Published in final edited form as:

Genet Epidemiol. 2010 February ; 34(2): 151–158. doi:10.1002/gepi.20445.

Detection of Parent-of-Origin Effects Using General Pedigree Data

Ji-Yuan Zhou^{1,2}, Jie Ding², Wing K. Fung¹, and Shili Lin^{2,*}

¹Department of Statistics and Actuarial Science, The University of Hong Kong, Hong Kong, China

²Department of Statistics, The Ohio State University, Columbus, Ohio

Abstract

Genomic imprinting is an important epigenetic factor in complex traits study, which has generally been examined by testing for parent-of-origin effects of alleles. For a diallelic marker locus, the parental-asymmetry test (PAT) based on case-parents trios and its extensions to incomplete nuclear families (1-PAT and C-PAT) are simple and powerful for detecting parent-of-origin effects. However, these methods are suitable only for nuclear families and thus are not amenable to general pedigree data. Use of data from extended pedigrees, if available, may lead to more powerful methods than randomly selecting one two-generation nuclear family from each pedigree. In this study, we extend PAT to accommodate general pedigree data by proposing the pedigree PAT (PPAT) statistic, which uses all informative family trios from pedigrees. To fully utilize pedigrees with some missing genotypes, we further develop the Monte Carlo (MC) PPAT (MCPPAT) statistic based on MC sampling and estimation. Extensive simulations were carried out to evaluate the performance of the proposed methods. Under the assumption that the pedigrees and their associated affection patterns are randomly drawn from a population of pedigrees with at least one affected offspring, we demonstrated that MCPPAT is a valid test for parent-of-origin effects in the presence of association. Further, MCPPAT is much more powerful compared to PAT for trios or even PPAT for all informative family trios from the same pedigrees if there is missing data. Application of the proposed methods to a rheumatoid arthritis dataset further demonstrates the advantage of MCPPAT.

Keywords

genomic imprinting; Monte Carlo sample; missing genotype; rheumatoid arthritis

INTRODUCTION

Genomic imprinting is an important epigenetic factor in complex traits study, which can lead to complete or partial deactivation of the maternally or paternally inherited copy of a gene. Some complex diseases, such as Beckwith–Wiedemann, Prader–Willi, and Angelman syndromes, have been found to be related to maternal imprinting [Falls et al., 1999; Ziegler and König, 2006]. For other complex diseases such as diabetes, hereditary paragangliomas, schizophrenia, intrauterine growth retardation, autism, neural tube defects, and obesity, imprinting effects are suspected or hypothesized to play an important role, although specific

© 2009 Wiley-Liss, Inc.

*Correspondence to: Department of Statistics, The Ohio State University, 1958 Neil Avenue, Columbus, OH 43210-1247. shili@stat.osu.edu .

WEB RESOURCES The URLs for data presented herein are as follows: MC-PDT, <http://www.stat.osu.edu/statgen/SOFTWARE/MC-PDT/>; R Project for Statistical Computing, <http://www.r-project.org/>.

genes have yet to be identified conclusively [Abel, 2004; Chatkupt et al., 1992; Dong et al., 2005; Rampersaud et al., 2007; Samaco et al., 2005; Temple et al., 1995; Van Schothorst et al., 1996].

Genomic imprinting has been largely examined in analyses of complex traits through testing for parent-of-origin effects of alleles [Hager et al., 2008]. The parental-asymmetry test (PAT) [Weinberg, 1999] with case-parents trio data and its extensions to nuclear families with missing parental genotypes (1-PAT and C-PAT) [Zhou et al., 2009] are simple and powerful for such a task. However, these methods are suitable only for nuclear families and thus are not amenable to general pedigree data. Using information from extended pedigrees, if available, may lead to more efficient and more powerful methods than randomly selecting one two-generation nuclear family from each pedigree. Furthermore, following the argument of Ding et al. [2006] and Ding and Lin [2008], making use of families with incomplete data based on Monte Carlo (MC) sampling is likely to result in a more powerful test than one only using the complete portion of the data if there is missing genotypes.

Motivated by the desire to make full use of available information as discussed above, in this study, we extend the PAT to accommodate general pedigree data by proposing the pedigree parental-asymmetry test (PPAT) statistic. This test uses all informative family trios (trios that have an affected offspring who is heterozygous and two parents with different genotypes) and accounts for the dependencies among the trios. By borrowing the idea from Ding et al. [2006], we further develop the Monte Carlo PPAT (MCPPAT) statistic to accommodate general pedigree data with missing genotypes based on MC sampling. By doing so, information from all affected offspring in a pedigree is fully utilized. We carried out extensive simulations to study the properties of the test statistics. Both PPAT and MCPPAT maintain the correct type I error rates. For MCPPAT, this conclusion is under the assumption that the pedigrees and affection patterns are randomly drawn from a population of pedigrees with at least one affected offspring. Furthermore, as anticipated, MCPPAT has much more power compared to PPAT if there is a substantial amount of missing data in the pedigrees, which in turn is more powerful than PAT using only one trio from each pedigree. We applied both MCPPAT and PPAT to a United Kingdom (UK) rheumatoid arthritis (RA) dataset to illustrate their utility for real data analysis. The results further substantiate our finding from the simulation that MCPPAT is advantageous over PPAT with pedigrees having some missing genotypes.

METHODS

BACKGROUND AND NOTATION

We consider a disease susceptibility locus (DSL) with disease allele D and normal allele d . Suppose the marker locus of interest has two alleles, M_1 and M_2 . Note that there are four haplotypes DM_1 , dM_1 , DM_2 , and dM_2 at the DSL and the marker locus jointly. Let $\phi_{D/D}$, $\phi_{D/d}$, $\phi_{d/D}$, and $\phi_{d/d}$ be the penetrances of genotypes D/D , D/d , d/D , and d/d respectively, where the allele before “/” is paternal and that after “/” is maternal. The parent-of-origin effects are measured by the degree of imprinting $I = (\phi_{D/d} - \phi_{d/D})/2$, i.e. half the difference of the two heterozygote penetrances. As such, $I = 0$ indicates no parent-of-origin effects. Furthermore, I ranges from $(\phi_{d/d} - \phi_{D/D})/2$ (complete paternal parent-of-origin effect) to $(\phi_{D/D} - \phi_{d/d})/2$ (complete maternal parent-of-origin effect) [Strauch et al., 2000].

For convenience, let 0, 1, and 2 represent the marker genotypes M_2M_2 , M_1M_2 , and M_1M_1 , respectively. We also let F , M , and C denote the genotypes of father, mother, and child, respectively, and therefore F , M , and C take possible values of 0, 1, or 2. Throughout this study, we assume that mating symmetry holds and there is no maternally mediated genotype

effect. To test for parent-of-origin effects, we also assume that the marker locus and DSL are in linkage disequilibrium (LD).

PAT FOR CASE-PARENTS TRIOS

We begin by describing PAT for detecting parent-of-origin effects [Weinberg, 1999]. For N case-parents trios, Weinberg [1999] proposed the following PAT to test for parent-of-origin effects:

$$\text{PAT} = \frac{N_{F>M,C=1} - N_{F<M,C=1}}{\sqrt{N_{F>M,C=1} + N_{F<M,C=1}}} = \frac{\sum S}{\sqrt{\sum S^2}},$$

where $N_{F>M,C=1}$ ($N_{F<M,C=1}$) is the number of case-parents trios with heterozygous child in which the father carries more (fewer) copies of marker allele M_1 than the mother, i.e. the number of heterozygous children who inherit marker allele M_1 from the father (mother). We can rewrite PAT as in the second equality, where the summation is over all informative family trios, and $S = 1(-1)$ if the father has more (less) copies of allele M_1 than the mother. It turns out that the square of PAT is a McNemar's test.

PPAT FOR GENERAL PEDIGREE DATA

Suppose we have a collection of pedigrees each with at least one affected nonfounder. To test for parent-of-origin effects, we propose the PPAT as follows. Instead of using only one case-parents trio from each pedigree as in PAT, all the informative family trios from each pedigree will be used in PPAT. For a dataset with N pedigrees each having n_j informative

trios, $j = 1, \dots, N$, let $Q_j = \sum_{i=1}^{n_j} S_{ji}$, where S_{ji} is the PAT statistic for trio i in pedigree j . Then the PPAT statistic for the whole dataset is defined as

$$\begin{aligned} T &= \frac{\sum_{j=1}^N Q_j}{\sqrt{\sum_{j=1}^N Q_j^2}} = \frac{\sum_{j=1}^N \sum_{i=1}^{n_j} S_{ji}}{\sqrt{\sum_{j=1}^N \left(\sum_{i=1}^{n_j} S_{ji} \right)^2}} \\ &= \frac{\sum_{j=1}^N \sum_{i=1}^{n_j} S_{ji}}{\sqrt{\sum_{j=1}^N \left(\sum_{i=1}^{n_j} S_{ji}^2 + \sum_{i \neq k} S_{ji} S_{jk} \right)}}, \end{aligned} \quad (1)$$

which takes the dependencies among the S_{ji} 's within the same pedigree into account as we elaborate in the following. Under the null hypothesis of no parent-of-origin effects, each Q

has an expectation of zero (i.e. $E(Q_j) = 0, j = 1, 2, \dots, N$) and hence $E\left(\sum_{j=1}^N Q_j\right) = 0$. Then

$$\begin{aligned} \text{Var}\left(\sum_{j=1}^N Q_j\right) &= \sum_{j=1}^N \text{Var}(Q_j) = \sum_{j=1}^N E(Q_j^2) = E\left(\sum_{j=1}^N Q_j^2\right) \\ &= E\left[\sum_{j=1}^N \left(\sum_{i=1}^{n_j} S_{ji}\right)^2\right] \\ &= E\left[\sum_{j=1}^N \left(\sum_{i=1}^{n_j} S_{ji}^2 + \sum_{i \neq k} S_{ji} S_{jk}\right)\right]. \end{aligned}$$

The first equality in the above formula holds under the assumption that all N families (extended pedigrees) are independent, whereas the second equality is true because $E(Q_j) = 0$

under the null hypothesis. Therefore, $\sum_{j=1}^N \left(\sum_{i=1}^{n_j} S_{ji}^2 + \sum_{i \neq k} S_{ji} S_{jk} \right)$ provides an unbiased estimator for the variance of $\sum_{j=1}^N Q_j$, justifying the statistic defined in (1). As can be seen from the variance estimator, there are cross-product terms between the trios from the same pedigree, which account for the dependencies among the trios. Further, when the number of families, N , is sufficiently large, the PPAT statistic as defined in (1) follows a standard normal distribution approximately.

MCPPAT WHEN THERE ARE MISSING GENOTYPES

When the genotypes of some individuals in a pedigree are missing, the above PPAT only uses the subset of trios that have complete information and discards those with incomplete data. This practice may lead to a power loss (as we will see from the simulation results below). To make full use of all available data, we propose the following MCPPAT statistic that will include all affected offspring in the analysis. We let Q denote the contribution from a pedigree to the statistic defined in (1) assuming there are complete data. Here the subscript that indexes the family is suppressed to simplify notation without causing ambiguity. When some of the individuals' genotypes are missing, Q is not computable. As such, we replace Q by its conditional expectation, Q_{MC} , given the observed genotypes, G_0 . That is,

$$Q_{MC} = E[Q|G_0] = E[Q(G_m, G_0, A)|G_0], \quad (2)$$

where $Q(G_m, G_0, A)$ in the second equality of Equation (2) is an expanded notation of Q to explicitly show its dependence on the observed genotypes G_0 and the missing genotypes G_m , and A is the collection of observed phenotypes (disease affection statuses). However, evaluation of the expectation is usually not computationally feasible due to the large number of summations over all sets of possible genotypes G_m given G_0 . Instead, we estimate Q_{MC} based on a MC simulation scheme. Specifically, we draw independent samples G_{mk} , $k = 1; \dots, K$, from $P(G_m|G_0)$, which can be accomplished efficiently based on the peeling algorithm using the SLINK software [Weeks et al., 1990]. The desired statistic is estimated as follows:

$$Q_{MC} \approx \frac{1}{K} \sum_{k=1}^K Q(G_{mk}, G_0, A).$$

The MCPPAT statistic T_{MC} is then calculated, as before, but substituting each Q in Equation (1) by the corresponding Q_{MC} . If we assume that all the pedigrees are drawn from a certain underlying population and treat A as random, it can be shown that the expectation $E(Q_{MC})$ is equal to 0 when there are no parent-of-origin effects. The proof is similar to that for $E(D_{MC})$ in Ding et al. [2006] and we omit it here for brevity.

SIMULATION STUDY

SETTINGS

We study the size and power of the proposed tests under various settings of parameter values. Since the premise of the tests for parent-of-origin effects is under the presence of association, we assume that the disease locus is in complete linkage but in LD with the single nucleotide polymorphism (SNP) marker locus. The haplotype frequencies of all settings are given in Table IA. Note that, in all settings, the disease allele D is more likely to be associated with marker allele M_1 .

Three sets of homozygote penetrances $\phi_{D/D}$ and $\phi_{d/d}$ are used, leading to nine different combinations together with the haplotype frequencies (Table IA). For each specification of $\phi_{D/D}$ and $\phi_{d/d}$ in Table IA, we consider three parent-of-origin effect models for setting the two heterozygote penetrances $\phi_{D/d}$ and $\phi_{d/D}$: no parent-of-origin effects, incomplete maternal parent-of-origin effect, and complete maternal parent-of-origin effect (Table IB); the settings therein refer to the settings in Table IA. Because of the symmetry of the test statistics for paternal and maternal parent-of-origin effects, it is not necessary to consider the corresponding paternal parent-of-origin effect models. Note that the no parent-of-origin effects models are for studying the type I error rates of the tests, while the other two are for the study of power.

Each simulated dataset consists of 100 three-generation 10-member pedigrees with the same structure as shown in Figure 1. We first generated the founders' haplotypes at the marker and the DSL based on the specified haplotype frequencies. Then, haplotypes of the descendants were generated assuming no recombination. We assigned the affection status of each individual according to his/her genotype at the DSL and the four penetrances $\phi_{D/D}$, $\phi_{D/d}$, $\phi_{d/D}$, and $\phi_{d/d}$. A pedigree was ascertained if at least one of the nonfounders was affected. The genotypes of some of the individuals will be removed to assess its effects on the various test statistics as detailed below. Also, model parameters, such as the penetrances, that were used to generate the data were not used in our analysis as the tests are nonparametric.

For each model setting (a total of three sets of nine settings as specified in Table I) investigated, 20,000 replicates were simulated. Fifty MC samples of missing genotypes were generated for each replicate using the SLINK software [Weeks et al., 1990]. Either the true marker allele frequencies or those estimated from the genotyped founders in each replicate were used in the MC sampling where applicable. MCPPAT and PPAT were performed using our software package MC-PDT based on the R environment (R Project for Statistical Computing). We investigated the actual type I error rates corresponding to three nominal significance levels: 5, 1, and 0.5%. Recall that our proposed parent-of-origin tests are only applied to markers that have been implicated to be associated with the trait of interest, and therefore, the number of tests to be performed is usually much smaller than that for a genome wide association study. In some situations, perhaps only a handful of candidate loci may be considered. Such a consideration is reflected in our selection of the nominal levels.

For assessment of the proposed tests (PPAT and MCPPAT) and for comparison with PAT, we considered the following seven (four PPAT, two MCPPAT, and one PAT) statistics. The first PPAT is for complete data without any missing genotype, which can be regarded as the gold standard, and is denoted as $PPAT_{full}$. The second PPAT ($PPAT_{1,5}$) is based on the data after removing the genotypes of individuals 1 and 5 (using informative trio data from the nuclear family consisting of individuals 4, 6, 9, and 10). Two other PPAT statistics, one assuming only data for individual 1 was missing ($PPAT_1$) and the other assuming only individual 5 was missing ($PPAT_5$), were also considered. Two MCPPAT versions (both assuming the genotypes of individuals 1 and 5 were missing) were investigated, the first using the true marker allele frequencies ($MCPPAT_T$) and the second using the estimated frequencies ($MCPPAT_E$). The PAT statistic used only one of the trios from the nuclear families including individuals 4, 6, 9, and 10; this statistic can be regarded as the baseline.

SIZES OF MCPPAT AND PPAT

We first investigated the type I error rates for the nine no parent-of-origin effects model and haplotype frequency (LD pattern) combinations. For all statistics, the nominal type I error rates were controlled to be at 5, 1, or 0.5%, while their actual sizes are given in Table II. As can be seen from the table, the sizes of the two MCPPAT versions ($MCPPAT_T$ and

MCPAT_E) and all four versions of PPAT (PPAT_{full}, PPAT_{1,5}, PPAT₁, and PPAT₅) are generally quite close to the corresponding nominal levels, which verifies the validity of the proposed MCPAT and PPAT empirically.

POWER COMPARISON OF MCPAT, PPAT, AND PAT

Two sets of nine settings each, one for complete and the other for incomplete maternal parent-of-origin effects, were used to study the power of the tests. Figure 2 plots the estimated powers (based on 20,000 replicates) of the seven statistics against the nine settings. Indeed, the gold standard (PPAT_{full}) achieves the highest power in all nine settings. All the other five statistics based on either MCPAT or PPAT also have higher power than PAT, the baseline. Since PPAT_{1,5} used only data from the nuclear family comprised of individuals 4, 6, 9, and 10, its power gain over PAT (which used one of the trios from this family) is moderate. However, PPAT₁ and PPAT₅, which assumed one more individual's genotype was observed, led to much more substantial power increase. Although the powers for PPAT₁ and PPAT₅ are close, it appears that the three-generation data may be a bit more informative than the two sibships with common grandparents. MCPAT using data assuming genotypes of two individuals were missing, irrespective of using the true allele frequencies or the estimated ones, closely follow behind the gold standard. They even have more power than PPAT₁ and PPAT₅ that assumed only one individual's genotype was missing. These results demonstrate that, through MC sampling to infer the missing genotypes, much of the missing information can be recovered. Furthermore, the statistic can be practically used, as MCPAT_E based on the estimated frequencies has the correct type I error rates and similar powers as MCPAT_T that used the true frequencies. By comparing the plots in the left column vs. those in the right column of Figure 2, the powers of the proposed statistics under the complete parent-of-origin effect model are higher than the corresponding ones under the incomplete parent-of-origin effect model, as expected. Furthermore, power decreases as the nominal level is decreased, and the effect is more profound for the incomplete model.

To understand the pattern of power increase of MCPAT and PPAT over PAT, we examined MCPAT_E as an example and calculated the number of affected offspring over all replicates and all nine settings. Recall that PAT only used one trio with an affected child while MCPAT_E used all affected offspring. As such, more power may be gained with MCPAT if there are more affected offspring. Indeed, Figure 3, which plots the increase of the absolute value of the MCPAT_E statistic over that of PAT vs. the number of affected offspring, shows a linear relationship. Similar results for the other PPAT and MCPAT tests were observed (results not shown).

APPLICATION TO UK RA DATA

We applied the MCPAT and PPAT to the RA data from the United Kingdom (UK) [John et al., 2004]. These data were made available through the Genetic Analysis Workshop 15 [Amos et al., 2007]. There are 175 families comprised of 903 individuals in this dataset, which were genotyped at 10,156 SNP markers over the 22 autosomes; no marker data on the X nor the Y chromosomes were available. There is a substantial amount of missing genotypes in the data. For example, for chromosomes 1–3, only 15.6 families have complete genotype information on the average for each SNP. As such, we expect a substantial gain in information for MC-based procedures that make use of information from all affected individuals and partial genotypes.

Specifically, MCPAT and PPAT were used to test for parent-of-origin effects in the presence of association. That is, we first used the Monte Carlo pedigree disequilibrium test (MCPDT) [Ding et al., 2006] to identify SNPs that are associated with RA and then

performed parent-of-origin effects tests focusing on only those SNPs. A total of 146 SNPs were identified by MCPDT at the 1% significance level. Because we treated the association test as a preliminary step, we opted for a less stringent criterion to reduce the chance of missing any potential SNPs in the subsequent analysis for parent-of-origin effects. The results of MCPPAT and PPAT tests on the 146 SNPs are given in Table III. MCPPAT identified three of the SNPs (SNP76924 on Chromosome 7, SNP615817 and SNP615818 on Chromosome 13) at the significance level of 0.01, while the *P*-values for the 146 SNPs all exceed 0.01 based on PPAT. Because MCPPAT and PPAT are shown to have similar and correct type I error rates (from our simulation study), the results from this real data analysis demonstrate substantial gains in information through making use of partially genotyped families.

DISCUSSION

In summary, we extend the PAT to accommodate general pedigree data for testing parent-of-origin effects in the presence of association. Our simulation study showed that both PPAT and MCPPAT indeed control the size well under the null hypothesis of no parent-of-origin effects. Our simulation results also demonstrated that, as expected, they are much more powerful than PAT when data from larger pedigrees are available. Furthermore, MCPPAT can be more powerful than PPAT even when MCPPAT uses data with more missing information than PPAT. It is also practically feasible, and can recapture most of the missing information, at least for the settings that we have examined.

Like any other method, MCPPAT has its own drawbacks. If a dataset only contains a small number of genotyped founders, the frequency estimate may not be reliable. However, our limited simulation study seems to indicate that the results are not overly sensitive to the minor perturbation of the allele frequencies. Two crucial assumptions are needed for MCPPAT to be a valid test. One is the ascertainment criterion: in the data collection process, the requirement for a family to be included is at least one affected nonfounder. However, such ascertainment scheme appears to be easily satisfied for a genetic epidemiology study that collects pedigree data. Further, we expect MCPPAT to remain a reasonable test even with minor violation of the ascertainment scheme, following the work of Ding et al. [2006] and Ding and Lin [2008]. Another assumption of MCPPAT is that the underlying population is homogenous. When estimating allele frequencies in MCPPAT, it is assumed that all pedigrees in the dataset are from the same population with a single set of frequencies for each marker. If population stratification exists but the subpopulations are similar except for the allele frequencies of some test markers, the overall effect of ignoring population structure may be minimal [Ding, 2008]. However, if the subpopulations are very different in some other aspects (e.g. level of missingness), the impact of assuming a single set of allele frequencies would be much larger and therefore further work is warranted to address this issue. It is worth further emphasizing, though, that PPAT does not require these two assumptions, and thus it is robust against population stratification and can be much more powerful than analyses using just nuclear family data. Finally, we caution that our conclusions are mainly based on the results from our simulation study, although we expect the general trends to hold regardless of the particular data.

Acknowledgments

We thank the Editor-in-Chief and two reviewers for their insightful and helpful suggestions, which have led to improvement of the presentation. The Genetic Analysis Workshops are supported by NIH grant R01 GM031575 and AR44422.

Contract grant sponsor: Hong Kong RGC CERG Research Grant; Contract grant number: HKU 702207P; Contract grant sponsor: NIH; Contract grant numbers: 5R01HG002657; R01 GM031575; AR44422.

REFERENCES

- Abel KM. Fetal origins of schizophrenia: testable hypotheses of genetic and environmental influences. *Br J Psychiatry*. 2004; 184:383–385. [PubMed: 15123500]
- Amos C, Chen W, Remmers E, Siminovitch K, Seldin M, Criswell L, Lee A, John S, Shephard N, Worthington J, Cornelis F, Plenge R, Begovich A, Dyer T, Kastner D, Gregersen P. Data for genetic analysis workshop (GAW) 15 problem 2, genetic causes of rheumatoid arthritis and associated traits. *BMC Proc Suppl*. 2007; 1:S3.
- Chatkupt S, Lucek PR, Koenigsberger MR, Johnson WG. Parental sex effect in spina bifida: a role for genomic imprinting? *Am J Med Genet*. 1992; 44:508–512. [PubMed: 1442897]
- Ding, J. PhD Dissertation. The Ohio State University; 2008. Monte Carlo pedigree disequilibrium test with missing data and population structure.
- Ding J, Lin S. XMCPDT does have correct type I error rate. *Am J Hum Genet*. 2008; 82:528–530. [PubMed: 18252235]
- Ding J, Lin S, Liu Y. Monte Carlo pedigree disequilibrium test for markers on the X chromosome. *Am J Hum Genet*. 2006; 79:567–573. [PubMed: 16909396]
- Dong CH, Li WD, Geller F, Lei L, Li D, Gorlova OY, Hebebrand J, Amos CI, Nicholls RD, Price RA. Possible genomic imprinting of three human obesity-related genetic loci. *Am J Hum Genet*. 2005; 76:421–437. [PubMed: 15657873]
- Falls JG, Pulford DJ, Wylie AA, Jirtle RL. Genomic imprinting: implications for human disease. *Am J Pathol*. 1999; 154:635–647. [PubMed: 10079240]
- Hager R, Cheverud JM, Wolf JB. Maternal effects as the cause of parent-of-origin effects that mimic genomic imprinting. *Genetics*. 2008; 78:1755–1762. [PubMed: 18245362]
- John S, Shephard N, Liu G, Zeggini E, Cao M, Chen W, Vasavda N, Mills T, Barton A, Hinks A, Eyre S, Jones KW, Ollier W, Silman A, Gibson N, Worthington J, Kennedy GC. Whole-genome scan, in a complex disease, using 11,245 single-nucleotide polymorphisms: comparison with microsatellites. *Am J Hum Genet*. 2004; 75:54–64. [PubMed: 15154113]
- Rampersaud E, Morris RW, Weinberg CR, Speer MC, Martin ER. Power calculations for likelihood ratio tests for offspring genotype risks, maternal effects, and parent-of-origin (POO) effects in the presence of missing parental genotypes when unaffected siblings are available. *Genet Epidemiol*. 2007; 31:18–30. [PubMed: 17096358]
- Samaco RC, Hogart A, LaSalle JM. Epigenetic overlap in autism-spectrum neurodevelopmental disorders: MECP2 deficiency causes reduced expression of UBE3A and GABRB3. *Hum Mol Genet*. 2005; 14:483–492. [PubMed: 15615769]
- Strauch K, Fimmers R, Kurz T, Deichmann KA, Wienker TF, Baur MP. Parametric and nonparametric multipoint linkage analysis with imprinting and two-locus-trait models: application to mite sensitization. *Am J Hum Genet*. 2000; 66:1945–1957. [PubMed: 10796874]
- Temple IK, James RS, Crolla JA, Sitch FL, Jacobs PA, Howell WM, Betts P, Baum JD, Shield JP. An imprinted gene(s) for diabetes? *Nat Genet*. 1995; 9:110–112. [PubMed: 7719335]
- van Schothorst EM, Jansen JC, Bardoel AF, van der Mey AG, James MJ, Sobol H, Weissenbach J, Van Ommen GJ, Cornelisse CJ, Devilee P. Confinement of PGL, an imprinted gene causing hereditary paragangliomas, to a 2-cM interval on 11q22-q23 and exclusion of DRD2 and NCAM as candidate genes. *Eur J Hum Genet*. 1996; 4:267–273. [PubMed: 8946172]
- Weeks DE, Ott J, Lathrop GM. SLINK: a general simulation program for linkage analysis. *Am J Hum Genet*. 1990; 47:A204.
- Weinberg CR. Methods for detection of parent-of-origin effects in genetic studies of case-parents triads. *Am J Hum Genet*. 1999; 65:229–235. [PubMed: 10364536]
- Zhou JY, Hu YQ, Lin S, Fung WK. Detection of parent-of-origin effects based on complete and incomplete nuclear families with multiple affected children. *Hum Hered*. 2009; 67:1–12. [PubMed: 18931505]
- Ziegler, A.; König, IR. *A Statistical Approach to Genetic Epidemiology: Concepts and Applications*. Wiley-VCH; Weinheim: 2006.

Fig. 1. Pedigree structure used for the simulation study. Individuals 1 and/or 5's genotypes were assumed missing for the analysis based on incomplete data.

Fig. 2.

Power comparisons of the proposed statistics under (a) complete parent-of-origin effect model and at nominal 5% level, (b) incomplete parent-of-origin effect model and at nominal 5% level, (c) complete parent-of-origin effect model and at nominal 1% level, (d) incomplete parent-of-origin effect model and at nominal 1% level, (e) complete parent-of-origin effect model and at nominal 0.5% level, (f) incomplete parent-of-origin effect model and at nominal 0.5% level. PPAT_{full} is for complete pedigree data with no missing genotype in each pedigree; MCPPAT_T and MCPPAT_E are based on the true and estimated allele frequencies, respectively, for incomplete pedigree data (removing individuals 1 and 5); PPAT₁ and PPAT₅ are for the families with either individual 1 or individual 5 missing without using MC samples; PPAT_{1,5} employs informative family trios only from the nuclear family with individuals 4, 6, 9 and 10 (removing individuals 1 and 5); PAT is based on trios randomly selected from the nuclear family of individuals 4, 6, 9 and 10.

Fig. 3. Difference between the absolute values of MCPPAT_E and PAT vs. the number of affected offspring under (a) complete parent-of-origin effect model and (b) incomplete parent-of-origin effect model.

TABLE I

Simulation settings^a

A. Haplotype frequencies and homozygote penetrances settings									
Setting	Haplotype frequencies				Penetrances				
	DM_1	dM_1	DM_2	dM_2	$\phi_{D/D}$	$\phi_{d/d}$	$\phi_{D/d}$	$\phi_{D/d}$	$\phi_{d/d}$
1	0.2	0.0	0.1	0.7	0.390	0.260			
2	0.3	0.1	0.0	0.6	0.390	0.260			
3	0.3	0.0	0.0	0.7	0.390	0.260			
4	0.2	0.0	0.1	0.7	0.440	0.240			
5	0.3	0.1	0.0	0.6	0.440	0.240			
6	0.3	0.0	0.0	0.7	0.440	0.240			
7	0.2	0.0	0.1	0.7	0.580	0.180			
8	0.3	0.1	0.0	0.6	0.580	0.180			
9	0.3	0.0	0.0	0.7	0.580	0.180			

B. Heterozygote penetrances settings for three parent-of-origin effect models									
Setting	Parent-of-origin effect models				Complete				
	$\phi_{D/d}$	$\phi_{d/D}$	$\phi_{D/d}$	$\phi_{d/D}$	$\phi_{D/d}$	$\phi_{d/D}$	$\phi_{D/d}$	$\phi_{d/D}$	$\phi_{D/d}$
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	0.380	0.380	0.530	0.230	0.580	0.180			

^aA and B combined lead to three sets of nine settings for our simulation study. The first set assumes no parent-of-origin effects for studying the type I error rates. The second and third sets entertain the incomplete and complete parent-of-origin effect models, respectively, for the study of the power of the proposed statistics.

TABLE II

Type I error rates (%) based on 20,000 replicates

Type I error rates									
Setting	Complete data		Incomplete data						
	PPAT _{full}	MCPPAT _T	MCPPAT _E	PPAT _{1,5}	PPAT ₁	PPAT ₅	PAT		
A. At nominal 5% level									
1	4.96	4.99	4.97	5.32	4.95	4.92	5.01		
2	4.80	4.84	4.93	5.21	4.86	4.61	5.41		
3	4.98	5.18	5.20	5.16	5.05	4.84	5.21		
4	4.96	5.00	5.04	5.29	4.86	4.96	5.32		
5	5.05	4.94	4.94	4.97	4.90	5.01	5.14		
6	4.94	5.01	5.00	5.01	4.70	5.04	5.50		
7	5.08	5.14	5.19	5.03	4.99	4.94	5.03		
8	5.06	5.03	4.99	5.17	4.95	5.08	5.33		
9	4.94	4.97	4.97	4.97	5.02	4.91	5.21		
B. At nominal 1% level									
1	0.86	0.89	0.92	0.70	0.87	0.78	0.76		
2	0.91	0.80	0.81	0.95	0.84	0.83	0.90		
3	0.95	0.98	0.96	0.87	0.88	0.93	0.86		
4	0.94	1.03	1.01	0.79	0.81	0.99	0.85		
5	0.97	0.83	0.91	0.82	0.91	0.87	0.81		
6	0.92	0.87	0.89	0.97	0.75	0.90	0.94		
7	0.97	1.05	1.02	0.81	0.89	0.90	0.84		
8	1.01	1.01	1.00	0.88	0.92	1.11	0.82		
9	1.01	0.95	0.97	0.87	0.93	0.98	0.88		
C. At nominal 0.5% level									
1	0.405	0.430	0.410	0.360	0.395	0.365	0.440		
2	0.400	0.415	0.400	0.475	0.345	0.350	0.505		
3	0.385	0.430	0.415	0.420	0.440	0.335	0.435		
4	0.470	0.465	0.485	0.445	0.325	0.395	0.560		
5	0.455	0.470	0.490	0.405	0.385	0.410	0.435		

Setting	Type I error rates									
	Complete data					Incomplete data				
	PPAT _{full}	MCPAT _T	MCPAT _E	PPAT _{I,S}	PPAT _I	PPAT _S	PAT			
6	0.415	0.435	0.430	0.470	0.365	0.435	0.515			
7	0.475	0.510	0.510	0.390	0.445	0.340	0.450			
8	0.475	0.505	0.465	0.395	0.400	0.510	0.435			
9	0.540	0.460	0.430	0.425	0.435	0.475	0.435			

TABLE III

Results of application to rheumatoid arthritis data^a

A. Contingency table showing MCPPAT and PPAT results in the presence of association at 1% level (i.e. $P_{\text{MCPDT}} < 0.01$)			
	$P_{\text{PPAT}} < 0.01$	$P_{\text{PPAT}} \geq 0.01$	Total
$P_{\text{MCPPAT}} < 0.01$	0	3	3
$P_{\text{MCPPAT}} \geq 0.01$	0	143	143
Total	0	146	146

B. P -values of MCPDT, PPAT, and MCPPAT of the three SNPs for which $P_{\text{MCPPAT}} < 0.01$			
SNP name	P_{MCPDT}	P_{PPAT}	P_{MCPPAT}
SNP76924	0.0058	0.0833	0.0067
SNP615817	0.0026	0.0143	0.0015
SNP615818	0.0058	0.0343	0.0010

^a P_{MCPDT} , P_{PPAT} , P_{MCPPAT} are the P -values of MCPDT, PPAT, and MCPPAT, respectively.