

## ORIGINAL ARTICLE

# A powerful association test for qualitative traits incorporating imprinting effects using general pedigree data

Ji-Yuan Zhou<sup>1</sup>, Hai-Qiang He<sup>1</sup>, Xiao-Ping You<sup>1</sup>, Shao-Zhan Li<sup>1</sup>, Ping-Yan Chen<sup>1</sup> and Wing Kam Fung<sup>2</sup>

For qualitative traits and diallelic marker loci, the pedigree disequilibrium test (PDT) based on general pedigrees and its extension (Monte Carlo PDT (MCPDT)) for dealing with missing genotypes are simple and powerful tests for association. There is an increasing interest of incorporating imprinting into association analysis. However, PDT and MCPDT do not take account of the information on imprinting effects in the analysis, which may reduce their test powers when the effects are present. On the other hand, the transmission disequilibrium test with imprinting (TDTI\*) combines imprinting into the mapping of association variants. However, TDTI\* only accommodates two-generation nuclear families and thus is not suitable for extended pedigrees. In this article, we first extend PDT to incorporate imprinting and propose PDTI for complete pedigrees (no missing genotypes). To fully utilize pedigrees with missing genotypes, we further develop the Monte Carlo PDTI (MCPDTI) statistic based on Monte Carlo sampling and estimation. Both PDTI and MCPDTI are derived in a two-stage framework. Simulation study shows that PDTI and MCPDTI control the size well under the null hypothesis of no association and are more powerful than PDT and TDTI\* (based on a sample of nuclear families randomly selecting from pedigrees) when imprinting effects exist.

*Journal of Human Genetics* (2015) 60, 77–83; doi:10.1038/jhg.2014.109; published online 18 December 2014

## INTRODUCTION

Imprinting, also known as parent-of-origin effects, is a genetic phenomenon that a certain gene is differentially expressed between paternal and maternal alleles, which is an important epigenetic factor in the study of complex traits.<sup>1,2</sup> Currently, some researches have demonstrated that some complex diseases are related to imprinting effects, such as Beckwith–Wiedemann, Prader–Willi and Angelman syndromes.<sup>3–5</sup> Morison *et al.*<sup>6</sup> constructed an imprinted gene database to collect genes that show imprinting effects (<http://igc.otago.ac.nz>). For diallelic marker loci, Zhou *et al.*<sup>7</sup> proposed the pedigree parental-asymmetry test (PPAT) and Monte Carlo (MC) PPAT (MCPAT) based on general pedigree data to test for parent-of-origin effects. PPAT is generally suitable for complete pedigrees (without any missing genotypes), while MCPAT can deal with pedigree data with missing genotypes (incomplete pedigrees) based on MC sampling and estimation procedures, which can recapture most of the missing genotypes based on the observed genotypes in the pedigrees. On the other hand, in the literature, some parametric approaches have been developed to test for imprinting. Yang and Lin<sup>8,9</sup> proposed two likelihood methods for simultaneously detecting imprinting and maternally mediated effects in pedigrees. Lin *et al.*<sup>10</sup> and Ainsworth

*et al.*<sup>11</sup> suggested two multi-locus likelihood approaches for assessing parent-of-origin effects by using mothers and their offspring.

Currently, there has been an increasing interest of incorporating imprinting into association analysis. For a diallelic marker locus, Xia *et al.*<sup>12</sup> developed the transmission disequilibrium test with imprinting (TDTI\*) for qualitative traits, which takes imprinting effects into account. Xia *et al.*<sup>13</sup> further extended TDTI\* to tackle quantitative traits and proposed the Q-C-TDTI test statistic. However, both TDTI\* and Q-C-TDTI only accommodate two-generation nuclear families and thus are not suitable for extended pedigrees. On the other hand, the pedigree disequilibrium test (PDT) based on general pedigree data and its extension MCPDT based on MC sampling and estimation for dealing with missing genotypes are simple and powerful tests for association in the presence of linkage.<sup>14,15</sup> However, PDT and MCPDT do not take into consideration the information on imprinting effects in the analysis, which may reduce their test powers when the effects are present.

In this article, we first extend PDT to take imprinting effects into account and propose two new test statistics  $PDT_m$  and  $PDT_p$ , the maternal and paternal versions of PDT, based on complete pedigrees for qualitative traits. Then we suggest the PDTI statistic, which is

<sup>1</sup>State Key Laboratory of Organ Failure Research, Department of Biostatistics, School of Public Health and Tropical Medicine, Southern Medical University, Guangzhou, China and

<sup>2</sup>Department of Statistics and Actuarial Science, The University of Hong Kong, Hong Kong, China

Correspondence: Professor J-Y Zhou, State Key Laboratory of Organ Failure Research, Department of Biostatistics, School of Public Health and Tropical Medicine, Southern Medical University, No. 1838, North Guangzhou Avenue, Guangzhou 510515, China.

E-mail: zhoujiyuan5460@hotmail.com

or Professor WK Fung, Department of Statistics and Actuarial Science, The University of Hong Kong, Pokfulam Road, Hong Kong, China.

E-mail: wingfung@hku.hk

Received 30 July 2014; revised 19 November 2014; accepted 24 November 2014; published online 18 December 2014

derived in a two-stage test process for detecting association in the presence of linkage. In the first stage, we use the existing imprinting test PPAT to detect if there is paternal, maternal or no imprinting effects at the marker locus under study. In the second stage, we test for association by selecting an appropriate statistic among the three test statistics PDT<sub>m</sub>, PDT and PDT<sub>p</sub> according to the findings of PPAT. To make full use of pedigrees with some missing genotypes, we further extend PDTI and develop the Monte Carlo PDTI (MCPDTI) statistic based on MC sampling and estimation, which utilizes MCPAT in the first stage to test for imprinting and then conducts association analysis in the second stage by choosing a suitable statistic among the three test statistics MCPDT<sub>m</sub> (maternal version of MCPDT), MCPDT and MCPDT<sub>p</sub> (paternal version of MCPDT) according to the findings of MCPAT. As MC sampling and estimation require marker allele frequencies, so we assume that the population under study is homogeneous for MCPDTI. Simulation study under various settings is conducted to evaluate the validity and performance of the proposed tests. Simulation results show that PDTI and MCPDTI control the size well under the null hypothesis of no association. When the Hardy–Weinberg equilibrium in the population is not severely violated, MCPDTI still controls the type I error rates well under the null hypothesis of no association from our simulation study. Further, power comparison demonstrates that PDTI and MCPDTI are more powerful than PDT and TDTI\* (based on a sample of nuclear families, where we randomly select one from each pedigree) when imprinting effects exist. On the other hand, PDTI and MCPDTI have almost the same performance as PDT in the absence of imprinting, which are still much more powerful than TDTI\*.

## MATERIALS AND METHODS

### Background and notations

Suppose that we collect a sample of  $N$  pedigrees with the  $i$ th pedigree having  $n_i$  nuclear families,  $i = 1, 2, \dots, N$ . For the  $j$ th nuclear family in the  $i$ th pedigree, assume that there are  $n_{1ij}$  case-parent trios (consisting of an affected child and his/her parents) and  $n_{2ij}$  discordant sib pairs (DSPs, being composed of one affected and one unaffected sibs),  $i = 1, 2, \dots, N$ ;  $j = 1, 2, \dots, n_i$ . Consider a diallelic marker locus with two alleles  $M_1$  and  $M_2$ . Let  $F_{ij}$ ,  $M_{ij}$  and  $C_{ijk}$  denote the number of copies of  $M_1$  for the father, the mother and the  $k$ th child (affected or unaffected) in the  $j$ th nuclear family of the  $i$ th pedigree, respectively. As such,  $F_{ij}$ ,  $M_{ij}$  and  $C_{ijk}$  take possible values of 0, 1 or 2 for genotypes  $M_2M_2$ ,  $M_1M_2$  or  $M_1M_1$ . We also consider a disease susceptibility locus (DSL) with disease allele  $D$  and normal allele  $d$ . To distinguish the paternal and maternal alleles, we introduce four ordered genotypes  $D/D$ ,  $D/d$ ,  $d/D$  and  $d/d$  at the DSL, of which the allele before the slash is paternal and the allele after the slash is maternal. Further, the corresponding penetrances are denoted by  $f_2$ ,  $f_1$ ,  $f_0$  and  $f_0$ , respectively. When the Mendel's law holds,  $f_1 = f_0$ , which means that there is no parent-of-origin effects; otherwise, there may be parent-of-origin effects at the DSL. The degree of imprinting is defined as  $I = (f_1 - f_0)/2$ , which is used to measure the imprinting effects. If association exists, the marker allele  $M_1$  is assumed to be in positive linkage disequilibrium (LD) with the disease. To this end,  $I > 0$  ( $I < 0$ ) suggests the presence of maternal (paternal) imprinting effect.<sup>16</sup> Further, we assume mating symmetry in the parental generation,<sup>17,18</sup> that the population under study is homogeneous and that there are no maternally mediated effects at the DSL.

### Existing PDT and PPAT statistics for complete pedigree data

We begin by describing the existing PDT (for testing for association) and PPAT (for the detection of parent-of-origin effects) statistics. For the  $j$ th nuclear family in the  $i$ th pedigree, let  $s_{ij} = \sum_{k=1}^{n_{1ij}} (T_{ijk} - NT_{ijk}) + \sum_{l=1}^{n_{2ij}} (A_{ijl} - U_{ijl})$  and  $s_{l ij} = \sum_{k=1}^{n_{1ij}} (I_{F_{ij} > M_{ij}, C_{ijk}=1} - I_{F_{ij} < M_{ij}, C_{ijk}=1})$ ,  $i = 1, 2, \dots, N$ ;  $j = 1, 2, \dots, n_i$ , where  $T_{ijk}$  and  $NT_{ijk}$  represent the numbers of  $M_1$  being transmitted and not being transmitted from the heterozygous parents in each case-parent trio, respectively;  $A_{ijl}$  and  $U_{ijl}$  denote the numbers of  $M_1$  in the affected and

unaffected sibs for each DSP, respectively;  $I_{\{\text{comparison statement}\}}$  is 1 if the comparison statement is true and is 0 otherwise. Taking the summation of the above statistics over all the nuclear families and all the pedigrees, we obtain  $s = \sum_{i=1}^N \sum_{j=1}^{n_i} s_{ij}$  and  $s_l = \sum_{i=1}^N \sum_{j=1}^{n_i} s_{l ij}$ , and then the PDT and PPAT statistics can be formulated as  $\text{PDT} = s / \sqrt{\text{Var}_0(s)}$ ,  $\text{PPAT} = s_l / \sqrt{\text{Var}_0(s_l)}$ , where  $\text{Var}_0(s) = \sum_{i=1}^N (\sum_{j=1}^{n_i} s_{ij})^2$  and  $\text{Var}_0(s_l) = \sum_{i=1}^N (\sum_{j=1}^{n_i} s_{l ij})^2$  are the unbiased estimates of the variances of  $s$  and  $s_l$  under the null hypotheses of no association and no imprinting, respectively.<sup>7,15</sup>

### PDTI for complete pedigree data

PDT is a powerful method for detecting association. However, it does not take into account the information on parent-of-origin effects in the analysis, which may suffer from much power loss when imprinting effects are present. To this end, following the idea of constructing TDTI\* for nuclear families,<sup>12</sup> we propose the PDTI statistic for complete pedigrees in the following, which is based on the existing PDT, and two statistics PDT<sub>m</sub> and PDT<sub>p</sub> proposed in this article, to incorporate parent-of-origin effects into association study. The statistics PDT<sub>m</sub> and PDT<sub>p</sub>, the maternal and paternal versions of PDT, can be obtained by replacing  $s_{ij}$  involved in PDT by  $s_{mij}$  and  $s_{pij}$ , respectively, where

$$s_{mij} = \sum_{k=1}^{n_{1ij}} (T_{mijk} - NT_{mijk}) + \sum_{l=1}^{n_{2ij}} (A_{mijl} - U_{mijl}),$$

$$s_{pij} = \sum_{k=1}^{n_{1ij}} (T_{pijk} - NT_{pijk}) + \sum_{l=1}^{n_{2ij}} (A_{pijl} - U_{pijl});$$

$T_{mijk}$  and  $NT_{mijk}$  ( $T_{pijk}$  and  $NT_{pijk}$ ) are the numbers of  $M_1$  being transmitted and not being transmitted from the heterozygous mother (father), respectively;  $A_{mijl}$  and  $U_{mijl}$  ( $A_{pijl}$  and  $U_{pijl}$ ) are the numbers of the maternal (paternal)  $M_1$  allele in the affected and unaffected sibs for a DSP, respectively. The calculation of  $T_{mijk} - NT_{mijk}$  ( $T_{pijk} - NT_{pijk}$ ) is direct, while the calculation of  $A_{mijl} - U_{mijl}$  ( $A_{pijl} - U_{pijl}$ ) is a little more complicated as we need to determine which allele in each member of a DSP is maternal and which one is paternal. To do so, we need to simultaneously consider two members in each DSP together with their parents.

For easy description here, we ignore the subscripts for indexing the DSP, nuclear family and pedigree and simply use  $A - U$ ,  $A_m - U_m$  and  $A_p - U_p$  to replace  $A_{ijl} - U_{ijl}$ ,  $A_{mijl} - U_{mijl}$  and  $A_{pijl} - U_{pijl}$ , respectively. Note that  $A - U = 0$  involved in PDT when two members in a DSP have the same genotypes, which does not provide information to association. As such,  $A_m - U_m = 0$  and  $A_p - U_p = 0$  under this situation. Furthermore, if the mating type is  $M_1M_2 \times M_1M_2$ , and one of two members in a DSP is homozygous and the other is heterozygous (for example, the affected and unaffected sibs, respectively, have genotypes  $C_A = M_1M_1$  and  $C_U = M_1M_2$ ), then the allele  $M_1$  in the unaffected sib may come from either the father or the mother. Thus, we assume that both  $A_m - U_m$  and  $A_p - U_p$  are 0.5 for this situation, following the idea from Li et al.<sup>19</sup> Similarly, both  $A_m - U_m$  and  $A_p - U_p$  are taken to be 0.5 for  $C_A = M_1M_2$  and  $C_U = M_2M_2$ ,  $-0.5$  for  $C_A = M_1M_2$  and  $C_U = M_1M_1$  and  $C_A = M_2M_2$  and  $C_U = M_1M_2$ . Table 1 shows the values of  $A_m - U_m$  and  $A_p - U_p$  based on all the 29 possible genotype combinations of a DSP and their parents, together with the corresponding conditional joint probabilities  $w_r$ 's given that two children are a DSP,  $r = 1, 2, \dots, 29$ . It is shown in Supplementary Information that  $E(A_m - U_m) = 0$  and  $E(A_p - U_p) = 0$  under the null hypothesis of no association.

Let  $s_m = \sum_{i=1}^N \sum_{j=1}^{n_i} s_{mij}$  and  $s_p = \sum_{i=1}^N \sum_{j=1}^{n_i} s_{pij}$ . Thus,  $E(s_m) = 0$  and  $E(s_p) = 0$  under the null hypothesis of no association. As such, the following PDT<sub>m</sub> and PDT<sub>p</sub> statistics are valid in testing for association

$$\text{PDT}_m = \frac{s_m}{\sqrt{\text{Var}_0(s_m)}}, \quad \text{PDT}_p = \frac{s_p}{\sqrt{\text{Var}_0(s_p)}}, \quad (1)$$

where  $\text{Var}_0(s_m) = \sum_{i=1}^N (\sum_{j=1}^{n_i} s_{mij})^2$  and  $\text{Var}_0(s_p) = \sum_{i=1}^N (\sum_{j=1}^{n_i} s_{pij})^2$  are, respectively, the unbiased estimates of the variances of  $s_m$  and  $s_p$  under the null hypothesis of no association, which have taken into account the dependencies among all the case-parent trios, the dependencies among all the DSPs and the dependencies between the case-parent trios and the DSPs within each pedigree (see Supplementary Information). As can be seen from the variance estimates, both the PDT<sub>m</sub> and PDT<sub>p</sub> statistics asymptotically follow a standard normal

**Table 1** Values of  $A_m - U_m$  and  $A_p - U_p$  for all the 29 possible genotype combinations of a DSP and their parents, together with the corresponding conditional joint probabilities  $w_r$ 's given that two children are a DSP

r	Mating type <sup>a</sup>	C <sub>A</sub> <sup>b</sup>	C <sub>U</sub> <sup>c</sup>	A <sub>m</sub> - U <sub>m</sub>	A <sub>p</sub> - U <sub>p</sub>	w <sub>r</sub>
1	M <sub>1</sub> M <sub>1</sub> × M <sub>1</sub> M <sub>1</sub>	M <sub>1</sub> M <sub>1</sub>	M <sub>1</sub> M <sub>1</sub>	0	0	w <sub>1</sub>
2	M <sub>1</sub> M <sub>1</sub> × M <sub>1</sub> M <sub>2</sub>	M <sub>1</sub> M <sub>1</sub>	M <sub>1</sub> M <sub>1</sub>	0	0	w <sub>2</sub>
3	M <sub>1</sub> M <sub>1</sub> × M <sub>1</sub> M <sub>2</sub>	M <sub>1</sub> M <sub>1</sub>	M <sub>1</sub> M <sub>2</sub>	1	0	w <sub>3</sub>
4	M <sub>1</sub> M <sub>1</sub> × M <sub>1</sub> M <sub>2</sub>	M <sub>1</sub> M <sub>2</sub>	M <sub>1</sub> M <sub>1</sub>	-1	0	w <sub>4</sub>
5	M <sub>1</sub> M <sub>1</sub> × M <sub>1</sub> M <sub>2</sub>	M <sub>1</sub> M <sub>2</sub>	M <sub>1</sub> M <sub>2</sub>	0	0	w <sub>5</sub>
6	M <sub>1</sub> M <sub>1</sub> × M <sub>2</sub> M <sub>2</sub>	M <sub>1</sub> M <sub>2</sub>	M <sub>1</sub> M <sub>2</sub>	0	0	w <sub>6</sub>
7	M <sub>1</sub> M <sub>2</sub> × M <sub>1</sub> M <sub>1</sub>	M <sub>1</sub> M <sub>1</sub>	M <sub>1</sub> M <sub>1</sub>	0	0	w <sub>7</sub>
8	M <sub>1</sub> M <sub>2</sub> × M <sub>1</sub> M <sub>1</sub>	M <sub>1</sub> M <sub>1</sub>	M <sub>1</sub> M <sub>2</sub>	0	1	w <sub>8</sub>
9	M <sub>1</sub> M <sub>2</sub> × M <sub>1</sub> M <sub>1</sub>	M <sub>1</sub> M <sub>2</sub>	M <sub>1</sub> M <sub>1</sub>	0	-1	w <sub>9</sub>
10	M <sub>1</sub> M <sub>2</sub> × M <sub>1</sub> M <sub>1</sub>	M <sub>1</sub> M <sub>2</sub>	M <sub>1</sub> M <sub>2</sub>	0	0	w <sub>10</sub>
11	M <sub>1</sub> M <sub>2</sub> × M <sub>1</sub> M <sub>2</sub>	M <sub>1</sub> M <sub>1</sub>	M <sub>1</sub> M <sub>1</sub>	0	0	w <sub>11</sub>
12	M <sub>1</sub> M <sub>2</sub> × M <sub>1</sub> M <sub>2</sub>	M <sub>1</sub> M <sub>1</sub>	M <sub>1</sub> M <sub>2</sub>	0.5	0.5	w <sub>12</sub>
13	M <sub>1</sub> M <sub>2</sub> × M <sub>1</sub> M <sub>2</sub>	M <sub>1</sub> M <sub>1</sub>	M <sub>2</sub> M <sub>2</sub>	1	1	w <sub>13</sub>
14	M <sub>1</sub> M <sub>2</sub> × M <sub>1</sub> M <sub>2</sub>	M <sub>1</sub> M <sub>2</sub>	M <sub>1</sub> M <sub>1</sub>	-0.5	-0.5	w <sub>14</sub>
15	M <sub>1</sub> M <sub>2</sub> × M <sub>1</sub> M <sub>2</sub>	M <sub>1</sub> M <sub>2</sub>	M <sub>1</sub> M <sub>2</sub>	0	0	w <sub>15</sub>
16	M <sub>1</sub> M <sub>2</sub> × M <sub>1</sub> M <sub>2</sub>	M <sub>1</sub> M <sub>2</sub>	M <sub>2</sub> M <sub>2</sub>	0.5	0.5	w <sub>16</sub>
17	M <sub>1</sub> M <sub>2</sub> × M <sub>1</sub> M <sub>2</sub>	M <sub>2</sub> M <sub>2</sub>	M <sub>1</sub> M <sub>1</sub>	-1	-1	w <sub>17</sub>
18	M <sub>1</sub> M <sub>2</sub> × M <sub>1</sub> M <sub>2</sub>	M <sub>2</sub> M <sub>2</sub>	M <sub>1</sub> M <sub>2</sub>	-0.5	-0.5	w <sub>18</sub>
19	M <sub>1</sub> M <sub>2</sub> × M <sub>1</sub> M <sub>2</sub>	M <sub>2</sub> M <sub>2</sub>	M <sub>2</sub> M <sub>2</sub>	0	0	w <sub>19</sub>
20	M <sub>1</sub> M <sub>2</sub> × M <sub>2</sub> M <sub>2</sub>	M <sub>1</sub> M <sub>2</sub>	M <sub>1</sub> M <sub>2</sub>	0	0	w <sub>20</sub>
21	M <sub>1</sub> M <sub>2</sub> × M <sub>2</sub> M <sub>2</sub>	M <sub>1</sub> M <sub>2</sub>	M <sub>2</sub> M <sub>2</sub>	0	1	w <sub>21</sub>
22	M <sub>1</sub> M <sub>2</sub> × M <sub>2</sub> M <sub>2</sub>	M <sub>2</sub> M <sub>2</sub>	M <sub>1</sub> M <sub>2</sub>	0	-1	w <sub>22</sub>
23	M <sub>1</sub> M <sub>2</sub> × M <sub>2</sub> M <sub>2</sub>	M <sub>2</sub> M <sub>2</sub>	M <sub>2</sub> M <sub>2</sub>	0	0	w <sub>23</sub>
24	M <sub>2</sub> M <sub>2</sub> × M <sub>1</sub> M <sub>1</sub>	M <sub>1</sub> M <sub>2</sub>	M <sub>1</sub> M <sub>2</sub>	0	0	w <sub>24</sub>
25	M <sub>2</sub> M <sub>2</sub> × M <sub>1</sub> M <sub>2</sub>	M <sub>1</sub> M <sub>2</sub>	M <sub>1</sub> M <sub>2</sub>	0	0	w <sub>25</sub>
26	M <sub>2</sub> M <sub>2</sub> × M <sub>1</sub> M <sub>2</sub>	M <sub>1</sub> M <sub>2</sub>	M <sub>2</sub> M <sub>2</sub>	1	0	w <sub>26</sub>
27	M <sub>2</sub> M <sub>2</sub> × M <sub>1</sub> M <sub>2</sub>	M <sub>2</sub> M <sub>2</sub>	M <sub>1</sub> M <sub>2</sub>	-1	0	w <sub>27</sub>
28	M <sub>2</sub> M <sub>2</sub> × M <sub>1</sub> M <sub>2</sub>	M <sub>2</sub> M <sub>2</sub>	M <sub>2</sub> M <sub>2</sub>	0	0	w <sub>28</sub>
29	M <sub>2</sub> M <sub>2</sub> × M <sub>2</sub> M <sub>2</sub>	M <sub>2</sub> M <sub>2</sub>	M <sub>2</sub> M <sub>2</sub>	0	0	w <sub>29</sub>

Abbreviation: DSP, discordant sib pair.

<sup>a</sup>The first genotype is the father's and the second one is the mother's.

<sup>b</sup>C<sub>A</sub> denotes the genotype of the affected sib in a DSP.

<sup>c</sup>C<sub>U</sub> denotes the genotype of the unaffected sib in a DSP.

distribution under the null hypothesis of no association when the sample size is large.

Based on PDT<sub>m</sub>, PDT and PDT<sub>p</sub>, the proposed PDTI statistic is formulated as

$$PDTI = PDT_m I_{\{PPAT < -z_{\alpha_I/2}\}} + PDTI_{\{PPAT \leq z_{\alpha_I/2}\}} + PDT_p I_{\{PPAT > z_{\alpha_I/2}\}},$$

where  $\alpha_I$  is the prespecified significance level for the imprinting test PPAT, and  $z_{\alpha_I/2}$  is the upper quantile of the standard normal distribution corresponding to the probability of  $\alpha_I/2$ . The formula of the PDTI statistic demonstrates that, in the presence of paternal (maternal) imprinting effect, the PDT<sub>m</sub> (PDT<sub>p</sub>) statistic would be chosen to conduct the detection of association. Otherwise, PDT will be used to test for the association.

In the absence of association, the PDT<sub>m</sub>, PDT, PDT<sub>p</sub> and PPAT statistics asymptotically follow a standard normal distribution  $N(0, 1)$ , respectively. However, they are not independent, and the covariance between PDT<sub>m</sub> and PPAT under the null hypothesis of no association can be estimated as

$$C\hat{ov}_0(PDT_m, PPAT) = \frac{C\hat{ov}_0(s_m, s_I)}{\sqrt{V\hat{ar}_0(s_m)}\sqrt{V\hat{ar}_0(s_I)}},$$

where  $C\hat{ov}_0(s_m, s_I) = \sum_{i=1}^N \sum_{j=1}^{n_i} \sum_{k=1}^{n_i} s_{mij} s_{Iik}$ , which is the unbiased estimate of the covariance between  $s_m$  and  $s_I$  under the null hypothesis of no association (see Supplementary Information). Also, the covariance between PDT and PPAT, and that between PDT<sub>p</sub> and PPAT, can be estimated in a similar way. Denote the estimates of the 2 × 2 variance-covariance matrices of the statistics

(PDT<sub>m</sub>, PPAT), (PDT, PPAT) and (PDT<sub>p</sub>, PPAT) by  $\hat{\Sigma}_m$ ,  $\hat{\Sigma}$  and  $\hat{\Sigma}_p$ , respectively. Under the null hypothesis  $H_0$  of no association, the asymptotic distributions of these three pairs of statistics are bivariate normal distributions with zero mean vectors and the variance-covariance matrices given above. Let  $t$  be the observed value of PDTI and denote the  $P$ -value of the test for association by  $P = P_{H_0}(PDTI > |t|)$ . Then, based on the asymptotic bivariate normal distributions, we obtain

$$P = \int_{R(y)} \int_{-\infty}^{-z_{\alpha_I/2}} \frac{1}{2\pi|\hat{\Sigma}_m|^{1/2}} \exp\left\{-\frac{1}{2}(x, y)\hat{\Sigma}_m^{-1}(x, y)'\right\} dx dy + \int_{R(y)} \int_{-z_{\alpha_I/2}}^{z_{\alpha_I/2}} \frac{1}{2\pi|\hat{\Sigma}|^{1/2}} \exp\left\{-\frac{1}{2}(x, y)\hat{\Sigma}^{-1}(x, y)'\right\} dx dy + \int_{R(y)} \int_{z_{\alpha_I/2}}^{+\infty} \frac{1}{2\pi|\hat{\Sigma}_p|^{1/2}} \exp\left\{-\frac{1}{2}(x, y)\hat{\Sigma}_p^{-1}(x, y)'\right\} dx dy,$$

where  $R(y) = \{y: -\infty < y < -|t|\} \cup \{y: |t| < y < \infty\}$ . Moreover, we have also developed a method on how to find the critical value of PDTI, which can be used to determine whether or not to reject the null hypothesis (see Supplementary Information).

Although we introduce the PDTI method under the assumption that  $M_1$  is in positive LD with the disease allele  $D$ , PDTI remains valid for the negative LD between  $M_1$  and  $D$ . Note that if  $M_1$  is in positive LD with  $D$ , the positive (negative) values of PPAT indicate maternal (paternal) imprinting effect. Conversely, if  $M_1$  is in negative LD with  $D$  (that is,  $M_2$  is positively associated with  $D$ ), the positive (negative) values of PPAT show paternal (maternal) imprinting effect. However, similar to TDTI\*, the PDTI statistic keeps unchanged irrespective of the positive or negative LD, which signifies the validity of PDTI for practical use (the proof omitted for brevity).

### MCPDTI for incomplete pedigree data

In the case in which there are missing genotypes in pedigrees, we extend PDTI and propose MCPDTI by using MC sampling, which may recapture most information on missing genotypes based on the observed genotypes. The existing MCPDT and MCPDAT statistics are the MC versions of PDT and PPAT, respectively. Now, we turn to introduce the MCPDT<sub>m</sub> and MCPDT<sub>p</sub> statistics, the MC versions of PDT<sub>m</sub> and PDT<sub>p</sub>, respectively.

For easy description here, we delete the subscripts for indexing the nuclear family and pedigree. Let  $s_m$  denote the contribution from a pedigree to the PDT<sub>m</sub> statistic in Equation (1). Let  $G_o$  and  $G_u$  be the sets of the observed and unobserved (missing) genotypes in the pedigree, respectively, and  $S$  be the observed phenotypes (disease affection statuses). For the case with missing data,  $s_m$  is unavailable, while its conditional expectation given  $G_o$  can be estimated by  $s_{mMC}$ ,<sup>7,14</sup> which is based on the MC simulation scheme and can be formulated as follows,  $s_{mMC} \approx \frac{1}{X} \sum_{x=1}^X s_m(G_{ux}, G_o, S)$ , where  $G_{ux}$ 's are independent samples randomly drawn from  $P(G_u|G_o)$ ,  $x = 1, 2, \dots, X$ , which can be efficiently accomplished based on the peeling algorithm using the SLINK software.<sup>20</sup> Subsequently, MCPDT<sub>m</sub> can be calculated by using Equation (1) with  $s_m$  replaced by  $s_{mMC}$ . Similarly, MCPDT<sub>p</sub> can be given by replacing  $s_p$  by  $s_{pMC}$ .

Note that if all the pedigrees are assumed to be drawn from a certain underlying population and  $S$  is treated as random, the expectations of  $s_{mMC}$  and  $s_{pMC}$  are 0 under the null hypothesis of no association. The idea of the corresponding proof is similar to Zhou et al.<sup>7</sup> and Ding et al.,<sup>14</sup> and hence we omit it here for brevity. As such, the MCPDTI statistic can be formulated by

$$MCPDTI = MCPDT_m I_{\{MCPDAT < -z_{\alpha_I/2}\}} + MCPDTI_{\{MCPDAT \leq z_{\alpha_I/2}\}} + MCPDT_p I_{\{MCPDAT > z_{\alpha_I/2}\}}.$$

Obviously, MCPDTI shares some common properties with PDTI. First of all, MCPDTI is not influenced by positive or negative LD between  $M_1$  and  $D$ . Next, MCPDAT and three MCPDT-type statistics are not independent. Hence, to clarify the asymptotic null distribution of MCPDTI, the derivation of their covariances among them is required. On the other hand, note that the MC scheme relies on marker allele frequencies. This means that MCPDTI needs the assumption of the population under study being homogeneous. The derivation

of the critical value of MCPDTI is similar to that of PDTI (omitted here for brevity).

## RESULTS

### Settings

In this section, a simulation study is conducted to investigate the performance of the proposed PDTI and MCPDTI statistics and to make the power comparison of PDTI and MCPDTI with the existing PDT and TDTI\*. A homogeneous population is considered. We simulate one marker and one DSL with the recombination fraction being fixed at 0.001. Three sets of haplotype frequencies for haplotypes  $DM_1$ ,  $dM_1$ ,  $DM_2$  and  $dM_2$  are used to simulate the test powers: {0.13, 0.02, 0.12, 0.73}, {0.22, 0.03, 0.03, 0.72} and {0.23, 0.12, 0.02, 0.63}, with the allele frequencies of  $M_1$  being taken as 0.15, 0.25 and 0.35, respectively. The allele frequency of disease allele  $D$  is fixed at 0.25. For studying the empirical size of the proposed test statistics under the null hypothesis of no association, four haplotype frequencies for haplotypes  $DM_1$ ,  $dM_1$ ,  $DM_2$  and  $dM_2$  are just taken as the products of two allele frequencies on each haplotype, respectively. For example, if the allele frequency of  $M_1$  is 0.15, then the frequency of haplotype  $DM_1$  is  $0.25 \times 0.15 = 0.0375$ . However, the recombination fraction between the marker and the DSL is still fixed to be 0.001. Further, note that no association means no parent-of-origin effects. To this end, we simulate the empirical size of the proposed test statistics under no association and no parent-of-origin effect model (see below).

We consider three sets of two homozygote penetrances  $f_2$  and  $f_0$ , {0.39, 0.26}, {0.44, 0.24} and {0.48, 0.22}, which are very similar to those in Ding *et al.*<sup>14</sup> For each set of homozygote penetrances, we further consider three parent-of-origin effect models by setting the

various values of  $f_{10}$  and  $f_{01}$ : no parent-of-origin effect model, incomplete parent-of-origin effect model, and complete parent-of-origin effect model. Therefore, for each parent-of-origin-effect model, three sets of haplotype frequencies and three sets of homozygote penetrances lead to nine simulation settings, which are listed in Table 2.

Figure 1 gives three types of common pedigree structures used in our simulation study: (a) two-generation family with 5 individuals, (b) three-generation family with 10 individuals, and (c) four-generation family with 12 individuals. To investigate the performance of the proposed tests under different sample sizes, 50 and 100 pedigrees under each pedigree structure are simulated in each replicate, which leads to the total sample size being 150 and 300, respectively. Note that here a pedigree was ascertained if at least one of the nonfounders was affected. For MCPDTI, 50 MC samples of missing genotypes are generated for each replicate using the SLINK software.<sup>20</sup> Both the true marker allele frequencies or those estimated from the genotyped founders in each replicate are used in the MC sampling where applicable.

To evaluate the performance of our proposed tests and compare PDTI and MCPDTI with the existing PDT and TDTI\* tests, we consider the following five statistics: PDTI, PDT, TDTI\*, MCPDTI<sub>T</sub>, and MCPDTI<sub>E</sub>. The first three statistics PDTI, PDT and TDTI\* are obtained based on complete pedigree data without any missing genotypes. PDTI can be regarded as a golden standard for the association study incorporating parent-of-origin effects. Note that PDT does not take account of parent-of-origin effects, and hence its powers will not be influenced by different degrees of imprinting (that is, different parent-of-origin effect models). As such, here PDT can be considered as a reference line. Note that TDTI\* can only accommodate two-generation nuclear family data. So, in our simulation, TDTI\* is based on a sample of nuclear families, where only one nuclear family is randomly chosen from a pedigree. Further, MCPDTI<sub>T</sub> and MCPDTI<sub>E</sub> are on the basis of incomplete pedigree data with genotypes of individual 1 in two-generation families, individuals 1, 4 and 5 in three-generation pedigrees and individuals 1 and 3 in four-generation pedigrees being removed from complete pedigree data; MCPDTI<sub>T</sub> and MCPDTI<sub>E</sub> are based on the true and estimated marker allele frequencies, respectively.

Ten thousand replicates under each simulation setting mentioned above are simulated to assess the empirical size and powers of the proposed tests at the significance levels of 5% and 1%. The corresponding significance level of the imprinting test in the first stage is taken to be the same as the association test in the second stage. The size and power results at the 1% significance level can be seen in Supplementary Table S1 and Supplementary Figure S1, respectively.

To have a better understanding of the performance of PDTI and MCPDTI statistics, we conduct two additional simulation studies as shown in Supplementary Information. One is based on the settings with the less frequency of disease allele  $D$  (being fixed at 0.1) and the family sample consisting of 150 and 200 pedigrees under each pedigree structure, which leads to 450 and 600 pedigrees in total, respectively. For detailed simulation settings, see Supplementary Table S2. The other simulation study is based on the settings with the frequency of disease allele  $D$  being taken as 0.1, 0.15 and 0.2 under Hardy–Weinberg disequilibrium at a candidate locus, where the inbreeding coefficient  $\rho$  is taken to be 0.03 and 0.05. The family sample in each replicate consists of 100 pedigrees under each pedigree structure, where the number of pedigrees is 300 in total. The detailed simulation settings can be found in Supplementary Tables S3 and S4.

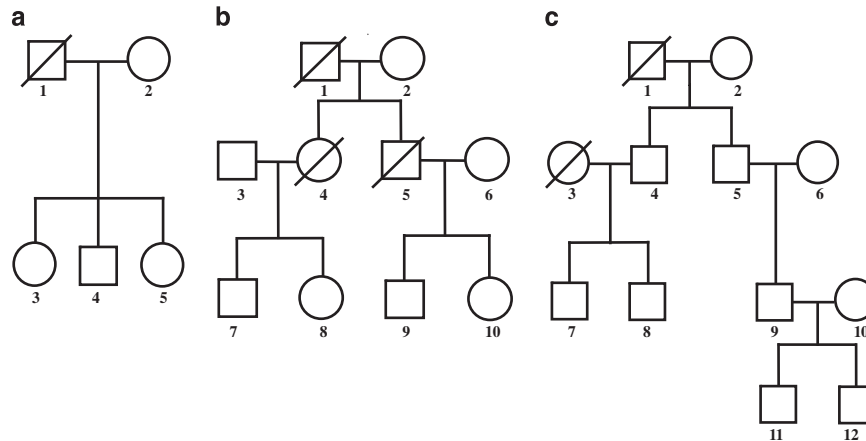
**Table 2 Simulation settings<sup>a</sup>**

A. Haplotype frequencies and homozygote penetrances						
Setting	Haplotype frequencies				Penetrances	
	$DM_1$	$dM_1$	$DM_2$	$dM_2$	$f_2$	$f_0$
1	0.13	0.02	0.12	0.73	0.390	0.260
2	0.23	0.12	0.02	0.63	0.390	0.260
3	0.22	0.03	0.03	0.72	0.390	0.260
4	0.13	0.02	0.12	0.73	0.440	0.240
5	0.23	0.12	0.02	0.63	0.440	0.240
6	0.22	0.03	0.03	0.72	0.440	0.240
7	0.13	0.02	0.12	0.73	0.480	0.220
8	0.23	0.12	0.02	0.63	0.480	0.220
9	0.22	0.03	0.03	0.72	0.480	0.220

B. Heterozygote penetrances for three parent-of-origin effect models						
Setting	Parent-of-origin effect models					
	No		Incomplete		Complete	
	$f_{10}$	$f_{01}$	$f_{10}$	$f_{01}$	$f_{10}$	$f_{01}$
1,2,3	0.325	0.325	0.370	0.280	0.390	0.260
4,5,6	0.340	0.340	0.420	0.260	0.440	0.240
7,8,9	0.350	0.350	0.460	0.240	0.480	0.220

<sup>a</sup>A and B combined lead to three sets of nine settings to study the simulated power. The first set assumes the absence of parent-of-origin effects, and the second and third sets are the incomplete and complete parent-of-origin effect models, respectively.



**Figure 1** Pedigree structure used for the simulation studies. (a) Nuclear family; (b) three-generation family; (c) four-generation family. Genotypes of individual 1 in nuclear family, individuals 1, 4 and 5 in three-generation family and individuals 1 and 3 in four-generation family are assumed to be missing for the analysis based on incomplete data.

**Table 3** Type I error rates (in percentage (%)) of PDTI, PDT, TDTI\* and MCPDTI under nine simulation settings based on 150 and 300 pedigrees at nominal 5% significance level, respectively

Setting	Complete data			Incomplete data	
	PDTI	PDT	TDTI*	MCPDTI <sub>T</sub>	MCPDTI <sub>E</sub>
<i>Based on 150 pedigrees</i>					
1	5.33	4.95	5.03	5.15	5.22
2	4.93	4.66	4.71	4.92	4.97
3	4.97	4.70	4.57	4.69	4.58
4	5.18	5.02	5.10	5.14	5.33
5	4.91	4.74	4.49	4.62	4.93
6	5.09	4.91	5.25	4.99	5.18
7	5.30	5.12	5.11	5.08	5.24
8	5.01	4.74	4.96	5.10	5.21
9	5.24	5.11	5.35	5.16	5.40
<i>Based on 300 pedigrees</i>					
1	5.19	4.87	4.86	5.07	5.42
2	4.89	4.79	5.25	5.28	5.40
3	5.27	4.99	5.05	5.32	5.44
4	5.37	5.14	5.26	5.36	5.38
5	5.51	5.35	5.25	5.16	5.36
6	4.74	4.52	4.79	4.71	4.88
7	5.16	4.92	5.21	5.09	4.97
8	4.99	4.77	5.05	4.73	4.91
9	5.32	5.01	5.28	5.22	5.19

Abbreviations: MCPDTI<sub>E</sub>, Monte Carlo PDT with imprinting based on the estimated marker allele frequencies; MCPDTI<sub>T</sub>, Monte Carlo PDT with imprinting based on the true marker allele frequencies; PDT, pedigree disequilibrium test; PDTI, pedigree disequilibrium test with imprinting; TDTI\*, transmission disequilibrium test with imprinting.

The simulation results are shown in Supplementary Tables S5–S8 and Supplementary Figures S2–S5.

**Size and powers of PDTI, PDT, TDTI\* and MCPDTI**

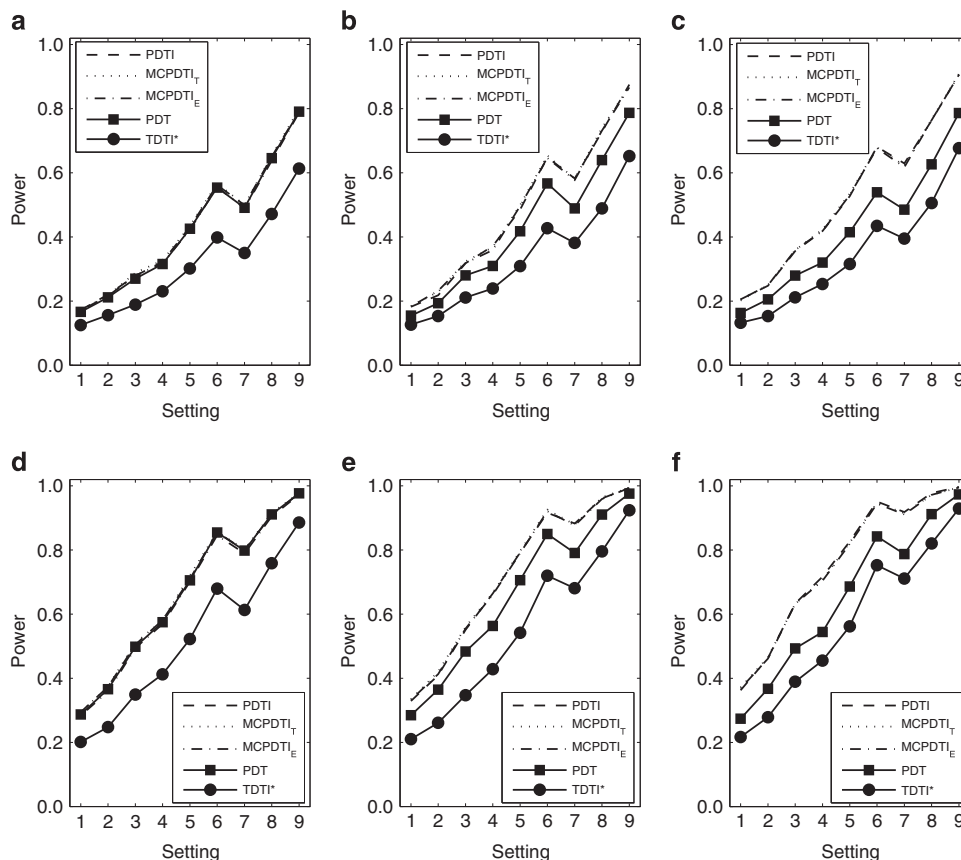
Table 3 displays the empirical type I error rates of PDTI, PDT, TDTI\* and MCPDTI under nine simulation settings based on 150 and 300 pedigrees at the nominal 5% significance level, respectively. For 10 000 simulation replicates, the s.e. for the type I error rate estimate is  $(0.05 \times 0.95/10\ 000)^{1/2} = 2.18 \times 10^{-3}$  when the nominal level is fixed at

5%. Note that in Table 3, there are altogether 54 observed type I error rates corresponding to the proposed tests (PDTI, MCPDTI<sub>T</sub> and MCPDTI<sub>E</sub>), and 2 out of the 54 are greater than the upper limit of the interval  $(0.04564, 0.05436)$ , that is,  $0.05 \pm 2 \times \text{s.e.}$  Thus, generally speaking, the tests maintain the type I error rates satisfactorily.

Figure 2 gives the simulated powers of PDTI, MCPDTI<sub>T</sub>, MCPDTI<sub>E</sub>, PDT and TDTI\* under nine simulation settings based on 150 and 300 pedigrees at the 5% significance level for no, incomplete and complete parent-of-origin effect models, respectively. We can see from each subplot of the figure that MCPDTI, by the MC sampling scheme, can recapture most information on the missing genotypes and has similar performance to PDTI based on complete pedigree data. MCPDTI based on the estimated marker allele frequencies (MCPDTI<sub>E</sub>) almost has the same power as that based on the true marker allele frequencies (MCPDTI<sub>T</sub>). When there are no parent-of-origin effects, with the haplotype frequencies, penetrances and sample size being fixed (for each simulation setting in Figures 2a and d), the powers of PDTI, MCPDTI<sub>T</sub> and MCPDTI<sub>E</sub> are very close to that of PDT and are all much higher than that of TDTI\*, which only accommodates nuclear family data. However, for each simulation setting with incomplete or complete parent-of-origin effects (Figures 2b, c, e or f), PDTI, MCPDTI<sub>T</sub> and MCPDTI<sub>E</sub> are all more powerful than both PDT and TDTI\*. The power of PDT based on pedigrees is much larger than TDTI\*, even though PDT does not take account of imprinting but TDTI\* does.

Furthermore, when the sample size is fixed for no, incomplete and complete parent-of-origin effect models (among Figures 2a, b and c or among Figures 2d, e and f), the power of PDT (the reference base) is almost not affected by the degree of imprinting. However, PDTI, MCPDTI and TDTI\* under the incomplete and complete parent-of-origin effect models are more powerful, compared with no parent-of-origin effect model. The power difference between our proposed test PDTI/MCPDTI<sub>T</sub>/MCPDTI<sub>E</sub> and PDT is larger and the gap in power between PDT and TDTI\* is smaller, by comparing the power results under the complete parent-of-origin effect model with the incomplete model. The powers based on 300 pedigrees are higher than those based on 150 pedigrees (Figure 2; d vs a, e vs b and f vs c), and the corresponding power improvement of PDTI and MCPDTI is more significant compared with that of PDT, as expected.

It is also shown in each subplot of Figure 2 that, with the sample size and the penetrances being fixed, the powers of all the five test



**Figure 2** Powers of PDTI, MCPDTI, PDT and TDTI\* under nine simulation settings at the 5% significance level. (a) One hundred and fifty pedigrees and no parent-of-origin effect model; (b) 150 pedigrees and incomplete parent-of-origin effect model; (c) 150 pedigrees and complete parent-of-origin effect model; (d) 300 pedigrees and no parent-of-origin effect model; (e) 300 pedigrees and incomplete parent-of-origin effect model; (f) 300 pedigrees and complete parent-of-origin effect model.

statistics become larger when the simulation settings change from settings 1 to 3 (settings 4 to 6 or settings 7 to 9), due to the increase of the LD coefficient for three sets of haplotype frequencies. As for settings 1, 4 and 7 (or settings 2, 5 and 8/settings 3, 6 and 9), which are based on the same set of haplotype frequencies but different relative risks of two homozygote penetrances ( $f_2/f_0 = 1.50, 1.83$  and  $2.18$ ), all the statistics have higher and higher power.

## DISCUSSION

In this article, we have extended PDT and MCPDT to PDTI and MCPDTI to incorporate parent-of-origin effects for association analysis based on general pedigree data. Simulation study is carried out to investigate the validity and performance of the proposed PDTI and MCPDTI tests and to compare PDTI and MCPDTI with the existing PDT and TDTI\* for various sets of haplotype frequencies, parent-of-origin effect models, significance levels and sample sizes. Simulation results show that the proposed PDTI and MCPDTI control the size well under the null hypothesis of no association.

As for power comparison of PDTI and MCPDTI with PDT (ignoring the information on imprinting effects) and TDTI\* (only accommodating two-generation nuclear families), under no parent-of-origin effect model, PDTI and MCPDTI have similar performance, which are comparable to PDT. Under incomplete parent-of-origin effect model, PDTI and MCPDTI, by using the information on the parent-of-origin effects and the extended pedigree data, outperform PDT and TDTI\*. Note that in our simulation study, MCPDTI for

incomplete pedigree data is even more powerful than PDT for complete pedigree data, although the power improvement is not so significant. However, as the degree of imprinting increases, under the complete parent-of-origin effect model, the powers of PDTI and MCPDTI are much higher than that of PDT and TDTI\*. Moreover, with the sample size increasing, the improvement in power of our proposed methods over PDT is more distinct. Even though the frequency of the disease allele is smaller (being taken as 0.1), our methods are still more powerful than PDT and TDTI\* in the presence of imprinting effects (see Supplementary Information). On the other hand, MCPDTI can recapture most of the missing genotypes based on the observed genotypes by the MC sampling and estimation. Our software MCPDTI is freely available at <http://www.echobelt.org/web/UploadFiles/MCPDTI.html>, which is implemented in R (<http://www.r-project.org/>).

In the calculation process of the proposed PDTI and MCPDTI test statistics, we first divide each extended pedigree in the collected sample into several nuclear families (two-generation families). For each nuclear family, PDTI and MCPDTI will use all the case-parent trios and all the DSPs in it. As such, the alleles of both parents will repeatedly be used in different case-parent trios, the alleles of some affected children may repeatedly be used in both case-parent trios and DSPs and the alleles of some unaffected children may repeatedly be utilized in different DSPs. Further, for different nuclear families from this extended pedigree, there may also be the repeated use of the alleles of some family members who can be the children of the elder

generation and the parents of the younger generation. Thus the repeated use of these family members mentioned above will lead to the dependencies among the case-parent trios, the dependencies among the DSPs and the dependencies between the case-parent trios and the DSPs within each pedigree. However, it should be emphasized that PDTI and MCPDTI have taken account of all the dependencies (see Supplementary Information).

Although our simulation study is conducted under settings with positive LD coefficient between marker allele  $M_1$  and disease allele  $D$ , PDTI and MCPDTI are still suitable for situations where  $M_1$  is in negative LD with  $D$ . On the other hand, like other methods, the proposed PDTI and MCPDTI methods have their own limitations. PDTI is not applicable to the situation where there are some genotypes missing, and thus we recommend MCPDTI for this case to conduct association analysis. Moreover, MCPDTI needs the assumption of the underlying population being homogeneous. However, it still controls the size well if the violation of Hardy–Weinberg equilibrium is not so severe in the population (see Supplementary Information). In addition, although our simulation study is conducted by fixing the recombination fraction at 0.001, the proposed PDTI and MCPDTI are still more powerful than the existing PDT and TDTI\* for the complete LD case and when the recombination fraction is 0 (see Supplementary Tables S9 and S10). Finally, we should mention that, in the data collection process of our simulation study, the requirement for a pedigree to be included is at least one affected nonfounder, while it is not necessary to require that each nuclear family within a large pedigree has at least one affected child. In practice, the pedigrees with more affected individuals are more likely to be recruited. This ascertainment bias due to the difference in the number of affected individuals in pedigree is not considered in the present study. How the ascertainment bias affects the proposed test statistics is our future work.

#### CONFLICT OF INTEREST

The authors declare no conflict of interest.

#### ACKNOWLEDGEMENTS

We thank the Associate Editor and two reviewers for their insightful and helpful suggestions, which have led to improvement of the presentation. This work was supported by the National Natural Science Foundation of China (grants 81373098 and 81072386), Science and Technology Planning Project of Guangdong Province, China (grant 2013B021800038), the National and Guangdong University Students' Innovation and Enterprise Training Project

of China (grants 201312121014 and 1212113029), grant from School of Public Health and Tropical Medicine of Southern Medical University, China (grant GW201422) and the Hong Kong RGC GRF Research Grant (HKU 766511M). We are also grateful to Dr Jie Ding who shares his R code with us.

- 1 Pfeifer, K. Mechanisms of genomic imprinting. *Am. J. Hum. Genet.* **67**, 777–787 (2000).
- 2 Reik, W. & Walter, J. Genomic imprinting: parental influence on the genome. *Nat. Rev. Genet.* **2**, 21–32 (2001).
- 3 Falls, J. G., Pulford, D. J., Wylie, A. A. & Jirtle, R. L. Genomic imprinting: implications for human disease. *Am. J. Pathol.* **154**, 635–647 (2001).
- 4 Ziegler, A., König, I. R. & Pahlke, F. *A Statistical Approach to Genetic Epidemiology: Concepts and Applications, with an E-Learning Platform* (Wiley-Vch, Weinheim, Germany, 2010).
- 5 Zhou, J. Y., Mao, W. G., Li, D. L., Hu, Y. Q., Xia, F. & Fung, W. K. A powerful parent-of-origin effects test for qualitative traits incorporating control children in nuclear families. *J. Hum. Genet.* **57**, 500–507 (2012).
- 6 Morison, I. M., Paton, C. J. & Cleverley, S. D. The imprinted gene and parent-of-origin effect database. *Nucl. Acids Res.* **29**, 275–276 (2001).
- 7 Zhou, J. Y., Ding, J., Fung, W. K. & Lin, S. Detection of parent-of-origin effects using general pedigree data. *Genet. Epidemiol.* **34**, 151–158 (2010).
- 8 Yang, J. & Lin, S. Likelihood approach for detecting imprinting and in utero maternal effects using general pedigrees from prospective family-based association studies. *Biometrics* **68**, 477–485 (2012).
- 9 Yang, J. & Lin, S. Robust partial likelihood approach for detecting imprinting and maternal effects using case-control families. *Ann. Appl. Stat.* **7**, 249–268 (2013).
- 10 Lin, D., Weinberg, C. R., Feng, R., Hochner, H. & Chen, J. A multi-locus likelihood method for assessing parent-of-origin effects using case-control mother-child pairs. *Genet. Epidemiol.* **37**, 152–162 (2013).
- 11 Ainsworth, H. F., Unwin, J., Jamison, D. L. & Cordell, H. J. Investigation of maternal effects, maternal-fetal interactions and parent-of-origin effects (imprinting), using mothers and their offspring. *Genet. Epidemiol.* **35**, 19–45 (2011).
- 12 Xia, F., Zhou, J. Y. & Fung, W. K. A powerful approach for association analysis incorporating imprinting effects. *Bioinformatics* **27**, 2571–2577 (2011).
- 13 Xia, F., Zhou, J. Y. & Fung, W. K. Powerful tests for association on quantitative trait loci incorporating imprinting effects. *J. Hum. Genet.* **58**, 384–390 (2013).
- 14 Ding, J., Lin, S. & Liu, Y. Monte Carlo pedigree disequilibrium test for markers on the X chromosome. *Am. J. Hum. Genet.* **79**, 567–573 (2006).
- 15 Martin, E. R., Monks, S. A., Warren, L. L. & Kaplan, N. L. A test for linkage and association in general pedigrees: the pedigree disequilibrium test. *Am. J. Hum. Genet.* **67**, 146–154 (2000).
- 16 Strauch, K., Fimmers, R., Kurz, T., Deichmann, K. A., Wienker, T. F. & Baur, M. P. Parametric and nonparametric multipoint linkage analysis with imprinting and two-locus-trait models: application to mite sensitization. *Am. J. Hum. Genet.* **66**, 1945–1957 (2000).
- 17 Hu, Y. Q., Zhou, J. Y. & Fung, W. K. An extension of the transmission disequilibrium test incorporating imprinting. *Genetics* **175**, 1489–1504 (2007).
- 18 Weinberg, C. R. Methods for detection of parent-of-origin effects in genetic studies of case-parents triads. *Am. J. Hum. Genet.* **65**, 229–235 (1999).
- 19 Li, X., Sui, Y., Liu, T., Wang, J., Li, Y. & Lin, Z. *et al.* A model for family-based case-control studies of genetic imprinting and epistasis. *Brief. Bioinform.* **15**, 1069–1079 (2013).
- 20 Weeks, D. E., Ott, J. & Lathrop, G. M. SLINK: a general simulation program for linkage analysis. *Am. J. Hum. Genet.* **47**, A204 (1990).

Supplementary Information accompanies the paper on Journal of Human Genetics website (<http://www.nature.com/jhg>)