

METHODOLOGY ARTICLE

Open Access



Generalized disequilibrium test for association in qualitative traits incorporating imprinting effects based on extended pedigrees

Jian-Long Li^{1,2}, Peng Wang¹, Wing Kam Fung^{3*} and Ji-Yuan Zhou^{1*} 

Abstract

Background: For dichotomous traits, the generalized disequilibrium test with the moment estimate of the variance (GDT-ME) is a powerful family-based association method. Genomic imprinting is an important epigenetic phenomenon and currently, there has been increasing interest of incorporating imprinting to improve the test power of association analysis. However, GDT-ME does not take imprinting effects into account, and it has not been investigated whether it can be used for association analysis when the effects indeed exist.

Results: In this article, based on a novel decomposition of the genotype score according to the paternal or maternal source of the allele, we propose the generalized disequilibrium test with imprinting (GDTI) for complete pedigrees without any missing genotypes. Then, we extend GDTI and GDT-ME to accommodate incomplete pedigrees with some pedigrees having missing genotypes, by using a Monte Carlo (MC) sampling and estimation scheme to infer missing genotypes given available genotypes in each pedigree, denoted by MCGDTI and MCGDT-ME, respectively. The proposed GDTI and MCGDTI methods evaluate the differences of the paternal as well as maternal allele scores for all discordant relative pairs in a pedigree, including beyond first-degree relative pairs. Advantages of the proposed GDTI and MCGDTI test statistics over existing methods are demonstrated by simulation studies under various simulation settings and by application to the rheumatoid arthritis dataset. Simulation results show that the proposed tests control the size well under the null hypothesis of no association, and outperform the existing methods under various imprinting effect models. The existing GDT-ME and the proposed MCGDT-ME can be used to test for association even when imprinting effects exist. For the application to the rheumatoid arthritis data, compared to the existing methods, MCGDTI identifies more loci statistically significantly associated with the disease.

Conclusions: Under complete and incomplete imprinting effect models, our proposed GDTI and MCGDTI methods, by considering the information on imprinting effects and all discordant relative pairs within each pedigree, outperform all the existing test statistics and MCGDTI can recapture much of the missing information. Therefore, MCGDTI is recommended in practice.

Keywords: Generalized disequilibrium test, Genomic imprinting, Monte Carlo sampling, Qualitative trait

* Correspondence: wingfung@hku.hk; zhoujiyuan5460@hotmail.com

³Department of Statistics and Actuarial Science, The University of Hong Kong, Hong Kong, China

¹State Key Laboratory of Organ Failure Research, Ministry of Education, and Guangdong Provincial Key Laboratory of Tropical Disease Research, Department of Biostatistics, School of Public Health, Southern Medical University, Guangzhou, China

Full list of author information is available at the end of the article



Background

Genomic imprinting is an important epigenetic phenomenon in studying complex traits, where the expression levels of certain genes rely on their parental origin [1–3]. Morison et al. [4, 5] constructed an imprinted gene and parent-of-origin effect database to collect genes that show imprinting effects, which has been updated by Glaser et al. [6] to include parental origin of de novo mutations. Furthermore, some researches have demonstrated that genomic imprinting plays an important role in several human genetic diseases such as Beckwith-Wiedemann syndrome, Silver-Russell syndrome, pseudohypoparathyroidism and transient neonatal diabetes mellitus [7–10].

For a diallelic marker locus, there have been many family-based methods to test for the association between genotype scores and dichotomous traits [11–15]. Among them, the generalized disequilibrium test with the moment estimate of the variance (GDT-ME) [15] is a powerful method, which is the generalization of the traditional transmission disequilibrium test [11] by using the genotype differences between all discordant relative pairs (including those beyond first-degree relatives) within a family. Currently, there has been increasing interest of incorporating imprinting to improve the test power of association analysis. However, GDT-ME does not take imprinting effects into account, and it has not been investigated whether it can be used for association analysis when the effects indeed exist. On the other hand, Xia et al. [16] developed the transmission disequilibrium test with imprinting for qualitative traits based on two-generation nuclear families, while it is not suitable for extended pedigrees. As such, the pedigree disequilibrium test with imprinting (PDTI) and its extension Monte Carlo (MC) PDTI (MCPDTI) to accommodate pedigrees with missing genotypes were proposed to test for association, which consider the influence of imprinting on association study [17]. However, they only utilize the genotype differences between all first-degree relative pairs in a family, which may reduce their test powers if ignoring the information on the genotype differences between beyond first-degree relatives.

To incorporate imprinting effects into association analysis, in this article, we develop a novel decomposition of the genotype score of each individual according to the paternal or maternal source of the allele. Based on these paternal and maternal allele scores, we propose the generalized disequilibrium test with imprinting (GDTI) for association for complete pedigrees without any missing genotypes. Then, borrowing the idea of Zhou et al. [18] and Ding et al. [19], we further extend GDTI and GDT-ME to accommodate incomplete pedigrees where the genotypes of some individuals in pedigrees are missing, based on a MC sampling and

estimation scheme to infer the missing genotypes given the observed genotypes in each pedigree, which are denoted by MCGDTI and MCGDT-ME, respectively. Advantages of the proposed GDTI and MCGDTI test statistics over existing methods are demonstrated by simulation studies under various simulation settings and by application to the rheumatoid arthritis (RA) dataset [20]. Simulation results show that the proposed GDTI, MCGDTI and MCGDT-ME control the type I error rates well under the null hypothesis of no association and no imprinting. The existing GDT-ME and the proposed MCGDT-ME can be used to test for association even when imprinting effects exist. MCGDTI can recapture much of the missing information. Further, the proposed tests outperform the existing methods under complete, incomplete and no imprinting effect models. For the real data application, compared to the existing methods, MCGDTI identifies more loci statistically significantly associated with RA after Bonferroni correction.

Methods

Notations

Suppose a diallelic marker locus with alleles M_1 and M_2 , and three possible genotypes are respectively M_2M_2 , M_1M_2 and M_1M_1 . We consider a disease susceptibility locus with the disease allele D and the normal one d , and the corresponding ordered genotypes are D/D , D/d , d/D and d/d with penetrances f_2 , f_{10} , f_{01} and f_0 , respectively. $f_{10}=f_{01}$ indicates no imprinting effects at the disease susceptibility locus. Further, the coefficient of linkage disequilibrium (LD) between alleles M_1 and D is taken as $LD = P(DM_1) - P_D P_{M_1}$, where $P(DM_1)$ is the frequency of haplotype DM_1 , and P_D and P_{M_1} are the allele frequency of D and M_1 , respectively. Suppose that we collect n independent pedigrees. Within the i^{th} pedigree which contains N_i family members ($i=1, 2, \dots, n$), without loss of generality, we assume that the first A_i individuals are affected and the other $U_i = N_i - A_i$ members are unaffected. Let Y_{ij} be the disease status of the j^{th} individual in the i^{th} pedigree ($i=1, 2, \dots, n$; $j=1, 2, \dots, N_i$), i.e. $Y_{ij}=1$ (0) denotes that the individual is affected (unaffected).

Existing generalized disequilibrium test with moment estimate of variance

We begin by describing the existing GDT-ME test [15]. For convenience, we define the genotype score X_{ij} by the number of allele M_1 in the genotype of the j^{th} individual in the i^{th} pedigree, i.e. $X_{ij}=0, 1$ and 2 for the genotypes M_2M_2 , M_1M_2 and M_1M_1 , respectively. As such, the logistic regression model is

$$\log \frac{P(Y_{ij} = 1)}{1 - P(Y_{ij} = 1)} = \beta_0 + \beta_1 X_{ij}, \tag{1}$$

where β_0 is the intercept, and β_1 is the regression coefficient; Y_{ij} is the disease status of the j^{th} individual in the i^{th} pedigree. Then, the GDT-ME test statistic can be expressed as follows, which is used to model the association between the disease status and X_{ij} :

$$\begin{aligned} \text{GDT-ME} &= \frac{\sum_{i=1}^n S_i}{\sqrt{\sum_{i=1}^n S_i^2}} \\ &= \frac{\sum_{i=1}^n \sum_{j=1}^{A_i} \sum_{k=A_i+1}^{N_i} (X_{ij} - X_{ik}) \frac{1}{N_i}}{\sqrt{\sum_{i=1}^n \left(\sum_{j=1}^{A_i} \sum_{k=A_i+1}^{N_i} (X_{ij} - X_{ik}) \frac{1}{N_i} \right)^2}}, \end{aligned} \tag{2}$$

where $S_i = \sum_{j=1}^{A_i} \sum_{k=A_i+1}^{N_i} (X_{ij} - X_{ik}) \frac{1}{N_i}$ is the score of the i^{th} pedigree and $\sum_{i=1}^n S_i^2$ is an unbiased moment estimate of the variance of $\sum_{i=1}^n S_i$. The variance of $\sum_{i=1}^n S_i$ can also be estimated based on the information on kinship coefficients when identity by descent (IBD) is unknown [15]. For convenience, we denote the corresponding test statistic by GDT in this article.

GDTI for complete pedigree data

Although GDT-ME is a powerful association test and is robust to population stratification (PS) [15], it does not take the information on imprinting effects into consideration. In this article, we are going to investigate whether GDT-ME can be used to test for association when there are imprinting effects. Moreover, we propose the following generalized disequilibrium test incorporating imprinting effects (GDTI). Note that in GDT-ME, the genotype score X_{ij} is coded as the counts of allele M_1 for the j^{th} individual in the i^{th} pedigree, i.e.

$$X_{ij} = \begin{cases} 0, & M_2M_2 \\ 1, & M_1M_2 \\ 2, & M_1M_1 \end{cases}$$

To incorporate the information on imprinting effects into analysis, we divide the X_{ij} into two parts, $X_{ij}^{(p)}$ and $X_{ij}^{(m)}$, according to the paternal or maternal source of the allele, where $X_{ij} = X_{ij}^{(p)} + X_{ij}^{(m)}$, and $X_{ij}^{(p)}$ and $X_{ij}^{(m)}$ are respectively coded as follows:

$$X_{ij}^{(p)} = \begin{cases} 0, & \text{if the individual's genotype is } M_2M_2, \text{ which indicates that} \\ & \text{one of two } M_2\text{'s came from father, or } M_1M_2 \text{ with } M_1 \\ & \text{coming from mother} \\ 0.5, & \text{if the individual's genotype is } M_1M_2, \text{ but it is not sure} \\ & \text{which allele came from father (i.e. the } M_1M_2 \text{ founders,} \\ & \text{or the } M_1M_2 \text{ nonfounders with both parents being} \\ & \text{heterozygous)} \\ 1, & \text{if the individual's genotype is } M_1M_1, \text{ or } M_1M_2 \text{ with } M_1 \\ & \text{coming from father} \end{cases}$$

and

$$X_{ij}^{(m)} = \begin{cases} 0, & \text{if the individual's genotype is } M_2M_2, \text{ which} \\ & \text{indicates that one of two } M_2\text{'s came from} \\ & \text{mother, or } M_1M_2 \text{ with } M_1 \text{ coming from father} \\ 0.5, & \text{if the individual's genotype is } M_1M_2, \text{ but it is not} \\ & \text{sure which allele came from mother} \\ & \text{(i.e. the } M_1M_2 \text{ founders, or the } M_1M_2 \text{ nonfounders} \\ & \text{with both parents being heterozygous)} \\ 1, & \text{if the individual's genotype is } M_1M_1, \text{ or } M_1M_2 \\ & \text{with } M_1 \text{ coming from mother} \end{cases}$$

We call $X_{ij}^{(p)}$ and $X_{ij}^{(m)}$ the paternal allele score and the maternal allele score, respectively. So, we use the following logistic regression to model the association between the disease status Y_{ij} and the allele scores $X_{ij}^{(p)}$ and $X_{ij}^{(m)}$:

$$\log \frac{P(Y_{ij} = 1)}{1 - P(Y_{ij} = 1)} = \beta_0 + \beta_p X_{ij}^{(p)} + \beta_m X_{ij}^{(m)},$$

where β_0 is the intercept, and β_p and β_m are the regression coefficients; β_p is used to describe the effect of allele M_1 coming from his (her) father, and β_m measures the effect of allele M_1 whose parental origin is his (her) mother. The null hypothesis $H_0: \beta_p = \beta_m = 0$ denotes no association and no imprinting; $\beta_p = \beta_m \neq 0$ indicates that the association exists while there are no imprinting effects, and the logistic regression model can be reduced to the model of GDT-ME (Equation (1)); $\beta_p \neq \beta_m$ represents that both association and imprinting effects exist. As such,

$$P(Y_{ij} = 1) = \frac{\exp(\beta_0 + \beta_p X_{ij}^{(p)} + \beta_m X_{ij}^{(m)})}{1 + \exp(\beta_0 + \beta_p X_{ij}^{(p)} + \beta_m X_{ij}^{(m)})}.$$

Note that the disease statuses of all the family members in each pedigree are uncorrelated, conditional on their own genotypes at the marker locus. Then, the likelihood that the first A_i individuals are affected, conditional on the fact that there are A_i affected individuals in total in the i^{th} pedigree, is (the detailed derivation refers to Additional file 1: Appendix):

$$P\left(\sum_{j=1}^{A_i} Y_{ij} = A_i \mid \sum_{j=1}^{N_i} Y_{ij} = A_i\right) = \frac{P\left(\sum_{j=1}^{A_i} Y_{ij} = A_i\right)}{\sum_{s_l} P\left(\sum_{j \in s_l} Y_{ij} = A_i\right)}$$

$$= \frac{\exp\left\{\frac{1}{U_i} \sum_{j=1}^{A_i} \sum_{k=A_i+1}^{N_i} \left[\left(X_{ij}^{(p)} - X_{ik}^{(p)}\right) \beta_p + \left(X_{ij}^{(m)} - X_{ik}^{(m)}\right) \beta_m \right]\right\}}{\sum_{s_l} \exp\left\{\frac{1}{U_i} \sum_{j \in s_l} \sum_{k=A_i+1}^{N_i} \left[\left(X_{ij}^{(p)} - X_{ik}^{(p)}\right) \beta_p + \left(X_{ij}^{(m)} - X_{ik}^{(m)}\right) \beta_m \right]\right\}},$$

where s_l 's are all the possible combination that A_i out of N_i individuals are affected by shuffling the affection statuses of all the N_i individuals in the i^{th} pedigree; s_l is the l^{th} possible combination; $U_i = N_i - A_i$ is the number of unaffected individuals in the i^{th} pedigree. As such, the log-likelihood function for the i^{th} pedigree is

$$l_i = \frac{1}{U_i} \sum_{j=1}^{A_i} \sum_{k=A_i+1}^{N_i} \left[\left(X_{ij}^{(p)} - X_{ik}^{(p)}\right) \beta_p + \left(X_{ij}^{(m)} - X_{ik}^{(m)}\right) \beta_m \right]$$

$$- \log \sum_{s_l} \exp\left(\frac{1}{U_i} \sum_{j \in s_l} \sum_{k=A_i+1}^{N_i} \left[\left(X_{ij}^{(p)} - X_{ik}^{(p)}\right) \beta_p + \left(X_{ij}^{(m)} - X_{ik}^{(m)}\right) \beta_m \right]\right).$$

Under the null hypothesis of no association ($H_0: \beta_p = \beta_m = 0$), the score test statistic for testing for association incorporating imprinting effects is formulated as follows (the details see Additional file 1: Appendix),

$$\text{GDTI} = \left(\sum_{i=1}^n D_{i1} \quad \sum_{i=1}^n D_{i2} \right) \cdot \left(\sum_{i=1}^n I_{i11} \quad \sum_{i=1}^n I_{i12} \right)^{-1} \cdot \left(\sum_{i=1}^n D_{i1} \right), \quad (3)$$

where $\sum_{i=1}^n D_{i1}$ and $\sum_{i=1}^n D_{i2}$ are the scores of β_p and β_m , respectively;

$\left(\sum_{i=1}^n I_{i11} \quad \sum_{i=1}^n I_{i12} \right)$ is the observed Fisher's information matrix of β_p and β_m ;

$$D_{i1} = \frac{1}{N_i} \sum_{j=1}^{A_i} \sum_{k=A_i+1}^{N_i} \left(X_{ij}^{(p)} - X_{ik}^{(p)} \right), \quad D_{i2} = \frac{1}{N_i} \sum_{j=1}^{A_i} \sum_{k=A_i+1}^{N_i} \left(X_{ij}^{(m)} - X_{ik}^{(m)} \right),$$

$$I_{i11} = \left. \frac{\partial^2 l_i}{\partial \beta_p^2} \right|_{\beta_p=0, \beta_m=0}$$

$$= \frac{\left(\frac{1}{U_i}\right)^2 \sum_{s_l} \left[\sum_{j \in s_l} \sum_{k=A_i+1}^{N_i} \left(X_{ij}^{(p)} - X_{ik}^{(p)} \right) \right]^2}{\binom{N_i}{A_i}}$$

$$- \left(\frac{A_i}{U_i N_i}\right)^2 \left[\sum_{j=1}^{A_i} \sum_{k=A_i+1}^{N_i} \left(X_{ij}^{(p)} - X_{ik}^{(p)} \right) \right]^2,$$

$$I_{i22} = \left. \frac{\partial^2 l_i}{\partial \beta_m^2} \right|_{\beta_p=0, \beta_m=0}$$

$$= \frac{\left(\frac{1}{U_i}\right)^2 \sum_{s_l} \left[\sum_{j \in s_l} \sum_{k=A_i+1}^{N_i} \left(X_{ij}^{(m)} - X_{ik}^{(m)} \right) \right]^2}{\binom{N_i}{A_i}},$$

$$- \left(\frac{A_i}{U_i N_i}\right)^2 \left[\sum_{j=1}^{A_i} \sum_{k=A_i+1}^{N_i} \left(X_{ij}^{(m)} - X_{ik}^{(m)} \right) \right]^2$$

and

$$I_{i12} = I_{i21} = \left. \frac{\partial^2 l_i}{\partial \beta_p \partial \beta_m} \right|_{\beta_p=0, \beta_m=0}$$

$$= \frac{\left(\frac{1}{U_i}\right)^2 \sum_{s_l} \left\{ \left[\sum_{j \in s_l} \sum_{k=A_i+1}^{N_i} \left(X_{ij}^{(p)} - X_{ik}^{(p)} \right) \right] \left[\sum_{j \in s_l} \sum_{k=A_i+1}^{N_i} \left(X_{ij}^{(m)} - X_{ik}^{(m)} \right) \right] \right\}}{\binom{N_i}{A_i}}$$

$$- \left(\frac{A_i}{U_i N_i}\right)^2 \left[\sum_{j=1}^{A_i} \sum_{k=A_i+1}^{N_i} \left(X_{ij}^{(p)} - X_{ik}^{(p)} \right) \right] \left[\sum_{j=1}^{A_i} \sum_{k=A_i+1}^{N_i} \left(X_{ij}^{(m)} - X_{ik}^{(m)} \right) \right].$$

GDTI asymptotically follows a chi-square distribution with the degrees of freedom being 2, under the null hypothesis of no association and no imprinting. It is noted from the above that the scores D_{i1} and D_{i2} evaluate the differences in paternal allele scores and maternal allele scores, respectively, for all discordant relative pairs in a pedigree, thus utilizing information beyond first-degree relative pairs. This is in contrast to other association testing methods under imprinting (e.g. PDTI), where extended pedigrees are considered as multiple nuclear families, and so information is not fully utilized.

MCGDTI and MCGDT-ME for incomplete pedigree data

When the genotypes of some individuals in a pedigree are missing, GDTI cannot be used directly. Therefore, in presence of missingness, we extend GDTI and propose MCGDTI based on a MC sampling and estimation process, which may recapture most information on missing genotypes based on the observed genotypes. Specifically, we replace D_{i1} , D_{i2} , I_{i11} , I_{i12} , I_{i21} and I_{i22} in GDTI by their conditional expectations, D_{i1MC} , D_{i2MC} , I_{i11MC} , I_{i12MC} , I_{i21MC} and I_{i22MC} , given the observed genotypes, G_o , where $T_{MC} = E(T(G_m, G_o, A) | G_o)$ for some statistic T , G_m is the set of missing genotypes; A is the collection of the observed phenotypes (disease affection statuses); $T(G_m, G_o, A)$ is the expanded notation of T to explicitly show its dependences on the missing genotypes G_m , the observed genotypes G_o and the observed phenotype collection A . Following Zhou et al. [18] and Ding et al. [19], we estimate D_{i1MC} , D_{i2MC} , I_{i11MC} , I_{i12MC} , I_{i21MC} and I_{i22MC} based on a MC simulation scheme. Specifically, if we set the MC size to be K , then we draw independent sample G_{mk} , $k = 1, 2, \dots, K$, from $P(G_m | G_o)$,

which can be accomplished efficiently based on the peeling algorithm using the SLINK software [21]. The statistic D_{i1MC} can be estimated by $\widehat{D}_{i1MC} = \frac{1}{K} \sum_{k=1}^K D_{i1}(G_{mk}, G_o, A)$. D_{i2MC} , I_{i11MC} , I_{i12MC} , I_{i21MC} and I_{i22MC} can be similarly estimated by \widehat{D}_{i2MC} , \widehat{I}_{i11MC} , \widehat{I}_{i12MC} , \widehat{I}_{i21MC} and \widehat{I}_{i22MC} , respectively. Then, the MCGDTI statistic is calculated after replacing D_{i1} , D_{i2} , I_{i11} , I_{i12} , I_{i21} and I_{i22} in Equation (3) by the corresponding \widehat{D}_{i1MC} , \widehat{D}_{i2MC} , \widehat{I}_{i11MC} , \widehat{I}_{i12MC} , \widehat{I}_{i21MC} and \widehat{I}_{i22MC} values, respectively. MCGDTI has an asymptotic chi-square distribution with the degrees of freedom being 2 under the null hypothesis.

Earlier studies showed that the transmission disequilibrium test can be employed for association analysis even when there are imprinting effects [16], and we find out that GDT-ME can also be used for such a purpose (see simulation studies later). In this connection, for incomplete pedigree data, we extend GDT-ME without considering imprinting effects and propose MCGDT-ME to test for association based on the MC sampling and estimation scheme. Being similar to MCGDTI, the MCGDT-ME statistic can be calculated, as before, but substituting each S_i in Equation (2) by $S_{iMC} = \frac{1}{K} \sum_{k=1}^K S_i(G_{mk}, G_o, A)$, i.e. $MCGDT-ME = \sum_{i=1}^n S_{iMC} / \sqrt{\sum_{i=1}^n S_{iMC}^2}$. MCGDT-ME follows a standard normal distribution approximately under the null hypothesis of no association.

Simulation settings

In this section, to explore the performance of the proposed GDTI, MCGDTI and MCGDT-ME statistics and compare the powers of GDTI, MCGDTI and MCGDT-ME with the existing MCPDTI, GDT-ME and GDT, we conduct the following simulation studies. We consider a homogeneous population. The marker locus and the disease susceptibility locus are in complete linkage. Three groups of haplotype frequencies for haplotypes DM_1 , dM_1 , DM_2 and dM_2 are considered to simulate the powers: LD1: {0.13, 0.02, 0.12, 0.73}, LD2: {0.23, 0.12, 0.02, 0.63} and LD3: {0.22, 0.03, 0.03, 0.72}, where the frequency P_{M_1} of marker allele M_1 for each group is 0.15, 0.35 and 0.25 with the frequency P_D of the disease allele D being fixed at 0.25, and the corresponding LD values are 0.092,5, 0.142,5 and 0.157,5, respectively. To investigate the empirical type I error rates under the null hypothesis of no association, the frequencies of four haplotypes are taken as the product of two allele frequencies on each haplotype, respectively. For example, when $P_{M_1} = 0.15$, the frequency of haplotype DM_1 is $P(DM_1) = 0.15 \times 0.25 = 0.0375$.

Three sets of two homozygote penetrances f_2 and f_0 for genotypes D/D and d/d , {0.390, 0.260}, {0.440, 0.240}

and {0.480, 0.220}, are investigated with the corresponding relative risk ($RR=f_2/f_0$) being 1.500, 1.833 and 2.182, respectively, which are similar to those in Ding et al. [19]. For each set of homozygote penetrances, three imprinting effect models by setting the various values of f_{10} and f_{01} are considered: no, incomplete and complete imprinting effect models. For no imprinting effect model, we set $f_1 = f_{10} = f_{01} = (f_2 + f_0)/2$. Note that no association implies no imprinting effects. So, we simulate the type I error rates of the proposed test statistics only under no association and no imprinting. Tables 1 and 2 give the simulation settings for studying the empirical size and the test power, respectively.

In addition, three types of pedigree structure are considered in our simulation study. The pedigree structures are shown in Fig. 1: (a) two-generation family with 5 individuals, (b) three-generation pedigree with 10 individuals, and (c) four-generation pedigree with 12 individuals. In each replicate, we simulate 30 (50) pedigrees under each pedigree structure and the resulting total sample size is 90 (150). Here the ascertainment scheme for a pedigree to be included is that there is at least one affected nonfounder in the pedigree. For MCGDTI, MCGDT-ME and MCPDTI, 50 MC samples of missing genotypes are generated for each replicate with use of the SLINK software [21]. In the MC sampling process, both the true marker allele frequencies and those estimated from the genotyped founders in each replicate are used.

For assessing the performance of the proposed tests (GDTI, MCGDTI and MCGDT-ME) and for comparing with the existing GDT-ME and GDT without considering imprinting effects [15], and MCPDTI with incorporating imprinting [17], we consider the following 9 tests. GDTI is based on complete data assuming no missing genotypes. The other 8 tests are for incomplete data, after the removal of the 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. MCGDTI_T, MCGDT-ME_T and

Table 1 Simulation settings for estimating size

Setting	P_{M_1}	f_2	f_1	f_0	RR
1	0.15	0.390	0.325	0.260	1.500
2	0.15	0.440	0.340	0.240	1.833
3	0.15	0.480	0.350	0.220	2.182
4	0.35	0.390	0.325	0.260	1.500
5	0.35	0.440	0.340	0.240	1.833
6	0.35	0.480	0.350	0.220	2.182
7	0.25	0.390	0.325	0.260	1.500
8	0.25	0.440	0.340	0.240	1.833
9	0.25	0.480	0.350	0.220	2.182

Table 2 Simulation settings for estimating power

A Haplotype frequencies					
LD setting	DM ₁	dM ₁	DM ₂	dM ₂	LD
LD1	0.130	0.020	0.120	0.730	0.0925
LD2	0.230	0.120	0.020	0.630	0.1425
LD3	0.220	0.030	0.030	0.720	0.1575
B Penetrances and imprinting effect models					
Imprinting effect model	f ₂	f ₁₀	f ₀₁	f ₀	RR
No	0.390	0.325	0.325	0.260	1.500
	0.440	0.340	0.340	0.240	1.833
	0.480	0.350	0.350	0.220	2.182
Incomplete	0.390	0.370	0.280	0.260	1.500
	0.440	0.420	0.260	0.240	1.833
	0.480	0.460	0.240	0.220	2.182
Complete	0.390	0.390	0.260	0.260	1.500
	0.440	0.440	0.240	0.240	1.833
	0.480	0.480	0.220	0.220	2.182

MCPDTI_T are on the basis of the true marker allele frequencies, while MCGDTI_E, MCGDT-ME_E and MCPDTI_E are based on the estimated marker allele frequencies. GDT-ME and GDT are also considered for incomplete data. Under each simulation setting, 10,000 replicates are simulated and the significance level is set at 1%. All the simulations are implemented by using the R software (version 3.4.1) [22].

Results

Size and power

Under 9 simulation settings given in Table 1, the empirical type I error rates of GDTI, MCGDTI_T, MCGDTI_E, MCGDT-ME_T, MCGDT-ME_E, GDT-ME, GDT, MCPDTI_T and MCPDTI_E are demonstrated in Table 3, based on 90 and 150 pedigrees at the 1% significance level, respectively. It is shown in Table 3 that the size of all the methods is generally close to the nominal level 1% under the null

hypothesis of no association and no imprinting, irrespective of different sample sizes. Thus, our proposed GDTI, MCGDTI_T, MCGDTI_E, MCGDT-ME_T and MCGDT-ME_E test statistics are valid for testing association.

Figures 2, 3 and 4 give the simulated powers of GDTI, MCGDTI_T, MCGDTI_E, MCGDT-ME_T, MCGDT-ME_E, GDT-ME, GDT, MCPDTI_T and MCPDTI_E based on 150 pedigrees at the 1% significance level under complete, incomplete and no imprinting effect models for different LD and RR values, respectively. The first 5 statistics are proposed tests, while the remaining four are existing tests. Additional file 1: Figures S1 - S3 show the corresponding simulated powers of all the methods based on 90 pedigrees. From the figures, we find that the powers of MCGDTI, MCGDT-ME and MCPDTI based on the true marker allele frequencies are very close to those based on the estimated marker allele frequencies (MCGDTI_T vs MCGDTI_E, MCGDT-ME_T vs MCGDT-ME_E, and MCPDTI_T vs MCPDTI_E), respectively. MCGDTI_T and MCGDTI_E can recapture much of the missing information, which are a little less powerful than GDTI for complete pedigree data. The existing MCPDTI test performs the worst even though it is constructed for testing association when imprinting effects are taken into consideration. On the other hand, MCGDT-ME, GDT-ME and GDT, though without accounting for imprinting, can be used for testing association even when imprinting effects exist. Moreover, they outperform MCPDTI substantially. It is probably due to the fact that MCGDT-ME, GDT-ME and GDT consider genotype differences between all discordant relative pairs, thus utilizing much more information than first-degree relative pairs used by MCPDTI. In Fig. 2 under complete imprinting effect model, when the LD and RR values are fixed, the proposed GDTI (assuming the data are complete) and MCGDTI statistics have higher powers than all the other test statistics. GDT (based on the IBD information) has better performance than GDT-ME, which is the result similar to that in Chen et al. [15].

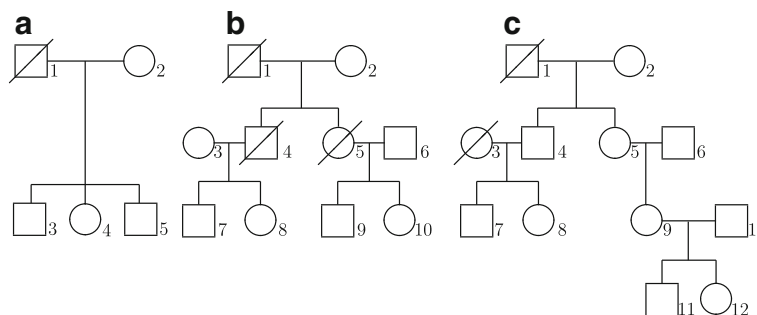


Fig. 1 Pedigree structures for the simulation studies. **a** Two-generation family. **b** Three-generation pedigree. **c** Four-generation pedigree. Genotypes of individual 1 in two-generation family, individuals 1, 4 and 5 in three-generation pedigree and individuals 1 and 3 in four-generation pedigree are assumed to be missing for the analysis based on incomplete data

Table 3 Empirical size (in percentage (%)) of GDTI, MCGDTI, MCGDT-ME, GDT-ME, GDT and MCPDTI^a

Setting	Complete data	Incomplete data							
	GDTI	MCGDTI _T	MCGDTI _E	MCGDT-ME _T	MCGDT-ME _E	GDT-ME	GDT	MCPDTI _T	MCPDTI _E
Based on 90 pedigrees									
1	0.98	1.04	1.10	0.95	0.99	0.82	0.85	0.80	0.84
2	1.14	1.12	1.14	1.12	1.06	0.82	0.87	0.66	0.74
3	0.99	1.05	0.99	0.91	0.94	0.80	0.87	0.81	0.86
4	0.85	0.97	1.05	0.98	0.91	0.74	0.90	0.91	0.91
5	0.96	1.11	1.11	0.96	0.88	0.87	1.03	0.73	0.81
6	1.11	1.13	1.14	0.84	0.81	0.98	1.05	0.90	0.89
7	1.13	0.85	0.96	0.96	0.88	0.70	0.82	0.94	1.05
8	1.09	1.11	1.13	0.86	0.86	0.72	0.76	0.83	0.86
9	1.11	1.14	1.13	1.05	1.12	0.82	0.89	1.04	1.00
Based on 150 pedigrees									
1	1.00	0.93	0.94	0.94	0.93	0.93	0.95	0.90	0.98
2	1.10	1.05	1.04	0.98	0.98	0.98	1.09	0.88	0.93
3	1.05	1.09	1.10	0.97	0.99	1.04	1.04	0.89	0.92
4	0.97	0.95	0.96	0.81	0.83	0.87	0.89	0.99	1.13
5	1.13	1.19	1.17	1.01	0.93	0.90	1.05	0.92	0.93
6	1.15	1.15	1.14	0.98	0.95	0.82	0.83	0.98	1.07
7	1.06	0.98	1.02	1.05	1.04	0.85	0.86	1.02	1.03
8	0.97	0.93	0.93	0.95	0.99	0.76	0.92	0.95	0.96
9	1.07	1.02	1.01	0.86	0.85	0.93	1.09	1.05	1.03

^aThe simulations are conducted under the null hypothesis of no association and no imprinting based on 10,000 replicates for 90 and 150 pedigrees at 1% significance level, respectively

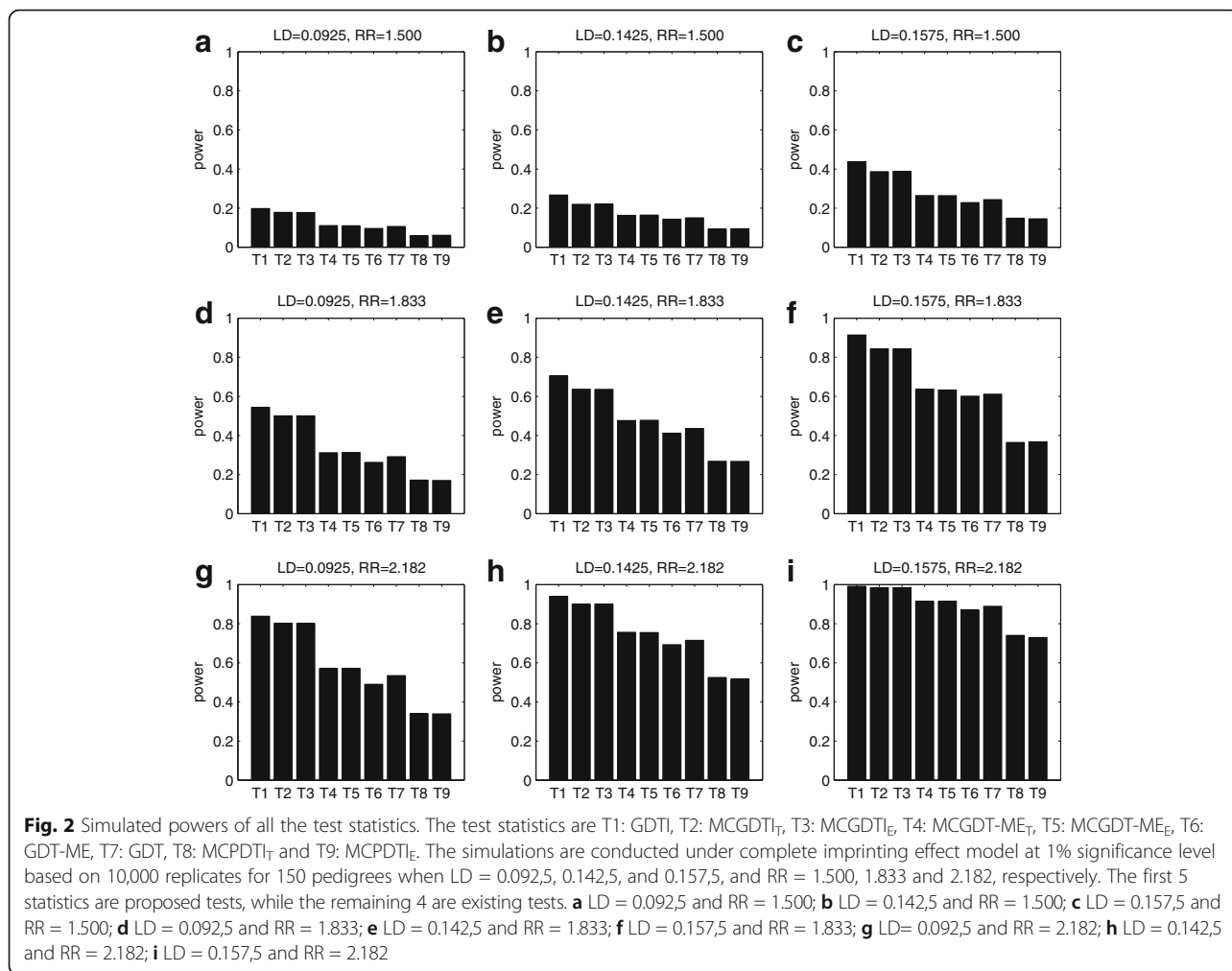
When the LD value changes from 0.092,5 to 0.157,5 and RR is unchanged, or the LD value is fixed and RR increases from 1.500 to 2.182, all the powers become larger and larger. The results in Fig. 3 under incomplete imprinting effect model are similar to those in Fig. 2. Figure 4 shows the performance of various tests under the no imprinting effect model. The proposed MCGDT-ME outperforms all the existing methods. MCGDTI is a bit less powerful than MCGDT-ME, as expected, and it has a similar performance to GDT-ME and GDT. By comparing the results in Figs. 2, 3 and 4, we find that when the imprinting effect model changes from complete model to incomplete one (i.e. the degree of imprinting effects decreases), the powers of the GDTI and MCGDTI are smaller and smaller. GDTI and MCGDTI attain the least powers under the no imprinting effect model. Finally, the powers of all the methods based on 150 pedigrees are higher than those based on 90 pedigrees (Fig. 2 vs Additional file 1: Figure S1, Fig. 3 vs Additional file 1: Figure S2, and Fig. 4 vs Additional file 1: Figure S3), respectively.

Application to RA data

We apply our proposed methods to the RA dataset from North American Rheumatoid Arthritis Consortium [20],

which is made available from Genetic Analysis Workshop 15 [23]. It has been approved by the providers of the RA data. In this dataset, a total of 757 pedigrees and 8017 individuals were collected, and 5407 autosomal single nucleotide polymorphisms (SNPs) were used. It should be noted that the genotypes of about 80% individuals are missing at these SNPs and thus the proposed MCGDTI (not GDTI) and MCGDT-ME methods are applied. To compare the performance of the proposed tests with the existing methods, we also implement the GDT-ME, GDT and MCPDTI methods in this real data analysis. On the other hand, note that there are 73 pedigree members with unknown affection statuses in this dataset. In addition, we use the existing Monte Carlo pedigree parental-asymmetry test (MCPPAT) to test if imprinting is present [18].

We use the following quality control rules to filter the data. First, a pedigree to be included has at least one affected nonfounder. Second, we delete pedigrees with stepfamilies. Finally, if the proportion of the individuals with missing genotypes among all the members in a pedigree is more than 50% based on the first SNP on Chromosome 1, then we exclude this pedigree. This can avoid the large variability on estimation created by pedigrees with high proportions of missingness. To this end,



we get 246 pedigrees with 1109 individuals. Among them, there are 11 individuals with the affection statuses being unavailable and we treat them as unaffected. We use all the available individuals (1992 individuals) in this dataset to estimate the marker allele frequencies, not just using the available founders, due to the large proportion of the individuals with missing genotypes in this dataset. Then, we calculate the values and the corresponding p-values of all the test statistics based on the estimated allele frequencies and 246 selected pedigrees. The significance level is fixed at $\alpha = 5\%$, and Bonferroni correction would test each individual hypothesis at the significance level of $\alpha' = 0.05/5407 = 9.247,3 \times 10^{-6}$, based on 5407 SNPs. The MC size for MCGDTI, MCGDT-ME, MCPDTI and MCPPAT is set to be 50.

The corresponding results of MCGDTI and MCGDT-ME at the significance level of $\alpha = 5\%$, with Bonferroni correction based on the p-values of these methods are shown in Table 4. From the table, MCGDTI identifies 3 SNPs statistically significantly associated with RA, which cannot be found by MCGDT-ME. Further, the 3 SNPs

identified by MCGDTI cannot be detected by GDT-ME, GDT and MCPDTI, and the corresponding contingency tables are the same as Table 4, which are not shown for brevity. The results from this real data application demonstrate a gain in information through incorporating imprinting effects (compared to MCGDT-ME), through making use of partially genotyped pedigrees (compared to GDT-ME and GDT), and through including the genotype differences between beyond first-degree relatives (compared to MCPDTI). In addition, we list the p-values of the association tests MCGDTI, MCGDT-ME, GDT-ME, GDT, MCPDTI and the imprinting test MCPPAT at these 3 SNPs in Additional file 1: Table S1. From the p-values of MCPPAT in this table, there are statistically significant imprinting effects at the 3 SNPs on RA, which may be why MCGDTI is more powerful than the other test statistics.

Discussion

In this article, based on a novel decomposition of the genotype score of an individual according to the paternal

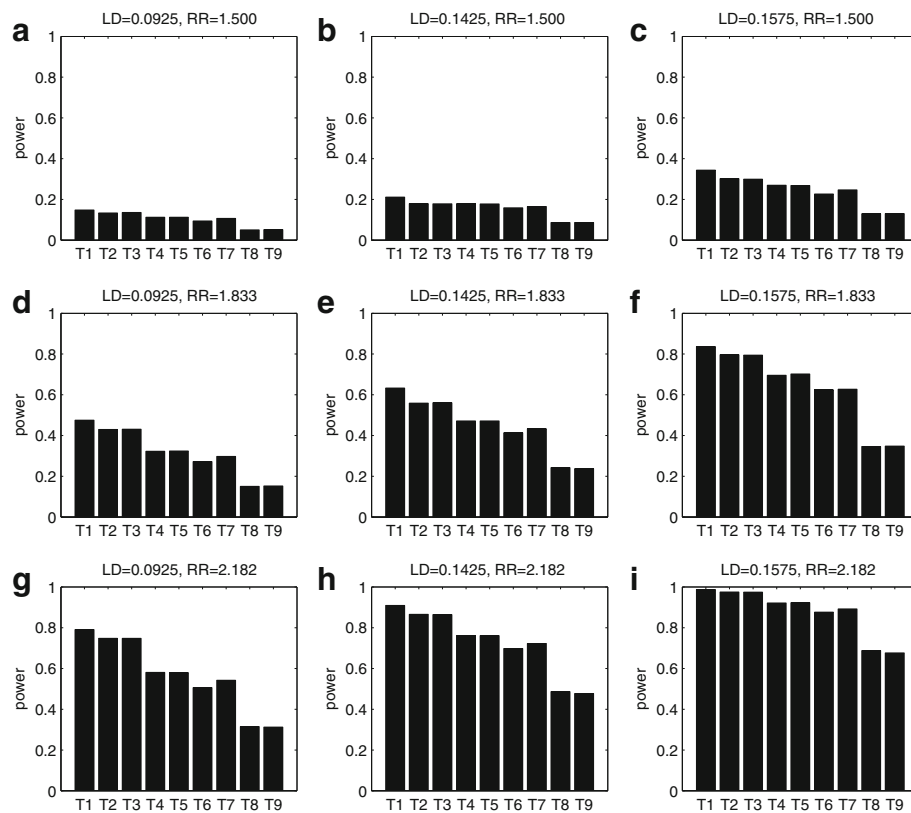


Fig. 3 Simulated powers of all the test statistics. The test statistics are T1: GDTI, T2: MCGDTI_T, T3: MCGDTI_E, T4: MCGDT-ME_T, T5: MCGDT-ME_E, T6: GDT-ME, T7: GDT, T8: MCPDTI_T and T9: MCPDTI_E. The simulations are conducted under incomplete imprinting effect model at 1% significance level based on 10,000 replicates for 150 pedigrees when LD = 0.092,5, 0.142,5, and 0.157,5, and RR = 1.500, 1.833 and 2.182, respectively. The first 5 statistics are proposed tests, while the remaining 4 are existing tests. **a** LD = 0.092,5 and RR = 1.500; **b** LD = 0.142,5 and RR = 1.500; **c** LD = 0.157,5 and RR = 1.500; **d** LD = 0.092,5 and RR = 1.833; **e** LD = 0.142,5 and RR = 1.833; **f** LD = 0.157,5 and RR = 1.833; **g** LD = 0.092,5 and RR = 2.182; **h** LD = 0.142,5 and RR = 2.182; **i** LD = 0.157,5 and RR = 2.182

or maternal source of an allele, we develop the GDTI test to test for association incorporating imprinting for complete pedigrees without missing genotypes. Then, using a MC sampling and estimation scheme, we extend GDTI and GDT-ME, and respectively develop MCGDTI and MCGDT-ME to deal with incomplete pedigrees, in which some individuals' genotypes are unavailable. Compared to PDTI and MCPDTI, GDTI and MCGDTI make use of the genotype differences between all discordant relative pairs, including beyond first-degree relatives. Simulation results indicate that GDTI, MCGDTI and MCGDT-ME control the size well under the null hypothesis of no association and no imprinting. As for the simulated powers, under complete and incomplete imprinting effect models, our proposed GDTI and MCGDTI methods by considering the information on imprinting effects and all discordant relative pairs outperform all the existing test statistics and MCGDTI can recapture much of the missing information. The application to the RA dataset also demonstrates the advantage of MCGDTI over other methods. Further, in this article,

we demonstrate that, the existing GDT-ME and the proposed MCGDT-ME, although not constructed under imprinting, can be used for testing association even when the effects exist. Moreover, we propose the MCGDT-ME test to handle incomplete pedigree data with missing genotypes, and the test is found to perform better than GDT-ME in simulation studies.

One of the major reasons for using within-family tests (e.g. GDT-ME and GDT) for association is their robustness to PS. On the other hand, note that MCGDTI, MCGDT-ME and MCPDTI need the MC sampling and estimation scheme to infer missing genotypes in pedigrees, which requires these pedigrees from a homogenous population. To investigate the performance of the proposed test statistics in the presence of PS, we consider a population consisting of two subpopulations and conduct the following simulation study. The parameters are set to be the same as those in Chen et al. [15]. Specifically, suppose that a disease susceptibility locus and a marker locus are in complete linkage but in linkage equilibrium and both allele frequencies P_D and

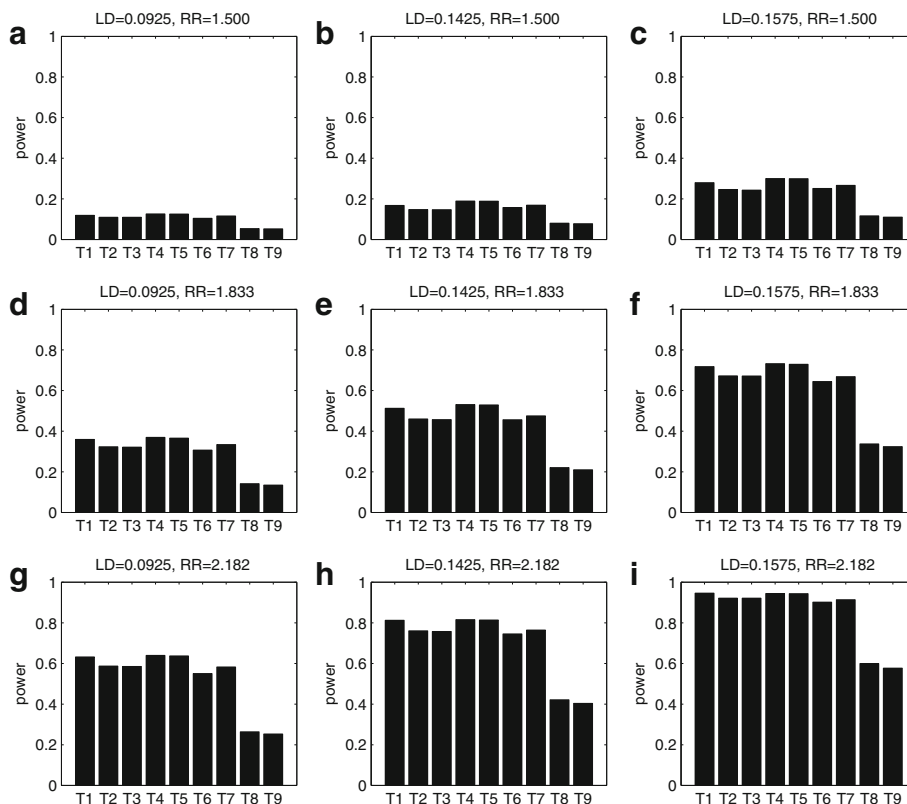


Fig. 4 Simulated powers of all the test statistics. The test statistics are T1: GDTI, T2: MCGDT_T, T3: MCGDT_E, T4: MCGDT-ME_T, T5: MCGDT-ME_E, T6: GDT-ME, T7: GDT, T8: MCPDT_T and T9: MCPDT_E. The simulations are conducted under no imprinting effect model at 1% significance level based on 10,000 replicates for 150 pedigrees when LD = 0.0925, 0.1425, and 0.1575, and RR = 1.500, 1.833 and 2.182, respectively. The first 5 statistics are proposed tests, while the remaining 4 are existing tests. **a** LD = 0.0925 and RR = 1.500; **b** LD = 0.1425 and RR = 1.500; **c** LD = 0.1575 and RR = 1.500; **d** LD = 0.0925 and RR = 1.833; **e** LD = 0.1425 and RR = 1.833; **f** LD = 0.1575 and RR = 1.833; **g** LD = 0.0925 and RR = 2.182; **h** LD = 0.1425 and RR = 2.182; **i** LD = 0.1575 and RR = 2.182

P_{M_1} are taken to be 0.1 (0.5) in the first (second) subpopulation. The penetrances f_2, f_{10}, f_{01} and f_0 of genotypes $D/D, D/d, d/D$ and d/d are assumed to be 0.45, 0.30, 0.30 and 0.20 in both subpopulations, respectively. In MCGDTI, MCGDT-ME and MCPDTI, the allele frequency P_{M_1} is estimated by genotyped founders from all the collected pedigrees, by assuming that they came from a single population, which may cause biases in the estimation of P_{M_1} . Two simulation scenarios of pedigree structure or level of genotypic missingness are considered. In the first scenario, 150 pedigrees (50 two-generation families, 50 three-generation pedigrees and 50 four-

generation pedigrees with the pedigree structures listed in Fig. 1) are sampled from each subpopulation and the only difference between two subpopulations is allele frequencies P_D and P_{M_1} . In the second scenario, 200 pedigrees (100 two-generation families and 100 three-generation pedigrees with the pedigree structures listed in Fig. 1) are simulated from the first subpopulation and 100 four-generation pedigrees with the pedigree structure listed in Fig. 1 are generated from the second subpopulation, where these two subpopulations are very different from each other in pedigree structure and level of

Table 4 Contingency table showing MCGDTI and MCGDT-ME results of application to RA data at $\alpha' = 9.247,3 \times 10^{-6a}$

	$P_{MCGDT-ME} < \alpha'$	$P_{MCGDT-ME} \geq \alpha'$	Total
$P_{MCGDTI} < \alpha'$	0	3	3
$P_{MCGDTI} \geq \alpha'$	0	5404	5404
Total	0	5407	5407

^a P_{test} denotes the p-value of the test

Table 5 Empirical size (in percentage (%)) of GDTI, MCGDTI, MCGDT-ME, GDT-ME, GDT and MCPDTI in the presence of population stratification^a

Scenario	Complete data		Incomplete data				
	GDTI		MCGDTI	MCGDT-ME	GDT-ME	GDT	MCPDTI
1	1.14		1.16	1.04	1.14	1.16	1.69
2	1.07		1.01	0.97	1.03	1.09	1.24

^aThe simulations are conducted under the null hypothesis of no association and no imprinting based on 10,000 replicates at 1% significance level

genotypic missingness. Then, the resulting total sample size of pedigrees is 300 for each simulation scenario. Other simulation settings are the same as those in the Simulation settings subsection. The simulated size results of GDTI, MCGDTI, MCGDT-ME, GDT-ME, GDT and MCPDTI are shown in Table 5. From the table, we find that all the proposed test statistics control the size well under the PS models, while the size of the existing MCPDTI test is a little inflated.

Just like the genotypes of some members in the collected pedigrees may be missing, it is also common in practice that the affection statuses of some individuals in the pedigrees may be unavailable. As mentioned in the real data application subsection, one way to deal with these individuals is to treat them as unaffected. To investigate if this influences the validity of the proposed test statistics, we conduct a few simulation studies. The simulation results show that the proposed methods are still valid to test for association by handling the missing affection status in this way (data not shown). However, this may impact their test powers under alternative hypotheses and we will carry out some simulation studies to check it in our future work.

Like other methods, our proposed GDTI and MCGDTI have their own limitations. In this article, we only consider using an empirical moment estimate based on large sample theory to estimate the variances of the numerators of GDTI and MCGDTI, while we do not propose the corresponding tests based on the variance estimates from the IBD information. This is because even though the IBD information between two alleles for the pair of allele scores $(X_{ij}^{(p)}, X_{ik}^{(p)})$, $(X_{ij}^{(p)}, X_{ik}^{(m)})$, $(X_{ij}^{(m)}, X_{ik}^{(p)})$ or $(X_{ij}^{(m)}, X_{ik}^{(m)})$ of the j^{th} and k^{th} individuals in the i^{th} pedigree is obtained, two allele scores in this pair may be different from each other for GDTI and MCGDTI and thus we cannot estimate the corresponding variance based on the IBD information, which is different from GDT (the details refer to Appendix B in Chen et al. [15]). For example, we consider a two-generation family in which the genotypes of the unaffected parents and the affected child are M_1M_2 , M_1M_2 and M_1M_1 , respectively. Then, when we compare the allele scores of the unaffected father and the affected child, the allele scores of the father and the child are respectively $X_F^{(p)} = X_F^{(m)} = 0.5$ and $X_C^{(p)} = X_C^{(m)} = 1$, which are different from each other. Fortunately, from our simulation study, MCGDTI for incomplete pedigrees even has the similar power to GDT under the no imprinting effect model, and is more powerful than GDT under the imprinting effect models.

We should mention that, because of utilizing the genotype differences between all discordant relative pairs, the requirement for a pedigree to be included is that this pedigree should have at least one affected

and one unaffected individuals. In addition, GDTI and MCGDTI do not take account of the covariates in analysis, which may cause the dependences between individuals within a family, even though under the null hypothesis of no association. This may be handled from the quasi-likelihood for a conditional logistic regression model [15, 24, 25]. So, our future work is to incorporate the covariates into GDTI and MCGDTI.

Conclusions

Under complete and incomplete imprinting effect models, our proposed GDTI and MCGDTI methods, by considering the information on imprinting effects and all discordant relative pairs within each pedigree, outperform all the existing test statistics and MCGDTI can recapture much of the missing information. Therefore, MCGDTI is recommended in practice.

Additional files

Additional file 1: Appendix. Construction of the GDTI test statistic.

Table S1. P-values of the test statistics applied to RA data at 3 SNPs with $P_{MCGDTI} < 9.247,3 \times 10^{-6}$. **Figures S1 - S3.** Simulated powers of all the test statistics. The test statistics are T1: GDTI, T2: MCGDTI_T, T3: MCGDTI_E, T4: MCGDT-ME_T, T5: MCGDT-ME_E, T6: GDT-ME, T7: GDT, T8: MCPDTI_T and T9: MCPDTI_E. The simulations are conducted under complete, incomplete and no imprinting effect models at 1% significance level based on 10,000 replicates for 90 pedigrees when LD = 0.092,5, 0.142,5, and 0.157,5, and RR = 1.500, 1.833 and 2.182, respectively. The first 5 statistics are proposed tests, while the remaining 4 are existing tests. (PDF 76 kb)

Abbreviations

GDT: Generalized disequilibrium test with the variance estimated based on the information on kinship coefficients when identity by descent is unknown; GDTI: Generalized disequilibrium test with imprinting; GDT-ME: Generalized disequilibrium test based on the moment estimate of the variance; IBD: Identity by descent; LD: Linkage disequilibrium; MC: Monte Carlo; MCGDTI: Monte Carlo GDTI; MCGDT-ME: Monte Carlo GDT-ME; MCPDTI: Monte Carlo pedigree disequilibrium test with imprinting; MCPPAT: Monte Carlo pedigree parental-asymmetry test; PDTI: Pedigree disequilibrium test with imprinting; PS: Population stratification; RA: Rheumatoid arthritis; RR: Relative risk; SNP: Single nucleotide polymorphism

Acknowledgments

The authors thank the reviewer for helpful comments that greatly improve the presentation of the article. The authors thank the Genetic Analysis Workshops for providing the RA data, which were supported by the National Institutes of Health grant R01 GM031575. The RA data were gathered with the support of grants from the National Institutes of Health grants N01-AR-2-2263 and R01-AR-44422, and the National Arthritis Foundation.

Funding

This work was supported by the National Natural Science Foundation of China grants 81,373,098, 81,773,544 and 81,573,207, Science and Technology Planning Project of Guangdong Province of China grant 2013B021800038 and the Hong Kong RGC GRF Research Grant 17,301,715.

Availability of data and materials

The dataset supporting the conclusions of this article is from North American Rheumatoid Arthritis Consortium, which is made available from Genetic Analysis Workshop 15 (<http://www.gaworkshop.org/>). Our software MCGDTI

is freely available at <http://www.echobelt.org/web/UploadFiles/MCGDTI.html>, which is implemented in R (<http://www.r-project.org/>).

Authors' contributions

JLL, PW, WKF and JYZ all contributed to the study design, analytical preparation and the writing of the manuscript. JLL and PW performed the simulation studies. JLL, WKF and JYZ analyzed the data and revised the manuscript. All authors have read and approved the final manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹State Key Laboratory of Organ Failure Research, Ministry of Education, and Guangdong Provincial Key Laboratory of Tropical Disease Research, Department of Biostatistics, School of Public Health, Southern Medical University, Guangzhou, China. ²State Key Laboratory of Ophthalmology, Zhongshan Ophthalmic Center, Sun Yat-sen University, Guangzhou, China. ³Department of Statistics and Actuarial Science, The University of Hong Kong, Hong Kong, China.

Received: 26 March 2017 Accepted: 4 October 2017

Published online: 16 October 2017

References

- Martienssen RA, Colot V. DNA methylation and epigenetic inheritance in plants and filamentous fungi. *Science*. 2001;293(5532):1070–4.
- Feil R, Berger F. Convergent evolution of genomic imprinting in plants and mammals. *Trends Genet*. 2007;23(4):192–9.
- Peters J. The role of genomic imprinting in biology and disease: an expanding view. *Nat Rev Genet*. 2014;15(8):517–30.
- Morison IM, Paton CJ, Cleverley SD. The imprinted gene and parent-of-origin effect database. *Nucleic Acids Res*. 2001;29(1):275–6.
- Morison IM, Paton CJ, Cleverley SD. The imprinted gene and parent-of-origin effect database. 2001. <http://igc.otago.ac.nz>. Accessed 26 Mar 2017.
- Glaser RL, Ramsay JP, Morison IM. The imprinted gene and parent-of-origin effect database now includes parental origin of de novo mutations. *Nucleic Acids Res*. 2006;34(Suppl 1):D29–31.
- Scharfmann R, Shield JPH. Development of the pancreas and neonatal diabetes. 1st ed. Switzerland: Karger; 2007.
- Falls JG, Pulford DJ, Wylie AA, Jirtle RL. Genomic imprinting: implications for human disease. *Am J Pathol*. 1999;154(3):635–47.
- Ziegler A, König IR, Pahlke F. A statistical approach to genetic epidemiology: concepts and applications, with an E-learning platform. 2nd ed. Germany: Wiley-VCH; 2010.
- Zhou JY, Mao WG, Li DL, YQ H, Xia F, Fung WK. A powerful parent-of-origin effects test for qualitative traits incorporating control children in nuclear families. *J Hum Genet*. 2012;57(8):500–7.
- Spielman RS, McGinnis RE, Ewens WJ. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet*. 1993;52(3):506–16.
- Horvath S, Xu X, Laird NM. The family based association test method: strategies for studying general genotype-phenotype associations. *Eur J Hum Genet*. 2001;9(4):301–6.
- Martin ER, Monks SA, Warren LL, Kaplan NL. A test for linkage and association in general pedigrees: the pedigree disequilibrium test. *Am J Hum Genet*. 2000;67(1):146–54.
- Laird NM, Horvath S, Xu X. Implementing a unified approach to family-based tests of association. *Genet Epidemiol*. 2000;19(Suppl 1):S36–42.
- Chen WM, Manichaikul A, Rich SS. A generalized family-based association test for dichotomous traits. *Am J Hum Genet*. 2009;85(3):364–76.
- Xia F, Zhou JY, Fung WK. A powerful approach for association analysis incorporating imprinting effects. *Bioinformatics*. 2011;27(18):2571–7.
- Zhou JY, He HQ, You XP, Li SZ, Chen PY, Fung WK. A powerful association test for qualitative traits incorporating imprinting effects using general pedigree data. *J Hum Genet*. 2015;60(2):77–83.
- Zhou JY, Ding J, Fung WK, Lin S. Detection of parent-of-origin effects using general pedigree data. *Genet Epidemiol*. 2010;34(2):151–8.
- Ding J, Lin S, Liu Y. Monte Carlo pedigree disequilibrium test for markers on the X chromosome. *Am J Hum Genet*. 2006;79(3):567–73.
- Amos CI, Chen WW, Remmers E, Siminovich KA, Seldin MF, Criswell LA, et al. Data for genetic analysis workshop (GAW) 15 problem 2, genetic causes of rheumatoid arthritis and associated traits. *BMC Proc*. 2007;1(Suppl 1):S3.
- Ott J, Lathrop GM. SLINK: a general simulation program for linkage analysis. *Am J Hum Genet*. 1990;47:A204.
- Team RC. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2013. <http://www.r-project.org>. 2017.
- Genetic Analysis Workshop. 1982. <https://www.gaworkshop.org>. Accessed 26 Mar 2017.
- Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika*. 1986;73(1):13–22.
- Liang KY, Pulver AE. Analysis of case-control/family sampling design. *Genet Epidemiol*. 1996;13(3):253–70.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

