

유전자형별 상대 위험도를 이용한 유전자-유전자간 상호작용 탐색

정지원¹ · 이재용² · 이석훈³ · 박미라⁴

¹(주)BIOTOXTECH, ²을지대학교 생리학 및 생물리학교실, ³충남대학교 통계학과
⁴을지대학교 예방의학교실

(2011년 7월 접수, 2011년 9월 채택)

요약

최근 유전학에서 주요 목표중 하나는 복합질환에 영향을 미치는 유전적 요인을 찾아내는 것이다. 유전자좌간의 상호작용이 있을 때에는 단일 유전자좌 분석으로는 이러한 목표를 달성하기 어려우므로, 유전자-유전자간 상호작용이나 유전자-환경인자간 상호작용분석을 고려할 필요가 있다. 자주 사용되는 MDR(multifactor dimensionality reduction)방법은 데이터를 고위험군과 저위험군으로 각각 병합하여 사용하므로 특정 유전자형에서 차이가 나는 경우에는 이를 찾아내기 어렵다. 본 연구에서는 이러한 점을 보완하도록 유전자형 조합에서의 대조군과 질환군의 상대 위험도를 이용하여 유전자-유전자간 상호작용을 탐색하는 방법을 제안하였다. MDR 공개데이터와 8가지 유전모형으로부터 생성한 모의자료의 분석을 통해 방법의 유용성을 확인하였다.

주요어: 유전자-유전자 상호작용, 상대위험도, 유전모형, MDR.

1. 서론

어떤 질병에 영향을 미치는 유전적 요인을 찾는 것은 현대 유전학의 주요 목표 중 하나이다. 그간 하나의 유전자좌가 질병에 미치는 영향을 확인하는 단일 유전자좌(single locus)방법들이 많이 적용되어 왔으나, 심혈관장애나 당뇨와 같은 복합 질환(complex disease)에서는 이러한 방법이 적절하지 않은 경우가 많다 (Heidema 등, 2006). 이때에는 여러 유전자간 상호작용이나 유전자와 환경 인자와의 상호작용에 의해 질병이 발생할 수 있으며, 여러 요인간의 상호작용을 찾고 해석하는 통계적 방법이 필요하다. 이러한 문제를 해결하기 위한 방법으로 로지스틱 회귀분석과 같이 회귀모형에 근거한 방법이나 정보이론에 의한 방법, MDR(multifactor dimensionality reduction) 등의 차원축소방법, machine learning방법 등이 적용되고 있다 (Cordell, 2009).

사례-대조연구에서의 상호작용탐색방법으로 대표적인 비모수적 방법인 MDR의 경우, 방법론이 단순하고 고차원의 상호작용에 대해서도 쉽게 확장할 수 있으며, 유전모형을 가정하지 않는다는 장점이 있다. 반면, 사례군과 대조군에서의 도수만으로 저위험군과 고위험군을 구분함으로써 유의성에 대한 검증 없이 데이터를 지나치게 단순화하며, 각 유전자형별 위험도는 고려하지 않고 셀을 병합하여 전반적인 연

이 논문은 2010년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (No.2010-0012133).

⁴교신저자: (301-832) 대전광역시 중구 계룡로 771번길, 을지대학교 예방의학교실, 교수.

E-mail: mira@eulji.ac.kr

관만을 보게 되는 단점이 있다. MDR방법에 대한 다양한 측면에서의 후속연구가 이어졌다. Chung 등 (2007)은 질병위험에 대한 새로운 척도로서 MDR절차에 추가적으로 오즈비와 신뢰구간을 구하여 위험요인의 중요도를 평가하는 ORMDR방법을 제안하였다. Lee 등 (2007)은 고차원으로 갈수록 빈 셀이나 빈도가 적은 셀에 적용가능한 LM MDR을 제안하였으며, Lou 등 (2007)과 Lou 등 (2008)은 각각 공변량이 있는 경우에 적용가능한 GMDR과 이를 가족 자료로 확장시킨 PGMDR을 발표하였다. 한편 Namkung 등 (2009)은 결측자료에 대한 보정을 한 MDR방법을 발표한 바 있다.

본 연구에서는 각 유전자에서 유전자형별로 상대위험도(relative risk)를 구하고, 그에 대한 검정을 통해 의미 있는 차이를 보이는 유전자형 조합을 찾는 방법론을 고려했다. 2절에서 MDR방법론과 제안된 방법을 설명하고, 3절에서 예제 데이터와 8가지 유전모형에 따르도록 생성된 모의자료에 적용한 결과를 MDR과 비교하여 유용성을 확인하였고, 4절에서 방법의 장단점을 논의하였다.

2. 유전자-유전자간 상호작용 분석

2.1. MDR 기반방법

Ritchie 등 (2001)에 의해 제안된 MDR(multifactor dimensionality reduction)방법은 사례-대조연구에서의 유전자-유전자간 상호작용을 탐색하는 방법이다. 알고리즘은 다음과 같다.

- (단계 1) 자료를 10등분하여 9개를 훈련세트(training set)로 정하고, 나머지 1개를 검증세트(test set)로 정한다.
- (단계 2) k -차원 상호작용을 보기 위해서는 전체 SNP에서 k 개의 SNP를 선택하여 k 차의 조합을 만든다. 이 조합에 대한 분할표를 작성한 후 각 셀의 대조군과 사례군의 비를 고려하여 고위험군 또는 저위험군으로 분류한다.
- (단계 3) 모든 유전자 조합에 대해 위에서 분류한 2개의 위험군과 질병 유무로 2×2 분할표를 작성한다. 훈련 정확도(training accuracy)가 가장 큰 조합을 선택하여 최적모형으로 선택한다.
- (단계 4) 선택된 모형을 검증세트에 적용시켜 검증 정확도(testing accuracy)를 계산한다.
- (단계 5) 위 과정을 10번 반복하여 교차검증을 통해 최적모형이 선택된 횟수를 계산한다. 이를 CVC(cross validation consistency)라 하며 이를 최대화하는 모형을 찾는다.

이 방법은 가장 높은 상호작용을 보이는 하나의 유전자조합을 찾게 되며, 유의성여부를 검증하지는 않는다. 또한 단순히 각 셀에서의 사례군과 대조군의 비를 비교하여 위험군을 결정하게 되므로 각 셀에서의 사례군과 대조군의 비가 전체 데이터에서의 비와 비슷하거나 사례군과 대조군의 수가 적을 경우에는 잘못된 결과를 가져오기 쉽다.

Chung 등 (2007)은 이러한 점을 보완하기 위해 질병위험도의 새로운 척도로서 MDR절차에 추가적으로 오즈비와 신뢰구간을 구하여 위험요인의 중요도를 평가하는 ORMDR방법을 제안하였다. 구체적인 절차는 다음과 같다.

- (단계 1) MDR의 단계1)-단계5)를 실행하여 최적의 SNP조합을 찾는다.
- (단계 2) 최적의 SNP조합에서 각 유전자형의 조합에 따른 오즈비(odds ratio)를 계산한다.
- (단계 3) 붓스트랩 데이터를 생성하여 오즈비의 경험적 분포를 구하고 신뢰구간을 계산하여 유의성을 판정한다.

이 방법은 오즈비를 제공함으로써 연관성의 척도를 제시하고 유의성 검증을 할 수 있다는 점에서 기존의 MDR보다 강점을 가지고 있다. 그러나 1차적으로 MDR을 수행하여 뽑힌 SNP조합에 대해서만 각 유전자형별로 연관성의 척도와 유의성을 제공하게 되므로, 모든 셀에 대한 고위험군/저위험군의 분류와 그 분류에 따른 통합에 의해 최적의 SNP조합이 판정되는 절차는 그대로 따르게 된다. 유전연구에 있어서는 통합했을 때의 전반적인 유의성이 아니라 특정 유전자형조합에서만 매우 유의하게 나타나는 경우가 발생할 수 있으며, 이 SNP조합을 찾아내는 것은 매우 의미가 있다. MDR이나 ORMDR은 절차가 단순한 대신, 이러한 조합을 찾아내는 데는 약점이 있다. 다음 소절에서는 유전자형별 상대위험도를 이용하여 각 유전자형별로 유의성을 찾아내는 방법을 제안하기로 한다.

2.2. 유전자형별 상대위험도를 이용한 방법

여기서는 ORMDR같이 자료를 통합하여 SNP조합을 선택한 후 유전자형별로 유의성을 판정하는 top-down 방식이 아니라, k 차 SNP조합의 각 유전자형별로 유의한 셀이 있는 SNP조합을 찾아내는 bottom-up방식을 고려하였다. 알고리즘은 다음과 같다.

(단계 1) k -차원 상호작용을 보기 위해서 전체 SNP에서 k 개의 SNP를 선택하여 k 차의 조합을 만들고, 각각의 유전자형에 따라 세분화하여 분할표를 만든다.

(단계 2) 분할표의 각 셀마다 대조군과 사례군의 수를 세고, 이를 전체의 대조군과 사례군의 수로 나눈 뒤, 다음 같이 셀 i 에서의 상대위험도(relative risk) RR_i 를 구한다. 여기서 n_i 와 m_i 는 각각 셀 i 에서의 사례군과 대조군 수이며, n_{total} 과 m_{total} 은 각각 사례군과 대조군의 총도수이다.

$$RR_i = \frac{n_i/n_{total}}{m_i/m_{total}} = \frac{\hat{\pi}_1}{\hat{\pi}_2}.$$

(단계 3) 상대위험도에 자연로그를 취한 $\log RR$ 은 근사적으로 정규분포를 따르게 되며, 이의 표준오차 추정치는 다음과 같이 계산할 수 있다. 이를 사용하여 각각의 상대위험도가 1인지에 대한 통계적 가설 검정을 각 셀에 대해 수행한다.

$$\hat{\sigma}(\log RR) = \left(\frac{1 - \hat{\pi}_1}{\hat{\pi}_1 n_{total}} + \frac{1 - \hat{\pi}_2}{\hat{\pi}_2 n_{total}} \right)^{\frac{1}{2}}.$$

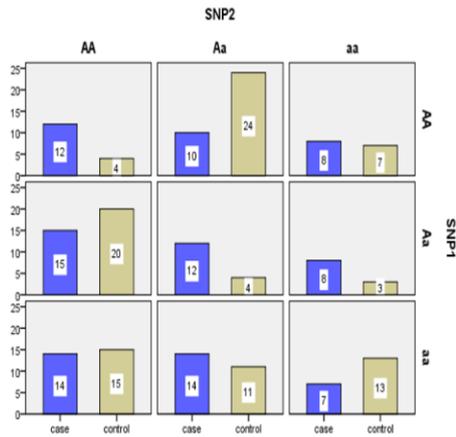
이 방법을 RRIE(relative risk interaction explorer)라고 하자. RRIE는 모든 셀에 대하여 통계적 검정을 통하여 유전자-유전자간 상호작용을 찾는 방법으로, 유전자형별로 구성된 분할표의 각 셀에서 대조군과 질환군의 차이가 나는 경우에 상호작용을 찾을 수 있다. 그림 2.1은 $k = 2$ 일 때 계산된 데이터 형태를 예로 보여준다. 이 때 각 셀의 대조군과 질환군의 수가 0인 경우에는 0.5로 대체하여 계산한다.

이 방법을 통해 유의한 결과가 나온 유전자-유전자 조합에 대해 셀 단위가 아닌 SNP단위로 전반적인 순위를 매기고 싶다면 단계 2에서 각각 계산된 상대위험도를 다음과 같이 종합한 OIE(overall interaction explorer)값을 사용할 수 있다.

$$OIE = \sum_{i=1}^z \left(\frac{\ln RR_i}{\hat{\sigma}_i} \right)^2,$$

여기서, RR_i 는 셀 i 에서의 상대위험률이고, $\hat{\sigma}_i$ 는 각 셀 i 에서의 $\log RR_i$ 의 표준 오차 추정치이며, z 는 셀의 총 개수이다.

고차원의 유전자조합을 고려할 때 각 셀에서의 관측수가 매우 적은 경우가 발생하여 접근 표준오차를 이용하기 어렵게 된다. 이 경우에는 점근적인 방법 대신 Zelen (1971)이나 Agresti (1992)와 같은 소표본에서의 추론방법을 고려할 수 있다.



(a) 2차원 data(대조군, 질환군 각 100명씩)

	AA	Aa	aa	
AA	RR_1 (= $\frac{12/100}{4/100} = 3.0$)	RR_2 (= 0.42)	RR_3 (= 1.14)	AA
Aa	RR_4 (= 0.75)	RR_5 (= 3.0)	RR_6 (= 2.67)	Aa
aa	RR_7 (= 0.93)	RR_8 (= 1.27)	RR_9 (= 0.54)	aa

(b) 상대위험도 계산 예

그림 2.1. 2차원 상호작용을 고려할 대의 분할표와 상대위험도의 예

표 3.1. RRIE방법을 통한 모형 선택과 MDR과의 비교

차원	유리한 모형수		최적모형	RR(유의확률)	MDR 최적모형(CVC)
	(1)	(2)			
1	3	1	$X1(aa)$	0.54(0.01)	$X1(9/10)$
			$X6(aa)$	0.46(0.01)	
			$X1(AA)$	1.34(0.02)	
2	50	7	$X5 \times X7(aa \times Aa)$	3.44(0.00)	$X1 \times X8(8/10)$
			$X1 \times X8(aa \times Aa)$	0.05(0.00)	
			$X1 \times X7(AA \times Aa)$	1.96(0.00)	
3	548	40	$X1 \times X6 \times X8(AA \times Aa \times Aa)$	4.00(0.00)	$X1 \times X6 \times X8(10/10)$
			$X1 \times X6 \times X8(Aa \times AA \times Aa)$	3.36(0.00)	
			$X1 \times X6 \times X8(Aa \times Aa \times Aa)$	0.13(0.00)	

(1): $\alpha = 0.05$, (2): $\alpha = 0.01$

3. 결과

3.1. 데이터분석

위에서 제안된 방법을 MDR 공개자료(<http://www.multifactor dimensionality reduction.org>)에 적용하여 MDR결과와 비교하였다. 이 데이터는 200명의 대조군과 200명의 질환군으로 이루어진 자료로서 20개의 SNP로 구성되어 있다.

각 SNP들은 3개의 유전자형을 갖는데 AA를 0, Aa를 1, aa를 2로 나타내었고, 대조군은 0, 질환군은 1로 나타내었다. 표 3.1은 1, 2, 3차원의 RRIE방법을 통하여 선택한 유전자형 조합과 MDR 결과를 비교한 표이다.

RRIE 선택 모형 수는 유의수준 0.05와 0.01에서 유의한 것으로 판단되는 유전자형 조합의 수이다. 먼저 MDR에서 선택된 최적모형과의 비교를 위하여 RRIE방법에서 유의확률이 가장 작은 모형순으로 3개를 선택하여 MDR과 비교하였다. RRIE방법은 하나의 SNP조합에 대해서 각 셀에 따라 별도의 검

표 3.2. OIE방법을 통한 모형 선택과 MDR과의 비교

차원	조합	OIE	MDR
1	X1	10.5147	X1
	X6	6.6800	
2	X1 × X8	25.7541	X1 × X8
	X1 × X6	24.8533	
3	X1 × X6 × X8	143.7531	X1 × X6 × X8

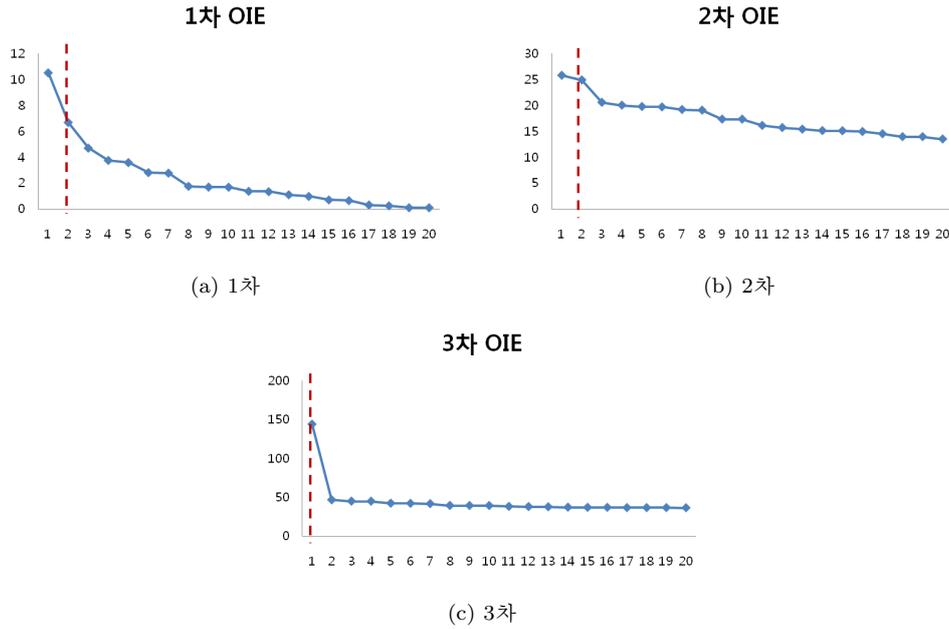


그림 3.1. 상위 20개의 유전자형 조합에 대한 1, 2, 3차 OIE 선택

정을 수행할 수 있고, 각각의 상대 위험도를 통하여 셀마다 대조군과 질환군의 비가 어느 정도 차이가 나는지에 대한 정보도 얻을 수 있다. 1차원의 경우 MDR에서는 X1이 최적모형으로 뽑히며, RRIE에서는 X1의 aa와 AA, X6의 aa가 유의한 것으로 나타났다. X1의 aa는 상대위험도가 0.54로서 대조군에 비해 사례군의 도수가 낮으며 AA의 경우에는 1.34로서 대조군의 도수가 사례군에 비해 낮은 것을 알 수 있다. 2차원에서는 RRIE의 경우 X5와 X7, X1과 X8의 상호작용이 매우 유의한 것으로 나타났으며, MDR에서는 X1과 X8의 상호작용이 최적으로 선택되었다. 3차원 상호작용에서는 두 방법 모두 X1, X6, X8의 상호작용이 가장 큰 것으로 나타났다.

표 3.2는 RRIE방법을 통해 걸러진 조합들을 대상으로 OIE방법을 통하여 선택한 유전자형 조합이다. OIE방법은 각 셀들을 종합적으로 고려한 방법이므로 MDR의 결과와 유사하다는 것을 알 수 있다.

그림 3.1은 OIE에 대한 scree plot이다. 이 그림을 통해 의미 있는 조합의 수를 선택할 수 있다.

3.2. 모의실험

세 방법을 8가지의 잘 알려진 2차원 유전모형에 대하여 모의실험을 통해 비교해 보았다. 그림 3.2의

Model1 (RR)		SNP 2			Model2 (RD)		SNP 2			Model3 (1R)		SNP 2			Model4 (T)		SNP 2			
		BB	Bb	bb			BB	Bb	bb			BB	Bb	bb			BB	Bb	bb	
S N P 1	AA	L	L	L	AA	L	L	L	AA	L	L	L	AA	L	L	L	AA	L	L	L
	Aa	L	L	L	Aa	L	L	L	Aa	L	L	L	Aa	L	L	H	Aa	L	L	H
	aa	L	L	H	aa	L	H	H	aa	H	H	H	aa	L	H	H	aa	L	H	H

Model5 (Mod)		SNP 2			Model6 (DD)		SNP 2			Model7 (XOR)		SNP 2			Model8 (Diagonal)		SNP 2			
		BB	Bb	bb			BB	Bb	bb			BB	Bb	bb			BB	Bb	bb	
S N P 1	AA	L	L	L	AA	L	L	L	AA	L	L	H	AA	L	L	H	AA	L	L	H
	Aa	L	L	H	Aa	L	H	H	Aa	L	L	H	Aa	L	H	L	Aa	L	H	L
	aa	H	H	H	aa	L	H	H	aa	H	H	L	aa	H	L	L	aa	H	L	L

그림 3.2. 2인자 유전모형

표 3.3. 오즈비에 따른 고위험군과 저위험군의 질병확률

자료유형	고위험군 오즈	저위험군 오즈	질병확률	
			고위험군	저위험군
A	1.2	1	6/11	1/2
B	1.5	1	3/5	1/2
C	2.0	1	2/3	1/2
D	6.0	1	6/7	1/2

8가지 모형에서 각 셀을 고위험군(H), 저위험군(L)으로 분류해 놓은 그림이다. 모형 1은 jointly recessive-recessive모형(RR), 모형 2는 jointly recessive-dominant모형(RD), 모형 3은 single gene recessive모형(1R), 모형 4는 threshold모형(T), 모형 5는 modifying effective모형(Mod), 모형 6은 jointly dominant-dominant모형(DD), 모형 7은 exclusive OR모형(XOR), 모형 8은 diagonal모형(Diagonal)이다 (Dong 등, 2008).

모의실험 자료의 생성은 다음과 같다. 먼저 SNP1과 SNP2를 하디-와인버그 평형(Hardy-Weinberg equilibrium)에 맞춰 AA 유전자형과 aa 유전자형의 비율은 0.25로, Aa 유전자형의 비율은 0.5의 확률을 갖는 삼항 분포를 따르도록 1000명의 자료를 생성한다. SNP1과 SNP2 조합의 유전자형별 분할표에서 각 셀을 유전모형에 맞도록 고위험군과 저위험군으로 구분한다. 분할표의 각 셀에서 표 3.3의 질병확률에 따라 고위험군과 저위험군의 질병여부를 생성한다. 유형 A는 고위험군의 오즈비가 1.2로 낮은 경우이며, 유형 B는 1.5, 유형 C는 2, 유형D는 오즈비가 6으로 매우 높은 자료이다. 나머지 SNP3에서 SNP10은 각각에 대해 AA 유전자형과 aa 유전자형의 비율은 0.25로, Aa 유전자형의 비율은 0.5의 확률을 갖는 삼항 분포를 따르도록 생성한다. 이 과정을 반복하여 각 모형당 100개씩의 자료를 만들었다.

표 3.4는 모의실험 결과로서 각 유전모형 및 자료유형별로 세 가지 방법에서 최적모형으로 SNP1과 SNP2를 선택한 횟수를 보여준다. 이 때 RRIE는 유의확률이 가장 작은 셀을 갖는 조합을, OIE는 OIE값이 가장 큰 값을 갖는 조합을 최적 선택모형으로 하여 MDR과 비교하였다.

표 3.4. 유형별 (SNP1, SNP2)쌍을 최적모형으로 선택한 횟수

유전모형	유형A			유형B			유형C			유형D		
	R	O	M	R	O	M	R	O	M	R	O	M
1	4	3	1	12	3	2	18	6	2	97	19	15
2	5	4	1	14	8	6	42	25	24	92	38	63
3	7	4	1	11	5	5	12	8	10	12	12	22
4	7	3	4	15	25	27	47	79	85	937	100	100
5	5	3	3	18	25	27	31	75	82	44	100	100
6	2	5	6	9	28	37	16	86	91	20	100	100
7	12	6	6	21	35	35	50	95	95	98	100	100
8	8	6	8	34	42	46	79	93	96	100	100	100

R: RRIE, O: OIE, M: MDR

표 3.5. RRIE에서 유의수준별 (SNP1, SNP2)쌍의 선택 횟수

유전모형	유형A		유형B		유형C		유형D	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
1	45	45	56	56	82	82	100	100
2	48	48	79	79	99	99	100	100
3	47	47	84	84	97	97	100	100
4	53	53	88	88	99	99	100	100
5	51	51	90	90	99	99	100	100
6	53	53	92	92	100	100	100	100
7	54	54	87	87	99	99	100	100
8	55	55	89	89	100	100	100	100

(1): $\alpha = 0.05$, (2): $\alpha = 0.01$

유형 A나 유형B와 같이 고위험군의 오즈비도 낮은 경우에는 세 방법 모두 최적모형을 잘 찾지 못하며, 오즈비가 커지면 커질수록 모든 유전모형에서 RRIE, OIE, MDR방법 모두 최적모형으로 선택하는 횟수가 많아진다. 유전모형 1, 2와 같이 분할표의 적은 수의 셀에서 대조군과 질환군의 차이가 나는 경우에는 전체 셀을 종합적으로 보는 MDR방법은 최적조합을 잘 찾지 못하는 반면 RRIE방법은 최적모형으로 선택한 횟수가 많으며, 오즈비가 클수록 이런 현상은 뚜렷이 나타남을 알 수 있다. 하지만 고위험군의 셀이 늘어나면 RRIE방법보다는 MDR방법이 최적모형을 SNP1과 SNP2의 조합을 선택하는 횟수가 늘어나게 됨을 알 수 있다. 그리고 상대위험도를 이용한 종합적인 방법인 OIE는 MDR과 비슷한 양상을 보이지만 1번과 2번 유전모형과 같이 적은 수의 셀에서 차이가 나는 경우에는 MDR보다 약간 더 선택하는 횟수가 많아졌다.

표 3.5는 RRIE방법을 각각의 유의수준 하에서 SNP1과 SNP2 조합이 유의하게 나타나는 횟수를 정리한 표이다. 고위험군과 저위험군의 오즈비가 별 차이가 없는 1번 유형에서도 SNP1, SNP2 조합이 유의하게 나오는 횟수가 많음을 알 수 있고, 오즈비가 커질수록 대부분의 모형에서 유의하게 나타나는 것을 확인할 수 있다.

4. 결론

유전자-유전자간 상호작용을 찾는 방법으로 자주 사용되고 있는 MDR은 유전모형에 무관한 비모수적 방법으로 알려져 있지만, 전체 데이터를 고위험군과 저위험군으로 단순분류함으로써 일부 셀에서만 큰 차이를 보이는 경우에는 이를 찾아낼 수 없다. 본 연구에서는 이러한 단점에 대한 대안으로서 각 유전

자형별로 상대위험도를 계산하는 RRIE방법을 제안하였다. 또한 이의 정보를 각 유전자조합별로 통합하는 OIE측도를 제시하였다. 모의실험결과 최적의 유전자쌍 하나만을 선택하는 것으로 비교했을 때 유전자형별 분할표의 적은 셀에서 고위험률이 발생한 경우에는 RRIE방법을 사용하는 것이 매우 유용하였다. 따라서 jointly recessive-recessive모형(RR)과 jointly recessive-dominant모형(RD)의 경우에는 RRIE방법이 더 유용하다고 생각된다. 분할표의 여러 셀에서 높은 오즈비가 발견될 때에는 MDR방법이 더 유용한 것으로 나타났다. 하지만 MDR이 최적의 한 쌍을 선택하는 방법인데 반해 RRIE방법은 유의한 쌍들을 모두 찾아내게 되므로, 실제 방법을 적용했을 때에는 유의한 유전자쌍을 식별하는데 있어 그다지 불리하지 않다. 또한 유전자형별 분할표에서 계산된 상대위험도를 셀의 표본 수에 의해 보정하는 효과를 가지므로 각 셀의 상대위험률의 정보에 차이를 두었다는 장점이 있다. 그러나 검정횟수가 늘어나면서 민감도가 높은 대신 특이도가 낮아지게 되므로 유의수준의 보정이 필요하다.

RRIE방법의 단점으로 생각할 수 있는 것은 하나의 유전자쌍에 대해서 분할표의 각 셀에 대해 각각 적용을 해야 하므로 계산횟수가 늘어난다는 것이다. 특히 다수의 유전인자를 동시에 고려해야 하는 경우, k 개 유전인자가 있다면 한 조합당 3^k 개의 상대위험도와 유의확률을 고려해야 하는 불편함이 있다. 하지만 이의 계산이 비교적 단순하며, MDR에서도 3^k 개의 셀을 고위험군/저위험군으로 판정한 다음 분류오류를 계산하고 이 과정을 다시 10번 반복적용(10-fold 방식의 경우)하게 되므로 MDR보다 효율성이 떨어진다고 할 수 없다. 따라서 유전자-유전자간 상호작용을 탐색하는데 있어서 RRIE방법의 사용이 유용할 것으로 생각된다.

참고문헌

- Agresti, A. (1992). A survey of exact inference for contingency tables, *Statistical Science*, **7**, 131–153.
- Chung, Y., Lee, S. Y., Elston, R. C. and Park, T. (2007). Odds ratio based multifactor-dimensionality reduction method for detecting gene-gene interactions, *Bioinformatics*, **23**, 71–76.
- Cordell, H. J. (2009). Detecting gene-gene interactions that underlie human diseases, *Nature Reviews Genetics*, **10**, 392–404.
- Dong, C., Chu, X., Wang, Y., Jin, L., Shi, T., Huang, W. and Li, Y. (2008). Exploration of gene-gene interaction effects using entropy-based methods, *European Journal of Human Genetics*, **16**, 229–235.
- Heidema, A. G., Boer, J. M. A., Nagelkerke, N., Mariman, E. C. M., Van der, A, D. L. and Feskens, E. J. M. (2006). The challenge for genetic epidemiologists: How to analyze large numbers of SNPs in relation to complex diseases, *BMC Genetics*, **7**, 1–15.
- Lee, S. Y., Chung, Y., Elston, R. C., Kim, Y. and Park, T. (2007). Log-linear model-based multifactor dimensionality reduction method to detect gene-gene interactions, *Bioinformatics*, **23**, 2589–2595.
- Lou, X. Y., Chen, G. B., Yan, L., Ma, J. Z., Zhu, J., Elston, R. C. and Li, M. D. (2007). A generalized combinatorial approach for detecting gene-by-gene and gene-by-environment interactions with application to nicotine dependence, *The American Journal of Human Genetics*, **80**, 1125–1137.
- Lou, X. Y., Chen, G. B., Yan, L., Ma, J. Z., Mangold, J. E., Zhu, J., Elston, R. C. and Li, M. D. (2008). A combinatorial approach to detecting gene-gene and gene-environment interactions in family studies, *The American Journal of Human Genetics*, **83**, 457–467.
- Namkung, J., Elston, R. C., Yang, J. M. and Park, T. (2009). Identification of gene-gene interactions in the presence of missing data using the multifactor dimensionality reduction method, *Genetic Epidemiology*, **33**, 646–656.
- Ritchie, M. D., Hahn, L. W., Roodi, N., Bailey, L. R., Dupont, W. D., Parl, F. F. and Moore, J. H. (2001). Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer, *The American Society of Human Genetics*, **69**, 138–147.
- Zelen, M. (1971). The analysis of several 2×2 contingency tables, *Biometrika*, **58**, 129–137.

Exploration of the Gene-Gene Interactions Using the Relative Risks in Distinct Genotypes

Ji-Won Jung¹ · Jaeyong Yee² · Suk-hoon Lee³ · Mira Park⁴

¹Biotoxtech, Inc.; ²Department of Physiology and Biophysics, Eulji University

³Department of Statistics, Chungnam National University

⁴Department of Preventive Medicine, Eulji University

(Received July 2011; accepted September 2011)

Abstract

One of the main objects of recent genetic studies is to understand genetic factors that induce complex diseases. If there are interactions between loci, it is difficult to find such associations through a single-locus analysis strategy. Thus we need to consider the gene-gene interactions and/or gene-environment interactions. The MDR(multifactor dimensionality reduction) method is being used frequently; however, it is not appropriate to detect interactions caused by a small fraction of the possible genotype pairs. In this study, we propose a relative risk interaction explorer that detects interactions through the calculation of the relative risks between the control and disease groups from each genetic combinations. For illustration, we apply this method to MDR open source data. We also compare the MDR and the proposed method using the simulated data eight genetic models.

Keywords: Gene-gene interaction, relative risk, genetic model, MDR.

This Research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the ministry of Education, Science and Technology (2010-0012133).

⁴Corresponding author: Professor, Department of Preventive Medicine, Eulji University, 771 Kyeryong-ro, Chung-Gu, Daejeon, Korea. E-mail: mira@eulji.ac.kr