Original Paper

Human Heredity

Hum Hered 2011;71:196–208
DOI: 10.1159/000328006

Received: January 30, 2011
Accepted after revision: April 4, 2011
Published online: July 20, 2011

Methods for Detecting Interactions between Imprinted Genes and Environmental Exposures Using Birth Cohort Designs with Mother-Offspring Pairs

Shuang Wang[a] Zhaoxia Yu[d] Rachel L. Miller[b, c] Deliang Tang[b] Frederica P. Perera[b]

[a]Department of Biostatistics, and [b]Columbia Center for Children's Environmental Health, Mailman School of Public Health, Columbia University, and [c]Division of Pulmonary, Allergy and Critical Care Medicine, Columbia University College of Physicians and Surgeons, New York, N.Y., and [d]Department of Statistics, University of California at Irvine, Irvine, Calif., USA

Key Words
Mother-offspring pairs · Genetic imprinting · igene-environment interaction

Abstract
Genomic imprinting is a form of epigenetic regulation in mammals in which the same allele of a gene is expressed differently depending on the parental origin of the allele. Traditionally, the detection of imprinted genes that affect complex diseases has been focused on linkage designs with pedigrees or case-parent designs with case-parent trios. In the past two decades, the birth cohort design with mother-offspring pairs has been applied to understand better the effect of environmental influences during pregnancy and beginning of life on the growth and development of children. No work has been done on the detection of imprinted genes using birth cohort designs. Moreover, although the importance of imprinting has been well recognized, no study has looked at how environmental exposures modify the effects of imprinted genes. In this study, we show that the proposed imprinting test using the birth cohort design with mother-offspring pairs is an efficient test for testing the interactions between imprinted genes and environmental exposures. Through extensive simulation studies and a real data application, the proposed imprinting test has demonstrated much improved power in detecting gene-environment interactions than that of a test assuming the Mendelian dominant model when the true underlying genetic model is imprinting.

Copyright © 2011 S. Karger AG, Basel

Introduction

Genomic imprinting is a form of epigenetic regulation in mammals in which the same allele of a gene is expressed differently according to parental origin of the allele [1]. Imprinted genes are thought to influence the development of the placenta, fetal and infant growth through affecting the maternal allocation of resources and the transfer of nutrients to the fetus and the newborn from the mother [1, 2]. Aberrant imprinting disturbs development and causes various preclinical and disease syndromes. Thus, the study of imprinting effects provides new insights into epigenetic gene modifications during development. A census of imprinted genes in 2005 suggested that approximately 41 genes in 16 chromosomal regions are imprinted in humans [3, 4, see also http://www.genenames.org/data/hgnc_data.php?hgnc_id=5379]. At pres-

KARGER
Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com

© 2011 S. Karger AG, Basel

Accessible online at:
www.karger.com/hhe

Shuang Wang, PhD
Department of Biostatistics, Mailman School of Public Health
Columbia University, 722 West 168th Street, Room 630
New York, NY 10032 (USA)
Tel. +1 212 342 4165, E-Mail sw2206@columbia.edu

ent, approximately 60 imprinted genes are known in humans, but there may be up to 1,000 imprinted genes [5]. A large number of complex diseases are known to have parent-of-origin effects, including asthma, autism, type I and type II diabetes, Alzheimer disease and schizophrenia [5]. More recently, Kong et al. [6] identified 5 single nucleotide polymorphisms (SNPs) to have parental-origin-specific associations with breast cancer, basal-cell carcinoma and type II diabetes using 38,167 Icelanders.

Traditionally, the detection of imprinted genes in complex diseases has been focused on linkage designs with pedigrees or case-parent designs with case-parent trios when parent-of-origin information is available in offspring [7–15]. To improve the health and well-being of children and to understand the effect of environmental influences during pregnancy and beginning of life on the growth and development of children, a number of longitudinal birth cohort designs with mother-offspring pairs have been applied in the past two decades. In most birth cohort designs, pregnant women are recruited into the study and environmental exposures during pregnancy are collected. Newborns are followed up after birth until school age for health and development information. Children's environmental exposures at the beginning of life also are collected. Usually only mothers and offspring are collected in birth cohort designs without the father, providing partial parent-of-origin information. No work has been done on the detection of main genetic imprinted effects using birth cohort designs with mother-offspring pairs. Most recently, Ainsworth et al. [16] investigated imprinting effects using affected offspring/control offspring and their mothers.

Although there is still controversy about the nature of the primary imprints, it was found that DNA methylation was a key molecular mechanism of imprinting. DNA methylation marks the imprinted genes differently in egg and sperm, and inheritance of these epigenetic marks may lead to differential gene expression [17–19]. Although the importance of genetic imprinting has been well recognized, no study to date has looked at how environmental exposures modify the effects of imprinted genes. This is especially important as the environment is known to play a key role in risks of almost all complex diseases through its profound role in determining the epigenetic alterations occurring in an individual's genome. Moreover, environmentally induced modifications to the epigenome that occur during the epigenetic reprogramming in early development tend to affect fetal growth and behavior after birth and are likely to have long-term health consequences [2, 20, 21].

In this study, we show that the birth cohort design with mother-offspring pairs is an efficient design for testing the interactions between imprinted genes and environmental exposures. We developed a new method that uses a newly proposed maternal and paternal imprinting coding to detect how prenatal environmental exposures or environmental exposures at beginning of life modify the effect of imprinted genes on child illness, such as asthma, using birth cohort designs with mother-offspring pairs. We denote such interactions between imprinted genes and environmental exposures as igene-environment interactions. We proposed to choose the optimal genetic models by cross-validation procedures that minimize the prediction errors. Through extensive simulation studies and an application on childhood asthma using a birth cohort study being conducted by the Columbia Center for Children's Environmental Health (CCCEH) in New York City, we demonstrated the feasibility and power of the proposed imprinting tests using mother-offspring pairs. Although we concentrated on the interactions between imprinted genes and environmental exposures, the proposed method is readily applied to detect main effects of imprinted genes or interactions between imprinted genes with mother-offspring pairs.

Material and Methods

We assume a bi-allelic marker with a high-risk allele A and a low-risk allele a, that have frequencies p and $1 - p$, respectively. Here the high-risk allele is the minor allele in the cohort. Assuming Hardy-Weinberg equilibrium (HWE) and random mating, the joint distribution of the 7 possible mother-offspring genotype combinations are displayed in table 1. Let A_m and a_m be the two alleles inherited from the mother, and A_f and a_f be the two alleles inherited from the father. There are 4 possible imprinted genotypes in the offspring, A_mA_f, A_ma_f, A_fa_m, and a_ma_f. Without paternal genotype information, the parent-of-origin information of each offspring genotype can still be fully determined except for cases when both mother and offspring genotypes are heterozygous (table 2). In this case, imprinted genotypes in the offspring can be either A_ma_f or A_fa_m. Given that both mother and offspring genotypes are heterozygous, we can derive the probability of the father having one type of genotype as follows:

$$\Pr(G_p = AA \mid G_m = Aa, G_c = Aa) = p^2,$$
$$\Pr(G_p = Aa \mid G_m = Aa, G_c = Aa) = 2p(1 - p),$$
$$\Pr(G_p = aa \mid G_m = Aa, G_c = Aa) = (1 - p)^2.$$

These happen to be the genotype frequencies in the general population following HWE. Note that in designs with case-parent trios, trios with both parents and the diseased child heterozygous are uninformative in terms of parent-of-origin information and are usually omitted from the analysis [7]. As a comparison, table 2

Table 1. Joint probabilities of mother-offspring pair genotypes

	Offspring genotype		
	AA	Aa	aa
Maternal genotype			
AA	p^3	$p^2(1-p)$	0
Aa	$p^2(1-p)$	$p(1-p)$	$p(1-p)^2$
Aa	0	$p(1-p)^2$	$(1-p)^3$

p is the minor allele frequency.

Table 2. Offspring parent-of-origin information

a For mother-offspring pairs

	Offspring genotype		
	AA	Aa	aa
Maternal genotype			
AA	A_mA_f	A_ma_f	
Aa	A_mA_f	A_ma_f / A_fa_m	a_ma_f
aa		A_fa_m	a_ma_f

b For parent-offspring trios

	Offspring genotype		
	AA	Aa	aa
Maternal genotype × paternal genotype			
AA × AA	A_mA_f		
AA × Aa	A_mA_f	A_ma_f	
AA × aa		A_ma_f	
Aa × AA	A_mA_f	A_fa_m	
Aa × Aa	A_mA_f	A_ma_f / A_fa_m	a_ma_f
Aa × aa		A_ma_f	a_ma_f
aa × AA		A_fa_m	
aa × Aa		A_fa_m	a_ma_f
aa × aa			a_ma_f

Blanks are not applicable.

Table 3. Maternal imprinting coding and paternal imprinting coding for offspring genotypes

Imprinting coding	Offspring imprinted genotype			
	A_mA_f	A_ma_f	A_fa_m	a_ma_f
Maternal	1	0	1	0
Paternal	1	1	0	0

also displays offspring parent-of-origin information when both maternal and paternal genotypes are available. With mother-offspring pairs, the offspring parent-of-origin information is ambiguous when both mother and offspring genotypes are heterozygous. We deal with such mother-offspring pairs with ambiguous parent-of-offspring information in two ways and will compare their performance. In the first way to deal with ambiguous mother-offspring pairs, we partially recover parent-of-origin information by assigning each pair pseudo-paternal genotype AA with a probability p^2, and pseudo-paternal genotype aa with a probability $(1-p)^2$, and by omitting such mother-offspring pairs with probability $2p(1-p)$. With pseudo-paternal genotypes AA and aa assigned for one pair, the parent-of-origin information of heterozygous offspring with a heterozygous mother can be obtained from table 2b. In the second way to deal with ambiguous mother-offspring pairs, we simply omit such pairs. We note the possibilities of misclassifying imprinting information with partially recovered parent-of-origin information and will compare the performance of the two methods in terms of type I error rates and power.

We propose a maternal imprinting coding and a paternal imprinting coding for offspring genotype to capture the maternal or paternal imprinting effect (table 3). Maternal imprinting refers to the situation when the allele of a particular gene inherited from the mother is transcriptionally silent while the paternally inherited allele is active. Paternal imprinting is the opposite – when the paternally inherited allele is silenced while the maternally inherited allele is active. Thus, when a genetic locus is a maternally imprinted locus, we code offspring imprinted genotypes 'A_mA_f' and 'A_fa_m' as 1, and offspring imprinted genotypes 'A_ma_f' and 'a_ma_f' as 0. When a genetic locus is a paternally imprinted locus, we code offspring imprinted genotypes 'A_mA_f' and 'A_ma_f' as 1, and offspring imprinted genotypes 'A_fa_m' and 'a_ma_f' as 0. With this maternal/paternal imprinting coding, we can apply the following logistic model on offspring disease status to detect interactions between imprinted genes and environment, i.e. igene-environment interactions:

$$\text{logit}(D_{o_i} = 1 \mid G_{o_i}, E_i) = \beta_0 + \beta_E \mathrm{E}_i + \beta_{G_o^i} G_{o_i}^i + \beta_{G_o^i \times E} \mathrm{E}_i \times G_{o_i}^i,\ i = 1, \ldots, n,$$

where E is the binary prenatal or postnatal environmental exposure status of the offspring; $G_{o_i}^i$ is offspring imprinting coding at the genetic locus of interest; and D_o is the offspring disease status. The regression coefficient $\beta_{G_o^i \times E}$ measures how environment modifies the genetic imprinting effect. By testing the regression coefficient $\beta_{G_o^i \times E}$ equal to zero or not using a likelihood ratio test with 1 degree of freedom, we could test the corresponding interactions. We will compare tests assuming the maternal or paternal imprinting genetic models to the test assuming the Mendelian dominant genetic model (coding $G_o = 1$ for offspring genotypes AA and Aa, $G_o = 0$ for offspring genotypes aa). We propose to use the cross-validation procedure that minimizes the prediction errors to choose the best genetic model among the maternal imprinting model, the paternal imprinting model, and the dominant model. More specifically, for the purpose of training and testing, we will perform 10-fold cross-validation, splitting the mother-offspring pair samples into 10 sets of approximately equal size. We then calculate the prediction error of the disease status for each offspring. Note that the model proposed can be readily adapted

for detection of main imprinting effects or interactions between imprinted genes using mother-offspring pairs.

In simulation settings with different parameter sets, the intercept β_0 is not a free parameter, but is determined by the population prevalence of the studied outcome in the offspring cohort, pD_o. This approach is similar to the procedure used in our previous work on gene-gene and gene-environment interactions [22–24]. With a prefixed population prevalence, we can obtain β_0 by solving the following equation:

$$\text{Population prevalence of offspring in a certain age range} = pD_o$$
$$= \Pr\left(D_o = 1\right) = \sum_{G_p, G_m, G_o, E} \Pr\left(D_o = 1 | G_p, G_m, G_o, E\right) \Pr\left(G_p, G_m, G_o, E\right)$$
$$= \sum_{G_p, G_m, G_o, E} \left(\Pr\left(D_o = 1 | G_o, G_p, G_m, E\right) \Pr\left(G_o | G_p, G_m\right) \Pr\left(G_p, G_m\right) \Pr(E)\right),$$

where G_p, G_m, and G_o are the genotypes of father, mother and offspring; D_o is the offspring disease status; and E is the environmental exposure status of the offspring. The same logistic model is used to model offspring penetrance $\Pr(D_o = 1 \mid G_o, G_p, G_m, E)$ with imprinting effects.

Simulation Studies: Type I Error and Power

Here, using simulations with a wide range of parameter settings, we compared the performance of the proposed imprinting tests to that of the test assuming the Mendelian dominant genetic model. To simplify the presentation, we denote the proposed maternal imprinting test (i.e. using the maternal imprinting coding) with partially recovered parent-of-origin information for ambiguous pairs as mPair, and the corresponding paternal imprinting test as pPair. For the imprinting tests omitting ambiguous mother-offspring pairs, we denote the maternal imprinting test as mPair.sub and the paternal imprinting test as pPair.sub. For the test assuming the Mendelian dominant model, we denote it as Dom. As a comparison, in all simulation settings, we compare the performance of the proposed imprinting tests to detect igene-environment interactions with mother-offspring pairs to that of the test with parent-offspring trios (we denote the maternal imprinting test as mTrio and the paternal imprinting test as pTrio).

Simulation Parameters

To evaluate type I error rates and power, the total sample size was fixed at N = 1,000 mother-offspring pairs. The frequency of the environmental exposure was set at 0.2. Population prevalence was set at two different levels, pD_o = 10 and 20%. We also considered different minor allele frequencies (MAF), p = 0.2, 0.3, and 0.4. The main offspring imprinting effect and the main environmental exposure effect were fixed at $OR_{G_o^i}$ = 1.5 and OR_E = 2.0, respectively. Different levels of the offspring igene-environment interaction effects were considered, $OR_{G_o^i \times E}$ = 1.0, 1.5, 2.0, 3.0, 4.0.

Simulation Setup

Each simulated study consisted of N mother-offspring pairs. We first simulated N maternal genotypes and paternal genotypes based on the population allele frequencies and the assumptions of HWE and random mating. We then generated offspring genotypes based on Mendelian transmission and simulated parental genotypes. When conducting tests involving mother-offspring pairs, paternal genotypes were discarded. Environmental exposures of offspring were generated based on a binomial distribution with pre-specified proportions of exposure. Under our selected parameter settings, for a mother-offspring pair, two sets of offspring outcome status were generated, either based on the offspring penetrance when the true underlying genetic model is maternal imprinting $\Pr(D_o = 1 \mid G_o^{i_m}, E)$ or based on the offspring penetrance when the true underlying genetic model is paternal imprinting $\Pr(D_o = 1 \mid G_o^{i_p}, E)$. The tests of interest (the proposed imprinting tests and the test assuming the Mendelian dominant genetic model) were performed using the simulated data and the procedures were repeated 1,000 times to evaluate type I error rates and power.

To evaluate the type I error rates of the proposed imprinting tests, data were generated under the null hypothesis of no interaction between imprinted genes and environmental exposure, H_0: $\beta_{G_o^i \times E} = 0$, i.e. H_0: $OR_{G_o^i \times E} = 1$. Type I error rates of the tests of interest were estimated by the proportion of times that the null hypothesis of no interaction between genetic polymorphisms and the environmental exposure was rejected.

Results

Type I Error

Table 4 displays the type I error rates to detect gene-environment interactions with the proposed imprinting tests and the test assuming the Mendelian dominant genetic model when the true underlying disease model is either the maternal imprinting genetic model or the paternal imprinting genetic model. We considered different MAF and population prevalence. In all scenarios, the nominal type I error rate 0.05 was well controlled for all tests examined.

Power

Table 5 displays power to detect gene-environment interactions with different simulation parameter settings when the true underlying disease model is either maternal imprinting or paternal imprinting. The power results were also plotted in figure 1 for better visualization. Table 6 shows the corresponding cross-validation results.

It is clear that when the true underlying genetic model is either maternal imprinting or paternal imprinting, the proposed imprinting tests using mother-offspring pairs that assume the correct imprinting effect consistently have increased power compared to the test assuming the Mendelian dominant genetic model on all scenarios considered (fig. 1). The power improvement is bigger when the true underlying genetic model is maternal imprinting rather than paternal imprinting and the power patterns differ depending on the true disease imprinting model and parameter settings.

When the true underlying disease model is maternal imprinting and the imprinting tests agree with the true

Table 4. Type I error rates to detect gene-environment interactions at the 0.05 significance level with the proposed imprinting tests and the test assuming the Mendelian dominant genetic model when the underlying true genetic model is either maternal imprinting or paternal imprinting

Pop. prev.	MAF	True disease model	mTrio	pTrio	mPair	pPair	mPair.sub	pPair.sub	Dom
0.1	0.2	I_m	0.047	0.038	0.046	0.032	0.051	0.023	0.047
		I_p	0.028	0.047	0.029	0.038	0.021	0.025	0.035
	0.3	I_m	0.052	0.048	0.047	0.049	0.045	0.028	0.058
		I_p	0.037	0.042	0.046	0.033	0.047	0.037	0.044
	0.4	I_m	0.039	0.047	0.040	0.041	0.035	0.046	0.041
		I_p	0.051	0.051	0.050	0.057	0.050	0.059	0.047
0.2	0.2	I_m	0.049	0.049	0.049	0.041	0.050	0.025	0.056
		I_p	0.036	0.045	0.033	0.039	0.033	0.038	0.047
	0.3	I_m	0.044	0.045	0.051	0.062	0.050	0.045	0.058
		I_p	0.053	0.051	0.051	0.052	0.054	0.055	0.048
	0.4	I_m	0.044	0.037	0.045	0.038	0.043	0.040	0.041
		I_p	0.046	0.044	0.049	0.054	0.046	0.046	0.046

Pop. prev. (population prevalence) was set at $pD_o = 0.1$ and 0.2; MAF was set at 0.2, 0.3, and 0.4; and environment exposure was set at 20%. The total sample size was fixed at $N = 1{,}000$ mother-offspring pairs. The simulation procedure was repeated 1,000 times.

I_m = Scenarios when the underlying true disease model is maternal imprinting; Ip = scenarios when the underlying true disease model is paternal imprinting.

model, tests using mother-offspring pairs have similar power to tests using parent-offspring trios when MAF is small. More specifically, compared to the mTrio test, the power loss of the mPair and mPair.sub tests is only about 1–2%. For example, when population prevalence is 20%, environmental exposure is 20%, MAF is 0.2, and the true underlying disease model is maternal imprinting, to detect an interaction of $OR_{G_o^i \times E} = 3$, mTrio test gives 62% power, mPair test gives 59% power, mPair.sub test gives 60% power, while Dom test gives only 39% power. When MAF is 0.4, the power loss increases to about 10–20% of the power of the mTrio test. For example, when population prevalence is 10%, environmental exposure is 20%, and MAF is 0.4, to detect an interaction of $OR_{G_o^i \times E} = 3$, mTrio test gives 41% power, mPair test gives 34% power, mPair.sub test gives 36% power, while Dom test gives only 19% power. This is expected as the percentage of heterozygous mother-offspring pairs increases as MAF increases. The power difference between the mPair test with partially recovered parent-of-origin information for heterozygous mother-offspring pairs and the mPair.sub test omitting such pairs can almost be neglected in all scenarios, with the mPair.sub test actually having slightly higher power than the mPair test when MAF is large. This is also expected because when MAF is large, the percentage of heterozygous paternal genotype is also large among those heterozygous mother-offspring pairs. Thus not much parent-of-origin information can be recovered and at the same time, some misclassification is possible (table 5 and fig. 1).

Cross-validation results suggest that among the 1,000 simulations, the majority of time the correct genetic model was chosen (table 6). Table 6 displays the number of times a model was chosen because of having the smallest mean prediction errors among the 5 models out of the 1,000 simulations. As expected, the correctly chosen percentage increases as the effect size of the igene-environment interaction increases. However, caution needs to be taken with applying the proposed method to detect interactions between the paternal imprinting effect and environment when the igene-environment interaction effect is small and MAF is low. For example, when population prevalence is 10%, MAF is 0.2, $OR_{G_o^i \times E} = 1.5$, and the underlying genetic model is paternal imprinting, 47.6% of the simulations chose the wrong genetic model. When the effect of the igene-environment interaction is modest to large, more than 90% of the simulations chose the correct genetic model to detect interactions between the maternal imprinting effect and environment. We also notice that in all simulation scenarios considered, the cross-val-

Table 5. Power to detect gene-environment interactions

Pop. prev.	MAF	True disease model	igene-environment interaction	mTrio	pTrio	mPair	pPair	mPair.sub	pPair.sub	Dom
0.1	0.2	I_m	$OR_{G_o^i m \times E} = 4.0$	0.603	0.036	0.587	0.048	0.589	0.049	0.405
			$OR_{G_o^i m \times E} = 3.0$	0.440	0.041	0.424	0.047	0.427	0.050	0.287
			$OR_{G_o^i m \times E} = 2.0$	0.218	0.044	0.207	0.048	0.208	0.040	0.141
			$OR_{G_o^i m \times E} = 1.5$	0.108	0.045	0.103	0.041	0.104	0.026	0.084
		I_p	$OR_{G_o^i m \times E} = 4.0$	0.038	0.596	0.046	0.513	0.076	0.316	0.385
			$OR_{G_o^i m \times E} = 3.0$	0.041	0.428	0.040	0.357	0.063	0.210	0.262
			$OR_{G_o^i m \times E} = 2.0$	0.046	0.210	0.037	0.177	0.046	0.109	0.124
			$OR_{G_o^i m \times E} = 1.5$	0.035	0.118	0.034	0.097	0.036	0.060	0.075
	0.3	I_m	$OR_{G_o^i m \times E} = 4.0$	0.610	0.042	0.551	0.062	0.556	0.070	0.384
			$OR_{G_o^i m \times E} = 3.0$	0.450	0.041	0.411	0.057	0.416	0.060	0.276
			$OR_{G_o^i m \times E} = 2.0$	0.237	0.045	0.203	0.054	0.222	0.049	0.135
			$OR_{G_o^i m \times E} = 1.5$	0.104	0.045	0.093	0.042	0.097	0.036	0.080
		I_p	$OR_{G_o^i m \times E} = 4.0$	0.045	0.618	0.079	0.525	0.102	0.479	0.386
			$OR_{G_o^i m \times E} = 3.0$	0.049	0.456	0.067	0.389	0.079	0.366	0.257
			$OR_{G_o^i m \times E} = 2.0$	0.039	0.252	0.049	0.202	0.055	0.165	0.119
			$OR_{G_o^i m \times E} = 1.5$	0.041	0.125	0.043	0.101	0.045	0.095	0.069
	0.4	I_m	$OR_{G_o^i m \times E} = 4.0$	0.562	0.061	0.468	0.078	0.493	0.087	0.276
			$OR_{G_o^i m \times E} = 3.0$	0.412	0.060	0.339	0.074	0.358	0.081	0.194
			$OR_{G_o^i m \times E} = 2.0$	0.209	0.053	0.182	0.046	0.183	0.057	0.093
			$OR_{G_o^i m \times E} = 1.5$	0.094	0.049	0.098	0.050	0.095	0.058	0.059
		I_p	$OR_{G_o^i m \times E} = 4.0$	0.053	0.564	0.089	0.484	0.090	0.503	0.274
			$OR_{G_o^i m \times E} = 3.0$	0.052	0.433	0.075	0.352	0.071	0.371	0.199
			$OR_{G_o^i m \times E} = 2.0$	0.051	0.229	0.066	0.181	0.057	0.192	0.102
			$OR_{G_o^i m \times E} = 1.5$	0.061	0.101	0.060	0.103	0.054	0.103	0.064
0.2	0.2	I_m	$OR_{G_o^i m \times E} = 4.0$	0.818	0.053	0.793	0.062	0.797	0.115	0.540
			$OR_{G_o^i m \times E} = 3.0$	0.622	0.052	0.593	0.057	0.604	0.084	0.389
			$OR_{G_o^i m \times E} = 2.0$	0.323	0.054	0.295	0.041	0.303	0.057	0.192
			$OR_{G_o^i m \times E} = 1.5$	0.154	0.052	0.136	0.042	0.143	0.043	0.089
		I_p	$OR_{G_o^i m \times E} = 4.0$	0.056	0.800	0.064	0.714	0.095	0.530	0.525
			$OR_{G_o^i m \times E} = 3.0$	0.047	0.632	0.060	0.545	0.074	0.378	0.370
			$OR_{G_o^i m \times E} = 2.0$	0.044	0.322	0.049	0.266	0.056	0.195	0.173
			$OR_{G_o^i m \times E} = 1.5$	0.039	0.137	0.038	0.123	0.046	0.088	0.086
	0.3	I_m	$OR_{G_o^i m \times E} = 4.0$	0.849	0.058	0.811	0.099	0.823	0.134	0.579
			$OR_{G_o^i m \times E} = 3.0$	0.684	0.054	0.619	0.079	0.648	0.105	0.400
			$OR_{G_o^i m \times E} = 2.0$	0.356	0.055	0.317	0.061	0.324	0.073	0.192
			$OR_{G_o^i m \times E} = 1.5$	0.151	0.057	0.139	0.068	0.150	0.055	0.098
		I_p	$OR_{G_o^i m \times E} = 4.0$	0.065	0.852	0.102	0.764	0.144	0.728	0.595
			$OR_{G_o^i m \times E} = 3.0$	0.060	0.686	0.088	0.600	0.109	0.544	0.416
			$OR_{G_o^i m \times E} = 2.0$	0.055	0.382	0.056	0.304	0.072	0.293	0.205
			$OR_{G_o^i m \times E} = 1.5$	0.053	0.168	0.063	0.131	0.061	0.123	0.095
	0.4	I_m	$OR_{G_o^i m \times E} = 4.0$	0.859	0.064	0.774	0.123	0.808	0.140	0.536
			$OR_{G_o^i m \times E} = 3.0$	0.682	0.049	0.595	0.103	0.639	0.114	0.383
			$OR_{G_o^i m \times E} = 2.0$	0.341	0.050	0.296	0.060	0.315	0.072	0.185
			$OR_{G_o^i m \times E} = 1.5$	0.140	0.049	0.104	0.050	0.119	0.052	0.083
		I_p	$OR_{G_o^i m \times E} = 4.0$	0.061	0.862	0.137	0.791	0.138	0.796	0.543
			$OR_{G_o^i m \times E} = 3.0$	0.055	0.700	0.115	0.607	0.107	0.620	0.378
			$OR_{G_o^i m \times E} = 2.0$	0.052	0.367	0.078	0.304	0.079	0.311	0.185
			$OR_{G_o^i m \times E} = 1.5$	0.055	0.158	0.062	0.143	0.057	0.153	0.093

Pop. prev. (population prevalence) was set at $pD_o = 0.1$ and 0.2; MAF was set at 0.2, 0.3, and 0.4; and environment exposure was set at 20%. The total sample size was fixed at $N = 1{,}000$ mother-offspring pairs. The simulation procedure was repeated 1,000 times.

I_m = Scenarios when the underlying true disease model is maternal imprinting; Ip = scenarios when the underlying true disease model is paternal imprinting.

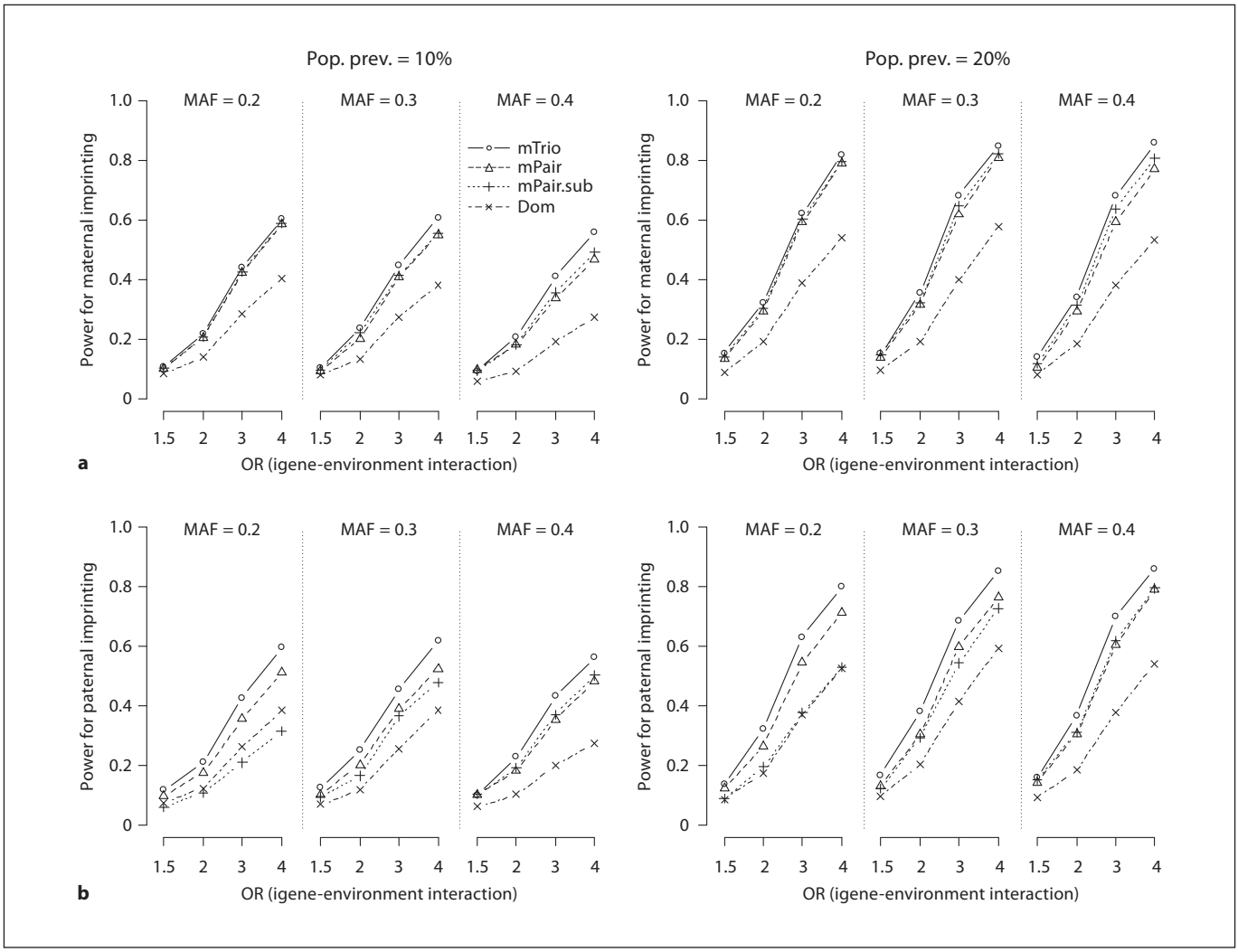

Fig. 1. Power to detect gene-environment interactions with the proposed imprinting tests and the test assuming the Mendelian dominant genetic model when the underlying true disease model is either maternal imprinting (**a**) or paternal imprinting (**b**). MAF was set at 0.2, 0.3, and 0.4; population prevalence (pop. prev.) was set at pD_o = 10 and 20%; environmental exposure was set at 20%; and the effects of gene-environment interactions were ranging from odds ratio (OR) 1.5 to 4.0.

idation procedures did not choose the dominant genetic model a single time; neither did they choose the imprinting methods with partially recovered parent-of-origin information. The reason that the mPair and pPair tests were not chosen might be due to the possible misclassification in recovered parent-of-origin information, which may lead to bigger prediction errors. Moreover, because the mean prediction errors of all tests are very small and close to each other, small increases in the individual prediction errors of those heterozygous offspring with heterozygous mothers may result in a much more prominent change in rankings of the 5 mean prediction errors.

When the true underlying disease model is paternal imprinting and the imprinting tests agree with the true model, the proposed pPair and pPair.sub tests with mother-offspring pairs lose about 20% of the power of the pTrio test with parent-offspring trios in all scenarios considered. As no paternal genotype information is available for mother-offspring pairs, it is expected that pPair or pPair.sub tests under the parental imprinting model would have lower power than that of mPair or mPair.sub tests under the maternal imprinting model. Comparisons between the power of the pPair test and the pPair.sub test suggests that partially recovering parent-of-origin infor-

Table 6. Cross-validation results to detect gene-environment interactions

Pop. prev.	MAF	True disease model	igene-environment interaction	mPair	pPair	mPair.sub	pPair.sub	Dom
0.1	0.2	I_m	$OR_{G_0^i m \times E} = 4.0$	0	0	950	50	0
			$OR_{G_0^i m \times E} = 3.0$	0	0	923	77	0
			$OR_{G_0^i m \times E} = 2.0$	0	0	820	180	0
			$OR_{G_0^i m \times E} = 1.5$	0	0	717	283	0
		I_p	$OR_{G_0^i m \times E} = 4.0$	0	0	210	790	0
			$OR_{G_0^i m \times E} = 3.0$	0	0	305	695	0
			$OR_{G_0^i m \times E} = 2.0$	0	0	428	572	0
			$OR_{G_0^i m \times E} = 1.5$	0	0	476	524	0
	0.3	I_m	$OR_{G_0^i m \times E} = 4.0$	0	0	979	21	0
			$OR_{G_0^i m \times E} = 3.0$	0	0	945	55	0
			$OR_{G_0^i m \times E} = 2.0$	0	0	829	171	0
			$OR_{G_0^i m \times E} = 1.5$	0	0	715	285	0
		I_p	$OR_{G_0^i m \times E} = 4.0$	0	0	80	920	0
			$OR_{G_0^i m \times E} = 3.0$	0	0	137	863	0
			$OR_{G_0^i m \times E} = 2.0$	0	0	257	743	0
			$OR_{G_0^i m \times E} = 1.5$	0	0	358	642	0
	0.4	I_m	$OR_{G_0^i m \times E} = 4.0$	0	0	966	34	0
			$OR_{G_0^i m \times E} = 3.0$	0	0	933	67	0
			$OR_{G_0^i m \times E} = 2.0$	0	0	844	156	0
			$OR_{G_0^i m \times E} = 1.5$	0	0	763	237	0
		I_p	$OR_{G_0^i m \times E} = 4.0$	0	0	38	962	0
			$OR_{G_0^i m \times E} = 3.0$	0	0	74	926	0
			$OR_{G_0^i m \times E} = 2.0$	0	0	166	834	0
			$OR_{G_0^i m \times E} = 1.5$	0	0	250	750	0
0.2	0.2	I_m	$OR_{G_0^i m \times E} = 4.0$	0	0	986	14	0
			$OR_{G_0^i m \times E} = 3.0$	0	0	976	24	0
			$OR_{G_0^i m \times E} = 2.0$	0	0	909	91	0
			$OR_{G_0^i m \times E} = 1.5$	0	0	811	189	0
		I_p	$OR_{G_0^i m \times E} = 4.0$	0	0	72	928	0
			$OR_{G_0^i m \times E} = 3.0$	0	0	126	874	0
			$OR_{G_0^i m \times E} = 2.0$	0	0	234	766	0
			$OR_{G_0^i m \times E} = 1.5$	0	0	318	682	0
	0.3	I_m	$OR_{G_0^i m \times E} = 4.0$	0	0	996	4	0
			$OR_{G_0^i m \times E} = 3.0$	0	0	987	13	0
			$OR_{G_0^i m \times E} = 2.0$	0	0	933	67	0
			$OR_{G_0^i m \times E} = 1.5$	0	0	859	141	0
		I_p	$OR_{G_0^i m \times E} = 4.0$	0	0	15	985	0
			$OR_{G_0^i m \times E} = 3.0$	0	0	54	946	0
			$OR_{G_0^i m \times E} = 2.0$	0	0	139	861	0
			$OR_{G_0^i m \times E} = 1.5$	0	0	238	762	0
	0.4	I_m	$OR_{G_0^i m \times E} = 4.0$	0	0	994	6	0
			$OR_{G_0^i m \times E} = 3.0$	0	0	983	17	0
			$OR_{G_0^i m \times E} = 2.0$	0	0	937	83	0
			$OR_{G_0^i m \times E} = 1.5$	0	0	879	121	0
		I_p	$OR_{G_0^i m \times E} = 4.0$	0	0	4	996	0
			$OR_{G_0^i m \times E} = 3.0$	0	0	12	987	0
			$OR_{G_0^i m \times E} = 2.0$	0	0	63	937	0
			$OR_{G_0^i m \times E} = 1.5$	0	0	130	870	0

Pop. prev. (population prevalence) was set at $pD_o = 0.1$ and 0.2; MAF was set at 0.2, 0.3, and 0.4; and environment exposure was set at 20%. The total sample size was fixed at $N = 1{,}000$ mother-offspring pairs. The simulation procedure was repeated 1,000 times.

I_m = Scenarios when the underlying true disease model is maternal imprinting; Ip = scenarios when the underlying true disease model is paternal imprinting.

mation can improve power when MAF is small (table 5 and fig. 1). We observed similar patterns of the results from the cross-validation procedures, although the percentage of choosing the correct genetic model is not as high as in cases when the underlying true genetic model is maternal imprinting (table 6).

When the imprinting tests do not agree with the true underlying imprinting model, i.e. when the true underlying disease model is maternal imprinting but the imprinting tests assume paternal imprinting or vise versa, no imprinting tests (m/pTrio, m/pPair, and m/pPair.sub) have any power. However, the test assuming the Mendelian dominant genetic model has some power (table 5).

We also conducted simulation studies with other frequencies of the environmental exposure such as 30 and 40%. Similar patterns in power results and cross-validation results are observed (data not shown).

Real Data Application

The proposed method was applied to detect gene-environment interactions in childhood asthma using the mother-offspring pair data from the birth cohort study currently being conducted by CCCEH in New York City. Eligible pregnant women entered the study at the beginning stage of their pregnancy.

During the 2nd or 3rd trimesters, the women carried a backpack containing a portable personal exposure air monitor during the day and kept it near the bed at night during a consecutive 48-hour period for polycyclic aromatic hydrocarbons (PAH) measurements, a widespread class of combustion-related pollutants commonly found in the air, food, and drinking water [25]. PAH bind covalently to DNA to form PAH-DNA adducts, a widely used indicator of DNA damage that has been associated with cancer [26–29]. Higher levels of cord PAH-DNA adducts have been associated with reduced scores on neurocognitive tests, alone or in combination of environmental tobacco smoke [30–32]. PAH-DNA adduct levels were measured in umbilical cord blood collected during delivery [33] using the method of Alexandrov et al. [34], which has a coefficient of variation of 12% and a lower limit of detection of 0.25 adducts per 10^8 nucleotides. Here we treated PAH-DNA adducts as a biomarker of exposure of PAH. Adducts were dichotomized into detectable/non-detectable, with the non-detectable rate being 30% in this cohort.

Childhood asthma was measured using a parental report of doctor-diagnosed asthma according to the validated Brief Respiratory Questionnaire [35] during a follow-up visit when the children were between 5 to 6 years old. As this is an ongoing study and not all children had turned 5 or 6, we use this data set to illustrate the feasibility and power of the proposed imprinting tests. In this subset, about 30% of the children have doctor-diagnosed asthma status at ages 5 or 6.

We selected 22 common genetic polymorphisms from candidate genes that play important roles in the metabolic activation and detoxification of PAH, and from candidate genes that are known to be related to the allergic immune system (table 7). None of the 22 SNPs tested are on the regions known to be imprinted for asthma. No multiple comparison adjustment was conducted as the purpose of the real data application is to demonstrate the feasibility and power of the proposed imprinting tests. Findings from this real data application need to be replicated with an independent study. The data set consists of 171 mother-offspring pairs that have a complete report of childhood asthma status and cord adduct levels. The actual sample size used for each test at each marker differs, as we observed different numbers of mother-offspring pairs with both heterozygous genotypes at different markers.

Table 8 presents the p values for gene-environment interactions of the 22 markers using both the proposed imprinting tests and the test assuming the Mendelian dominant model. p values <0.1 are highlighted (in bold). Also displayed are results from the 5-fold cross-validation procedures at those significant markers, where the test with the smallest mean prediction errors among the 5 is indicated. Four markers, *CYP1A1-78*, *CYP1B1-05*, *CYP1B1-06*, and *IL13-01*, significantly interact with the environmental exposure at the 0.1 significance level with the proposed imprinting tests. No significance was observed at these four markers with the test assuming the Mendelian dominant genetic model. Two markers, *CYP1A1-14* and *CYP1A1-15*, were observed to interact with the environmental exposure at the 0.1 significance level with both imprinting tests and the test assuming the dominant genetic model. However, marker *IL4R-06* has a p value of 0.09, assuming the dominant model, but the p values are all above 0.1 with tests assuming imprinting genetic models.

In terms of the cross-validation results, in the majority of cases, tests or category of tests (i.e. mPair and mPair.sub tests as one category, or pPair and pPair.sub tests as one category) that have the smallest mean prediction errors from the cross-validation procedures also have the smallest p values. Another observation from the cross-validation results is that tests omitting mother-offspring pairs both with heterozygous genotypes tend to have smaller prediction errors than tests using partially recovered parent-of-origin information, consistent with the simulation results. Again, this is probably due to the possibility of

Table 7. Chromosomal positions and gene locations of the 22 markers from the selected candidate genes

Gene	SNP	SNP rs number	Alleles	Chromosome	Position, bp
CYP1A1	CYP1A1-78	rs2198843	C/G	15	72,788,283
	CYP1A1-109	rs1456432	A/G	15	72,790,104
	CYP1A1-15	rs4646421	T/C	15	72,803,245
	CYP1A1-14	rs2606345	T/G	15	72,804,229
	CYP1A1-81	rs2472299	C/T	15	72,820,453
CYP1A2	CYP1A203	rs762551	A/C	15	72,828,970
	CYP1A212	rs2472304	A/G	15	72,831,291
	CYP1A252	rs4886406	A/C	15	72,844,256
CYP1B1	CYP1B1-66	rs162549	T/A	2	38,148,960
	CYP1B1-06	rs1056837	T/C	2	38,151,654
	CYP1B1-05	rs1056836	G/C	2	38,151,707
	CYP1B1-74	rs162560	A/G	2	38,153,019
	CYP1B1-04	rs10012	C/G	2	38,155,894
	CYP1B1-03	rs2617266	C/T	2	38,156,048
IL13	IL1301	rs20541	T/C	5	132,023,863
	IL1303	rs1800925	T/C	5	132,020,708
IL4	IL403	rs2070874	T/C	5	132,037,609
	IL408	rs2243251	G/A	5	132,037,686
	IL410	rs2243290	A/C	5	132,046,068
IL4R	IL4R01	rs1805010	A/G	16	27,263,704
	IL4R05	rs1805015	C/T	16	27,281,681
	IL4R06	rs1801275	A/G	16	27,281,901

misclassifying imprinting information, although both tests preserve the correct type I error rates. For example, at markers *CYP1B1-05* and *CYP1B1-06*, mPair.sub test has the smallest prediction errors among all 5 tests, mPair.sub test has the smallest p values, 0.071 and 0.091, respectively. Another example, at marker *CYP1A1-78*, pPair.sub test has the smallest prediction error among all 5 tests, and pPair test has the smallest p value of 0.098, that is slightly more significant than pPair.sub (second smallest p value = 0.111). We also note that for some markers, cross-validation results do not match with the p value results. For example, at marker *IL13-01*, mPair.sub test has the smallest prediction error among all 5 tests, but pPair test has the smallest p value of 0.042. This observation might be due to the small sample size, thus parameter estimates have a big variance in the cross-validation procedures.

Our simulation studies with a much larger sample size have suggested that the cross-validation procedures picked the correct genetic model most of the time in all parameter scenarios examined (table 6), although the second way of dealing with the heterozygous mother-offspring pairs (simply omitting them) is always preferred.

Discussion

In this study, we proposed an imprinting test with a newly proposed imprinting coding to detect interactions between environmental exposures and imprinted genes using mother-offspring pairs. Through extensive simulation studies, we showed that the birth cohort design with mother-offspring pairs is an efficient design for testing the interactions between imprinted genes and environmental exposures.

We observed that with mother-offspring pairs, the parent-of-origin information of each offspring can still be fully determined, except for the cases when both mother and offspring genotypes are heterozygous. In such a case, offspring imprinted genotypes can be either

Table 8. p value results of the interaction effects at 22 candidate markers with the proposed imprinting tests and the test based on the dominant genetic model

Marker	Proposed imprinting tests				Test assuming dominant model	CV
	mPair	mPair.sub	pPair	pPair.sub		
CYP1A1109	0.296	0.412	0.452	0.153	0.273	
CYP1A114	**0.002**	**0.035**	**0.032**	**0.006**	**0.003**	pPair.sub
CYP1A115	**0.063**	**0.083**	0.173	0.121	**0.079**	mPair.sub
CYP1A178	0.578	0.228	**0.098**	0.111	0.143	pPair.sub
CYP1A181	0.157	0.186	0.386	0.671	0.821	
CYP1A203	0.197	0.121	0.904	0.746	0.972	
CYP1A212	0.145	0.149	0.781	0.865	0.275	
CYP1A252	0.462	0.221	0.202	0.277	0.240	
CYP1B103	0.520	0.767	0.425	0.649	0.787	
CYP1B104	0.656	0.929	0.980	0.241	0.947	
CYP1B105	**0.071**	**0.071**	0.512	0.330	0.150	mPair.sub
CYP1B106	**0.132**	**0.091**	0.629	0.476	0.284	mPair.sub
CYP1B166	0.521	0.498	0.259	0.136	0.636	
CYP1B174	0.476	0.467	0.891	0.222	0.757	
IL1301	0.888	0.926	**0.042**	0.988	0.865	mPair.sub
IL1303	0.711	0.878	0.833	0.553	0.742	
IL403	0.462	0.616	0.483	0.152	0.924	
IL408	0.312	0.351	0.575	0.991	0.698	
IL410	0.498	0.422	0.318	0.236	0.569	
IL4R01	0.567	0.459	0.190	0.960	0.398	
IL4R05	0.879	0.934	0.161	0.850	0.378	
IL4R06	0.812	0.625	0.890	0.853	**0.090**	mPair.sub

For better visualization, only cross-validation (CV) results for those markers with p values <0.1 are displayed: the test with the smallest mean prediction errors among the 5 tests from the 5-fold CV procedures.

$A_m a_f$ or $A_f a_m$, i.e. the parent-of-origin information is ambiguous. We proposed two ways to deal with such ambiguous pairs, partially recovering parent-of-origin information or omitting such mother-offspring pairs, and compared their performance. The proposed imprinting tests mPair/pPair, and mPair.sub/pPair.sub, from both ways, respectively, correctly preserved the type I error rate, with the first way that partially recovers the parent-of-origin information being slightly more powerful in some scenarios. However, the second way of dealing with the heterozygous mother-offspring pairs (simply omitting them) is always preferred by the cross-validation procedures, as it leads to smaller mean prediction errors in all simulation scenarios. This might be due to the possibilities of misclassifying imprinting information when partially recovering parent-of-origin information. Thus, in real data analyses, we may simply omit heterozygous mother-offspring pairs.

The simulation results demonstrated the feasibility and power of the proposed imprinting tests over the test that assumes the Mendelian dominant genetic model. When the true underlying genetic model is either maternal imprinting or paternal imprinting, the proposed imprinting tests with mother-offspring pairs that assume the correct imprinting effect consistently have much higher power than the test assuming the Mendelian dominant genetic model. The power improvement is bigger when the true underlying genetic model is maternal imprinting rather than paternal imprinting. Also, when the true underlying disease model is maternal imprinting and the imprinting tests agree with the true model, imprinting tests using mother-offspring pairs has similar power to those using parent-offspring trios. When the true underlying disease model is paternal imprinting and the imprinting tests agree with the true model, imprinting tests using mother-offspring pairs lose about 20% of the power of the tests using parent-offspring trios.

In the application of our proposed methods to the questionnaire data on clinical childhood asthma, we successfully identified four markers that significantly inter-

act with the environmental exposure at the 0.1 significance level. In the majority of cases, tests or category of tests that have the smallest mean prediction errors from the cross-validation procedures also have the smallest p values. In this study, we concentrated only on the interactions between a single marker and an environmental exposure. Although both our simulation studies and the real data application suggested that omitting mother-offspring pairs with both heterozygous genotypes might be more preferred than partially recovering parent-of-origin information for such pairs, we may be able to recover better parent-of-origin information for mother-offspring pairs with both heterozygous genotypes if we could borrow parent-of-origin information from adjacent markers. We plan to explore this further as our future project.

Acknowledgements

Support from the National Institute of Environmental Health Sciences (grants P01 009600, R01 ES08977, R01 ES111158, R01 ES012468, R01 ES09089, and DISCOVER P50ES015905); the U.S. Environmental Protection Agency (grants R827027 and R8260901); the Herbert Irving Cancer Center; Core grant 5P30 CA 13696-23; the National Cancer Institute (grant R03 CA150140-01); the New York Community Trust, the Gladys & Roland Harriman Foundation, and Trustees of the Blanchette Hooker Rockefeller Fund is gratefully acknowledged. All authors declare no conflict of interest with this work.

References

1 Reik W, Walter J: Genomic imprinting: parental influence on the genome. Nat Rev Genet 2001;2:21–32.
2 Frost JM, Moore GE: The importance of imprinting in the human placenta. PLoS Genet 2010;6:e1001015.
3 Morison IM, Ramsay JP, Spencer HG: A census of mammalian imprinting. Trends Genet 2005;21:457–465.
4 Ober C, Loisel D, Gilad Y: Sex-specific genetic architecture of human disease. Nat Rev Genet 2008;9:911–922.
5 Horsthmeke B: Mechanisms of imprint dysregulation. Am J Med Genet C Semin Med Genet 2010;154C:321–328.
6 Kong A, Steinthorsdottir V, Masson G, Thorleifsson G, Sulem P, Besenbacher S, Jonasdottir A, Sigurdsson A, Kristinsson K, Jonasdottir A, Frigge M, Gylfason A, Olason P, Gudjonsson S, Sverrisson S, Stacey S, Sigurgeirsson B, Benediktsdottir K, Sigurdsson H, Jonsson T, Benediktsson R, Olafsson J, Johannsson O, Hreidarsson A, Sigurdsson G, the DIAGRAM Consortium, Ferguson-Smith A, Gudbjartsson D, Thorsteinsdottir U, Stefansson K: Parental origin of sequence variants associated with complex diseases. Nature 2009;462:868–874.
7 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.
8 Weinberg CR, Wilcox AJ, Lie RT: A log-linear approach to case-parent-triad data: assessing effects of disease genes that act either directly or through maternal effects and that may be subject to parental imprinting. Am J Hum Genet 1998;62:969–978.
9 Kistner EO, Infante-Rivard C, Weinberg CR: A method for using incomplete triads to test maternally mediated genetic effects and parent-of-origin effects in relation to a quantitative trait. Am J Epidemiol 2006;163: 255–261.
10 Knapp M, Strauch K: Affected-sib-pair test for linkage based on constraints for identical-by-descent distributions corresponding to disease models with imprinting. Genet Epidemiol 2004;26:273–285.
11 Strauch K, Baur MP: Parent-of-origin, imprinting, mitochondrial, and x-linked effects in traits related to alcohol dependence: presentation Group 18 of Genetic Analysis Workshop 14. Genet Epidemiol 2005;29 (suppl 1):S125–S132.
12 Feng R, Zhang H: A genomic imprinting test for ordinal traits in pedigree data. Genet Epidemiol 2008;32:132–142.
13 Shete S, Zhou X, Amos CI: Genomic imprinting and linkage test for quantitative-trait loci in extended pedigrees. Am J Hum Genet 2003;73:933–938.
14 Whittaker JC, Gharani N, Hindmarsh P, McCarthy MI: Estimation and testing of parent-of-origin effects for quantitative traits. Am J Hum Genet 2003;72:1035–1039.
15 Hanson RL, Kobes S, Lindsay RS, Knowler WC: Assessment of parent-of-origin effects in linkage analysis of quantitative traits. Am J Hum Genet 2001;68:951–962.
16 Ainsworth HF, Unwin J, Jamison DL, Cordell JH: Investigation of maternal effects, maternal-fetal interactions and parent-of-origin effects (imprinting), using mothers and their offspring. Genet Epidemiol 2011; 35:19–45.
17 McGrath J, Solter D: Completion of mouse embryogenesis requires both the maternal and paternal genomes. Cell 1984;37:179–183.
18 Surani MA, Barton SC, Norris ML: Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. Nature 1984;308:548–550.
19 DeChiara TM, Robertson EJ, Efstratidiadis A: Parental imprinting of the mouse insulin-like growth factor II gene. Cell 1991;64:849–859.
20 Groom A, Elliott HR, Embleton ND, Relton CL: Epigenetics and child health: basic principles. Arch Dis Child 2011;in press.
21 Perera F, Herbstman J: Prenatal environmental exposures, epigenetics, and disease. Reprod Toxicol 2011;in press.
22 Wang S, Zhao HY: Sample size needed to detect gene-gene interactions using association designs. Am J Epidemiol 2003;158:899–914.
23 Wang S, Zhao HY: Sample size needed to detect gene-gene interactions using linkage analysis. Ann Hum Genet 2007;71:828–842.
24 Wang S, Zheng T, Chanock S, Jedrychowski W, Perera FP: Methods for detecting interactions between genetic polymorphisms and prenatal environment exposure with a mother-child design. Genet Epidemiol 2010; 34:125–132.
25 International Agency for Research on Cancer: IARC Monographs on the Evaluation of the Carcinogenic Risk of Chemicals to Humans. vol 32. Polynuclear Aromatic Compounds. Part l. Chemical, Environmental, and Experimental Data. Lyon, World Health Organization, 1983, pp 1–453.

26 Bartsch H: DNA adducts in human carcinogenesis: etiological relevance and structure-activity relationship. Mutat Res 1996;340: 67–69.
27 Poirier MC, Beland FA: DNA adduct measurements and tumor incidence during chronic carcinogen exposure in animal models: implications for DNA adduct-based human cancer risk assessment. Chem Res Toxicol 1992;5:749–755.
28 Rybicki BA, Rundle A, Savera AT, Sankey SS, Tang D: Polycyclic aromatic hydrocarbon-DNA adducts in prostate cancer. Cancer Res 2004;64:8854–8859.
29 Tang D, Li TY, Liu JJ, Chen YH, Qu L, Perera FP: PAH-DNA adducts in cord blood and fetal and child development in a Chinese cohort. Environ Health Perspect 2006;114: 1297–1300.
30 Perera FP, Rauh V, Whyatt RM, Tsai WY, Tang D, Diaz D, Hoepner L, Barr D, Tu YH, Camann D, Kinney P: Effect of prenatal exposure to airborne polycyclic aromatic hydrocarbons on neurodevelopment in the first 3 years of life among inner-city children. Environ Health Perspect 2006;114: 1287–1292.
31 Perera FP, Tang D, Rauh V, Tu YH, Tsai WY, Becker M, Stein JL, King J, Del Priore G, Lederman SA: Relationship between polycyclic aromatic hydrocarbon-DNA adducts, environmental tobacco smoke, and child development in the World Trade Center cohort. Environ Health Perspect 2007;115:1497–1502.
32 Perera FP, Li Z, Whyatt R, Hoepner L, Wang S, Camann D, Rauh V: Prenatal polycyclic aromatic hydrocarbon exposure and child intelligence at age 5. Pediatrics 2009;124: e195–e202.
33 Perera FP, Tang D, Tu YT, Cruz LA, Borjas M, Bernert T, Whyatt RM: Biomarkers in maternal and newborn blood indicate heightened fetal susceptibility to procarcinogenic DNA damage. Environ Health Perspect 2004;112:1133–1136.
34 Alexandrov K, Rojas M, Geneste O, Castegnaro M, Camus AM, Petruzzelli S Giuntini C, Bartsch H: An improved fluorometric assay for dosimetry of benzo(*a*)pyrene diol-epoxide-DNA adducts in smokers' lung: comparisons with total bulky adducts and aryl hydrocarbon hydroxylase activity. Cancer Res 1992;52:6248–6253.
35 Chang C, Gauvey-Kern K, Johnson A, Kelvin EA, Chew GL, Perera F, Miller RL: Cord blood versus age 5 mononuclear cell proliferation on IgE and asthma. Clin Mol Allergy 2010;8:11.