

Testing Maternal-Fetal Genotype Incompatibility With Mother-Offspring Pair Data

Wenjiang J. Fu^{1,*}, Ming Li², Kelian Sun¹, Yuehua Cui³, Minping Qian⁴, Roberto Romero⁵

¹Department of Epidemiology and Biostatistics, Michigan State University, East Lansing, Michigan 48824 USA.

²Division of Biostatistics, Department of Pediatrics, University of Arkansas for Medical Sciences, Little Rock, AR, 72202.

³Department of Statistics and Probability, Michigan State University, East Lansing, Michigan 48824 USA.

⁴Department of Mathematical Science, Peking University, Beijing, China 100871.

⁵Perinatology Research Branch, NICHD/NIH, Hutzel Women's Hospital, Detroit, Michigan 48201 USA.

Abstract:

Maternal-fetal genotype (MFG) incompatibility has been shown to be strongly associated with a number of genetic diseases. Most current methods for the MFG test rely on parents-offspring trio genotype data and do not apply to mother-offspring pair data where paternal genotype is unavailable. In this paper, we developed a two-stage method for testing the allelic effect of given SNP through subgroup analysis of compatible MFG pairs and further tested the MFG incompatibility effect according to different scenarios determined by the allelic effect. Simulation studies demonstrated that this novel two-stage model is powerful in detecting the MFG incompatibility effect. This method can be implemented through publicly available statistical software, such as SAS, R, etc. We demonstrate with a case-control study of small for gestational age neonates that the method identified a SNP in the IGF2R gene with a significant allelic effect ($p = 0.037$), and a SNP in the IGF1 gene with a significant MFG incompatibility effect ($OR = 0.75$ and $p = 0.033$).

Corresponding author: Wenjiang Fu, PhD, Associate Professor, Department of Epidemiology and Biostatistics, Michigan State University, 909 Fee Road, East Lansing, Michigan 48824, Email: fuw@msu.edu

Running title: Test MFG Incompatibility among Mother-Child Paired

Keywords: Genotype-incompatibility; Mother-child pair; Small for gestational-age-neonates; Two-stage model.

Received: Nov 21, 2012; **Accepted:** Dec 21, 2012; **Published:** Jul 23, 2013

Academic Editor: Shanzhi Wang, Biochemistry department; Albert Einstein College of Medicine, NY, USA

Background

Single nucleotide polymorphisms (SNPs) have been shown with growing evidence to be associated with many diseases through polymorphisms at single locus or through haplotype at multiple loci (Zhang et al. 2004, Spinka et al. 2004, Cui et al. 2007).^{43,37,9} Pediatric studies have demonstrated strong association of phenotypes or childhood diseases with parental genotypes, offspring genotypes or both (Goddard et al. 2007 and references therein)¹⁴. In addition, research using both maternal and perinatal genotype data focusing on the interaction between the maternal and offspring genes has provided further evidence in gene-gene interaction, in particular, the effect of genotype incompatibility between the maternal and fetal genes has been found to play a critical role (Sinsheimer, Elston and Fu 2011). Hence searching for genes of incompatible maternal-fetal genotypes (MFG) is of particular interest.

Studies of the incompatibility of MFG can be traced back to as early as 1960s on hemolytic disease of newborn caused by anti-K antibody (Kulich V, Kout M. 1967)¹⁹ and neonatal thrombopenia feto maternal incompatibility in human leukocytes antigen (HL-A) system (Salet et al. 1973a and 1973b)^{33,34}. However, not until 1990s had the MFG received much attention (Hollister et al 1996). Although early studies introduced the MFG concept in human diseases research (Kulich V, Kout M. 1967, Salet et al. 1973a and 1973b, Stubbs, Ritvo and Mason-Brothers 1985)^{33,34,19,38}, statistical modeling of the MFG was not employed until 2002.

In a pioneering work, a MFG loglinear model was proposed to examine the MFG incompatibility effect by analyzing parents-offspring trio genotype data (Sinsheimer et al 2003)³⁶ and was applied to examine the association with the susceptibility of schizophrenia (Palmer et al. 2002)²⁹. The MFG loglinear model is based on Poisson regression, which divides the samples into different categories of all combinations of parents-offspring trio genotypes. The number of cases in each category is then modeled with loglinear model following a Poisson distribution. Without Hardy-Weinberg equilibrium (HWE) assumption, the expectation of the number of cases in each category was parameterized with corresponding mating rate, mother or child genetic effects as well as the MFG incompatibility effect. The parameters are then

estimated through the maximum likelihood approach and the inference is made accordingly.

The initial loglinear model for MFG test was improved in a series of articles. The correlation between siblings and relatives was incorporated (Kraft et al 2004)¹⁷. An exact MFG test was proposed to increase the statistical power of the likelihood ratio test by permuting the trio counts to generate a background distribution (Minassian et al 2005)²⁵. The MFG test has so far been applied to study a number of diseases, including RhD and HLA-B for schizophrenia, ABO for autism etc. (Palmer et al 2006, Zandi et al, 2006)^{30,42}. Recently, the MFG test was further extended for nuclear family data (Minassian et al 2006, Childs et al 2008).^{26,8}

The loglinear model for MFG test has been proven to be successful in a number of studies. In general, its requirement of parents-offspring trio genotype data, however, may limit its application to many studies where no paternal genotype data are available, although the statistical expectation-maximization (EM) algorithm or direct maximization of the observed loglikelihood (Misassian 2006, Hsieh 2006)^{26,30} can be implemented to deal with missing data when a portion of samples have missing paternal or maternal genotypes. Shi et al (2008)³⁵ showed that Mendelian inheritance, random mating and parental allelic exchangeability may be incorporated in the loglinear model simultaneously. Chen, Zheng, and Wilson (2009)⁷ took a retrospective likelihood approach to achieve the same effect of modeling family information. Recently, Chen et al (2012)⁶ further studied a semiparametric model to incorporate environmental factors. These models have been demonstrated with consistent and efficient estimation. However, due to the model complexity, the interpretation is not readily understood by many principal investigators and the model is not easy to implement by non-statistical professionals. Hence it is desirable and attractive to have a simple and yet efficient model to provide an easy-to-understand-and- implement procedure using currently available statistical software.

The fact that many perinatal studies with mother-offspring pairs do not have paternal genotype makes it desirable to extend the MFG test from parents-offspring trio genotype data to mother-offspring

pair genotype data. For this purpose, we extended the MFG test and developed a two-stage logistic regression model. This model conducts analysis in two steps. In the first step, it tests allelic effect of given SNP through a subgroup analysis using only mother-offspring pairs that had compatible MFG. It requires a moderately large sample size which allows subgroup analysis with sufficient power for testing the allelic effect and leads to several scenarios according to the allelic effect. In the second step, the MFG incompatibility effect is tested using all mother-offspring pairs through a logistic regression model accommodating the allelic effect with different scenarios. This two-stage model is different from the ones proposed by Li et al. (2009)²¹ and Ainsworth et al. (2011)¹. It enjoys meaningful interpretation of logistic regression and the allelic effects of SNPs, and can be implemented with common statistical software, such as SAS, R, etc. Statistical simulation studies demonstrate that this two-stage model achieves sufficient power to detect the MFG incompatibility effect. The method is also illustrated by an application to a case-control study of small for gestational age (SGA) neonates.

Materials and Method

Data

SGA Dataset

A total of 991 mother-offspring pairs (406 SGA cases and 585 controls) were included in this case-control study conducted in Chile. SGA was defined as birth weight below 10th percentile for the corresponding gestational age [Kramer 1987]¹⁸. Clinical and socio-demographic data were summarized in Table S1 (**Supplementary Material**), which shows no statistical significance between the two groups in maternal age, maternal BMI, smoking status and baby's gender. Parity was marginally significant with $p=0.048$. The control group had a larger mean gestational age and larger mean birth weight as expected.

SNP Data

Template DNA was obtained for genotyping through whole genome amplification of genomic DNA extracted from blood samples of patients following an automated DNA isolation protocol (BioRobot 9604, Qiagen, Valencia, CA, USA). Genotyping of single nucleotide polymorphisms (SNPs) was carried out using the massARRAYTM System (Sequenom Inc. San Diego,

CA, USA) by the high-throughput genotyping facility at Genaisance Inc (New Haven, CT, USA). SNP genotyping quality control was conducted through genotyping consistency of repeated samples. More than 1300 SNPs were genotyped in this candidate gene based case-control study for several phenotypes of pregnancy complication, including preeclampsia, SGA, etc. We focus on the IGF-I and IGF-II genes and their receptors only in this study.

A total of 44 SNPs were genotyped in IGF-I, IGF-II, IGF1R and IGF2R genes. Among them, 5 SNPs were found to be homozygous in the study population and 13 SNPs had one minor allele of frequency less than 0.01 and thus were excluded. HWE was examined in the combined mother and offspring control population and all of the 26 remaining SNPs satisfied the HWE with a cut-off p -value > 0.0001 . Ten SNPs out of these 26 remaining SNPs were found to have one minor allele with frequency equal to or smaller than 0.1 (homozygous genotype probability ≤ 0.01) and were excluded from the MFG incompatibility test. Thus the MFG incompatibility test was conducted on the remaining 16 SNPs.

Methods

We develop a two-stage model for testing the MFG incompatibility, and further conduct a simulation study.

The Two-Stage Method for MFG Incompatibility

Step 1. Determination of allelic effect

Step 1 determines the allelic effect between the two alleles and its scenario at each SNP. We first examine the minor allele frequency and the HWE for each SNP. Only those SNPs having a minor allele frequency ($p \geq 0.01$) and satisfying the HWE are kept for further study on the examination of allelic effect. The HWE is examined with a chi-square test in the combined mother and offspring control populations with a cut-off value of chi-square test p -value > 0.0001 .

To study the gene-gene interaction between the mother and her offspring at a given SNP, a composite SNP genotype of the mother and offspring pair is derived by combining the two. For example, if the mother genotype at a given SNP is 'AA' and the offspring genotype is 'AB', the composite genotype is 'AA_AB'. There are seven possible composite genotypes, see Table 1 for details. Among them, three are compatible:

Table 1. Composite genotypes of the mother-offspring pair

Mother genotype	Offspring genotype		
	AA	AB	BB
AA	AA_AA	AA_AB	n/a
AB	AB_AA	AB_AB	AB_BB
BB	n/a	BB_AB	BB_BB

Compatible genotypes are in bold letters.

'AA_AA', 'AB_AB' and 'BB_BB', and the rest are incompatible.

The allelic effect is examined through a subgroup analysis using only those mother-offspring pairs that had compatible genotypes, i.e. 'AA_AA', 'BB_BB', and 'AB_AB'. In doing so, the MFG incompatibility effect is not present and any significance is attributable to the allelic effect.

Step 2. Testing of MFG incompatibility

Step 2 tests the MFG incompatibility effect following each scenario determined in Step 1. A total of 4 scenarios were adopted accordingly to test the MFG incompatibility, see Table 2 for details of these scenarios and the parameterization of the MFG test. SNPs with a small frequency allele (allele frequency $p \leq 0.1$) are not tested for the MFG incompatibility effect.

Scenario 1. No allelic effect: $f(AA_AA) = f(AB_AB) = f(BB_BB)$, where function f represents the composite genotype effect.

A logistic regression model was then fitted to all mother-offspring pairs to test the association between the disease and the MFG incompatibility (1 for incompatible pairs and 0 for compatible) adjusting for environmental covariates.

Scenario 2. Allele A effect: $f(AA_AA) \neq f(AB_AB) = f(BB_BB)$, where the heterozygous genotype has the same effect as one homozygous but significantly different from the other.

To test the MFG incompatibility effect, a logistic regression model was fitted to all mother-offspring pairs adjusting for the individual 'AA' genotype effect β_{AA} of either mother or offspring genotype and the father's allele A contribution λ_A conditioning on the incompatible MFG.

Scenario 3. Heterozygosity effect: $f(AA_AA) = f(BB_BB) \neq f(AB_AB)$, where the two homozygous genotypes had the same effect but significantly different from the heterozygous.

To test the MFG incompatibility effect, a logistic regression model was fitted to all mother-offspring pairs adjusting for mother's 'AB' genotype effect β_{AB} and father's allele A contribution λ_A conditioning on the incompatible MFG.

Scenario 4. Unequal genotype effects (significantly different effects), including the following two cases.

4.1 Strong allelic interaction effect: $f(AB_AB) > f(AA_AA) > f(BB_BB)$ or

$f(AB_AB) < f(AA_AA) < f(BB_BB)$, where the heterozygous effect is the largest or smallest.

A logistic regression model was fitted to all mother-offspring pairs to test the MFG incompatibility adjusting for mother's AB genotype effect β_{AB} , mother's AA genotype effect β_{AA} , and father's allele A contribution λ_A conditioning on the incompatible MFG.

4.2 Additive or multiplicative allelic effect: $f(AA_AA) > f(AB_AB) > f(BB_BB)$, where the heterozygous effect is between the two homozygous effects.

To test the MFG incompatibility effect, a logistic regression model was fitted to all mother-offspring pairs adjusting for mother's allele A effect β_A and father's allele A contribution λ_A conditioning on the incompatible MFG.

It is worthwhile to note that for some SNPs, one minor allele may lead to small counts of the homozygous composite genotype of mother-offspring pairs of that allele, such as small counts of 'AA_AA' composite genotype when allele A has a small frequency. As a result, only a small number of samples may be observed

Table 2. Parameterization of logistic regression model testing MFG incompatibility

Scenario 1. $f(AA_AA) = f(AB_AB) = f(BB_BB)$: no allelic effect	
Model effect	Incompatible
$AB \rightarrow AB$	
$BB \rightarrow BB$	
$AA \rightarrow AA$	
$AA \rightarrow AB$	γ
$AB \rightarrow AA$	γ
$AB \rightarrow BB$	γ
$BB \rightarrow AB$	γ

γ is the MFG incompatibility effect.

Scenario 2. $f(AA_AA) \neq f(AB_AB) = f(BB_BB)$: allele A effect			
Model effect	Incompatible	Genotype	Father incompatibility
$AB \rightarrow AB$ & $BB \rightarrow BB$			
$AA \rightarrow AA$		β_{AA}	
$AA \rightarrow AB$	γ	β_{AA}	
$AB \rightarrow AA$	γ	β_{AA}	λ_A
$AB \rightarrow BB$	γ		
$BB \rightarrow AB$	γ		λ_A

γ is the MFG incompatibility effect. β_{AA} is the effect of genotype AA.

λ_A is the father's allele A effect given incompatibility.

Scenario 3. $f(AA_{AA}) = f(BB_{BB}) \neq f(AB_{AB})$: heterozygosity effect			
Model effect	Incompatible	Genotype	Father incompatibility
AA → AA & BB → BB			
AB → AB		β_{AB}	
AA → AB	γ		
AB → AA	γ	β_{AB}	λ_A
AB → BB	γ	β_{AB}	
BB → AB	γ		λ_A

γ is the MFG incompatibility effect. β_{AB} is the effect of mother's genotype AB.
 λ_A is the father's allele A effect given incompatibility

Scenrio 4.1: Strong interactive allele effect			
$f(AB) > f(AA) > f(BB)$ or $f(AB) < f(AA) < f(BB)$			
Model effect	Incompatible	genotype	Father incompatibility.
BB → BB			
AB → AB		β_{AB}	
AA → AA		β_{AA}	
AA → AB	γ	β_{AA}	
AB → AA	γ	β_{AB}	λ_A
AB → BB	γ	β_{AB}	
BB → AB	γ		λ_A

γ is the MFG incompatibility effect. β_{AB} is the effect of mother's genotype AB, and β_{AA} is the effect of mother's genotype AA. λ_A is the father's allele A effect given incompatibility.

Scenario 4.2: Additive or multiplicative allele effect $f(AA) > f(AB) > f(BB)$ or $f(AA) < f(AB) < f(BB)$			
Model effect	Incompatible	genotype	Father incompatibility.
BB → BB			
AB → AB		β_A	
AA → AA		$2\beta_A$ or β_{A2}	
AA → AB	γ	$2\beta_A$ or β_{A2}	
AB → AA	γ	β_A	λ_A
AB → BB	γ	β_A	
BB → AB	γ		λ_A

γ is the MFG incompatibility effect. β_A is the effect of mother's allele A.
 λ_A is the father's allele A effect given incompatibility.

in the compatible group with double homozygous minor allele. This often leads to non-significant MFG incompatibility in either Scenario 1 or 2 although it is difficult to distinguish between these two scenarios with a small frequency allele.

Simulation studies

We conducted simulations to study the power of the 2-stage test on the incompatibility for every scenario in Table 2. In each simulation, we generated 1000 subjects, each of which had a binary indicator for disease phenotypes (Y), MFG incompatibility effect (C), allelic effect (A) and father's allelic contribution effect given incompatibility (F). The genotype data were generated under the assumption of HWE. The logistic regression was applied to detect the incompatibility effect at the level of 0.05.

Data generation

We assumed the allele frequency p for allele A and $q=1-p$ for allele B. Table 3 provides the distribution of the MFG under the HWE. In all scenarios, p was generated from a uniform distribution $U[0, 1]$. The incompatibility index C was then generated from a Bernoulli distribution with $C = 1$ for incompatible MFG:

$$P(C=1) = p^2q + pq^2 + p^2q + pq^2 = 2pq,$$

$$P(C=0) = 1-2pq = p^2 + q^2$$

Other variables were generated according to the scenarios as follows.

No allelic effect. Two probabilities r_1 and r_2 for having disease with incompatible and compatible MFG, respectively, were generated from uniform distribution $U[0, 1]$, and 1000 subjects were generated with probabilities $Pr(Y=1/C=1) = r_1$, $Pr(Y=1/C=0) = r_2$.

Allele A effect. First, six probabilities $r_1, r_2, \dots, r_6 \sim U[0, 1]$ of having the disease were generated according to different values of C, A , and F , where F was generated with Bernoulli probabilities $Pr(F=1/C=1) = (p^2q + pq^2)/2pq = 0.5$ and $Pr(F=1/C=0) = 0$. A was the 'AA' genotype effect of the mother or the fetus or both, and was simulated by $Pr(A=1/F=1, C=1) = p$, $Pr(A=1/F=0, C=1) = p$, and $Pr(A=1/F=0, C=0) = p^3 / (p^2 + q^2)$. Y was the disease phenotype and was simulated with the probabilities r_1, \dots, r_6 as listed in Table 4.

Heterozygosity effect. First, six probabilities r_1, r_2, \dots, r_6 of having the disease were generated from uniform distribution $U[0, 1]$ according to different values of C, A, F . F was generated by Bernoulli trials with probabilities $Pr(F=1/C=1) = (p^2q + pq^2)/2pq = 0.5$ and $Pr(F=1/C=0) = 0$. A was the effect of genotype 'AB' in mother or fetus or both and was generated with $Pr(A=1/F=1, C=1) = p$, $Pr(A=1/F=0, C=1) = q$, and $Pr(A=1/F=0, C=0) = pq / (p^2 + q^2)$. Y was generated by Bernoulli trials with probabilities r_1, \dots, r_6 as listed in Table 4.

Strong allelic interaction effect. First, seven probabilities r_1, r_2, \dots, r_7 of having the disease were

Table 3. Distribution of maternal-fetal genotype under HWE with allele frequencies (p,q).

Mother genotype	Offspring genotype		
	AA	AB	BB
AA	p^3	p^2q	0
AB	p^2q	pq	pq^2
BB	0	pq^2	q^3
Pr(Y=1)	C	A	F

Table 4. Parameters in Simulation Studies

Pr(Y=1)	C	A	F
Models in scenarios 2 and 3.			
r1	1	1	1
r2	1	1	0
r3	1	0	1
r4	1	0	0
r5	0	1	1
r6	0	1	0
Models in scenario 4.			
r1	1	1	1
r2	1	1	0
r3	1	-1	1
r4	1	0	0
r5	0	1	0
r6	0	0	0
r7	0	1	0

generated from uniform distribution $U[0, 1]$ according to different values of C , A , F to accommodate three genotype effects β_{AB} , β_{AA} and β_{BB} . F was generated with the probabilities $Pr(F=1|C=0)=0$ and $Pr(F=1|C=1) = (p^2q+pq^2)/2pq = 0.5$. A was the additive effect of allele A and was generated with Bernoulli probabilities

$$Pr(A=1 | F=1, C=1) = p,$$

$$Pr(A=0 | F=1, C=1) = 0, Pr(A=-1 | F=1, C=1) = q,$$

$$Pr(A=1 | F=0, C=1) = q,$$

$$Pr(A=0 | F=0, C=1) = p, Pr(A=-1 | F=0, C=1) = 0,$$

$$Pr(A=1 | F=0, C=0) = pq/(p^2+q^2),$$

$$Pr(A=0 | F=0, C=0) = p^3/(p^2+q^2),$$

$$Pr(A=-1 | F=0, C=0) = q^3/(p^2+q^2),$$

where $A = 1$ for genotype 'AB' in mother or fetus, 0 for genotype 'AA', and -1 for genotype 'BB'. Y was generated by Bernoulli trials with probabilities r_1, \dots, r_7 as in Table 4.

Simulation Result

We fitted logistic regression model to the generated data with 1000 subjects and calculated the power of the test defined as out of the total number of simulations the number of times when the associated SNP was selected by statistical significance of the test. In scenario 1, the simulation was conducted with different odds ratio (OR) of disease between compatible and incompatible groups. The power of the study was calculated for each of the OR value 3, 2, 1, 1/2 or 1/3. The simulation was repeated 5,000 times for each value of OR. In scenarios 2-4, because of the random selection of the allelic effect and father effect, the OR was not controlled at first, but rather calculated after the allelic effect and father effect were generated. The ORs were categorized into different intervals of (0,1/3), [1/3,1/2), [1/2,2), [2,3) and [3,∞]. The simulation was repeated 50,000 times for each of these scenarios to accommodate the unspecified OR value. The power of the study was calculated for each OR interval. It was observed that the power of test increased with the OR as shown in Table 5 and Figure 1.

Further simulation was conducted to evaluate the specificity of the test. Ten SNPs were generated and the disease phenotype was generated only by SNP1 in Scenario 1, but independent of SNPs 2-9. The logistic regression model was fitted with 5000 repeats for each OR value, and the power was calculated for each SNP. The simulations demonstrated that the power increased

with the OR and that the false positive rate was about 1% (Table 6 and Fig. 2).

Application to a Case – Control Study of SGA

We demonstrate our method by studying the MFG incompatibility effect of SNPs on low birth weight in a case-control study of SGA neonates with candidate genes. While the definition of SGA can be found in the literature (Alkalay et al., 1998, Saenger et al., 2007)^{2,32}, the etiology remains complex. While it was suggested that most of the genetic influence originated in the fetal genome, other results (Ounsted et al. 1988)²⁸ also suggested that not only fetal genes but also genes regulating the maternal uterine environment could be important in determining the size at birth, especially for the SGA neonates.

Insulin-like growth factor (IGF) genes are known to have a major influence on fetal and postnatal growth, and their expression presents in all fetal tissues along the gestation process. On one hand, the cord concentrations of IGF1 and IGF2 have been shown to be highly correlated with birth weight (Ong et al., 2002)²⁷. On the other hand, while some polymorphisms in genes encoding IGF1, IGF2 and their respective receptors have been reported to be associated with birth weight (Vaessen et al. 2002)³⁹, but failed to be confirmed in other studies (Frayling et al. 2002)¹¹. Similarly, the reported association in studies with offspring population remains unconfirmed by one another. So far, no study has combined maternal and fetal genotypes together to study the association between SGA and the incompatibility of the mother and offspring genotypes. In this study, we choose the candidate genes (IGF1 and IGF2 and their receptors) to test the MFG incompatibility effect at each single SNP locus.

Results

Two-stage MFG incompatibility test

In the logistic regression analysis, mother's age (age), body mass index (BMI) and baby's sex (sex), were included as environmental covariates as suggested in the literature. Among the 16 SNPs considered for MFG incompatibility analysis, one SNP located at downstream 14463 intron21b of IGF2R gene was found to have a significant allele A effect (Table 7), where the composite genotype 'AA_AA' had a significantly different effect from 'AG_AG' and 'GG_GG' ($p = 0.01$) after Bonferroni's multiple-comparison adjustment. We also examined this

Figure 1. The power of the two-stage MFG test by the OR between the compatible and incompatible groups for scenarios 1 - 4.

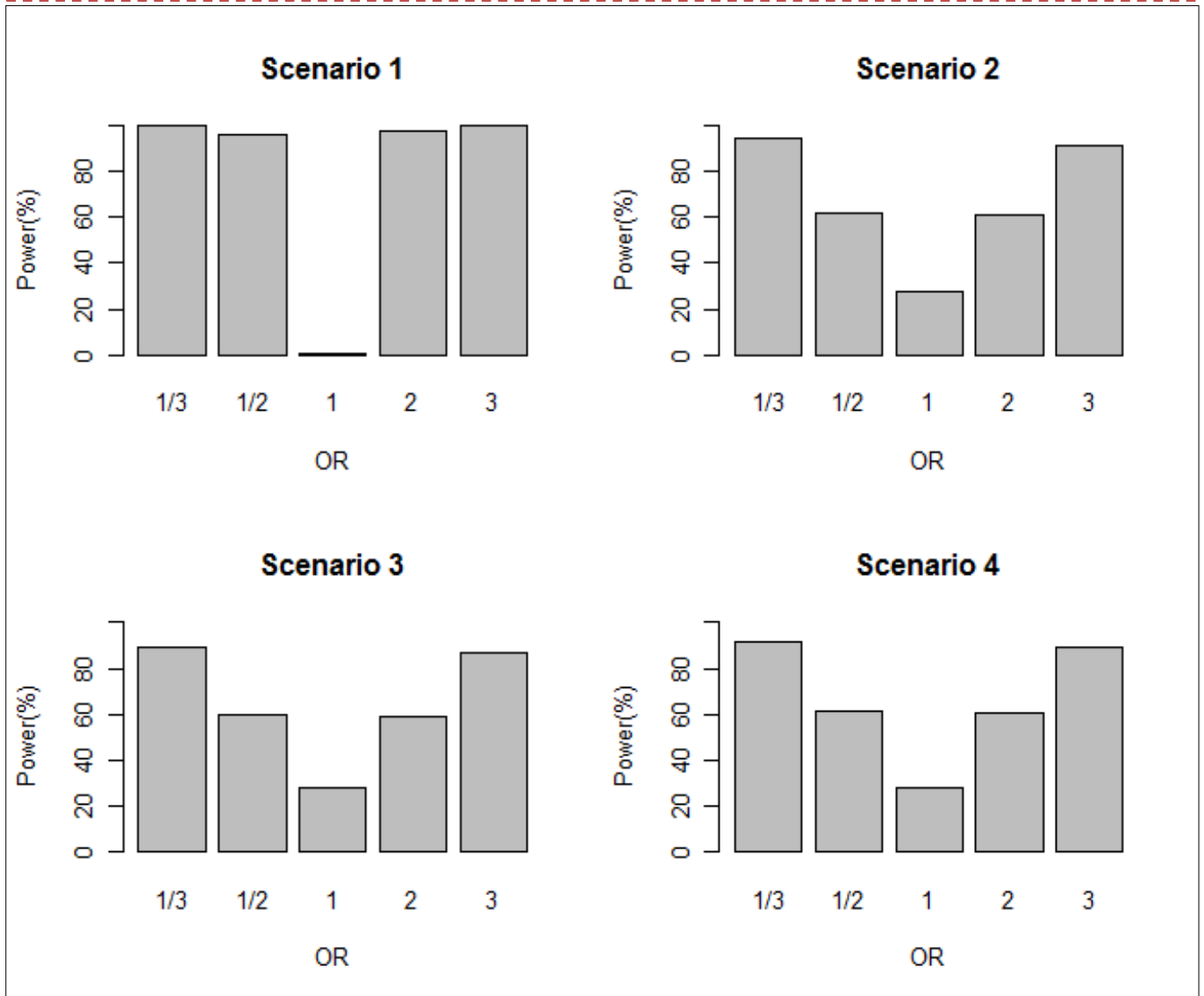


Figure 2. The specificity of the two-stage method for MFG test by OR.

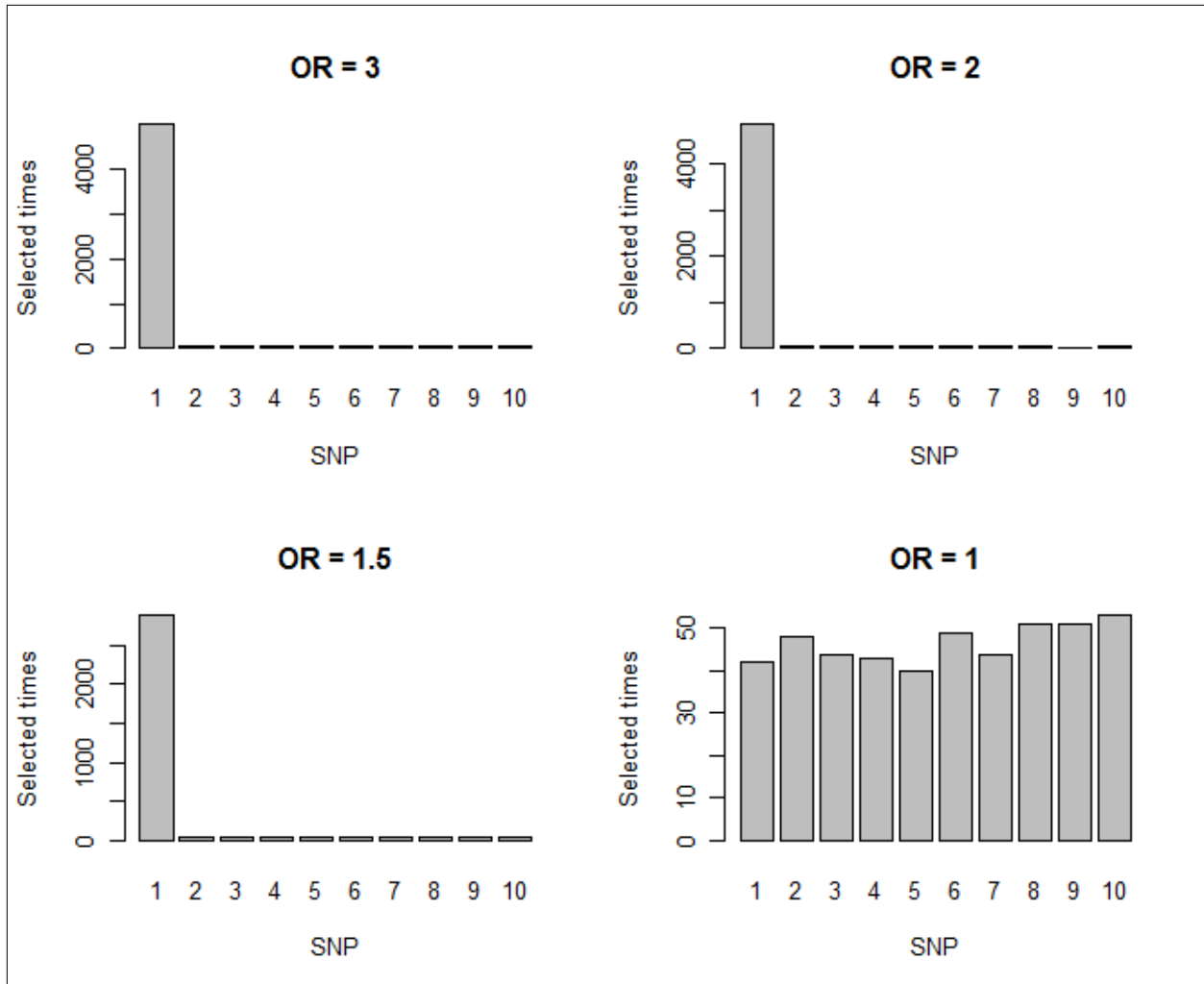


Table 5. Power of MFG incompatibility test by OR in Scenarios 2-4 simulation Odds Ratio (OR)

Scenario	(0,1/3) ^b	[1/3,1/2) ^b	[1/2,2) ^b	[2,3) ^b	[3,+∞) ^b
1	99.88% (4994/5000)	96.16% (4808/5000)	0.84% (42/5000)	97.6% (4880/5000)	99.98% (4999/5000)
2	94.62% (2234/2361)	61.84% (6112/9884)	27.67% (6747/24384)	60.61% (5702/9408)	91.19% (3614/3963)
3	88.89% (1665/1873)	59.46% (5574/9374)	27.71% (7384/26650)	58.72% (5452/9284)	86.98% (2452/2819)
4	91.17% (1486/1621)	61.25% (5565/9085)	27.63% (7661/27727)	60.40% (5516/9133)	89.11% (2169/2434)

^aPower was calculated as $n/N \times 100\%$: the number of times (n) incompatible MFG was selected by statistical significance out of the total number (N) of simulation runs.

^bFor Scenario 1, the OR was specified exactly 1/3,1/2, 1, 2, and 3.

Table 6. Power of MFG incompatibility test in Scenario 1 simulation based on 5000 repeats with 1000 subjects

SNP	OR=3	OR = 2	OR=1.5	OR=1
SNP1*	4999 (99.98%)	4880 (97.60%)	2881 (57.62%)	42 (0.84%)
SNP2	59 (1.18%)	48 (0.96%)	41 (0.82%)	48 (0.96%)
SNP3	53 (1.06%)	66 (1.32%)	39 (0.78%)	44 (0.88%)
SNP4	43 (0.86%)	39 (0.78%)	52 (1.04%)	43 (1.04%)
SNP5	41 (0.82%)	50 (1.00%)	43 (0.86%)	40 (0.80%)
SNP6	60 (1.20%)	56 (1.12%)	40 (0.80%)	49 (0.98%)
SNP7	60 (1.20%)	45 (0.90%)	44 (0.88%)	44 (0.88%)
SNP8	49 (0.98%)	53 (1.06%)	53 (1.00%)	51 (1.02%)
SNP9	50 (1.00%)	34 (0.68%)	38 (0.76%)	51 (1.02%)
SNP10	49 (0.98%)	44 (0.88%)	44 (0.88%)	53 (1.06%)

*SNP1 is the only SNP that is associated with the incompatible MFG in data generation.

Table 7. Test of allelic effect for each SNP

Polymorphism			Compatible Genotype	n*	Testing of Composite Genotype	p-value
SNP ID	Region	ATG				
41410456 G/A	IGF2R Intron 21b	14463	A/A_A/A	24	$f(A/A)=f(G/A)=f(G/G)$	0.005
			G/A_G/A	177	$f(A/A)=f(G/A)$	0.015
			G/G_G/G	320	$f(A/A)=f(G/G)$	0.008
					$f(G/A)=f(G/G)$	0.644

* Number of samples for each compatible genotype.

Table 9. List of Significant MFG Incompatibility Effect

Polymorphism		Effect Estimate	OR (95% CI)	p-value
SNP ID	Region (ATG)			
40925717 C/T	IGF1R Intron2a (1260)	-0.293	0.75 [0.57,0.98]	0.033

Table 8. SNP association analysis in single population

Polymorphism			Genotype (Samples)	Effect Estimate	OR (95% CI)	p-value
SNP ID	Region ATG	Overall P-value				
Mother Population						
42260934 A/T	IGF2R	0.037	A/A (250)	-0.343	0.71 [0.51, 0.98]	0.037
	In- tron17 11258		T/T (227) A/T (476)	-0.361 -		
41410456 G/A	IGF2R	0.037	A/A (72)	-0.609	0.54[0.31, 0.96]	0.034
	In- tron21 14463		G/G (484)	-0.258		
Offspring Population						
632197358 C/A	Intron 1 4560	0.040	A/A (0)	N/A	0.56 [0.31, 1.00]	0.040
			C/C (51) C/A (851)	-0.580 -		
41410456 G/A	IGF2R Intron21 14463	0.832	A/A (73)	-0.064	0.94 [0.55, 1.60]	0.816
			G/G (463) A/G (375)	0.058 -		

SNP in mother and offspring populations separately (Table 8) and found that this SNP had a significant genotype effect 'A/A' from 'A/G' in the mother population as well, but not significant in the offspring population. Further analysis did not reveal any MFG incompatibility effect of this SNP. We further studied the difference between the 'AA_AA' and 'AA_AG' composite genotypes. The p-value was 0.015, which suggested a significant effect of father's contribution of allele G given mother genotype 'AA' with the Bonferroni's correction. The remaining 15 SNPs had no significant allelic effect as shown in Table S2 [**Supplementary Material**]. Among them, one had significant MFG incompatibility effect (Table 9). This SNP is located at downstream 1260 intron2a of IGF1R gene and the incompatible MFG of the SNP decreased the risk for SGA with OR = 0.75 ($p=0.033$).

Discussion

Recent studies have shown that the MFG incompatibility is associated with several genetic phenotypes, including childhood autism, schizophrenia, etc. Former methods developed for the MFG incompatibility test are based on loglinear regression model on parents-offspring trio genotype data and have been successfully applied to a number of studies. However it is also common that paternal genotypes are unavailable in many studies of only mother-offspring pairs. Therefore, it is desirable to extend the MFG test from parents-offspring trio genotype data to mother-offspring pair genotype data. We have developed a novel two-stage MFG incompatibility test based on logistic regression model that examines the allelic effect of each SNP in the first step and the MFG incompatibility effect in the second step. We also have conducted simulation studies on the power of the test and applied this novel method to a study of the association between the development of SGA and the IGF genes and receptors.

Comparing with the former loglinear MFG incompatibility test, our novel method has the following features. 1). It does not require father's genotype data, thus is applicable to broad studies where father's genotype data are unavailable; 2). It tests the allelic effect first and classifies the SNPs into different scenarios by genetic effects of the alleles, which may provide insight on the genetic effects of the SNPs on the phenotypes under investigation; 3). The MFG

incompatibility test is conducted based on the allelic effect tested, is more specific for each allele and thus more powerful for each SNP; 4). The power of the MFG incompatibility test increases fast with the OR of the incompatible MFG effect while the false discovery rate is low around 0.01.

Conclusion

This novel two-stage method extends the MFG test from the parents-offspring trio genotype data to mother-offspring pair genotype data and thus makes the MFG test available through an easy-to-understand-easy-to-implement approach to many investigators in a wider range of studies who are not statistical professionals. The examination of the allelic effect may provide clues in the association between the SNP genotype and phenotype of interest, and is easy to interpret.

Acknowledgements

This work was partly supported by a contract from the NIH/NICHD to Michigan State University, and by NIH/NICHD intramural funding.

Conflict of Interest

The authors declare no conflict of interest.

Supplementary Data:

Supplementary Table 1:

Supplementary Table 2:

References

1. Ainsworth HF, Unwin J, Jamison DL, Cordell HJ. (2011) Investigation of maternal effects, maternal-fetal interactions and parent-of-origin effects (imprinting), using mothers and their offspring. *Genet Epidemiol.* 35(1): 19-45.
2. Alkalay AL, Graham Jr JM, Pomerance JJ: Evaluation of neonates born with intrauterine growth retardation: review and practice guidelines. *J Perinatol* 1998, 18, 142-151
3. Ashton IK, Eischenk I, McKenzie IZ: Insulin-like growth factors (IGF) I and II in human foetal plasma and relationship to gestation age and foetal size during midpregnancy. *Acta Endocrinol (Copenh)* 1985 110, 558-563

4. Baker J, Liu JP, Robertson EJ, Efstratiadis A: Role of insulin-like growth factors in embryonic and postnatal growth. *Cell* 1993 75, 73–82.
5. Bonapace G, Concolino D, Formicola S, Strisciuglio P: A novel mutation in a patient with insulin-like growth factor 1 (IGF-1) deficiency. *J Med Genet* 2003 40, 913–917.
6. Chen, J. Lin, D. Hochner, H. Semiparametric Maximum Likelihood Methods for Analyzing Genetic and Environmental Effects with Case-Control Mother–Child Pair Data, *Biometrics*, 68, 869–877.
7. Chen, J., Zheng, H., and Wilson, M. (2009). Likelihood ratio tests for maternal and fetal genetic effects on obstetric complications. *Genet. Epidemiology* 33, 526–538.
8. Childs, Ylisaukko-oja T, Turunen JA, Rehnström K, Peltonen L, Lange K, Palmer CGS, Sinsheimer JS: MFG Maximized: Testing for disease-related maternal-fetal genotype incompatibilities using the software package mendel. Abstract at Amer Soc. Hum. Genet. Conf. Philadelphia, 2008.
9. Cui, Y. W. Fu, K.L. Sun, R. Romero, and R. Wu: Mapping nucleotide sequences that encode complex binary disease traits with HapMap. *Current Genomics* 2007 8, 307-322.
10. Dunger DB, Petry CJ, Ong KK: Enetics of Size at birth. *Diabetes Care* 2007 30, S150-S155.
11. Frayling TM, Hattersley AT, McCarthy A, Holly J, Mitchell SM, Gloyn AL, Owen K, Davies D, Smith GD, Ben-Shlomo Y: A putative functional polymorphism in the IGF-I gene: association studies with type 2 diabetes, adult height, glucose tolerance, and fetal growth in U.K. populations. *Diabetes* 2002 51, 2313–2316.
12. Freemark M: Regulation of Maternal Metabolism by Pituitary and Placental Hormones: Roles in Fetal Development and Metabolic Programming. *Horm Res* 2006 65, 41-49.
13. Gluckman PD, Johnson-Barrett JJ, Butler JH, Edgar BW, Gunn TR: Studies of the insulin-like growth factor I and II by specific radioligands in umbilical cord blood. *Clin Endocrinol (Oxf)* 1983 19, 405–413.
14. Goddard, KA. Tromp, G. Romero, R. Olson, JM. Lu, Q. Xu, Z. Parimi, N. Nien, JK. Gomez, R. Behnke, E. Solari, M. Espinoza, J. Santolaya, J. Chaiworapongsa, T. Lenk, GM. Volkenant K, Anant MK, Salisbury BA, Carr J, Lee MS, Vovis GF, Kuivaniemi H: Candidate-gene association study of mothers with pre-eclampsia, and their infants, analyzing 775 SNPs in 190 genes. *Human Heredity*, 2007 63, 1-16.
15. Hattersley, Andrew. T. and John A: The fetal insulin hypothesis: an alternative explanation of the association of low birth weight with diabetes and cascular disease. *Lancet* 1999 353, 1789-1792.
16. Hayati AR, Cheah FC, Yong JF, Tan AE, Norizah, WM: The role of serum insulin-like growth factor I (IGF-I) in neonatal outcome. *J. Clin. Pathol.* 2004 57, 1299-1301.
17. Kraft P, Palmer CG, Woodward AJ, Turunen JA, Minassian S, Paunio T, Lonnqvist J, Peltonen L, Sinsheimer JS: RHD maternal-fetal genotype incompatibility and schizophrenia: extending the MFG test to include multiple siblings and birth order. *Eur J Hum Genet.* 2004;12, 192-198.
18. Kramer, M.S: Determinants of low birth weight: methodological assessment and meta-analysis. *Bull World health Org* 1987 65 pp. 663-737.
19. Kulich V, Kout M: Hemolytic disease of a newborn caused by anti-k antibody. *Cesk. Pediatr.* 1967; 22, 823-826.
20. Langford KS, Nicolaides KH, Jones J, Abbas A, McGregor AM, Miell JP: Serum insulin-like growth factor-binding protein-3 (IGFBP-3) levels and IGFBP-3 protease activity in normal, abnormal and multiple human pregnancy. *J Clin Endocrinol Metab* 1995 80, 21–27.
21. Li S, Lu Q, Fu W, Romero R, Cui Y. (2009) A regularized regression approach for dissecting genetic conflicts that increase disease risk in pregnancy. *Stat Appl Genet Mol Biol.* 8(1): Article 45.
22. Little RE , Sing CF: Genetic and environmental influences on human birthweight. *Am J Hum Genet* 1987 40, 512-526.
23. Ludwig T, Eggenschwiler J, Fisher P, D'Ercole AJ, Davenport ML, Efstratiadis A: Mouse mutants lacking the type 2 IGF receptor (IGF2R) are rescued from perinatal lethality in IGF2 and IGF1R null backgrounds. *Dev Biol* 1996 177, 517–535.

24. Magnus P: Caused of variation in birth weight: a study of offspring of twins. *Clin Genet* 1984 25, 15-24.
25. Minassian SL, Palmer CG, Sinsheimer JS: An exact maternal-fetal genotype incompatibility (MFG) test. *Genet Epidemiol.* 2005;28, 83-95.
26. Minassian SL, Palmer CG, Turunen JA, Paunio T, Lönnqvist J, Peltonen L, Woodward JA, Sinsheimer JS: Incorporating serotypes into family based association studies using the MFG test. *Ann Hum Genet.* 2006;70, 541-543.
27. Ong K, Kratzsch J, Kiess W, Costello M, Scott C, Dunger D: Size at birth and cord blood levels of insulin, insulin-like growth factor I (IGF-I), IGF-II, IGF-binding protein-1 (IGFBP-1), IGFBP-3, and the soluble IGF-II/mannose-6-phosphate receptor in term human infants: The ALSPAC Study Team: Avon Longitudinal Study of Pregnancy and Childhood. *J Clin Endocrinol Metab* 2000 85, 4266–426.
28. Ounsted M, Scott A, Moar VA: Constrained and unconstrained fetal growth: associations with some biological and pathological factors. *Ann Hum Biol* 1988 15, 119–129.
29. Palmer CG, Turunen JA, Sinsheimer JS, Minassian S, Paunio T, Lonnqvist J, Peltonen L, Woodward JA: RHD maternal-fetal genotype incompatibility increases schizophrenia susceptibility. *Am J Hum Genet.* 2002;71, 1312-9. Epub 2002 Nov 18
30. Palmer, C. G. S. Hsieh, H.J. Reed, E.F. Lonnqvist, J. Peltonen, L. Woodward, J.A. and Sinsheimer, J.S: HLA-B Maternal-Fetal Genotype Matching Increases Risk of Schizophrenia. *The American Journal of Human Genetics*, 2006 79, 710–715.
31. Reece EA, Wizniter A, Le E, Homko CJ, Behrman H, Spencer EM: The relation between human fetal growth and fetal blood levels of insulin-like growth factors I and II, their binding proteins, and receptors. *Obstet Gynecol* 1994 84, 88–95.
32. Saenger P, Czernichow P, Hughes L and Reiter EO: Small for Gestational Age: Short Stature and Beyond. *Endocrine Reviews* 2007 28, 219-251.
33. Salet J, Colombani J, Girot R, Beaufile F, Lejeune C, Fajgenbaum J: Neonatal thrombopenia with fetomaternal incompatibility in the HL-A system. *Rev Fr Transfus.* 1973b;16, 243-50.
34. Salet J, Colombani J, Fajgenbaum J, Beaufile F, Lejeune C, Girot R : Neonatal thrombopenia with fetomaternal incompatibility in the HL-A system. *Arch Fr Pediatr.* 1973b ;30, 533-40.
35. Shi, M., Umbach, D. M., Vermeulen, S., and Weinberg, C. R. (2008). Making the most of case-mother/control-mother studies. *Amer. J. Epidemiology* 168, 541–547.
36. Sinsheimer JS, Palmer CG, Woodward JA: Detecting genotype combinations that increase risk for disease: maternal-fetal genotype incompatibility test. *Genet Epidemiol.* 2003 Jan;24, 1-13.
37. Spinka, C. Carroll, R. and Chatterjee, L: Analysis of case-control studies of genetic and environmental factors with missing genetic information and haplotype-phase ambiguity. *Genetic Epidemiology*, 2004 29, 108-127.
38. Stubbs EG, Ritvo ER, Mason-Brothers A: Autism and shared parental HLA antigens. *Am Acad Child Psychiatry.* 1985;24, 182-185.
39. Vaessen N, Janssen JA, Heutink P, Hofman A, Lamberts SW, Oostra BA, Pols HA, van Duijn CM: Association between genetic variation in the gene for insulin-like growth factor-I and low birthweight. *Lancet* 2002 359, 1036-1037.
40. Verhaeghe J, Van Bree R, Van Herck E, Laureys J, Bouillon R, Van Assche FA: C-peptide, insulin-like growth factors I and II, and insulin-like growth factor binding protein-1 in umbilical cord serum: correlations with birthweight. *Am J Obstet Gynecol* 1993 169, 89–97.
41. World Health Organization: World health organization 2007. Geneva: World Health Organization.
42. Zandi, P.P. Kalaydjian, A. Avramopoulos, D. Shao, H. Fallin, M.D C.J: Rh and ABO maternal-fetal incompatibility and risk of autism. *American Journal of medical Genetics Part B Neuropsychiatric Genetics.* 2006 141b, 643-647.
43. Zhang, K. Qin, Z. Liu, J. Chen, T. Waterman, M. and Sun, F: Haplotype Block Partitioning and Tag SNP Selection Using Genotype Data and Their Applications to Association Studies. *Genome Research* 2004 14, 908-916.