

A model for family-based case–control studies of genetic imprinting and epistasis

Xin Li*, Yihan Sui*, Tian Liu*, Jianxin Wang, Yongci Li, Zhenwu Lin, John Hegarty, Walter A. Koltun, Zuoheng Wang and Rongling Wu

Submitted: 15th May 2013; Received (in revised form): 20th June 2013

Abstract

Genetic imprinting, or called the parent-of-origin effect, has been recognized to play an important role in the formation and pathogenesis of human diseases. Although the epigenetic mechanisms that establish genetic imprinting have been a focus of many genetic studies, our knowledge about the number of imprinting genes and their chromosomal locations and interactions with other genes is still scarce, limiting precise inference of the genetic architecture of complex diseases. In this article, we present a statistical model for testing and estimating the effects of genetic imprinting on complex diseases using a commonly used case–control design with family structure. For each subject sampled from a case and control population, we not only genotype its own single nucleotide polymorphisms (SNPs) but also collect its parents' genotypes. By tracing the transmission pattern of SNP alleles from parental to offspring generation, the model allows the characterization of genetic imprinting effects based on Pearson tests of a 2×2 contingency table. The model is expanded to test the interactions between imprinting effects and additive, dominant and epistatic effects in a complex web of genetic interactions. Statistical properties of the model are investigated, and its practical usefulness is validated by a real data analysis. The model will provide a useful tool for genome-wide association studies aimed to elucidate the picture of genetic control over complex human diseases.

Keywords: *genetic imprinting; epistasis; contingency table; single nucleotide polymorphism*

INTRODUCTION

Genetic imprinting arises from a gene when either its maternally or paternally derived copy is expressed while the other copy is silenced [1]. As a consequence of epigenetic marks due to differential DNA methylation during gametogenesis, genetic

imprinting has been found to play a pivotal role in regulating the formation, development, function and evolution of complex traits and diseases [2–5]. Tremendous efforts have been made to study the epigenetic and molecular mechanisms of this phenomenon [6], but the number and distribution of

Corresponding author. Rongling Wu, Center for Statistical Genetics, Pennsylvania State University, Hershey, PA 17033, USA. Tel.: +001-717-531-2037; Fax: +001-717-531-0480; E-mail: rwu@phs.psu.edu

*These authors contributed to this work equally.

Xin Li is a Master's student in bioinformatics in the Center for Computational Biology at Beijing Forestry University.

Yihan Sui is a fourth-year undergraduate in biological science in the College of Biological Science and Technology at Beijing Forestry University.

Tian Liu is a post-doctoral researcher in statistical genetics in the Center for Lifespan Psychology at Max Planck Institute for Human Development.

Jianxin Wang is a Professor in information science in the College of Information at Beijing Forestry University.

Yongci Li is an Associate Professor in statistics in the College of Science at Beijing Forestry University.

Zhenwu Lin is an Assistant Professor in surgery in the Department of Surgery at Penn State College of Medicine.

John Hegarty is an Assistant Professor in surgery in the Department of Surgery at Penn State College of Medicine.

Walter A. Koltun is a Professor in surgery in the Department of Surgery at Penn State College of Medicine.

Zuoheng Wang is an Assistant Professor in biostatistics in the Department of Biostatistics at Yale University.

Rongling Wu is a Professor in biostatistics and bioinformatics in the Center for Statistical Genetics at Pennsylvania State University.

imprinted genes and their epistatic interactions are poorly understood, thereby limiting our ability to estimate the effects of imprinting genes on complex traits or diseases. Several authors have used linkage analysis to identify the regions of the genome that contain imprinted sequence variants and further understand the epigenetic variation of complex traits extended the transmission disequilibrium test to test imprinting effects [7–12]. Howey and Cordell reviewed two packages of software for estimating maternal and imprinting effects based on child–parent configurations [13].

As one of the simplest and most powerful approaches for genome-wide identification of disease genes, association studies based on a case–control design have been widely used in human genetics [14]. This design allows the inheritance modes of genetic control to be tested and estimated. Statistical analyses based on a contingency table suggest that some genes function additively, whereas the others function in a dominant or overdominant manner [15]. A recent model by Liu *et al.* [16] integrates traditional quantitative genetic principles into a case–control design, producing a 2×2 contingency table, in which epistasis of different kinds and orders can be characterized. More recently, Zhang *et al.* [17] have embraced haplotype discovery procedures and Liu *et al.*'s model to identify genetic interactions between haplotypes located at different positions of the genome. Sui *et al.* [18] have incorporated DNA methylation mechanisms into Liu *et al.*'s model, allowing the effect of epigenetic marks on complex disease to be tested.

In this article, we describe a statistical model for estimating genetic imprinting based on a commonly used case–control design by implementing family structure. Unlike classical case–control studies in which random samples of cases and controls are drawn, respectively, from a natural population, a family-based case–control design includes genotyping both these random samples and their parents. By classifying cases and controls into different groups of genotypes at individual genes based on their own genotypes and their parents' genotypes, we can discern the discrepancy of alleles in their origin, which allows imprinting effects to be characterized. We performed simulation studies to test the statistical behavior of the model and validate its utilization. The new model was used to analyze a real data set collected for inflammatory bowel disease (IBD) [10], leading to the detection of significant imprinting effects at several loci.

BASIC MODEL

Population sampling

The study is based on a case–control study, in which a group of patients (cases) are randomly sampled from a natural population, matched with normal subjects (controls) of a similar size in terms of demographic factors. Genome-wide genotyping is conducted for both case and control groups. For each case and control, their maternal and paternal parents are also genotyped, no matter whether these parents have a disease or are normal. Thus, our case–control design is constructed using random samples of families rather than random samples of individuals from a population.

Consider a single nucleotide polymorphism (SNP) with two alleles, A and a , which generate three genotypes AA (coded as 2), Aa (coded as 1) and aa (coded as 0). Owing to sex-linked selection, the genotypes have different segregating proportions in a population of each sex. Let P_2, P_1, P_0 and Q_2, Q_1, Q_0 denote three genotype frequencies in the female and male population, respectively. A mating between the maternal and paternal parents forms nine different genotypic combinations (Table 1). If the population is at Hardy–Weinberg disequilibrium (HWD), in which the mating is not random, then the frequencies of genotypic combinations are expressed as the products of maternal and paternal genotype frequencies plus disequilibria, D_1, D_2, D_3 and D_4 . Let n_{ij} denote the size of the combination between maternal genotype i and paternal genotype j ($i, j = 2, 1, 0$), from which the genotype frequencies are estimated as

$$\begin{aligned} \hat{P}_2 &= \frac{1}{n..} (n_{22} + n_{21} + n_{20}), \quad \hat{P}_1 = \frac{1}{n..} (n_{12} + n_{11} + n_{10}), \\ \hat{P}_0 &= \frac{1}{n..} (n_{02} + n_{01} + n_{00}), \quad \hat{Q}_2 = \frac{1}{n..} (n_{22} + n_{12} + n_{02}), \\ \hat{Q}_1 &= \frac{1}{n..} (n_{21} + n_{11} + n_{01}), \quad \hat{Q}_0 = \frac{1}{n..} (n_{20} + n_{10} + n_{00}), \\ \hat{D}_1 &= \frac{n_{22}}{n..} - \hat{P}_2 \hat{Q}_2, \quad \hat{D}_2 = \frac{n_{21}}{n..} - \hat{P}_2 \hat{Q}_1, \\ \hat{D}_3 &= \frac{n_{12}}{n..} - \hat{P}_1 \hat{Q}_2, \quad \hat{D}_4 = \frac{n_{11}}{n..} - \hat{P}_1 \hat{Q}_1, \end{aligned} \quad (1)$$

where $n.. = \sum_{i=0}^2 \sum_{j=0}^2 n_{ij}$ is the total number of families sampled for cases or controls.

If there is a segregation distortion in a unisex population, the genotype frequencies are expressed as $P_2 = p^2 + D_f$, $P_1 = 2p(1-p) - 2D_f$ and $P_0 = (1-p)^2 + D_f$ for the female population and $Q_2 = q^2 + D_m$, $Q_1 = 2q(1-q) - 2D_m$ and $Q_0 = (1-q)^2 + D_m$ for the male population, respectively, where p and q are

Table 1: Mating types of three genotypes, their observations and frequencies under HWD or HWE and no distortion as well as Mendelian ratios of their offspring genotypes

Mating type	Observation	Frequency		Offspring genotype			
		HWD	Equilibrium	AA	A a	a A	aa
AA × AA	n_{22}	$P_2Q_2 + D_1$	p^2q^2	1	0	0	0
AA × Aa	n_{21}	$P_2Q_1 + D_2$	$2p^2q(1-q)$	$\frac{1}{2}$	$\frac{1}{2}$	0	0
AA × aa	n_{20}	$P_2Q_0 - D_1 - D_2$	$p^2(1-q)^2$	0	1	0	0
Aa × AA	n_{12}	$P_1Q_2 + D_3$	$2p(1-p)q^2$	$\frac{1}{2}$	0	$\frac{1}{2}$	0
Aa × Aa	n_{11}	$P_1Q_1 + D_4$	$4p(1-p)q(1-q)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$
Aa × aa	n_{10}	$P_1Q_0 - D_3 - D_4$	$2p(1-p)(1-q)^2$	0	$\frac{1}{2}$	0	$\frac{1}{2}$
aa × AA	n_{02}	$P_0Q_2 - D_1 - D_3$	$(1-p)^2q^2$	0	0	1	0
aa × Aa	n_{01}	$P_0Q_1 - D_2 - D_4$	$2(1-p)^2q(1-q)$	0	0	$\frac{1}{2}$	$\frac{1}{2}$
aa × aa	n_{00}	$P_0Q_0 + D_1 + D_2 + D_3 + D_4$	$(1-p)^2(1-q)^2$	0	0	0	1

Note: The parent-of-origin of the alleles for the heterozygous offspring is separated by a vertical line.

the frequency of allele A , and D_f and D_m are the coefficient of distortion in the female and male population, respectively. If the population is at Hardy-Weinberg equilibrium (HWE) and also there is no distortion, we express the genotype frequencies as p^2 or q^2 for AA , $2p(1-p)$ or $2q(1-q)$ for Aa , and $(1-p)^2$ or $(1-q)^2$ for aa (Table 1).

Whether the population is at HWE can be tested by formulating the hypotheses $H_0: D_1 = D_2 = D_3 = D_4 = 0$ versus H_1 : at least one of D 's $\neq 0$, from which the log-likelihood ratio test statistic is calculated and compared with a chi-square critical threshold with four degrees of freedom. The HWD test provides basic information about the mating behavior of the original population.

It is also interesting to investigate how the population has diversified between two sexes by testing $H_0: p = q$ and $D_f = D_m$ versus H_1 : at least one of equality does not hold. The sex-dependent differences in allele frequency and distortion coefficient can be tested separately. If these tests are significant, this indicates that sex-linked selection has played a pivotal role in affecting the genetic diversity and evolution of populations, which can be thought of an important cause of genetic imprinting.

Offspring genotyping

In Table 1, we also give segregation ratios of offspring genotypes from each mating type under Mendel's first law. As the genotypes of the parents are known, we can infer the maternal or paternal origin of the alleles in the offspring for all mating types, except for type $Aa \times Aa$ in which the parental

origin of the alleles for the heterozygous offspring is unknown. For example, type AA (maternal) \times Aa (paternal) produces offspring genotype AA and Aa , but it is clear that heterozygote Aa has allele A inherited from the maternal parent and allele a from the paternal parent. We use configuration $A|a$ to separate the parental origin of alleles with one at the left side of the vertical line being from the maternal and the other at the right side from the paternal. Type Aa (maternal) \times AA (paternal) produces genotypes AA and Aa , the latter of which should be configuration $a|A$. For type Aa (maternal) \times Aa (paternal), which produces genotypes AA , Aa and aa , we do not know whether a given offspring of Aa is $A|a$ or $a|A$.

Let n_{ijk} and m_{ijk} denote the observations of offspring genotype k ($k=2$ for genotype AA , 1 for Aa , 0 for aa) derived from the $i \times j$ mating type in the cases and controls, respectively. The total observations for the cases and controls are expressed as $n_{...} = \sum_{i=0}^2 \sum_{j=0}^2 \sum_{k=0}^2 n_{ijk}$ and $m_{...} = \sum_{i=0}^2 \sum_{j=0}^2 \sum_{k=0}^2 m_{ijk}$. It is not difficult to give the observations of offspring genotypes AA and aa in case and control groups, respectively, by summing the numbers of the corresponding genotypes over all possible mating types. The same thing can be done for two configurations $A|a$ and $a|A$, but for the offspring heterozygote Aa derived from $Aa \times Aa$, these two configurations, despite mixed, can take a half of n_{111} for each according to Mendel's segregation ratio. Table 2 presents a 2×4 contingency table having four offspring genotypes/configurations in cases and controls,

Table 2: Observations for four offspring genotypes/configurations in cases and controls

	AA	A a	a A	aa	Total
Case	$n_2 = n_{222} + n_{212} + n_{122} + n_{112}$	$n_1 = n_{211} + n_{201} + \frac{1}{2}n_{111} + n_{101}$	$n_{1'} = n_{121} + \frac{1}{2}n_{111} + n_{021} + n_{011}$	$n_0 = n_{110} + n_{100} + n_{010} + n_{000}$	R_{2+}
Control	$m_2 = m_{222} + m_{212} + m_{122} + m_{112}$ R_{+2}	$m_1 = m_{211} + m_{201} + \frac{1}{2}m_{111} + m_{101}$ R_{+1}	$m_{1'} = m_{121} + \frac{1}{2}m_{111} + m_{021} + m_{011}$ $R_{+1'}$	$m_0 = m_{110} + m_{100} + m_{010} + m_{000}$ R_{+0}	R_{1+}

which is generated using observations of offspring genotypes.

Test and estimation

Contingency table for genotypes and configurations

The derivation of our model starts with Table 2. Let R_{uv} denote the actual frequency of observations in the u th row ($u = 1$ for cases and 2 for controls) and v th column ($v = 2$ for AA, 1 for A|a, 1' for a|A and 0 for aa), let R_{u+} and R_{+v} denote column and row marginals, respectively, and let R_{++} denote the grand total. If the genotypes/configurations are independent of the disease, then the probability of cell uv is simply the product of the probabilities of marginal categories u and v , expressed as $r_{uv} = \frac{R_{u+}R_{+v}}{R_{++}}$.

A Pearson statistic for testing the genotype/configuration–disease association can be calculated by

$$\chi_G^2 = \sum_{u=1}^2 \sum_{v=0}^2 \frac{(r_{uv} - R_{uv})^2}{r_{uv}} \tag{2}$$

which follows a chi-square distribution with $(2-1) \times (4-1) = 3$ degrees of freedom. If $\chi_G^2 > \chi_{0.05(df=3)}^2$, this suggests that the genotypes/configurations are significantly associated with the disease. Otherwise, there is no significant association between the genotypes/configurations and the disease.

Testing imprinting effect

Based on Table 2, we will test various genetic effects, including additive genetic, dominant genetic and imprinting genetic effects. According to Liu *et al.*'s model, we construct a 2×2 contingency table to test the additive effect (a) as

	AA	aa
Case	n_2	n_0
Control	m_2	m_0

(3)

a 2×2 contingency table to test the dominant effect (d) as

	Aa	AA + aa
Case	$n_1 + n_{1'}$	$\frac{1}{2}(n_2 + n_0)$
Control	$m_1 + m_{1'}$	$\frac{1}{2}(m_2 + m_0)$

(4)

a 2×2 contingency table to test the imprinting effect (I) as

	A a	a A
Case	n_1	$n_{1'}$
Control	m_1	$m_{1'}$

(5)

Chi-square test statistics are calculated for each effect, i.e. χ_a^2 , χ_d^2 and χ_I^2 , based on the corresponding contingency tables (3)–(5). As cell counts in these tables are formed by a weighted combination of observed cell counts in Table 2, these test statistics may not obey a chi-square distribution with one degree of freedom. Liu *et al.* [16] have proved that under the null hypothesis, these test statistics are asymptotically smaller than $\chi_{0.05(df=1)}^2$. They further derived an approximation approach for determining the critical thresholds.

Estimating imprinting effects

The effect sizes of a , d and I are estimated by odd ratios (OR). A general form of contingency tables (3)–(5) is expressed as the joint distribution of two binary random variables X and Y , i.e.

	Y = 1	Y = 2
X = 1	N_1	N_2
X = 2	M_1	M_2

from which an OR is calculated as

$$OR = \frac{N_1 M_2}{N_2 M_1} \tag{6}$$

describing the extent to which the two variables are related through a particular effect. $OR = 1$ suggests no relationship. The confidence interval of the OR effect value is calculated as $e^{\ln \hat{OR} \pm z \cdot SE}$, where z is a standard normal deviate corresponding to the level of confidence ($z = 1.96, 2.576$ for 95% and 99% confidence, respectively), and SE is the standard deviation of OR calculated as $SE = \sqrt{\frac{1}{n_1} + \frac{1}{n_2} + \frac{1}{m_1} + \frac{1}{m_2}}$.

Missing data

In practice, it is possible that some individuals have no information about the genotype of their

Table 3: Frequencies of different mating types and Mendelian ratios of offspring genotypes when paternal information is missing

Mating Type	Observation	Frequency	Offspring genotype			
			AA	A a	a A	aa
AA × -	n_2	P_2	$P_2 + \frac{1}{2}P_1$	$\frac{1}{2}P_1 + P_0$	0	0
Aa × -	n_1	P_1	$\frac{1}{2}P_2 + \frac{1}{4}P_1$	$\frac{1}{4}P_1 + \frac{1}{2}P_0$	$\frac{1}{2}P_2 + \frac{1}{4}P_1$	$\frac{1}{4}P_1 + \frac{1}{2}P_0$
aa × -	n_0	P_0	0	0	$\frac{1}{2}P_1 + P_2$	$P_0 + \frac{1}{2}P_1$

Note: n_2 , n_1 , and n_0 are the cumulative numbers of families with maternal parent AA, Aa and aa, respectively. Genotype frequencies P_2 , P_1 and P_0 here are expressed in terms of maternal parents.

Table 4: Observations for four offspring genotype/configuration in cases and controls when paternal information is missing

	AA	A a	a A	aa	Total
Case	$n_2 = n_{2.2} + n_{1.2}$	$n_1 = n_{2.1} + (\frac{1}{4}P_1 + \frac{1}{2}P_0)n_{1.1}$	$n_{1'} = n_{0.1} + (\frac{1}{2}P_2 + \frac{1}{4}P_1)n_{1.1}$	$n_0 = n_{1.0} + n_{0.0}$	R_{2+}
Control	$m_2 = m_{2.2} + m_{1.2}$ R_{+2}	$m_1 = m_{2.1} + (\frac{1}{4}\hat{P}_1 + \frac{1}{2}\hat{P}_0)m_{1.1}$ R_{+1}	$m_{1'} = m_{0.1} + (\frac{1}{2}\hat{P}_2 + \frac{1}{4}\hat{P}_1)m_{1.1}$ $R_{+1'}$	$m_0 = m_{1.0} + m_{0.0}$ R_{+0}	R_{1+}

Note: $n_{i,k}$ and $m_{i,k}$ are the numbers of offspring genotype k from families with maternal parental genotype i in cases and controls, respectively. \hat{P}_2 , \hat{P}_1 and \hat{P}_0 are genotype frequencies of AA, Aa and aa in controls.

Table 5: Population genetic parameters used to simulate family sizes according to Table 1

Parameter	Case		Control	
	Female	Male	Female	Male
Mating disequilibrium				
D_1	0.04		0.05	
D_2	0.04		0.05	
D_3	0.04		0.05	
D_4	0.04		- ^a	
Situation 1				
Allele frequency (A)	0.3	0.5	0.5	0.3
Allele frequency (a)	0.7	0.5	0.5	0.7
Distortion D_f/D_m	0.05	0.10	0.10	0.05
Situation 2				
Allele frequency (A)	0.3	0.3	0.5	0.5
Allele frequency (a)	0.7	0.7	0.5	0.5
Distortion D_f/D_m	0.05	0.05	0.10	0.10

^aThis disequilibrium is not fixed, whose value is taken depending on the OR value.

parents. This missing problem does not affect the grouping of homozygote AA and aa but makes it difficult to distinguish between A|a and a|A. Later in the text, we describe a procedure for constructing a contingency table for four genotypes/configurations when one parent is missing. Consider a family-based control design in which the genotypes

of maternal parents are available but those of paternal parents are missing with data structure tabulated in Table 3. In this situation, a 2×4 contingency table, as shown in Table 4, can be generated from Table 3.

From Table 4, we can now test and estimate the additive, dominant and imprinting effects at a given SNP using the procedure described earlier in the text. As can be seen, even if the parental information of some families is missing, we can still conduct imprinting test.

Worked example

The model was used to analyze a real data derived from a case-control study with family structure. The study population includes 58 families from the Milton S Hershey Familial Inflammatory Bowel Disease Registry, from which 105 offspring affected with IBD (cases) and 139 matched offspring without IBD (controls) were sampled for SNP genotyping. A total of five SNPs (Arg39Gln, Glu514Gln, Pro979Leu, Gly1066Gly and Pro979Leu) were typed from the candidate gene GLD5 for IBD [19].

None of the five SNPs displays a significant overall genetic effect based on a chi-square test for a 2×4 contingency table as shown in Table 2. This table is

reorganized to a 2×2 contingency table based on additive (3), dominant (4) and imprinting effects (5). It was found that no significant additive and dominant effects exist to affect IBD. But tests for imprinting effects indicate that SNP Arg39Gln triggers a significant imprinting effect on IBD ($P=0.0476$), and SNP Gly1066Gly displays a marginally significant imprinting effect ($P=0.0608$). The OR values for these SNPs were estimated as 1.5 and 1.7, respectively, suggesting their importance in influencing IBD.

Simulation

We performed simulation studies to investigate the statistical properties of the model. A natural human population was mimicked, from which a panel of unrelated families (each including a male parent, a female parent, and one or more children) is randomly sampled. Given a total of 1000 subjects, the simulation considers two sampling strategies, 1000×1 (1000 families with a single child) and 200×5 (200 families with five children). For each sampling strategy, we considered two situations in which the same marker is segregating differently in populations of cases and controls but with or without difference between two sexes. The parameters that were used to simulate genotype frequencies of AA , Aa and aa and the size of each mating type for cases and controls are given in Table 5, from which the observations of offspring genotypes/configurations were obtained per Mendel's segregation ratio.

We will focus on the calculation of the power and false-positive rates of detecting the imprinting effect from the model. The count data in 2×2 contingency table (5) were simulated by assuming different ORs. The OR describes and quantifies the

extent to which the two variables are related. Values of $OR=1, 1.5, 2.0, 2.5$ are regarded as no effect, small effect, moderate effect and large effect, respectively. The count data were simulated under each of these values. To simulate the data under these OR constraints, we relax the disequilibrium D_4 of in the control population. In each situation, simulation was replicated 1000 times to estimate the power of imprinting detection by the new model.

Table 6 gives the results of power calculation from simulated data from 1000×1 and 200×5 sampling strategies. In general, power is not affected by sex-specific differences in population structure (allele frequencies and distortion degree) because the results are similar in two situations. The strategy of sampling more small families performs better than that of sampling less larger families, especially when the genetic effect is modest. The power to detect the dominance effect is generally larger than that to detect the additive and imprinting effects because more data are used in the former than latter. On the other hand, the model has a small false-positive rate (<0.06 or 0.05) when each of these effects does not exist actually. The main message from the aforementioned simulation is that, for a given sample size, many families each of small size are recommended as a better sampling strategy to detect the additive, dominant and imprinting effects.

INTERACTION MODEL TWO-LOCUS EFFECTS

The model can be extended to analyze two SNPs simultaneously that may interact with each other to determine the disease. Consider two SNPs, **A** (with

Table 6: Power calculation of imprinting detection under different sampling strategies

Strategy		Situation 1			Situation 2		
		<i>a</i>	<i>d</i>	<i>l</i>	<i>a</i>	<i>d</i>	<i>l</i>
200×5	OR=1.0	0.049	0.017	0.045	0.052	0.015	0.052
	OR=1.5	0.280	0.368	0.279	0.300	0.372	0.290
	OR=2.0	0.582	0.853	0.569	0.723	0.856	0.557
	OR=2.5	0.750	0.981	0.749	0.930	0.981	0.687
1000×1	OR=1.0	0.038	0.017	0.043	0.057	0.018	0.041
	OR=1.5	0.866	0.980	0.872	0.884	0.981	0.885
	OR=2.0	0.996		0.996			
	OR=2.5						

two alleles A , a) and \mathbf{B} (with two alleles B , b). At each SNP, genotypes/configurations are denoted as 2 for AA or BB , 1 for $A|a$ or $B|b$, 1' for $a|A$ or $b|B$ and 0 for aa or bb . The four haplotypes, AB , Ab , aB and ab , unite to produce 16 diplotypes with genotypic values defined as

$$AB|AB: \mu_{22} = \mu + a_1 + a_2 + i_{aa}$$

$$AB|Ab: \mu_{21} = \mu + a_1 + d_2 + I_2 + i_{ad} + i_{aI}$$

$$AB|aB: \mu_{12} = \mu + d_1 + I_1 + a_2 + i_{da} + i_{Ia}$$

$$AB|ab: \mu_{11} = \mu + d_1 + I_1 + d_2 + I_2 + i_{dd} + i_{dI} + i_{Id} + i_{II}$$

$$Ab|AB: \mu_{21'} = \mu + a_1 + d_2 - I_2 + i_{ad} - i_{aI}$$

$$Ab|Ab: \mu_{20} = \mu + a_1 - a_2 - i_{aa}$$

$$Ab|aB: \mu_{11'} = \mu + d_1 + I_1 + d_2 - I_2 + i_{dd} - i_{dI} + i_{Id} - i_{II}$$

$$Ab|ab: \mu_{10} = \mu + d_1 + I_1 - a_2 - i_{da} - i_{Ia}$$

$$aB|AB: \mu_{1'2} = \mu + d_1 - I_1 + a_2 + i_{da} - i_{Ia}$$

$$aB|Ab: \mu_{1'1} = \mu + d_1 - I_1 + d_2 + I_2 + i_{dd} + i_{dI} - i_{Id} - i_{II}$$

$$aB|aB: \mu_{02} = \mu - a_1 + a_2 - i_{aa}$$

$$aB|ab: \mu_{01} = \mu - a_1 + d_2 + I_2 - i_{ad} - i_{aI}$$

$$ab|AB: \mu_{1'1'} = \mu + d_1 - I_1 + d_2 - I_2 + i_{dd} - i_{dI} - i_{Id} + i_{II}$$

$$ab|Ab: \mu_{1'0} = \mu + d_1 - I_1 - a_2 - i_{da} + i_{Ia}$$

$$ab|aB: \mu_{01'} = \mu - a_1 + d_2 - I_2 - i_{ad} + i_{aI}$$

$$ab|ab: \mu_{00} = \mu - a_1 - a_2 + i_{aa}$$

where μ is the overall mean; a_1 , d_1 and I_1 are the additive, dominant and imprinting effects at SNP \mathbf{A} ; a_2 , d_2 and I_2 are the additive, dominant and imprinting effect at SNP \mathbf{B} ; i_{aa} , i_{ab} , i_{aI} , i_{da} , i_{db} , i_{dI} , i_{Ia} , i_{Id} and i_{II} are the additive \times additive, additive \times dominant, additive \times imprinting, dominant \times additive, dominant \times dominant, dominant \times imprinting, imprinting \times additive, imprinting \times dominant and imprinting \times imprinting interaction effects between the two SNPs. These parameters can be solved from the aforementioned group of equations as

$$\mu = \frac{1}{4}(\mu_{22} + \mu_{20} + \mu_{02} + \mu_{00}) \quad (7)$$

$$a_1 = \frac{1}{4}(\mu_{22} + \mu_{20} - \mu_{02} - \mu_{00})$$

$$d_1 = \frac{1}{4}(\mu_{12} - \mu_{22} - \mu_{20} + \mu_{10} + \mu_{1'2} - \mu_{02} + \mu_{1'0} - \mu_{00}) \quad (8)$$

$$I_1 = \frac{1}{4}(\mu_{12} + \mu_{10} - \mu_{1'2} - \mu_{1'0}) \quad (9)$$

$$a_2 = \frac{1}{4}(\mu_{22} - \mu_{20} + \mu_{02} - \mu_{00}) \quad (10)$$

$$d_2 = \frac{1}{4}(\mu_{21} - \mu_{22} + \mu_{21'} - \mu_{20} - \mu_{02} + \mu_{01} + \mu_{01'} - \mu_{00}) \quad (11)$$

$$I_2 = \frac{1}{4}(\mu_{21} - \mu_{21'} + \mu_{01} - \mu_{01'}) \quad (12)$$

$$i_{aa} = \frac{1}{4}(\mu_{22} - \mu_{20} - \mu_{02} + \mu_{00}) \quad (13)$$

$$i_{ad} = \frac{1}{4}(\mu_{21} - \mu_{22} + \mu_{21'} - \mu_{20} + \mu_{02} - \mu_{01} - \mu_{01'} + \mu_{00}) \quad (14)$$

$$i_{aI} = \frac{1}{4}(\mu_{21} - \mu_{21'} - \mu_{01} + \mu_{01'}) \quad (15)$$

$$i_{da} = \frac{1}{4}(\mu_{12} - \mu_{22} + \mu_{20} - \mu_{10} + \mu_{1'2} - \mu_{02} - \mu_{1'0} + \mu_{00}) \quad (16)$$

$$i_{dd} = \frac{1}{4}(\mu_{22} - \mu_{21} - \mu_{12} + \mu_{11} - \mu_{21'} + \mu_{20} + \mu_{11'} - \mu_{10} - \mu_{1'2} + \mu_{1'1} + \mu_{02} - \mu_{01} + \mu_{1'1'} - \mu_{1'0} - \mu_{01'} + \mu_{00}) \quad (17)$$

$$i_{dI} = \frac{1}{4}(\mu_{11} - \mu_{21} + \mu_{21'} - \mu_{11'} + \mu_{1'1} - \mu_{01} - \mu_{1'1'} + \mu_{01'}) \quad (18)$$

$$i_{Ia} = \frac{1}{4}(\mu_{12} - \mu_{10} - \mu_{1'2} + \mu_{1'0}) \quad (19)$$

$$i_{Id} = \frac{1}{4}(\mu_{11} - \mu_{12} + \mu_{11'} - \mu_{10} + \mu_{1'2} - \mu_{1'1} - \mu_{1'1'} + \mu_{1'0}) \quad (20)$$

$$i_{II} = \frac{1}{4}(\mu_{11} - \mu_{11'} - \mu_{1'1} + \mu_{1'1'}) \quad (21)$$

Later in the text, we describe a procedure for testing each of these effects in a family-based case-control design.

Family structure

We assume that natural selection has driven two sexes to be segregating differently. Let $p_{11} = p_1 p_2 + D_F$, $p_{10} = p_1(1 - p_2) - D_F$, $p_{01} = (1 - p_1)p_2 - D_F$ and $p_{00} = (1 - p_1)(1 - p_2) + D_F$ denote the frequencies of four haplotypes AB , Ab , aB and ab at two SNPs, \mathbf{A} and \mathbf{B} , in the female population, where p_1 and p_2 are the allele frequencies of A and B and D_F is the linkage disequilibrium between the two SNPs [20]. Similarly, we have parameters q_{11} , q_{10} , q_{01} , q_{00} , q_1 , q_2 and D_M for the male population.

The two SNPs under consideration produce nine genotypes $AABB$, $AABb$, $AAaa$, $AaBB$, $AaBb$, $Aabb$, $aaBB$, $aaBb$ and $aabb$ with frequencies denoted as P_{22} , P_{21} , P_{20} , P_{12} , P_{11} , P_{10} , P_{02} , P_{01} and P_{00} for the female population and Q_{22} , Q_{21} , Q_{20} , Q_{12} , Q_{11} , Q_{10} , Q_{02} , Q_{01} and Q_{00} for the male population. Under non-random mating, these genotypes will form 81 genotype-by-genotype combinations, which frequencies are expressed as the products of maternal genotype frequencies and paternal genotype frequencies plus 64 disequilibria. Using a similar framework in Table 1 and Equation (1), we can estimate sex-dependent genotype frequencies and 64 disequilibria.

If each parent for a mating combination is homozygous for both markers, their offspring will have one genotype. As long as one parent is heterozygous for one marker, the offspring will have two or more genotypes. For one marker, the offspring will have two or more genotypes. However, only when both markers are heterozygous for at least one parent, the genotype frequencies of offspring will be determined by the recombination fraction between the markers (θ). In the Supplementary Material, the structure and frequencies of maternal by paternal genotype combinations under random mating and their offspring genotype frequencies are given. For a double heterozygote $AaBb$, its observed genotype may be derived from two possible diploypes, $AB|ab$ or $Ab|aB$, with different formation frequencies. For the female population, haplotype frequencies produced by each of these two diploypes are expressed as

Diplotype	Prop.	AB	Ab	aB	ab
$AB ab$	ϕ	$\frac{1}{2}(1-\theta)$	$\frac{1}{2}\theta$	$\frac{1}{2}\theta$	$\frac{1}{2}(1-\theta)$
$Ab aB$	$1-\phi$	$\frac{1}{2}\theta$	$\frac{1}{2}(1-\theta)$	$\frac{1}{2}(1-\theta)$	$\frac{1}{2}\theta$
Overall	$\frac{1}{2}\omega_1$	$\frac{1}{2}\omega_2$	$\frac{1}{2}\omega_2$	$\frac{1}{2}\omega_1$	

with $\phi = p_{11}p_{00}/(p_{11}p_{00} + p_{10}p_{01})$ and $\omega_1 = (1-\theta)\phi + \theta(1-\phi)$ and $\omega_2 = \theta\phi + (1-\theta)(1-\phi)$ that sum to 1. Similarly, for males, we have the recombination fraction ϑ and proportions $\varphi = q_{11}q_{00}/(q_{11}q_{00} + q_{10}q_{01})$, and $\varpi_1 = (1-\vartheta)\varphi + \vartheta(1-\varphi)$ and $\varpi_2 = \vartheta\varphi + (1-\vartheta)(1-\varphi)$.

Let $n_{i_1i_2j_1j_2}$ denote the size of a particular mating type between maternal genotype i_1i_2 and paternal genotype j_1j_2 , with $i_1, j_1 = 2$ for AA , 1 for Aa and 0 for aa ; $i_2, j_2 = 2$ for BB , 1 for Bb and 0 for bb . Let $n_{i_1i_2j_1j_2k_1k_2}$ denote the number of offspring

genotype/configuration k_1k_2 in a family derived from maternal genotype i_1i_2 and paternal genotype j_1j_2 , with $k_1 = 2$ for AA , 1 for $A|a$, 1' for $a|A$ and 0 for aa , $k_2 = 2$ for BB , 1 for $B|b$, 1' for $b|B$ and 0 for bb . All possible mating types, their sizes and frequencies, and the frequencies of their offspring genotypes/configurations, can be tabulated as follows:

	$AABB \times$...	$AABB \times$...	$AaBb \times$...
	$AABB$		$AaBb$		$AaBb$	
	n_{2222}		n_{2211}		n_{1111}	
Offspring	$P_{22}Q_{22}$		$P_{22}Q_{11}$		$P_{11}Q_{11}$	
Config.	$+D_1$		$+D_5$		$+D_{37}$	
$AB AB$	1	...	$\frac{1}{2}\varpi_1$...	$\frac{1}{4}\omega_1\varpi_1$...
$AB Ab$	0	...	$\frac{1}{2}\varpi_2$...	$\frac{1}{4}\omega_1\varpi_2$...
$AB aB$	0	...	$\frac{1}{2}\varpi_2$...	$\frac{1}{4}\omega_1\varpi_2$...
$AB ab$	0	...	$\frac{1}{2}\varpi_1$...	$\frac{1}{4}\omega_1\varpi_1$...
$Ab AB$	0	...	0	...	$\frac{1}{4}\omega_2\varpi_1$...
$Ab Ab$	0	...	0	...	$\frac{1}{4}\omega_2\varpi_2$...
$Ab aB$	0	...	0	...	$\frac{1}{4}\omega_2\varpi_2$...
$Ab ab$	0	...	0	...	$\frac{1}{4}\omega_2\varpi_1$...
$aB AB$	0	...	0	...	$\frac{1}{4}\omega_2\varpi_1$...
$aB Ab$	0	...	0	...	$\frac{1}{4}\omega_2\varpi_2$...
$aB aB$	0	...	0	...	$\frac{1}{4}\omega_2\varpi_2$...
$aB ab$	0	...	0	...	$\frac{1}{4}\omega_2\varpi_1$...
$ab AB$	0	...	0	...	$\frac{1}{4}\omega_1\varpi_1$...
$ab Ab$	0	...	0	...	$\frac{1}{4}\omega_1\varpi_2$...
$ab aB$	0	...	0	...	$\frac{1}{4}\omega_1\varpi_2$...
$ab ab$	0	...	0	...	$\frac{1}{4}\omega_1\varpi_1$...

where $D_1, \dots, D_5, \dots, D_{37}, \dots$ are the coefficients of disequilibria.

A family design allows the simultaneous estimation of haplotype frequencies and the recombination fraction. By assuming HWD, Li and Wu [20] implemented a two-stage hierarchical EM algorithm to estimate haplotype frequencies ($p_{11}, p_{10}, p_{01}, p_{00}$) using parental mating type data ($n_{i_1i_2j_1j_2}$) at the upper level, and the recombination fraction (θ) using segregation data of offspring genotypes from each mating type ($n_{i_1i_2j_1j_2k_1k_2}$). Liu et al. [21] relaxed this assumption to allow the estimation of these parameters. These estimated parameters can be used to construct a contingency table

containing 16 diplotypes, with cells for cases expressed as

$$\begin{aligned}
 AB|AB \ n_{22} &= n_{222222} + \dots + \frac{1}{2}\omega_1 n_{221122} + \dots + \frac{1}{4}\omega_1 \varpi_1 n_{111122} + \dots \\
 AB|Ab \ n_{21} &= 0n_{222222} + \dots + \frac{1}{2}\omega_2 n_{221122} + \dots + \frac{1}{4}\omega_1 \varpi_2 n_{111122} + \dots \\
 AB|aB \ n_{21'} &= 0n_{222222} + \dots + \frac{1}{2}\omega_2 n_{221122} + \dots + \frac{1}{4}\omega_1 \varpi_2 n_{111122} + \dots \\
 AB|ab \ n_{20} &= 0n_{222222} + \dots + \frac{1}{2}\omega_1 n_{221122} + \dots + \frac{1}{4}\omega_1 \varpi_1 n_{111122} + \dots \\
 Ab|AB \ n_{12} &= 0n_{222222} + \dots + 0n_{221122} + \dots + \frac{1}{4}\omega_2 \varpi_1 n_{111122} + \dots \\
 Ab|Ab \ n_{11} &= 0n_{222222} + \dots + 0n_{221122} + \dots + \frac{1}{4}\omega_2 \varpi_2 n_{111122} + \dots \\
 Ab|aB \ n_{11'} &= 0n_{222222} + \dots + 0n_{221122} + \dots + \frac{1}{4}\omega_2 \varpi_2 n_{111122} + \dots \\
 Ab|ab \ n_{10} &= 0n_{222222} + \dots + 0n_{221122} + \dots + \frac{1}{4}\omega_2 \varpi_1 n_{111122} + \dots \\
 aB|AB \ n_{1'2} &= 0n_{222222} + \dots + 0n_{221122} + \dots + \frac{1}{4}\omega_2 \varpi_1 n_{111122} + \dots \\
 aB|Ab \ n_{1'1} &= 0n_{222222} + \dots + 0n_{221122} + \dots + \frac{1}{4}\omega_2 \varpi_2 n_{111122} + \dots \\
 aB|aB \ n_{1'1'} &= 0n_{222222} + \dots + 0n_{221122} + \dots + \frac{1}{4}\omega_2 \varpi_2 n_{111122} + \dots \\
 aB|ab \ n_{1'0} &= 0n_{222222} + \dots + 0n_{221122} + \dots + \frac{1}{4}\omega_2 \varpi_1 n_{111122} + \dots \\
 ab|AB \ n_{02} &= 0n_{222222} + \dots + 0n_{221122} + \dots + \frac{1}{4}\omega_1 \varpi_1 n_{111122} + \dots \\
 ab|Ab \ n_{01} &= 0n_{222222} + \dots + 0n_{221122} + \dots + \frac{1}{4}\omega_1 \varpi_2 n_{111122} + \dots \\
 ab|aB \ n_{01'} &= 0n_{222222} + \dots + 0n_{221122} + \dots + \frac{1}{4}\omega_1 \varpi_2 n_{111122} + \dots \\
 ab|ab \ n_{00} &= 0n_{222222} + \dots + 0n_{221122} + \dots + \frac{1}{4}\omega_1 \varpi_1 n_{111122} + \dots
 \end{aligned}
 \tag{22}$$

Similarly, the observations for controls can also be calculated, expressed as $m_{22}, m_{21}, m_{21'}, m_{20}, m_{12}, m_{11}, m_{11'}, m_{10}, m_{1'2}, m_{1'1}, m_{1'1'}, m_{1'0}, m_{02}, m_{01}, m_{01'}$ and m_{00} for these offspring diplotypes, respectively.

Testing and estimating genetic effects

From contingency table (22), we can formulate a number of tests for the additive, dominant, epistatic and imprinting effects and their mutual interaction effects. This can be done according to the procedure as described in Equations (3)–(5). First, we generate a 2×2 contingency table for 15 genetic effects based on the expressions of (7)–(21). Second, the Pearson test statistic is calculated to test the significance of each effect and OR calculated to estimate each effect from these 2×2 contingency tables. Testing the additive, dominant and imprinting effects at each SNP and the additive \times additive (i_{aa}), additive \times dominant (i_{ad}), additive \times imprinting (i_{ai}), dominant \times additive (i_{da}), dominant \times dominant (i_{dd}), dominant \times imprinting (i_{di}), imprinting \times additive (i_{ia}), imprinting \times dominant (i_{id}) and imprinting \times imprinting interaction effects (i_{II}) between the two SNPs can be conducted using 2×2 contingency tables (23)–(37), respectively, i.e.

a_1	AA	aa	(23)
Case	$n_{22} + n_{20}$	$n_{02} + n_{00}$	
Control	$m_{22} + m_{20}$	$m_{02} + m_{00}$	

d_1	AA	aa	(24)
Case	$n_{12} + n_{10} + n_{1'2} + n_{1'0}$	$n_{22} + n_{20} + n_{02} + n_{00}$	
Control	$m_{12} + m_{10} + m_{1'2} + m_{1'0}$	$m_{22} + m_{20} + m_{02} + m_{00}$	

i_1	AA	aa	(25)
Case	$n_{12} + n_{10}$	$\mu_{1'2} + n_{1'0}$	
Control	$m_{12} + m_{10}$	$m_{1'2} + m_{1'0}$	

a_2	AA	aa	(26)
Case	$n_{22} + n_{02}$	$n_{20} + n_{00}$	
Control	$m_{22} + m_{02}$	$m_{20} + m_{00}$	

d_2	AA	aa	(27)
Case	$n_{21} + n_{21'} + n_{01} + n_{01'}$	$n_{22} + n_{20} + n_{02} + n_{00}$	
Control	$m_{21} + m_{21'} + m_{01} + m_{01'}$	$m_{22} + m_{20} + m_{02} + m_{00}$	

i_2	AA	aa	(28)
Case	$n_{21} + n_{01}$	$n_{21'} + n_{01'}$	
Control	$m_{21} + m_{01}$	$m_{21'} + m_{01'}$	

i_{aa}	AA	aa	(29)
Case	$n_{22} + n_{00}$	$n_{20} + n_{02}$	
Control	$m_{22} + m_{00}$	$m_{20} + m_{02}$	

i_{ad}	AA	aa	(30)
Case	$n_{21} + n_{21'} + n_{02} + n_{00}$	$n_{22} + n_{20} + n_{01} + n_{01'}$	
Control	$m_{21} + m_{21'} + m_{02} + m_{00}$	$m_{22} + m_{20} + m_{01} + m_{01'}$	

i_{ai}	AA	aa	(31)
Case	$n_{21} + n_{01'}$	$n_{21'} + n_{01}$	
Control	$m_{21} + m_{01'}$	$m_{21'} + m_{01}$	

i_{da}	AA	aa	(32)
Case	$n_{12} + n_{20} + n_{1'2} + n_{00}$	$n_{22} + n_{10} + n_{02} + n_{1'0}$	
Control	$m_{12} + m_{20} + m_{1'2} + m_{00}$	$m_{22} + m_{10} + m_{02} + m_{1'0}$	

i_{dd}	AA	aa	(33)
Case	$n_{22} + n_{11} + n_{20} + n_{11} + n_{1'1} + n_{02} + n_{1'1'} + n_{00}$	$n_{21} + n_{12} + n_{21'} + n_{10} + n_{1'2} + n_{01} + n_{1'0} + \mu_{01}$	
Control	$m_{22} + m_{11} + m_{20} + m_{11'} + m_{1'1} + m_{02} + m_{1'1'} + m_{00}$	$m_{21} + m_{12} + m_{21'} + m_{10} + m_{1'2} + m_{01} + m_{1'0} + m_{01}$	

i_{di}	AA	aa	(34)
Case	$n_{11} + n_{21'} + n_{1'1} + n_{01'}$	$n_{21} + n_{11'} + n_{01} + n_{1'1'}$	
Control	$m_{11} + m_{21'} + m_{1'1} + m_{01'}$	$m_{21} + m_{11'} + m_{01} + m_{1'1'}$	

i_{ia}	AA	aa	(35)
Case	$n_{12} + n_{1'0}$	$n_{10} + n_{1'2}$	
Control	$m_{12} + m_{1'0}$	$m_{10} + m_{1'2}$	

i_{id}	AA	aa	(36)
Case	$n_{11} + n_{11'} + n_{1'2} + n_{1'0}$	$n_{12} + n_{10} + n_{1'1} + n_{1'1'}$	
Control	$m_{11} + m_{11'} + m_{1'2} + m_{1'0}$	$m_{12} + m_{10} + m_{1'1} + m_{1'1'}$	

i_{II}	AA	aa	(37)
Case	$n_{11} + n_{1'1'}$	$n_{11'} + n_{1'1}$	
Control	$m_{11} + m_{1'1'}$	$m_{11'} + m_{1'1}$	

The critical thresholds for the tests of the aforementioned effects can be obtained by an approximation approach derived by Liu *et al.* [16].

Computer simulation

We further simulated two SNPs to test the epistasis and imprinting effects and their interactions. We assume the two markers, with linkage disequilibrium 0.05 and recombination fraction 0.10, which are segregating with particular allele frequencies in cases and controls. Basic data structure was simulated using the tables given in the Supplementary Material. Here, we focus on the power calculation of additive \times dominant (i_{ad}), imprinting \times dominant (i_{id}) and imprinting \times imprinting effects (i_{ii}). Under $OR = 1.0$ and 2.0 , we simulated a 2×2 contingency table based on (33), (36) and (37), respectively. The allele frequencies of SNPs **A** and **B** are $P=0.6$ (allele *A*) and $q=0.5$ (allele *B*), respectively, in cases. The allele frequencies in controls were determined by adjusting the *OR* values. In both cases and controls, no sex-specific differences were assumed.

The model displays low false-positive rates under all sampling strategies. For a given sample size, more families each of small size are recommended because this strategy has better power than fewer large families (Table 7). It seems that imprinting \times dominant and imprinting \times imprinting effects can be easily detected. To well detect the additive \times dominant epistatic effect, more sample size (at least 2000×1) is needed. Overall, the model is powerful for detecting epistatic effects and imprinting effects as well as their interaction effects.

DISCUSSION

Given its significance and popularity, genetic imprinting has received tremendous attention in genetic and genomic studies [1,6]. Many studies have focused the molecular mechanisms of this phenomenon and its interplay with epigenetic marks [4]. Its genetic control as a complex trait has

also been studied using linkage analysis [7–11,22] and the transmission disequilibrium test [12]. Two packages of software have been written for imprinting estimation [13].

As a powerful genetic tool, genome-wide association studies (GWAS) have been widely used to study the genetic architecture of complex diseases [14]. Thus, the integration of imprinting effect into a GWAS setting should enhance our capacity to study the genetic control of this phenomenon. In this article, we have presented a simple χ^2 model for testing imprinting effects in a case-control design by incorporating classic quantitative genetic principles into this design. Quantitative genetics and case-control design were regarded as two different areas until several modeling works by Wu and group [15–18]. This study has for the first time integrated imprinting effects into a case-control study by genotyping information for the parents of each case and control. By tracing the parental transmission path of alleles [20], the model is able to separate the effect of the same allele when it comes from different parents.

An important advantage of this model is that it is statistically simple because all tests are reflected in a 2×2 contingency table. For this kind of table, we can easily test the association between the two variables, i.e. disease and epigenetic effect in this situation, and estimate the extent of this association using *OR*s. Conventional log-linear regression models can also be used to test imprinting effects, but it would be time-consuming or even computationally intractable when a number of effects are tested simultaneously [16]. During the model derivation, we assume that the population is at HWE and that sampled families are independent of each other. But in practice, these assumptions may not be true. For a non-equilibrium population, conceptual models have been derived to model allelic and genotype frequencies and zygotic disequilibria [22,23]. When families are related, their structure can be modeled through identical-by-descent alleles [24].

The most meritorious feature of this model lies in its ability to characterize the genetic interactions of imprinting loci distributed throughout the genome. Many imprinted loci have been thought to be highly interacted through the mediation of proteins, RNA and DNA [25–27]. It is possible that these interactions play an important role in the evolution of genetic coadaptation to changing environment [28]. With the availability of whole-genome genotyping and sequencing technologies, it has been

Table 7: Power calculation of epistatic and imprinting effect detection for two associated markers under different sampling strategies

Strategy		i_{ad}	i_{id}	i_{ii}
200×5	$OR = 1$	0.0016	0.0490	0.0450
	$OR = 2$	0.0298	0.6796	0.4140
1000×1	$OR = 1$	0.0020	0.0560	0.0380
	$OR = 2$	0.4340		0.9670
2000×1	$OR = 1$	0.0030	0.0470	0.0440
	$OR = 2$	0.8610		
3000×1	$OR = 1$	0.0040	0.0520	0.0410
	$OR = 2$	0.984		

possible to comprehend genetic control mechanisms of complex diseases and, ultimately, integrate genetic information into routine clinical therapies for disease treatment and prevention. The model presented here provides a powerful statistical and computational tool for detecting genes that determine complex diseases in an imprinting and interactive way. It will help to elucidate a detailed picture of the genetic architecture for developmental disorders.

SUPPLEMENTARY DATA

Supplementary data are available online at <http://bib.oxfordjournals.org/>.

Key Points

- Genetic imprinting is an important phenomenon thought to affect complex diseases in humans.
- Genetic imprinting may operate through interactions with other genetic effects, although this has not been fully explored.
- We describe a statistical model for detecting genetic imprinting and its interactions with additive, dominant and epistatic effects based on a simple 2×2 contingency table.
- By incorporating it into GWAS, this model will find its immediate implication for comprehending the genetic architecture of human diseases.

FUNDING

NSF/IOS-0923975 and UL1 TR000127 from the National Center for Advancing Translational Sciences (NCATS). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH.

References

1. Reik W, Walter J. Genomic imprinting: parental influence on the genome. *Nat Rev Genet* 2001;**2**:21–32.
2. Cattanach BM, Beechey CV, Peters J. Interactions between imprinting effects in the mouse. *Genetics* 2004;**168**:397–413.
3. Wilkinson LS, Davies W, Isles AR. Genomic imprinting effects on brain development and function. *Nat Rev Neurosci* 2007;**4**:1–19.
4. Barlow DP. Genomic imprinting: A mammalian epigenetic discovery model. *Annu Rev Genet* 2011;**45**:379–403.
5. Kärst S, Vahdati A, Brockmann G, *et al.* Genomic imprinting and genetic effects on muscle traits in mice. *BMC Genom* 2012;**13**:408.
6. Sha K. A mechanistic view of genomic imprinting. *Annu Res Genomics Hum Genet* 2008;**9**:197–216.
7. de Koning DJ, Rattink AP, Harlizius B, *et al.* Genome-wide scan for body composition in pigs reveals important role of imprinting. *Proc Natl Acad Sci USA* 2000;**97**:7947–50.
8. Liu T, Todhunter RJ, Wu S, *et al.* A random model for mapping imprinted quantitative trait loci in a structured pedigree: an implication for mapping canine hip dysplasia. *Genomics* 2007;**90**:276–84.
9. Cheverud JM, Hager R, Roseman C, *et al.* Genomic imprinting effects on adult body composition in mice. *Proc Natl Acad Sci USA* 2008;**105**:4253–8.
10. Wang CG, Wang Z, Luo JT, *et al.* A model for transgenerational imprinting variation in complex traits. *PLoS One* 2010;**5**:e11396.
11. Wang CG, Wang Z, Prows DR, *et al.* A computational framework for the inheritance of genomic imprinting for complex traits. *Brief Bioinform* 2012;**13**:34–45.
12. Xia F, Zhou JY, Fung WK. A powerful approach for association analysis incorporating imprinting effects. *Bioinformatics* 2011;**27**:2571–7.
13. Howey R, Cordell HJ. PREMIM and EMIM: tools for estimation of maternal, imprinting and interaction effects using multinomial modelling. *BMC Bioinform* 2012;**13**:149.
14. Manolio TA. Genomewide association studies and assessment of the risk of disease. *N Engl J Med* 2010;**363**:166–76.
15. Wang Z, Liu T, Lin Z, *et al.* A general model for multilocus epistatic interactions in case-control studies. *PLoS One* 2010;**5**:e11384.
16. Liu T, Thalamuthu A, Liu JJ, *et al.* Asymptotic distribution for epistatic tests in case-control studies. *Genomics* 2011;**98**:145–51.
17. Zhang L, Liu R, Wang Z, *et al.* Modeling haplotype-haplotype interactions in case-control genetic association studies. *Front Genet* 2012;**3**:2.
18. Sui YH, Wu WM, Wang Z, *et al.* A case-control design for testing and estimating epigenetic effects on complex diseases. *Brief Bioinform*. 2013 Jan 17. [Epub ahead of print].
19. Lin Z, Nelson L, Franke A, *et al.* OCTN1 variant L503F is associated with familial and sporadic inflammatory bowel disease. *J Crohns Colitis* 2010;**4**:132–8.
20. Li Q, Wu RL. A multilocus model for constructing a linkage disequilibrium map in human populations. *Stat Appl Genet Mol Biol* 2009;**8**:1–25.
21. Liu JY, Wang Z, Wang YQ, *et al.* Model and algorithm for linkage disequilibrium analysis in a non-equilibrium population. *Front Genet* 2012;**3**:78.
22. Cui YH, Lu Q, Cheverud JM, *et al.* Model for mapping imprinted quantitative trait loci in an inbred F₂ design. *Genomics* 2006;**87**:543–51.
23. Wu S, Yang J, Wu RL. Mapping quantitative trait loci in a non-equilibrium population. *Stat Appl Mol Genet Biol* 2010;**9**. Article 32.
24. Hill WG, Hernandez-Sanchez J. Prediction of multilocus identity by descent. *Genetics* 2007;**176**:2307–15.
25. Constanca M, Kelsey G, Reik W. Resourceful imprinting. *Nature* 2004;**432**:53–7.
26. Varrault A, Gueydan C, Delalbre A, *et al.* Zac1 regulates an imprinted gene network critically involved in the control of embryonic growth. *Dev Cell* 2006;**11**:711–22.
27. Sandhu KS. Systems properties of proteins encoded by imprinted genes. *Epigenetics* 2010;**5**:627–36.
28. Wolf JB. Evolution of genomic imprinting as a coordinator of coadapted gene expression. *Proc Natl Acad Sci USA* 2013;**110**:5085–90.