

Detection of Parent-of-Origin Effects in Complete and Incomplete Nuclear Families with Multiple Affected Children Using Multiple Tightly Linked Markers

Ji-Yuan Zhou, Shili Lin, Wing K. Fung and Yue-Qing Hu

Presenter: Jingyuan Yang
Statistical Genetics Journal Club
Department of Statistics
The Ohio State University

February 3, 2009

Background

Existing Methods

Proposed Method

Simulation

Background

Existing Methods

Proposed Method

Simulation

Introduction

Parent-of-origin effect \Leftrightarrow Genomic imprinting effect

Incomplete families: Nuclear families with missing data for one parent, but not both parents.

HAP-1-PAT, one parent is available for each family with an arbitrary number of affected children.

HAP-C-PAT: both complete and incomplete families.

Notation

- m tightly linked markers so that the recombination frequency between any pair can be assumed to be zero.
- Term any nonempty subset Q of the marker set $\{1, 2, \dots, m\}$ as a 'marker combination'. For a marker combination Q , let $\mathbf{H} = \{h_1, h_2, \dots, h_n\}$ be the set of all haplotypes (compatible with the observed genotype data) and $\Theta = \{\theta_1, \theta_2, \dots, \theta_n\}$ be the corresponding population haplotype frequencies.
- Suppose D and d are the disease and normal alleles at a disease susceptibility locus (DSL). Let P_D denote the frequency of disease allele D , and let $\phi_{D/D}$, $\phi_{D/d}$, $\phi_{d/D}$ and $\phi_{d/d}$ be the penetrances of genotypes D/D , D/d , d/D and d/d at the DSL, respectively, where the allele before $/$ is paternal and the allele after $/$ is maternal.

Notation (Cont.)

The parent-of-origin effects are measured by the **degree of imprinting** $I = (\phi_{D/d} - \phi_{d/D})/2$, which is half the difference of the two heterozygote penetrances.

- $I = 0$ indicating no parent-of-origin effects.
- $I = (\phi_{d/d} - \phi_{D/D})/2$ indicating complete paternal parent-of-origin effect
- $I = (\phi_{D/D} - \phi_{d/d})/2$ indicating complete maternal parent-of-origin effect

We also represent the genotype relative risks as $\gamma_2 = \phi_{D/D}/\phi_{d/d}$, $\gamma_{1p} = \phi_{D/d}/\phi_{d/d}$, and $\gamma_{1m} = \phi_{d/D}/\phi_{d/d}$.

Model Assumptions

- No maternally mediated genetic effects.
- Mating symmetry in the population is assumed.
- When some parental genotypes are missing, missing of a parental genotype is independent of his/her underlying genotype.

Background

Existing Methods

Proposed Method

Simulation

Parental Asymmetry Test (PAT)

Suppose the marker locus of interest has two alleles, M_1 and M_2 . For convenience, let 0, 1 and 2 represent the marker genotypes M_2M_2 , M_1M_2 and M_1M_1 , respectively. We also let F , M and C represent the genotypes of the father, mother and child, respectively, and therefore, F , M and C take possible values of 0, 1 or 2. For n_C genotypes case-parent triads

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

where $N_{F>M,C=1}$ is the number of case-parents triads with heterozygous child in which the father carries more (fewer) copies of marker allele M_1 than the mother.

1-PAT

For n_I case-parent pairs with known marker genotypes for the available parent and affected child (comprising n_M case-mother pairs and n_F case-father pairs),

$$1\text{-PAT} = \frac{\omega(N_{M<C,C=1} - N_{M>C,C=1}) + (1 - \omega)(N_{F>C,C=1} - N_{F<C,C=1})}{\sqrt{\omega^2(N_{M<C,C=1} + N_{M>C,C=1}) + (1 - \omega)^2(N_{F>C,C=1} + N_{F<C,C=1}) + (n_M + n_F)^{-1}(N_{M<C,C=1} - N_{M>C,C=1})(N_{F>C,C=1} - N_{F<C,C=1})}},$$

where the weight $\omega = n_F / (n_M + n_F)$. Under the null hypothesis of no parent-of-origin effects, the joint genotype distribution of the case-mother pair and that of the case-father pair are the same.

Therefore, $E[\omega N_{M<C,C=1}] = E[(1 - \omega) N_{F<C,C=1}]$ and

$E[\omega N_{M>C,C=1}] = E[(1 - \omega) N_{F>C,C=1}]$, and then $E[1\text{-PAT}] = 0$.

HAP-PAT

Suppose we have n_C independent genotyped case-parent triads. For each $h_i \in \mathbf{H}$, let t_{C_i1} and t_{C_i2} denote the numbers of heterozygous children who inherit haplotype h_i from the father and mother, respectively. Then

$$\text{HAP-PAT} = \frac{n-1}{n} \sum_{i=1}^n \frac{(t_{C_i1} - t_{C_i2})^2}{t_{C_i1} + t_{C_i2}}$$

Background

Existing Methods

Proposed Method

Simulation

With Complete Phased Hyplotypes

	Parental genotype	Child's genotype
Father Missing	(h_3, h_1)	(h_2, h_1)
	(h_4, h_2)	(h_1, h_3)
Mother Missing		(h_1, h_4)
		(h_5, h_1)
	(h_5, h_3)	(h_3, h_2)
	(h_2, h_1)	(h_5, h_1)
		(h_1, h_2)

With Complete Phased Haplotype

	Parental genotype	Child's genotype
Father Missing	(h_3, h_1)	(h_2, h_1)
	(h_4, h_2)	(h_1, h_3)
Mother Missing		(h_5, h_1)
	(h_5, h_3)	(h_3, h_2)
	(h_2, h_1)	(h_5, h_1)
		(h_1, h_2)

	t_{M_*1}	t_{M_*2}
t_{M_1*}	2	1
t_{M_2*}	1	0
t_{M_3*}	0	1
t_{M_4*}	0	1
t_{M_5*}	0	0

With Complete Phased Hyplotypes

	Parental genotype	Child's genotype
Father Missing	(h_3, h_1)	(h_2, h_1) (h_1, h_3)
	(h_4, h_2)	(h_1, h_4)
Mother Missing		(h_5, h_1)
	(h_5, h_3)	(h_3, h_2) (h_5, h_1)
	(h_2, h_1)	(h_1, h_2)

	t_{M_*1}	t_{M_*2}		t_{F_*1}	t_{F_*2}
t_{M_1*}	2	1	t_{F_1*}	1	2
t_{M_2*}	1	0	t_{F_2*}	0	2
t_{M_3*}	0	1	t_{F_3*}	1	0
t_{M_4*}	0	1	t_{F_4*}	0	0
t_{M_5*}	0	0	t_{F_5*}	2	0

Test Statistic

$$\text{HAP-1-PAT} = \frac{n-1}{n} \sum_{i=1}^n \frac{[\omega(t_{M_i1} - t_{M_i2}) + (1-\omega)(t_{F_i1} - t_{F_i2})]^2}{\omega^2(t_{M_i1} + t_{M_i2}) + (1-\omega)^2(t_{F_i1} + t_{F_i2})}$$

With complete phase information, we are able to construct two contingency tables. The total number of incomplete nuclear families is denoted by $n_I = n_M + n_F$. If there are n_j affected children in the h -th incomplete family, then we have $n_{CMP} = \sum_{j=1}^{n_M} n_j$ case-mother pairs and $n_{CFP} = \sum_{j=n_{M+1}}^{n_I} n_j$ case-father pairs.

$$\omega = n_{CFP} / (n_{CMP} + n_{CFP})$$

Permutation Procedure

- Note that, under the null hypothesis of no parent-of-origin effects, the joint genotype distribution of incomplete nuclear families with missing father and that of incomplete nuclear families with missing mother are the same.
- Without loss of generality, we assume that, in a sample with n_I incomplete nuclear families, the first n_M are families with missing fathers while the remaining $n_F = n_I - n_M$ are with missing mothers. We permute the order of the families to obtain a new sample with the same size n_I .
- Repeating this permutation procedure B times leads to a collection, $\{\text{HAP-1-PAT}_1^*, \dots, \text{HAP-1-PAT}_B^*\}$.

$$p\text{-value} = \frac{|\{b : 1 \leq b \leq B, \text{HAP-1-PAT}_b^* \geq \text{HAP-1-PAT}\}|}{B}$$

With Unphased Hyplotypes

For haplotype $h_i \in \mathbf{H}$, since the four quantities $t_{M_i,1}$, $t_{M_i,2}$, $t_{F_i,1}$ and $t_{F_i,2}$ in HAP-1-PAT cannot be directly calculated, we replace it by $E(t|\mathbf{G}) = \sum_Z [t(Z)P(Z|\mathbf{G})]$, where the summation is over all the possible haplotype explanations Z compatible with genotype \mathbf{G} .

Multiple Testing

The statistic HAP-1-PAT was to test for parent-of-origin effects for a given marker combination Q . But prior to any analysis it is usually unknown as to which marker combination provides the greatest power for testing for parent-of-origin effects.

HAP-1-PAT Q are correlated. We consider $H_0 = \bigcap_Q H_0^Q$, which is the global null hypothesis that there is no parent-of-origin effects. A small $p^{min} = \min_Q P_0^Q$ would provide evidence against the global null. To compare p^{min} to its sampling distribution under H_0 , we can recycle the permuted samples.

Mixture of Complete and Incomplete Families

Suppose we have a mixture of n_C complete families and n_I incomplete families (with n_M missing fathers and n_F missing mothers).

$$\text{HAP-C-PAT} = \frac{n-1}{n} \sum_{i=1}^n \frac{[(t_{C,i1} - t_{C,i2}) + \omega(t_{M,i1} - t_{M,i2}) + (1-\omega)(t_{F,i1} - t_{F,i2})]^2}{(t_{C,i1} + t_{C,i2}) + \omega^2(t_{M,i1} + t_{M,i2}) + (1-\omega)^2(t_{F,i1} + t_{F,i2})}$$

Background

Existing Methods

Proposed Method

Simulation

Simulation Settings

- Family number is fixed at 200.
- The relative risk with two copies of the disease allele, γ_2 is fixed at 2.
 - PEM0:
no parent-of-origin effect $\gamma_{1p}=1.5$, $\gamma_{1m} = 1.5$
 - PEM1:
incomplete maternal parent-of-origin effect $\gamma_{1p}=2$, $\gamma_{1m} = 1.5$
 - PEM0:
complete maternal parent-of-origin effect $\gamma_{1p}=2$, $\gamma_{1m} = 1$
- Disease allele frequency: either 0.25 or 0.5.
- Incomplete family rate τ ; Missing father rate β .
- 5 tightly linked SNPs.
 - 3 HD scenarios (A B C in Table 2) for the three core SNPs (2-4).

Population Structure Setting

- Homogeneous population model
- Population stratification demographic model
- Assortative mating demographic model