

환경적인 요인을 보정한 한우의 우수 유전자 조합 선별

이제영¹ · 진미현²

^{1,2}영남대학교 통계학과

접수 2012년 4월 10일, 수정 2012년 4월 24일, 게재확정 2012년 4월 30일

요약

인간의 질병과 가축의 경제적인 특성은 하나의 유전자가 아닌 여러 유전자의 상호작용의 영향을 더 많이 받는다고 알려져 있다. 본 논문에서는 유전적인 효과만을 밝혀내기 위해 선형회귀모형을 활용하여 환경적인 요인을 보정하고, 최근 한우의 맛과 육질에 영향을 준다고 밝혀진 단일염기다형성 5개 (Oh 등, 2011)를 이용해 한우의 경제형질에 영향을 주는 우수 유전자 조합을 선별하고 우수 유전자형을 밝힌다. 이때, 많은 유전자들 중에서 우수한 유전자를 찾기 위한 비모수적인 방법인 다중인자 차원 축소 방법을 이용하여 단일 유전자의 효과보다 상호작용의 효과가 한우의 경제형질에 더 많은 영향을 준다는 사실을 확인하였다.

주요용어: 다중인자차원축소, 단일염기다형성, 유전자 조합, 유전자형.

1. 서론

유전학에서는 인간의 질병 및 가축의 경제적인 특성과 연관된 유전자를 규명하는 것이 주된 관심중의 하나이다. 특히 인간의 질병 및 가축의 경제적인 특성은 단일 유전자의 영향보다는 유전자-유전자 간의 상호작용의 영향을 많이 받는다고 알려져 있다. 유전자의 상호작용을 계산하기 위해서 표준 통계적 모형을 사용해왔으나 유전자의 수가 늘어남에 따라 그 모형의 복잡도가 증가되거나 해석이 어려워져 사실상 해석이 불가능해지는 경우가 종종 발생한다. 그래서 제안된 방법이 다중인자 차원 축소 (multifactor dimensionality reduction, MDR; Richie 등, 2001)방법이다. 하지만 MDR 방법은 사례군-대조군으로 이루어진 이분형 데이터에서만 적용이 가능하고 연속형 데이터에는 적용이 불가능하다는 한계가 있다. 연속형 데이터에도 MDR방법을 적용하기 위해 CART (classification and regression tree)방법을 활용한 확장된 다중인자 차원 축소 (expanded multifactor dimensionality reduction, E-MDR; Lee 등, 2008)방법, SVM (support vector machine)을 이용한 다중인자 차원 축소 (support vector machine multifactor dimensionality reduction, SVM-MDR; Lee와 Lee, 2010)방법, 유전자 행렬 매핑을 활용한 우수 유전자 조합 선별 (Lee 등, 2010) 등이 개발되었다. 인간의 질병 중의 하나인 심장의 위험인자를 규명하기 위해 MDR방법을 적용 (Lee 등, 2011)하기도 하였으며, 본 연구는 가축의 경제적인 특성을 이용하여 우수 유전자를 규명하기 위해 한우자료에 MDR방법을 적용하였다. 또한 한우의 경제형질에 영향을 주는 유전적인 효과만을 살펴보기 위해, 최근 한우의 맛과 육질에 영향을 준다고 밝혀진 단일염기다형성 (SNP) 5개를 이용해서 (Oh 등, 2011), 환경적인 요인을 보정하고 유전적인 요인들만 분석에 이용하였다. 분석에 사용된 한우의 경제형질은 올레인산 (oleic acid; C18:1), 단일불포화지방산 (monounsaturated fatty acid; MUFA), 근내지방도 (marbling

¹ 교신저자: (712-749) 경북 경산시 대동 214-1, 영남대학교 통계학과, 교수. E-mail: jlee@yu.ac.kr

² (712-749) 경북 경산시 대동 214-1, 영남대학교 통계학과, 대학원생.

score; MS)이며 이들 세 경제형질에 영향을 미치는 우수 유전자를 찾아낸 후, 세 경제형질에서 공통적으로 나타난 유전자를 한우의 종합 경제형질에 영향을 주는 우수 유전자로 활용하도록 한다. 또한 선별된 우수 유전자 중에서 우수 유전자형 (genotype)을 찾아 소개한다.

본 연구는 다음과 같이 구성되었다. 2절에서는 환경적 요인을 보정한 우수 유전자 조합 선별을 위해 선형회귀모형 및 MDR방법과 그 절차를 소개하고 3절에서는 우수 유전자 조합을 선별하기 위한 통계적 방법을 한우 데이터에 적용한 결과를 제시하였다. 4절에서는 3절에서 선별된 우수 유전자 조합에서 우수 유전자형을 선별한 결과를 소개하고 5절에서는 연구의 결과를 요약한다.

2. 우수 유전자 조합 선별을 위한 통계적 방법

2절에서는 우수 유전자 조합을 선별하기 위해 사용한 통계적 방법들에 대해서 살펴보고 각 방법의 특징과 절차를 소개한다. 또한 우수 유전자형 선별의 결과를 검증하기 위해 사용한 순열 검증 방법도 함께 소개한다.

2.1. 보정된 선형 회귀 모형

유전분석에서 어느 개체의 표현형은 그 개체의 유전적인 효과와 환경적인 효과에 의해 결정이 된다. 즉, 개체의 표현형은 다음과 같이 나타낼 수 있다.

$$P = E + G \quad (2.1)$$

P 는 개체의 표현형이고 E 는 환경적인 효과, G 는 유전적인 효과이다. 그래서 한우의 경제형질에 대한 연구에서는 다음과 같은 선형 회귀모형을 사용한다 (Casas 등, 2005).

$$y_{ijk} = \mu + farm_i + \beta * age + SNPs_j + e_{ijk} \quad (2.2)$$

$$(P) \quad (E) \quad (G)$$

$$i = 1, \dots, f, j = 1, \dots, m, k = 1, \dots, n$$

여기서 y_{ijk} 는 한우의 경제형질이고 $farm_i$ 는 장소 (17군데)의 고정효과이며 β 는 일령에 대한 회귀계수, $SNPs_j$ 는 SNP의 고정효과, e_{ijk} 는 $N(0, \sigma^2)$ 인 확률변수이다. 그러나 우리가 관심을 가지고 있는 부분은 한우의 경제형질에 영향을 주는 유전적인 효과를 밝혀내는 것이기 때문에 환경적인 요인을 보정한 아래의 (2.3)의 모형을 연구에 이용한다 (Matsuhashi 등, 2011).

$$y_{ijk} - (farm_i + \beta * age) = \mu + SNPs_j + e_{ijk} \quad (2.3)$$

$$(P) \quad (E) \quad (G)$$

$$i = 1, \dots, f, j = 1, \dots, m, k = 1, \dots, n.$$

이러한 환경적 요인을 보정한 선형회귀모형의 연구 분석을 위하여 Minitab 버전 16을 활용하였다.

2.2. 다중인자 차원 축소 MDR 방법

인간의 질병 유무와 같은 이분형 자료에 대한 유전자의 상호작용 효과를 찾기 위해 MDR 방법이 제시되었으며, MDR 방법은 일반화된 선형 모형의 전통적 통계기법과는 달리 모수에 대한 추정과 모형에 대한 가정이 필요하지 않은 비모수적 방법이다. 실험군-대조군의 비율을 통해 독립변수를 고위험군과 저위험군으로 분류한 뒤 목표변수에 대한 오분류율을 비교하는 방법이며 아래는 MDR 방법의 과정을 나타낸 것이다.

- 단계 1. 데이터를 크기가 동일한 10개의 셋으로 나눈다. 그 중 9개를 학습용 자료, 나머지 하나를 검증용 자료로 설정한다.
- 단계 2. 선택된 SNP조합에서 SNP의 각 수준을 기초로 한 개체들을 각 셀에 기술한다. 예를 들어서 $k=2$ 일 경우, SNP는 3개의 수준으로 되어있으므로 $3^2 = 9$ 개의 셀을 가진다. 각각 9개의 셀에 실험군-대조군의 도수를 기술한다.
- 단계 3. 실험군-대조군의 비를 구하여 1보다 크거나 같으면 고위험군, 1보다 작으면 저위험군으로 정한다. 예를 들어 실험군=3, 대조군=5 인 경우 그 비는 0.6이므로 그 셀은 저위험군으로 정한다.
- 단계 4. 학습용 자료의 모든 셀에서 잘못 분류된 비율인 학습용 자료의 오분류율과 검증용 자료를 이용하여 잘못 예측된 분류 비율인 검증용 자료의 오분류율을 구한다.
- 단계 5. 단계 1에서 정의한 10쌍의 모든 데이터셋에 대해서 위의 과정을 반복하여 학습용 자료의 오분류율의 평균과 검증용 자료의 오분류율의 평균을 구한다.
- 단계 6. 구해진 학습용 자료의 오분류율의 평균과 검증용 자료의 오분류율의 평균값을 비교해서 가장 낮은 조합을 우수 SNP조합으로 선정한다.

2.3. 순열 검정

MDR 방법에 의해 선별된 유전자 조합과 유전자형이 통계적으로 유의성을 가지는지 알아보기 위해 t -검정과 순열 검정 (Good, 2000)의 p 값을 사용한다. 본 논문에서는 10,000번을 시행하여 다음의 절차에 따라 순열 검정을 하였다.

- 단계 1. 검정하고자 하는 가설 설정 - 각 방법에 의해 선별된 우수 유전자 조합의 유전형 이 한우의 경제형질에 유의한 영향을 미친다.
- 단계 2. 통계량과 기각역 설정 - 분석에 사용할 통계량으로 높은 그룹과 낮은 그룹으로 이분화했을 때 높은 그룹의 평균을 선택한다. 선정된 높은 그룹의 평균이 그룹 간 데이터를 서로 바꾸었을 때보다 높다면 그 그룹이 경제형질에 영향력이 있다고 판단한다.
- 단계 3. 기존 관측치의 통계량 계산 - 선별된 높은 그룹의 평균을 계산한다.
- 단계 4. 관측치의 재배열과 재배열 후의 통계량 계산 - 두 그룹의 데이터를 n 개만큼 랜덤 추출하여 그룹을 상호 변경한 후 각 높은 그룹의 평균을 구한다. 이 과정을 10,000번 반복한다.
- 단계 5. 결론 - 각 평균을 내림차순으로 정렬한 후 p 값을 구한다.

위의 순열 검정을 통해 우수 유전자 조합에서 발견한 우수 유전자형이 한우의 경제형질에 미치는 영향은 4절에서 살펴본다.

3. MDR 방법 적용 및 결과

3.1. 실험자료

이 데이터는 경북지역에서 자란 18 아비로부터 얻어진 513두의 한우로 구성되었다. 한우의 맛에 영향을 준다고 알려져 있는 올레인산 (oleic acid; C18:1)과 단일불포화지방산 (monounsaturated fatty acid; MUFA), 한우의 육질에 영향을 준다고 알려진 근내지방도 (marbling score; MS)를 분석에 사용 (Lee 등, 2011)하였다. 환경적인 요인을 보정한 각 경제형질의 평균과 표준편차를 표 3.1

에 나타내었다. 단일염기다형성 (single nucleotide polymorphism; SNP)은 최근 한우의 맛과 육질에 영향을 준다고 밝혀진 지방산합성효소 (fatty acid synthase; FASN)의 g.12870T>C, g.13126T>C, g.15532C>A, g.16907T>C, g.17921G>A로 5개 SNP를 본 연구에 이용하였다 (Oh 등, 2011).

표 3.1 환경적인 요인을 보정한 경제형질의 평균과 표준편차

경제형질	내용	N	평균	표준편차
C18:1	올레인산	513	44.30	2.12
MUFA	단일불포화지방산	513	53.50	2.34
MS	근내지방도	513	5.43	1.42

3.2. MDR 방법을 활용한 유전자 조합 선별 결과

다음은 한우 데이터에 MDR방법을 적용시킨 결과이다. MDR과정에서 필요한 그룹 변수로는 한우의 경제형질인 올레인산, 단일불포화지방산, 근내지방도를 각각 고려한 값으로써 데이터 마이닝 기법 중의 하나인 K-평균 알고리즘을 이용하여 이분화한 값을 사용하였다. 이분화한 데이터의 평균과 표준편차는 표 3.2에 나타내었다. 표 3.3은 각 경제형질별로 단일SNP와 SNP조합의 학습용 자료의 오분류율과 검증용 자료의 오분류율을 나타낸 것이다. 가장 낮은 학습용 자료의 오분류율과 검증용 자료의 오분류율을 가지는 SNP를 진하게 나타내었으며 그 결과는 다음과 같다. 먼저 단일 SNP의 결과를 살펴보면 올레인산에서 가장 낮은 학습용 자료의 오분류율과 검증용 자료의 오분류율을 가지는 SNP로 g.13126T>C가 선별되었다. 단일불포화지방산에서 가장 낮은 값을 가지는 SNP도 g.13126T>C가 선별되었으며 근내지방도에서는 g.15532C>A가 가장 낮은 값을 가지는 것으로 나타났다. 다음으로 SNP조합의 결과를 살펴보면 올레인산에서는 (g.13126T>C, g.15532C>A) 조합이 가장 낮은 학습용 자료의 오분류율과 검증용 자료의 오분류율을 가지는 것으로 나타났다. 단일불포화지방산에서도 (g.13126T>C, g.15532C>A) 조합이 가장 우수한 조합으로 선별되었으며, 근내지방도에서는 (g.12870T>C, g.15532C>A) 조합이 선별되었다. 그리고 근내지방도에서 가장 낮은 학습용 자료의 오분류율과 검증용 자료의 오분류율을 가지는 (g.12870T>C, g.15532C>A) 조합은 올레인산과 단일불포화지방산에서 두 번째로 낮은 학습용 자료의 오분류율과 검증용 자료의 오분류율을 가지는 조합으로 나타났으며, 올레인산과 단일불포화지방산에서 가장 낮은 학습용 자료의 오분류율과 검증용 자료의 오분류율을 가지는 (g.13126T>C, g.15532C>A) 조합은 근내지방도에서 두 번째로 낮은 값을 가지는 것으로 나타났다. 따라서 (g.12870T>C, g.15532C>A) 조합과 (g.13126T>C, g.15532C>A) 조합은 한우의 맛과 육질에 영향을 주는 상위 2개의 우수 SNP 조합이라고 할 수 있다. 게다가 단일 SNP에서의 학습용 자료의 오분류율과 검증용 자료의 오분류율값보다 SNP조합에서의 그 값이 더 낮은 것으로 보아 SNP조합의 효과가 단일 유전자의 효과보다 한우의 경제형질에 더 많은 영향을 미친다는 사실을 알 수 있다.

표 3.2 이분화한 데이터의 평균과 표준편차

경제형질	그룹	N	평균	표준편차
C18:1	1	153	46.70	1.31
	2	360	43.28	1.49
MUFA	1	199	55.75	1.49
	2	314	52.08	1.52
MS	1	272	6.51	0.83
	2	241	4.21	0.84

표 3.3 단일 SNP와 SNP조합의 학습용 자료의 오분류율과 검증용 자료의 오분류율

SNP	C18:1		MUFA		MS	
	학습용자료의 오분류율	검증용자료의 오분류율	학습용자료의 오분류율	검증용자료의 오분류율	학습용자료의 오분류율	검증용자료의 오분류율
g.12870T>C	0.3582	0.3583	0.3635	0.3637	0.4302	0.4305
g.13126T>C	0.3419	0.3419	0.3469	0.3469	0.4436	0.4438
g.15532C>A	0.4087	0.4092	0.4308	0.4432	0.4115	0.4117
g.16907T>C	0.4128	0.4128	0.4217	0.4221	0.4570	0.4751
g.17921G>A	0.4198	0.4198	0.4297	0.4301	0.4593	0.4848
(g.12870T>C, g.13126T>C)	0.3382	0.3771	0.3458	0.3485	0.4250	0.4302
(g.12870T>C, g.15532C>A)	0.3145	0.3145	0.3307	0.3308	0.3600	0.3603
(g.12870T>C, g.16907T>C)	0.3567	0.3874	0.3594	0.3892	0.4119	0.4370
(g.12870T>C, g.17921G>A)	0.3558	0.3874	0.3617	0.3786	0.4121	0.4319
(g.13126T>C, g.15532C>A)	0.2981	0.2982	0.3141	0.3141	0.3657	0.3660
(g.13126T>C, g.16907T>C)	0.3400	0.3725	0.3455	0.3639	0.4230	0.4454
(g.13126T>C, g.17921G>A)	0.3401	0.3541	0.3446	0.3587	0.4233	0.4459
(g.15532C>A, g.16907T>C)	0.3986	0.4091	0.4205	0.4512	0.4111	0.4309
(g.15532C>A, g.17921G>A)	0.4043	0.4352	0.4250	0.4690	0.4113	0.4203
(g.16907T>C, g.17921G>A)	0.4108	0.4142	0.4198	0.4698	0.4502	0.4792

4. 우수 유전자형 선별 결과

표 4.1 우수 유전자형 선별과 순열검정 적용 및 결과

경제형질	SNP조합	우수 유전자형	N	평균±표준편차	t-검정 (p값)	순열검정 (p값)
C18:1	g.13126T>C	TT	141	45.58±1.86	<0.001	<0.001
		Others	372	43.82±2.02		
MUFA	g.13126T>C	TT	141	54.84±2.06	<0.001	<0.001
		Others	372	53.00±2.24		
MS	g.15532C>A	CA, AA	344	5.59±1.42	<0.001	<0.001
		Others	169	5.10±1.37		
C18:1	(g.13126T>C, g.15532C>A)	TTCA, TTAA, TCAA	75	46.23±2.04	<0.001	<0.001
		Others	438	43.97±1.96		
		CCCA, TCAA, CCAA	71	46.11±2.07		
C18:1	(g.12870T>C, g.15532C>A)	Others	442	44.01±1.99	<0.001	<0.001
		TTCC, TTCA, TTAA, TCAA, CCAA	190	54.74±2.04		
		Others	323	52.78±2.12		
MUFA	(g.13126T>C, g.15532C>A)	CCCC, CCCA, TTAA, TCAA, CCAA	185	54.68±2.03	<0.001	<0.001
		Others	328	52.84±2.24		
		CCCC, TCCA, CCCA, TTAA, TCAA, CCAA	338	5.73±1.34		
MS	(g.12870T>C, g.15532C>A)	Others	175	4.87±1.41	<0.001	<0.001
		TTCC, TTCA, TCCA, TTAA, TCAA, CCAA	337	5.71±1.35		
		Others	176	4.89±1.40		

3절에서는 각 경제형질별로 경제형질에 영향을 주는 우수한 단일 SNP와 SNP조합을 살펴보았다. 4절에서는 각 경제형질별로 선별된 우수 SNP내의 어떤 유전자형이 경제형질의 가치를 높이는지를 규명하기 위해 우수 유전자형을 선별하고 순열 검정을 실시하여 선별된 유전자형의 통계적 유의성을 살펴본다. 표4.1은 선별된 유전자형이 통계적으로 유의한지를 알아보기 위해 t-검정과 순열검정을 실시한 결과이다. 먼저 단일 SNP의 결과부터 살펴보면 올레인산과 단일불포화지방산에 영향을 주는 것으로 나타난 g.13126T>C에서는 TT가 우수 유전자형으로 선별되었고, 근내지방도에 영향을 주는 g.15532C>A에서는 CA와 AA가 우수 유전자형으로 선별되었다. 이들 선별된 유전자형은 t-검정과 순열검정 결과에서 p값이 0.0001에 근접하여 경제형질에 아주 유의한 영향을 미치는 것으로 나타났다. 올레인산에 영향을 주는 것으로 선별된 (g.13126T>C, g.15532C>A)조합에서는 평균이 46.23인 TTCA, TTAA, TCAA 유전자형이, (g.12870T>C, g.15532C>A)조합에서는 평균이 46.11인 CCCA, TCAA, CCAA 유전자형이 우수 유전자형으로 선별되었다. 단일불포화지방산에 영향을 주는

(g.13126T>C, g.15532C>A)조합에서는 평균이 54.74인 TTCC, TTCA, TTAA, TCAA, CCAA가, (g.12870T>C, g.15532C>A)조합에서는 평균이 54.68인 CCCC, CCCA, TTAA, TCAA, CCAA가 우수 유전자형으로 선별되었다. 근내지방도에서는 (g.12870T>C, g.15532C>A)조합에서 평균이 5.73인 CCCC, TCCA, CCCA, TTAA, TCAA, CCAA와 (g.13126T>C, g.15532C>A)조합에서 평균이 5.71인 TTCC, TTCA, TCCA, TTAA, TCAA, CCAA가 각각 우수 유전자형으로 선별되었다. 이들 선별된 우수 유전자형들 역시 t-검정과 순열검정 결과에서 p값이 거의 0.0001로써 한우의 종합적인 경제형질에 모두 유의한 영향을 주는 유전자형임을 확인하였다. 최종적으로 세 경제형질에서 공통으로 나타난 유전자형을 진하게 나타내었으며 그 결과는 다음과 같다. (g.13126T>C, g.15532C>A)조합에서는 TTCA, TTAA, TCAA가 공통 우수 유전자형으로 선별되었고, (g.12870T>C, g.15532C>A)조합에서는 CCCA, TCAA, CCAA가 공통 우수 유전자형으로 선별되었다. 따라서 선별된 이들 유전자형은 한우의 맛과 질을 높이는 유전자형이라 할 수 있다. 그리고 단일불포화지방산내에서 선별된 단일 SNP의 우수 유전자형을 가지는 그룹의 평균 54.84와 SNP조합에서의 선별된 우수 유전자형을 가지는 그룹의 평균 54.74, 54.68에서 뚜렷한 차이가 나타나지 않았지만 올레인산과 근내지방도의 단일 SNP에서 선별된 우수 유전자형을 가지는 그룹과 SNP조합에서 선별된 우수 유전자형을 가지는 그룹의 평균을 비교하면 SNP조합에서의 평균이 더 높음을 알 수 있다. 이 역시 단일 유전자의 효과보다 상호작용의 효과에 더 많은 영향을 받는다는 사실을 보여주고 있다.

5. 결론 및 토의

본 연구는 고품질의 소고기를 선호하는 소비자들이 늘어남에 따라 한우의 맛과 육질에 영향을 미치는 것으로 알려진 지방산합성효소 (FASN)에서의 5개 SNP를 이용하여 한우의 경제형질에 영향을 미치는 우수 유전자를 알아보고자 하였다. 특히 환경적인 요인을 보정하여 유전적인 요인만으로 한우 경제형질에 연관된 우수 유전자를 선별하였고 그 결과 올레인산과 단일불포화지방산에 영향을 주는 우수 단일 SNP는 g.13126T>C가 선별되었고, 근내지방도에 영향을 주는 우수 단일 SNP로는 g.15532C>A가 선별되었다. SNP조합에서는 (g.13126T>C, g.15532C>A)와 (g.12870T>C, g.15532C>A)조합이 한우의 종합적인 경제형질에 영향을 미치는 것으로 나타났다. 또한 이들 우수 유전자 중에서도 경제형질을 높이는 우수 유전자형을 선별한 결과 g.13126T>C에서는 TT, g.15532C>A에서는 CA, AA가 우수 유전자형으로 선별되었고, (g.13126T>C, g.15532C>A)조합에서는 TTCA, TTAA, TCAA가, (g.12870T>C, g.15532C>A)조합에서는 CCCA, TCAA, CCAA가 세 경제형질에서 공통적인 우수 유전자형으로 선별되었다. 선별된 우수 유전자형들이 한우의 경제형질에 유의한 영향을 주는지 알아보기 위해 t-검정과 순열검정을 적용한 결과 모두 통계적으로 유의성을 가지는 것으로 확인할 수 있었다.

한편 표 3.3에 나타난 학습용 자료의 오분류율과 검증용 자료의 오분류율의 값이 절반가량 동일하거나 아주 근소하게 차이가 남을 확인할 수가 있다. 학습용 자료의 오분류율과 검증용 자료의 오분류율은 각각 학습용 자료의 정확도와 검증용 자료의 정확도를 토대로 구해지는 값이며, 본 연구에서는 정확도를 실험군과 대조군의 수가 균등하지 않아 편향된 결과를 초래할 문제를 방지하기 위하여 균등화된 정확도 (Velez 등, 2007)를 사용하였다. 이 균등화된 정확도를 계산하는 과정에서 학습용 자료의 정확도와 검증용 자료의 정확도가 같게 나타나거나 근소한 차이를 보여 학습용 자료의 오분류율과 검증용 자료의 오분류율의 값도 같아지거나 근소한 차이가 나타난 것으로 생각된다. 실제 이러한 문제가 균등화된 정확도를 이용함으로써 생겨난 결과인지에 대한 좀 더 깊이 있는 연구가 필요한 것으로 생각된다.

참고문헌

- Casas, E., White, S. N., Riley, D. G., Smith, T. P. L., Breneman, R. A., Olson, T.A., Johnson, D. D., Coleman, S. W., Bennett, G. L. and Chase, C. C. (2005). Assessment of single nucleotide polymorphisms in genes residing on chromosomes 14 and 29 for association with carcass composition traits in *Bos indicus* cattle. *Journal of Animal Science*, **83**, 13-19.
- Good, P. (2000). *Permutation test : A practical guide to resampling methods for testing hypotheses*, Springer-Verlag, New York.
- Lee, J. Y., Kwon, J. C. and Kim, J. J. (2008). Multifactor dimensionality reduction(MDR) analysis to detect single nucleotide polymorphisms associated with a carcass trait in a Hanwoo population. *Asian-Australasian Journal of Animal Sciences*, **21**, 784-788.
- Lee, J. Y. and Lee, J. H. (2010). Support vector machine and multifactor dimensionality reduction for detecting major gene interactions of continuous data. *Journal of the Korean Data & Information Science Society*, **21**, 1271-1280.
- Lee, J. Y., Lee, J. H. and Lee, Y. W. (2010). Detection of major genotype combination by genotype matrix. *Journal of the Korean Data & Information Science Society*, **21**, 387-395.
- Lee, Y. S., Oh, D. Y. and Yeo, J. S. (2011). Study on identification of candidate DNA marker related with beef quality in QTL region of BTA 2 in Hanwoo population. *Journal of the Korean Data & Information Science Society*, **22**, 661-669.
- Matsuhashi. T., Maruyama. S., Uemoto. Y., Kobayashi. N., Mannen. H., Abe. T., Sakaguchi. S. and Kobayashi. E. (2011). Effects of bovine fatty acid synthase, stearyl-coenzyme A desaturase, sterol regulatory element-binding protein 1, and growth hormone gene polymorphisms on fatty acid composition and carcass traits in Japanese Black cattle. *Journal of Animal Science*, **89**, 12-22.
- Oh, D. Y., Lee, Y. S., La, B. M., Yeo, J. S., Chung, E. Y., Kim, Y. Y. and Lee, C. Y. (2011). Fatty acid composition of beef is associated with exonic nucleotide variants of the gene encoding FASN. *Molecular Biology Reports*, **39**, 4083-4090.
- 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. *American Journal of Human Genetics*, **69**, 138-147.
- Velez, D. R., White, B. C., Motsinger, A. A., Bush, W. S., Ritchie, M. D., Williams, S. M. and Moore, J. H. (2007). A balanced accuracy function for epistasis modeling in imbalanced datasets using multifactor dimensionality reduction. *Genetic Epidemiology*, **31**, 306-315.

Major gene interaction identification in Hanwoo by adjusted environmental effects

Jea-Young Lee¹ · Mi-Hyun Jin²

¹²Department of Statistics, Yeungnam University

Received 10 April 2012, revised 24 April 2012, accepted 30 April 2012

Abstract

Human diseases and livestock economic traits are not typically the result of variation of a single genetic locus, but are rather the result of interplay between interactions among multiple genes and a variety of environmental exposures. We have used linear regression model for adjusted environmental effects and multifactor dimensionality reduction (MDR) method to identify gene-gene interaction effect of statistical model in general. Of course, we use 5 SNPs (single nucleotide polymorphism) which were studied recently by Oh et al. (2011). We apply the MDR (multifactor dimensionality reduction) method on the identify major interaction effects of single nucleotide polymorphisms responsible for economic traits in a Korean cattle population.

Keywords: Gene interaction, genotype, multifactor dimensionality reduction, single nucleotide polymorphism.

¹ Corresponding author: Professor, Department of Statistics, Yeungnam University, Kyungsan 712-749, Korea. E-mail: jlee@yu.ac.kr

² Graduate student, Department of Statistics, Yeungnam University, Kyungsan 712-749, Korea.