.8			
Marker/disea	se-locus I	D patter	rn 4						
А	4.8	3.7	5.5	4.5	3.9	5.0			
В	5.6	5.3	4.5	4.8	4.6	4.5			
С	5.9	3.9	4.7	4.5	5.1	5.0			

3 show the results when the family samples are generated under population model HPM: actual power for either single-marker analysis or haplotype analysis is plotted against incomplete family rate. Figure 2 focuses on the same three (P_D , LD, HD) combinations as in figure 1 (only varying HD), while figure 3 concentrates on the four combinations with varying LD patterns. It is observed from these figures that the power for detecting parent-of-origin effects under PEM2 is higher than under PEM1 for each corresponding setting, consistent with the observations made from figure 1. In general, HAP-C-PAT, which uses data from all families, is more powerful than its counterpart HAP-PAT, which only uses data from complete families. By comparing the results in figure 3 for different marker/disease-locus LD patterns, we find that the powers of the tests would reduce when the degree of LD decreases and the other parameters remain the same. We can also see that the relative performance of single-marker analysis and haplotype analysis depends strongly on the LD pattern. For marker/disease-locus LD patterns 1 and 3 (the DSL is in LD only with single marker SNP 3), the single-marker analysis has almost the same power as

Table 5. Type I error rates/powers (in %) of HAP-C-PAT

Parent-	Analys	Analysis and incomplete family rate $ au$								
of-origin effects	Single-	marker	analysis	Haplot	Haplotype analysis					
model	100%	50%	0%	0% 100%		0%				
Population stratification demographic model										
PEM0	5.60	4.60	3.70	5.00	3.80	4.80				
PEM1	16.2	20.3	25.6	16.0	19.7	24.6				
PEM2	71.8	78.7	86.4	73.2	80.0	88.7				
Assortative mating demographic model										
PEM0	4.90	5.60	3.50	3.70	3.80	4.00				
PEM1	20.4	20.9	30.8	28.0	29.0	42.3				
PEM2	73.9	77.6	87.6	95.8	95.7	99.4				

the haplotype analysis (fig. 3a vs. b and e vs. f). While for marker/disease-locus LD patterns 2 and 4 (the DSL is in LD with haplotype 111), we observe obviously higher powers with haplotype analysis than with single-marker analysis (fig. 3c vs. d and g vs. h).

Similar conclusions can be drawn based on the other two populations models, PSM and AMM (table 5 rows PEM1 and PEM2). More specifically, since the data generated under PSM is with LD1 (single marker/diseaselocus complete LD) in the subpopulation with a larger proportion, it is not surprising to see that the powers with single-marker analysis are almost the same as with haplotype analysis. On the other hand, under AMM, the population is with LD2 (complete haplotype/disease association), and as such, the powers with haplotype analysis are obviously higher than with single-marker analysis.

Discussion

There is an increasing consensus among human geneticists that it would be feasible and economical to incorporate missing data into tests for parent-of-origin effects. For late-onset diseases, sibships with only one parent or no parents available, are common. Because one cannot study parent-of-origin effects when both parents are missing [10], in the current paper, we develop the HAP-1-PAT to test for parent-of-origin effects for incomplete nuclear families with only one parent available in the framework of haplotype analysis. We further propose the HAP-C-PAT to incorporate data from both complete and incomplete families, which is more powerful than either HAP-PAT or HAP-1-PAT alone by making full usage of genetic information. One alternative, and potentially more efficient statistic is the numerator of HAP-C-PAT alone, which we call HAP-C-PAT-NUM. In most of the cases investigated in this paper, we find that HAP-C-PAT-NUM is more powerful than HAP-C-PAT for haplotype analysis, while HAP-C-PAT-NUM is slightly less powerful than HAP-C-PAT in most of the cases for single-marker analysis. However, when the incomplete family rate $\tau = 100\%$, the average difference in power between HAP-C-PAT-NUM and HAP-C-PAT is minuscule (less than 2%).

Computational consideration is another issue worth commenting. Due to the Monte Carlo procedure, the estimation of haplotype frequencies, and the calculations of global *p* values, there may be a significant demand in the running time and storage. However, practices for reducing the computational demand can be implemented. For example, for a marker combination that does not comprise all the markers, the corresponding haplotype frequencies can be deduced from the haplotype frequencies of all markers. That is, one may first estimate the haplotype frequencies for all *m* markers by the EM algorithm. Then the haplotype frequencies of all marker combinations comprising less than *m* markers follow as appropriate marginal distributions. Furthermore, in the EM algorithm, the incomplete nuclear families with missing father and those with missing mother need not

It is observed from the results in tables 3-5 that HAP-PAT and HAP-C-PAT are valid tests when the missingness of parental genotype data is independent of parental affection status, even though the probability of missingness may be sex dependent. However, affected children inherit the disease allele more likely from a parent who is affected. To this end, we investigate how sex and affection status dependent willingness of parents to participate in a study may affect the performance of these two tests. Two missing scenarios are considered [11]: (a) the whole family tends to be missing when the affected parent is less willing to participate in the study, where the willingness is sex dependent, and (b) only parents less willing to participate in the study tend to be missing and such missingness is both sex and affection status dependent. Under the population stratification demographic model, we evaluate the actual sizes of HAP-PAT (only using the complete families) and HAP-C-PAT (using both the complete and incomplete families) under both missing scenarios. Our findings (results not shown) indicate that both HAP-C-PAT and HAP-PAT control the size well for different affection status- and sex-dependent preference scenarios, even when the affection status dependency is extreme.

Hum Hered 2009;67:116-127

be distinguished, and hence the algorithm can be greatly simplified. To reduce storage, each haplotype is saved as an integer number, rather than a vector of alleles. Finally, in calculating the global *p* values, we merely reassign the sexes of the parents but leave their genotypes unchanged for each permutation sample. This implies that we need not re-employ the EM algorithm to estimate the haplotype frequencies nor enumerate the haplotypes. In calculating the value of HAP-C-PAT for each permutation sample, we just exchange each heterozygous child's contribution to 'paternal column' and 'maternal column' for those families in which the parents' sexes are reversed, which makes the implementation simple and fast.

Throughout this paper, we assume that genotypes at all marker loci are available for all individuals with observed data. However, in real data analyses, it is not uncommon that for some individuals in the sample the genotype data at one or more of the studied marker loci are missing. To deal with this problem, Ding et al. [17] suggested to impute the allele for an individual from the observed alleles in his/her parent or child at the same locus. The feasibility of adopting this method for HAP-C-PAT and its potential effect on the performance will be investigated in a future study.

Generally, haplotype analysis has been shown to have distinct advantages over single-marker analysis in genetic study of common diseases. In this paper, we also assess the statistical performance of single-marker analysis and haplotype analysis, which appears to be strongly dependent on the marker/disease-locus LD pattern, similar to the results in Becker et al. [3]. Prior to analysis it is usually unknown whether the DSL is only in LD with a single marker or in LD with a multiple-marker haplotype, so in general the haplotype analysis for all marker combinations is recommended, especially when the causative variants are not investigated directly or when there are multiple disease-causing alleles or even interactions between them. However, if it is certain from other studies that the disease locus is in LD with a single marker, then single-marker analysis is preferable because of its equally good performance as haplotype analysis and of its ease in dealing with this situation.

In this paper, we assume that Hardy-Weinberg equilibrium (HWE) holds in estimating haplotype frequencies. Fallin and Schork [20] demonstrated that the EM algorithm is reasonably robust to departure from HWE, and there is no noticeable increase in errors even with extreme departure from HWE toward excess homozygosity. For the population stratification demographic model and assortative mating demographic model, we also find that the type I error rates of HAP-C-PAT stay close to the nominal even when HWE is violated.

Acknowledgments

We would like to thank an associate editor and two reviewers for their insightful and helpful suggestions which greatly improved our presentation. This work was partially supported by a Hong Kong RGC CERG Research Grant (HKU 702207P), the National Natural Science Foundation of China (10561008), and the National Institute of Health grant 5R01HG002657.

Software

A software for HAP-C-PAT is available. Please contact Wing K. Fung via e-mail at wingfung@hku.hk.

- 1 Weinberg CR, Wilcox AJ, Lie RT: A log-linear approach to case-parent-triad data: assessing effects of disease genes that act either directly or through maternal effects and that may be subject to parental imprinting. Am J Hum Genet 1998;62:969–978.
- 2 Weinberg CR: Methods for detection of parent-of-origin effects in genetic studies of case-parents triads. Am J Hum Genet 1999; 65:229-235.
- 3 Becker T, Baur MP, Knapp M: Detection of parent-of-origin effects in nuclear families using haplotype analysis. Hum Hered 2006; 62:64–76.
- 4 Dawson E, Abecasis GR, Bumpstead S, Chen Y, Hunt S, Beare DM, Pabial J, Dibling T, Tinsley E, Kirby S, Carter D, Papaspyridonos M, Livingstone S, Ganske R, Lõhmussaar E, Zernant J, Tõnisson N, Remm M, Mägi R, Puurand T, Vilo J, Kurg A, Rice K, Deloukas P, Mott R, Metspalu A, Bentley DR, Cardon LR, Dunham I: A first generation linkage disequilibrium map of human chromosome 22. Nature 2002;418:544–548.
- 5 Zhao H, Zhang S, Merikangas KR, Trixler M, Wildenauer DB, Sun F, Kidd KK: Transmission/disequilibrium tests using multiple tightly linked markers. Am J Hum Genet 2000;67:936–946.
- 6 Niu T, Qin Z, Xu X, Liu JS: Bayesian haplotype inference for multiple linked single- nucleotide polymorphisms. Am J Hum Genet 2002;70:157–169.
- 7 Yu Z, Schaid DJ: Sequential haplotype scan methods for association analysis. Genet Epidemiol 2007;31:553–564.
- 8 Huang BE, Amos CI, Lin DY: Detecting haplotype effects in genomewide association studies. Genet Epidemiol 2007;31:803–812.
- 9 Liu PY, Lu Y, Deng HW: Accurate haplotype inference for multiple linked single-nucleotide polymorphisms using sibship data. Genetics 2006;174:499-509.

- 10 Rampersaud E, Morris RW, Weinberg CR, Speer MC, Martin ER: Power calculations for likelihood ratio tests for offspring genotype risks, maternal effects, and parent-oforigin (POO) effects in the presence of missing parental genotypes when unaffected siblings are available. Genet Epidemiol 2007; 31:18–30.
- 11 Zhou JY, Hu YQ, Lin S, Fung WK: Detection of parent-of-origin effects based on complete and incomplete nuclear families with multiple affected children. Hum Hered 2008 (in press).
- 12 Becker T, Knapp M: A powerful strategy to account for multiple testing in the context of haplotype analysis. Am J Hum Genet 2004; 75:561–570.
- 13 Strauch K, Fimmers R, Kurz T, Deichmann KA, Wienker TF, Baur MP: Parametric and nonparametric multipoint linkage analysis with imprinting and two-locus-trait models: application to mite sensitization. Am J Hum Genet 2000;66:1945–1957.
- 14 Risch N, Merikangas K: The future of genetic studies of complex human diseases. Science 1996;273:1516–1517.
- 15 Hu YQ, Zhou JY, Sun F, Fung WK: The transmission disequilibrium test and imprinting effects test based on case-parent pairs. Genet Epidemiol 2007;31:273–287.
- 16 Allen AS, Rathouz PJ, Satten GA: Informative missingness in genetic association studies: case-parent designs. Am J Hum Genet 2003;72:671–680.
- 17 Ding X, Zhang Q, Flury C, Simianer H: Haplotype reconstruction and estimation of haplotype frequencies from nuclear families with only one parent available. Hum Hered 2006;62:12–19.
- 18 Becker T, Knapp M: Maximum-likelihood estimation of haplotype frequencies in nuclear families. Genet Epidemiol 2004;27:21– 32.
- 19 Sun FZ, Flanders WD, Yang QH, Zhao HY: Transmission/disequilibrium tests for quantitative traits. Ann Hum Genet 2000;64:555– 565.
- 20 Fallin D, Schork NJ: Accuracy of haplotype frequency estimation for biallelic loci, via the expectation-maximization algorithm for unphased diploid genotype data. Am J Hum Genet 2000;67:947–959.