

확장된 다중인자 차원축소(E-MDR) 알고리즘에 기반한 유전자 상호작용 효과 규명

이제영¹ · 이호근² · 이용원³

¹영남대학교 통계학과, ²영남대학교 통계학과, ³영남대학교 통계학과

(2009년 6월 접수, 2009년 10월 채택)

요약

인간의 질병 또는 가축의 경제적인 특성에 관한 유전자의 규명은 매우 중요한 관심사이며, 우리나라 축산업을 대표하는 한우의 유전자원 보존과 능력향상은 매우 중요한 과제이다. 이를 연구하기 위해 기존 EST-based SNP 연관 지도를 사용하여 발굴한 유전자로 연구되어왔으나 이는 통계학적 모델에 기반한 연관지도 작성법으로 실제 위치와는 차이가 있을 수 있다. 따라서 Lee (2009)에 의해 EST-based SNP 연관지도와 염기서열 분석으로 작성되어지는 Gene on sequence를 함께 고려하여 한우의 경제형질 연관 후보 DNA marker들이 발견되었다. 한편, 통계모형의 상호작용 효과를 고려할 때, 유전자와 같은 범주형 data에서 범주가 많을 경우 상호작용의 조합이 많아지므로 종종 모수들의 상호작용에 대한 해석과 모형을 결정하는 것이 어려울 수 있다. 그래서 비모수적인 방법으로 다중인자 차원축소방법(MDR)을 사용해왔으며, 사례-대조 데이터에만 적용가능 MDR방법을 연속형 데이터에도 적용하기 위해 CART알고리즘을 적용한 확장된 다중인자 차원축소방법(E-MDR)이 제안되었다. 본 연구에서는 새롭게 발견된 단일염기다형성(SNP)으로부터 E-MDR방법을 적용하여 한우의 경제형질(일당증체량, 근내지방도)에 영향을 주는 우수 유전자 단일염기다형성을 규명하였다.

주요어어: Expanded MDR, Single Nucleotide Polymorphism, 근내지방도, 일당증체량, 한우 경제형질.

1. 서론

인간의 질병 또는 가축의 경제적인 특성에 관한 유전자의 규명은 유전학에서 매우 중요한 관심사이다. 또한 우리나라 축산업과 축종을 대표하는 한우의 유전자원 보존과 유전능력 향상을 위한 기술 개발은 우리나라 축산업 발전을 위해 매우 중요한 과제라 할 수 있다. 이를 위한 많은 연구가 진행되어 왔으며, 그 중 Lee 등 (2008b)은 유전자의 통계학적 모델에 기반한 EST-based SNP 연관지도 (Snelling 등, 2005)를 사용하여 발굴한 유전자와 한우의 경제형질과의 관계를 연구하였다. 하지만 연관지도작성의 경우 통계학적 모델로서 두 marker들간의 재조합율을 바탕으로 작성되어지기 때문에 염기서열 분석으로 작성되어지는 Gene on sequence와 서로 다른 차이가 나타날 수 있을 것이다 (Lee, 2009). 따라서 Lee (2009)은 기존의 통계학적 모델에 기반한 EST-based SNP 연관지도와 물리적 위치에 기반한 Gene on sequence를 함께 고려하여 새로운 후보 유전자들을 발견하였다. 또한 일반적으로 인간의 질병과 가축의 경제적인 특성은 하나의 유전자가 아닌 여러 유전자의 상호작용으로 일어난다고 믿고 있으며, 현재까지 소에서는 도체형질(일당증체량, 근내지방도)과 연관이 있는 SNP marker들이 일반가축에서 평가되어지거나 적용되고 있다 (Barendse 등, 2004; Page 등, 2004). 따라서 본 연구에서는 이들에

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

의해 새롭게 발견된 SNPs들의 상호작용에 의한 경제형질(일당증체량, 근내지방도)과의 연관성을 분석하였다.

한편, 통계모형의 상호작용을 고려한 모형으로 선형모형 같은 표준 통계적 모형을 사용해왔으나 유전자의 상호작용과 같이 범주형으로 된 변수의 경우 유전자의 수가 많거나 범주가 많을 경우 상호작용의 조합이 많아지므로 종종 모수들의 상호작용에 대한 해석과 모형을 결정하는 것이 어려울 수 있다. 그래서 다중인자 차원 축소(multifactor dimensionality reduction; MDR)방법 (Ritchie 등, 2001; Chung 등, 2005), 조합 분할(combinatorial partition method; CPM)방법 (Nelson 등, 2001), 제한된 분할(restricted partition method; RPM)방법 (Culverhouse 등, 2004) 등이 여러 유전자에 대한 상호작용을 결정하는 방법들로 개발되었다. 특히 MDR방법은 상호작용에 대한 명확한 모형의 가정이 없는 비모수적인 방법으로 적당한 high-order 차수의 data로 복잡한 관계를 밝힐 수 있었다. 하지만 MDR방법은 case와 control의 비율을 통해 독립변수의 범주를 High집단과 Low집단으로 분류한 후 목표변수에 대한 오분류율을 비교하여 분석하는 방법으로 연속형 data에 대해 적용하는데 문제점이 발생한다. 그래서 CART(classification and regression tree)알고리즘을 활용하는 방법인 확장된 다중인자 차원 축소(expanded multifactor dimensionality reduction; E-MDR)방법이 Lee 등 (2008a)에 의해 제시되었다.

본 연구에서는 Lee (2009)에 의해 소개된 한우의 경제형질을 조절하는 새로운 후보 DNA marker를 선정하는 과정을 소개하고, 선정된 DNA marker들을 E-MDR방법에 적용하여 한우의 경제형질에서 육질에 해당하는 근내지방도(marbling score; MS)와 육량에 해당하는 일당증체량(average daily gain; ADG)에 영향을 주는 우수 유전자 단일염기다형성(SNPs)를 규명하고자 한다.

2. 경제형질을 조절하는 후보 DNA marker 발굴

Lee (2009)에 의해 소개된 한우의 수많은 DNA marker들 중 경제형질을 조절하는 후보 DNA marker를 선정하는 과정을 소개한다. 연구에 사용된 한우는 농협중앙회 한우개량사업소의 후대검정집단인 30차에서 35차 국가 후대검정우 집단 476두로 구성되어졌다. 국가 후대검정우 집단의 개체들은 국가 씨수소 선발검정에서 당대검정으로 선발된 50두의 후보씨수소를 한우 암소개량농가에서 교배하여 생산된 수송아지이다.

2.1. 경제형질 연관 후보 DNA marker 선정

한우 6번 염색체에서 Kim 등 (2003)이 보고한 3개의 QTL 중 BMS1242와 ILST035는 일당증체량, 등지방두께, 등심단면적과 근내지방도에서 LOD(logarithm of odds)값이 3.0이상으로 나타났으며, BM4311은 근내지방도에서 LOD값이 3.0이상으로 나타났다 (Kim 등, 2003). 따라서 경제형질에 연관되는 유전자로 판단되는 QTL 양쪽으로 10cM 정도의 microsatellite를 선정하여 영역으로 지정하면, 육질과 육량에 대한 QTL과 연관된 유전자가 존재할 것으로 판단하였다. 그에 따라 3개의 QTL에서 양쪽으로 10cM정도 떨어진 microsatellite를 선정하여 그림 2.1과 같이 2개의 QTL영역으로 지정하였다. 이러한 2개의 QTL 영역에서 EST-based SNP 연관지도 (Snelling 등, 2005)에서 총 33개의 SNPs를 확인하였다.

발굴된 33개 SNPs 중 후보 유전자로 판단되어지는 LOC534614 유전자내의 SNPs 20개를 발견하였고 (그림 2.1), 이 중 대립유전자의 빈도가 0.1 미만이거나 유전자형의 빈도가 치우친 SNP인 g.934425+29T, g.34425+19T>C, g.-8606+137C>T를 제외하고 분석하였다. 이들을 제외한 17개의 SNPs들 간에 강한 연관불평형을 구성하고 있다는 것은 유전자가 하나의 변이로 작용하는 것이 아니

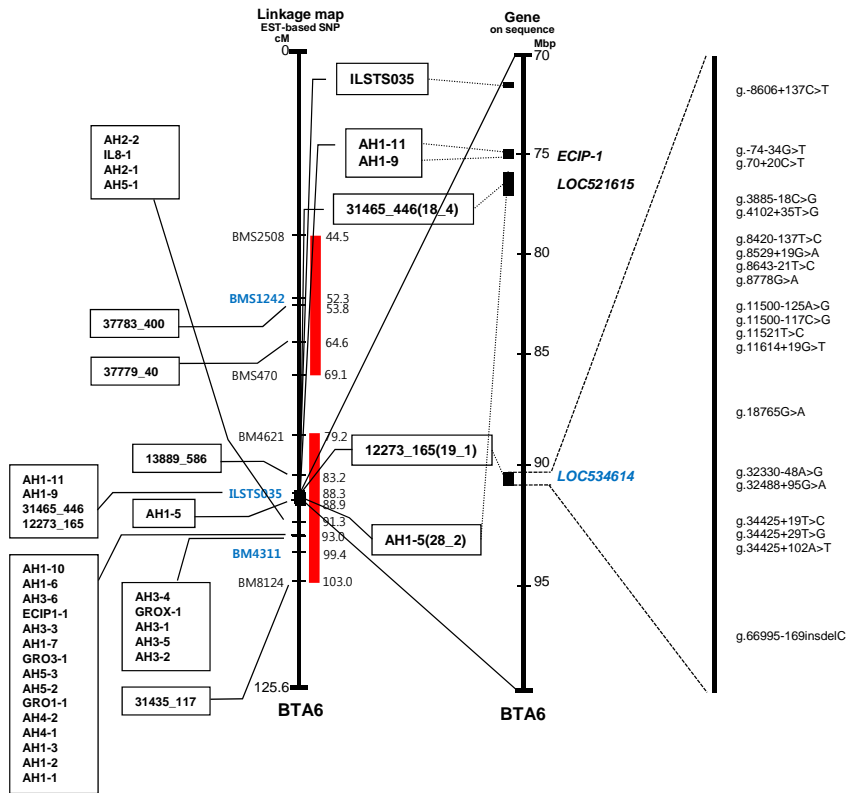


그림 2.1. Linkage map과 Gene on sequence에 기반한 SNPs의 발굴과정 (Lee, 2009)

라 여러개의 변이가 서로 조합되어 작용함을 의미한다. 또한 17개의 SNPs들의 상호작용을 모두 분석하는 것보다 htSNP(haplotype-tagging SNP)을 사용하여 분석하면 적은 수의 분석으로도 영향력을 확인할 수 있음을 나타내었으며, 그에 따라 6개의 SNPs(g.4102+36T>G, g.8778G>A, g.11500-117C>G, g.32330-48A>G, g.34425+102A>T, g.66995-169insdelC)를 확인 하였다. 이렇게 최종 선택된 6개의 SNPs (Lee, 2009)를 분석에 사용하였으며, 3장에서는 SNPs들의 경제형질에 대한 상호작용 효과를 분석하기 위한 E-MDR방법을 소개한다.

3. 분석 방법

3.1. MDR방법

MDR방법은 일반화된 선형 모형인 전통적인 통계 기법과는 달리 어떤 모수에 대한 추정과 genetic 모형의 가정을 요구하지 않는다. 즉, 상호작용에 대한 명확한 모형의 가정이 없는 비모수적인 방법으로 case와 control의 비율을 통해 독립변수의 범주를 High집단과 Low집단으로 분류한 후 목표변수에 대한 오분류율을 비교하여 분석하는 방법이다. 하지만 목표변수가 case와 control로 이루어진 데이터에 적용하는 방법으로 연속형 데이터에 대해 적용하는데 문제점이 발생한다. 3.2절에서 이 문제를 해결하기 위한 E-MDR방법과 과정에 대해 서술한다. 또한 데이터의 수는 419두로 SNP조합 marker의 각 수준이

최대 27개의 범주임을 감안할 때 각 cell의 빈도가 매우 낮거나 0이 될 가능성이 있다. 이를 해결하기 위해 Bootstrap sampling 방법 (Efron과 Tibshirani, 1993)을 사용하였다.

3.2. E-MDR방법

E-MDR방법은 case-control 데이터에만 적용 가능한 MDR방법의 문제점을 해결하기 위해 제안된 방법으로 CART알고리즘을 이용한 MDR방법의 확장이다 (Lee 등, 2008a). 3.1절에서 소개한 MDR과정은 다중범주를 high집단과 low집단으로 이분화하기 위해 목표변수가 case-control 형태가 되어야 한다. 따라서 목표변수가 연속형 데이터인 경우 이분화가 불가능하게 된다. E-MDR방법에서는 절차 3과 절차 4의 절차를 CART알고리즘을 통한 이분화 방법을 적용함으로써 연속형 자료인 경우에도 MDR방법을 적용할 수 있도록 제시한 기법이다. 다음의 그림 3.1과 절차들은 연속형 데이터에 대한 E-MDR방법의 분석과정을 보여준다. 또한 아래의 절차들을 SAS 9.13과 SAS Enterprise Miner 4.3프로그램을 이용하여 분석하였다.

- 절차 1. 데이터를 랜덤으로 10개의 같은 크기로 나눈다. 그리고 그중 9개를 training set으로 1개를 testing set으로 둔다.
- 절차 2. 모든(3개) SNP로부터 $k(1, 2, 3)$ 개의 SNP조합 중 하나를 선택한다.
- 절차 3. 선택된 SNP조합에서 SNP의 각각 수준을 기초로 한 개체들을 multifactor classes 또는 cells에 기술한다. 예를 들어서 $k = 2$ 일 경우 SNP는 3개의 수준으로 되어있다. 따라서 $3^2 = 9$ 개의 셀을 가진다. 각각 9개의 셀에 경제형질의 평균값을 기술한다.
- 절차 4. CART방법의 불순도 함수(impurity function)를 사용하여 이분화 한다. 모형에서 입력변수로는 상호작용을 나타내는 각 유전자형을 사용하며, 출력변수로는 경제형질을 사용한다. 분류 결과로 나온 그룹 중 평균이 높은 그룹이 high가 되고 평균이 낮은 그룹이 low가 된다.
- 절차 5. expanded MDR 모델에서 k -factor 조합에 대한 training set에서 식 3.1과 같이 ASE(average squared error; Lee 등 (2008a))를 구하여 ASE의 값이 가장 작은 SNP 조합 마커를 선정한다. 절차 6. MDR과 마찬가지로 나머지 1/10의 데이터인 testing set을 이용하여 P-ASE(prediction average squared error)를 구한다.

$$ASE = \frac{S_{high}}{N_{high}} + \frac{S_{low}}{N_{low}} \quad (3.1)$$

단, $S_{high} = \sum_{i=1}^n I_{high}(i) \sum_{j=1}^{N_i} (y_{ij} - \hat{y}_i)^2$, $S_{low} = \sum_{i=1}^n I_{low}(i) \sum_{j=1}^{N_i} (y_{ij} - \hat{y}_i)^2$,

$$N_{high} = \sum_{i=1}^n I_{high}(i) \cdot N_i, \quad N_{low} = \sum_{i=1}^n I_{low}(i) \cdot N_i,$$

$$I_{high}(i) = \begin{cases} 1, & i(\text{cell}) \in \text{high group}, \\ 0, & \text{o.w.}, \end{cases} \quad I_{low}(i) = \begin{cases} 1, & i(\text{cell}) \in \text{low group}, \\ 0, & \text{o.w.}, \end{cases}$$

(n : 범주의 수, N_i : i 범주의 관측치 수)

그 다음 위의 과정의 반복에서 나온 10개의 ASE와 P-ASE의 평균을 구해 그 값이 가장 낮은 것을 best n -factors 모형으로 정한다. MDR방법과 같이 앞에서 구한 각각의 ASE를 이용하여 CVC(cross validation consistency)를 구하는데 이것은 10번의 cross-validation을 시행할 때 각 시행에서 선택된 best model을 카운트 하는 것이다. 따라서 ASE와 P-ASE의 평균이 가장 낮고 CVC가 가장 높은 값이 best n -factors 모형이다.

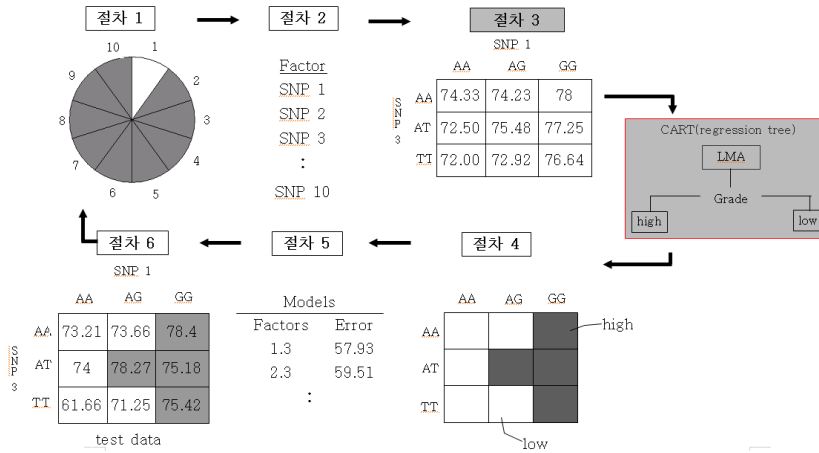


그림 3.1. Expanded MDR방법의 일반적인 절차 (Lee 등, 2008b)

또한 위의 E-MDR 절차를 통해 얻은 결과를 검증하기 위해 본 논문에서는 순열검정 (Good, 2000)을 Lee와 Lee (2009)에서 사용된 절차에 따라 시행하였다. 4장에서는 3장에서 살펴본 방법들을 적용한 결과를 나타낸다.

4. E-MDR방법의 적용 및 순열검정 결과

4.1. E-MDR방법의 적용 결과

일당증체량, 근내지방도에 대해 SNP의 조합에 대하여 E-MDR 과정을 10번 반복해서 나온 ASE와 P-ASE의 평균과 10번의 반복과정에서 나온 값을 기준으로 ASE와 P-ASE의 값이 가장 낮은 조합의 CVC를 구한 결과를 각각 표 4.1, 4.2와 같이 나타냈다.

표 4.1과 4.2를 통해 일당증체량과 근내지방도에서 2개의 SNPs 조합에 의한 결과는 g.4102+36T>G, g.11500-117C>G의 조합이 가장 좋은 것으로 나타났으며, 3개의 SNPs 조합에서는 g.4102+36T>G, g.11500-117C>G, g.32330-48A>G의 조합이 가장 좋은 것으로 나타났다.

4.2. 순열검정 결과

일당증체량, 근내지방도에 대해 각 선택된 모형의 유의성을 확인하기 위하여 순열검정('Good (2000)'이 제시한 절차를 통해)을 실시하여 표 4.3과 같이 나타냈다.

일당증체량의 경우 하나의 요인에 의한 효과인 g.32330-48A>G가 유의하지 않은 것으로 나타났으며, 2개의 요인에 의한 효과인 g.4102+36T>G, g.11500-117C>G의 조합과 3개의 요인에 의한 효과인 g.4102+36T>G, g.11500-117C>G, g.32330-48A>G의 조합이 유의하게 나타났으며, 근내지방도에서는 모든 형태에서 유의하게 나타나며, 일당증체량에서는 2개의 상호작용에 의한 차이가 가장 유의하게 나타났고, 근내지방도에서는 3개의 상호작용에 의한 차이가 가장 유의하게 나타났다. 따라서 한우의 경제형질인 일당증체량과 근내지방도에 영향을 주는 요인은 하나의 SNP에 대한 효과보다는 SNP의 조합에 의한 효과가 더 많다는 것을 검증하였으며, 많은 SNPs들의 조합 중 한우의 육질과 육량에 가장 많은 영향을 주는 SNPs 조합이 비록 2개의 상호작용에 의한 효과와 3개의 상호작용에 의한 효과로 다르게 나타났으나, 공통적으로 g.4102+36T>G, g.11500-117C>G를 포함하고 있으며, 근내지방도의 경

표 4.1. E-MDR방법에 의한 average ASE와 average P-ASE 결과

요인의 수	SNP 마커	일당증체량		근내지방도	
		ASE	P_ASE	ASE	P_ASE
1	g.4102+36T>G	0.008012	0.008024	16.5635	16.5796
	g.8778G>A	0.007998	0.008014	16.6286	16.6776
	g.11500-117C>G	0.008003	0.008017	16.5828	16.6066
	g.32330-48A>G	0.007958	0.007970	16.3377	16.3539
	g.34425+102A>T	0.008053	0.008067	16.6746	16.7223
	g.66995-169insdelC	0.008050	0.008062	16.6750	16.7203
2	g.4102+36T>G, g.8778G>A	0.007989	0.007903	16.5569	16.7062
	g.4102+36T>G, g.11500-117C>G	0.007796	0.007696	16.1373	16.3072
	g.4102+36T>G, g.32330-48A>G	0.007950	0.007878	16.3377	16.5118
	g.4102+36T>G, g.34425+102A>T	0.007978	0.007866	16.5635	16.7046
	g.4102+36T>G, g.66995-169insdelC	0.007983	0.007921	16.4827	16.6380
	g.8778G>A, g.11500-117C>G	0.007846	0.007747	16.2162	16.4331
	g.8778G>A, g.32330-48A>G	0.007958	0.007860	16.3377	16.5118
	g.8778G>A, g.34425+102A>T	0.007994	0.007929	16.5369	16.7091
	g.8778G>A, g.66995-169insdelC	0.007907	0.007801	16.6045	16.7864
	g.11500-117C>G, g.32330-48A>G	0.007852	0.007768	16.1622	16.3315
	g.11500-117C>G, g.34425+102A>T	0.007909	0.007819	16.2794	16.4872
	g.11500-117C>G, g.66995-169insdelC	0.007922	0.007821	16.3587	16.5240
	g.32330-48A>G, g.34425+102A>T	0.007958	0.007860	16.3377	16.5118
	g.32330-48A>G, g.34425+102A>T	0.007958	0.007860	16.3377	16.5118
g.34425+102A>T, g.66995-169insdelC	0.008043	0.007967	16.6284	16.8587	
3	g.4102+36T>G, g.8778G>A, g.11500-117C>G	0.007780	0.007679	16.1100	16.4079
	g.4102+36T>G, g.8778G>A, g.32330-48A>G	0.007887	0.007794	16.2519	16.3922
	g.4102+36T>G, g.8778G>A, g.34425+102A>T	0.007938	0.007858	16.4822	16.6888
	g.4102+36T>G, g.8778G>A, g.66995-169insdelC	0.007845	0.007746	16.4517	16.6007
	g.4102+36T>G, g.11500-117C>G, g.32330-48A>G	0.007755	0.007621	15.9860	16.1578
	g.4102+36T>G, g.11500-117C>G, g.34425+102A>T	0.007772	0.007700	16.0071	16.2227
	g.4102+36T>G, g.11500-117C>G, g.66995-169insdelC	0.007770	0.007685	16.0934	16.2656
	g.4102+36T>G, g.32330-48A>G, g.34425+102A>T	0.007944	0.007874	16.2741	16.4589
	g.4102+36T>G, g.32330-48A>G, g.66995-169insdelC	0.007876	0.007786	16.3357	16.5275
	g.4102+36T>G, g.34425+102A>T, g.66995-169insdelC	0.007929	0.007862	16.4677	16.6716
	g.8778G>A, g.11500-117C>G, g.32330-48A>G	0.007805	0.007745	16.0938	16.3804
	g.8778G>A, g.11500-117C>G, g.34425+102A>T	0.007837	0.007759	16.1596	16.3950
	g.8778G>A, g.11500-117C>G, g.66995-169insdelC	0.007811	0.007742	16.1677	16.3666
	g.8778G>A, g.32330-48A>G, g.34425+102A>T	0.007936	0.007867	16.2606	16.4510
	g.8778G>A, g.32330-48A>G, g.66995-169insdelC	0.007891	0.007779	16.3416	16.5314
	g.8778G>A, g.34425+102A>T, g.66995-169insdelC	0.007905	0.007829	16.4993	16.7549
	g.11500-117C>G, g.32330-48A>G, g.34425+102A>T	0.007837	0.007815	16.1210	16.3755
	g.11500-117C>G, g.32330-48A>G, g.66995-169insdelC	0.007808	0.007709	16.0591	16.2512
	g.11500-117C>G, g.34425+102A>T, g.66995-169insdelC	0.007869	0.007821	16.2347	16.4302
	g.32330-48A>G, g.34425+102A>T, g.66995-169insdelC	0.007940	0.007869	16.3260	16.5135

우 2개의 상호작용에 의한 효과와 3개의 상호작용에 의한 효과의 차이가 크지 않음을 알 수 있다. 즉, g.4102+36T>G와 g.11500-117C>G SNP의 상호작용이 한우의 육질과 육량에 가장 많은 영향을 준다는 것을 알 수 있다.

표 4.2. E-MDR방법에 의한 CVC 결과

요인의 수	SNP 마커	일당증체량	근내지방도
1	g.32330-48A>G	4	7
2	g.4102+36T>G, g.11500-117C>G	6	7
3	g.4102+36T>G, g.11500-117C>G, g.32330-48A>G	3	5

표 4.3. 각 형질에 대해 선택된 우수 SNPs 조합에 대한 순열검정 결과

우수 SNPs 조합	순열 검정 결과 (유의확률 값)	
	일당증체량(ADG)	근내지방도(MS)
g.32330-48A>G	0.10100	0.01411
g.4102+36T>G, g.11500-117C>G	0.00194	0.00424
g.4102+36T>G, g.11500-117C>G, g.32330-48A>G	0.00737	0.00367

5. 결론 및 토의

우리는 복합질병에 대한 위험 또는 가축의 경제특성의 여러 유전자에 관련된 polymorphism의 조합의 연관성을 연구하기위해 한우의 육량(일당증체량)과 육질(근내지방도)에 관련된 후보 DNA marker들의 조합을 연속형 데이터에 적용 가능한 E-MDR방법에 적용한 결과, 육질과 육량에 영향을 주는 SNPs조합이 동일함을 발견하였으며, 단일 SNP에 의한 영향보다 상호작용인 SNPs 조합에 의한 영향이 더 크다는 것을 증명하였다. 또한 일당증체량의 경우 g.4102+36T>G와 g.11500-117C>G SNP에 의한 조합이 가장 많은 영향을 주며, 근내지방도에서는 g.4102+36T>G와 g.11500-117C>G, g.32330-48A>G SNP에 의한 조합이 가장 많은 영향을 주는 것으로 나타났다. 하지만 근내지방도에서 2개의 SNP 조합에 의한 효과의 차이와 3개의 SNP 조합에 의한 효과의 차이가 크지 않음을 순열검정을 통해 알 수 있었다. 즉, 일당증체량과 근내지방도에 공통적으로 가장 많은 영향을 주는 조합은 g.4102+36T>G와 g.11500-117C>G SNP의 조합이라 할 수 있다. 따라서 이들 SNPs 조합들이 한우의 육질과 육량모두에 영향을 주는 형태라는 것이 규명되어졌다. 그러므로 g.4102+36T>G와 g.11500-117C>G SNP 조합의 상호작용 유전 정보를 분석하여 종합적인 경제형질을 개발하게 되면 한우를 고품질, 고능력화를 시킬 수 있을 것이다. 본 연구에서는 상호작용에 관련한 분석방법으로 E-MDR방법만을 사용하였다. 상호작용을 분석하는 다른 분석방법들을 통하여 결과를 비교하는 연구가 더 필요할 것이며, 또한 차후에 모의실험을 통한 E-MDR방법의 장점, 단점 및 특성을 연구하여 이와 같은 연구에서의 검증 능력을 확인하는 연구가 필요하다고 생각된다.

참고문헌

Barendse, W., Bunch, R., Thomas, M., Armitage, S., Baud, S. and Donaldson, N. (2004). The TG5 thyroglobulin gene test for a marbling quantitative trait loci evaluated in feedlot cattle, *Australian Journal of Experimental Agriculture*, **44**, 669-674.

Chung, Y. J., Lee, S. Y. and Park, T. S. (2005). Multifactor dimensionality reduction in the presence of missing observations, *Proceedings of the Autumn Conference Korea Statistical Society*, 31-36.

Culverhouse, R., Klein, T. and Shannon, W. (2004). Detecting epistatic interactions contributing to quantitative traits, *Genetic Epidemiology*, **27**, 141-152.

Efron, B. and Tibshirani, R. (1993). *An Introduction to the Bootstrap*, Chapman & Hall/CRC, New York.

Good, P. (2000). *Permutation Test: A Practical Guide to Resampling Methods for Testing Hypotheses*, Springer, New York.

Kim, J. W., Park, S. I. and Yeo, J. S. (2003). Linkage mapping and QTL on chromosome 6 in Hanwoo(Korean Cattle), *Asian-Australasian Journal of Animal Sciences*, **16**, 1402-1405.

- Lee, Y. S. (2009). Study on the identification of candidate genes and their haplotypes that are associated with growth and carcass traits in the QTL region of BTA6 in a Hanwoo population, *Ph. D. Thesis*, 1-94.
- Lee, J. Y., Kwon, J. C. and Kim, J. J. (2008a). 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, H. G. (2009). Multifactor dimensionality reduction(MDR) analysis by dummy variables, *The Korean Journal of Applied Statistics*, **22**, 435-442.
- Lee, Y. S., Lee, J. H., Lee, J. Y., Kim, J. J., Park, H. S. and Yeo, J. S. (2008b). Identification of candidate SNP (single nucleotide polymorphism) for growth and carcass traits related to QTL on chromosome 6 in Hanwoo (Korean cattle), *Asian-Australasian Journal of Animal Sciences*, **21**, 1703-1709.
- Nelson, M. R., Kardia, S. L. R., Ferrell, R. E. and Sing, C. F. (2001). A combinatorial partitioning method to identify multilocus genotypic partitions that predict quantitative trait variation, *Genome Research*, **11**, 458-470.
- Page, B. T., Casas, E., Quaas, R. L., Thallman, R. M., Wheeler, T. L., Shackelford, S. D., Koohmaraie, M., White, S. N., Bennett, G. L., Keele, J. W., Dikeman, M. E. and Smith, T. P. L. (2004). Association of markers in the bovine CAPN1 gene with meat tenderness in large crossbred populations that sample influential industry sires, *Journal of Animal Science*, **82**, 3474-3481.
- 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.
- Snelling, W. M., Casas, E., Stone, R. T., Keele, J. W., Harhay, G. P., Benett, G. L. and Smith, T. P. L. (2005). Linkage mapping bovine EST-based SNP, *BMC Genomics*, **6**, 74-84.

Study Gene Interaction Effect Based on Expanded Multifactor Dimensionality Reduction Algorithm

Jea-Young Lee¹ · Ho-Guen Lee² · Lee Yong-Won³

¹Department of Statistics, Yeungnam University; ²Department of Statistics, Yeungnam University

³Department of Statistics, Yeungnam University

(Received June 2009; accepted October 2009)

Abstract

Study the gene about economical characteristic of human disease or domestic animal is a matter of grave interest, preserve and elevation of gene of Korea cattle is key subject. Studies have been done on the gene of Korea cattle using EST-based SNP map, but it is based on statistical model, therefore there are difference between real position and statistical position. These problems are solved using both EST-based SNP map and Gene on sequence by Lee *et al.* (2009b). We have used multifactor dimensionality reduction(MDR) method to study interaction effect of statistical model in general. But MDR method cannot be applied in all cases. It can be applied to the only case-control data. So, method is suggested E-MDR method using CART algorithm. Also we identified interaction effects of single nucleotide polymorphisms(SNPs) responsible for average daily gain(ADG) and marbling score(MS) using E-MDR method.

Keywords: Average daily gain, expanded MDR, Hanwoo economic traits, marbling score, single nucleotide polymorphism.

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