

## 유전자 행렬 맵핑을 활용한 우수 유전자형 조합 선별

이제영<sup>1</sup> · 이종형<sup>2</sup> · 이용원<sup>3</sup>

<sup>123</sup>영남대학교 통계학과

접수 2010년 2월 28일, 수정 2010년 4월 29일, 게재확정 2010년 5월 6일

### 요약

인간의 질병 및 가축의 특성치에 관한 유전자 규명은 매우 중요한 과제이다. 유전자원 보존과 유전능력향상을 위한 기술 개발 역시 매우 중요한 관심사로서 이와 관련된 많은 연구들을 진행해왔다. 통계모형의 상호작용 효과를 분석하기 위해 로지스틱 회귀분석과 같은 전통적인 방법과 비모수적인 방법들이 개발되었지만 이들은 상호작용에 영향을 준 양적형질위치들의 하위 집단의 유전자형에 대해서 고려하지 않은 문제점이 있다. 따라서 많은 양적형질위치들을 한 번에 비교하여 특성치에 영향을 주는 양적형질위치의 상호작용과 그 하위집단인 유전자형을 규명하는 방법으로 유전자형 행렬 맵핑이 개발되었다. 본 연구에서는 EST-based SNP 연관지도에 의해 선정된 17개의 후보 단일염기다형성을 대상으로 유전자 행렬 맵핑을 활용하여 한우의 주요 경제형질인 일당증체량, 도체중, 등심단면적, 근내지방도에 영향을 주는 우수 유전자형 조합을 선별한다. 그리고 선별된 조합에 대해 유전자 행렬 맵핑 방법에서 적용되지 않은 순열검정을 도입함으로써 우수유전자형 조합에 대한 통계적인 유의성을 확인한다.

주요용어: 단일염기다형성, 유전자 행렬 맵핑, 한우 경제형질.

### 1. 서론

의학 및 생명공학의 발달과 더불어 유전학 역시 많은 발전을 하고 있다. 유전자원 보존과 유전능력 향상을 위한 기술 개발은 더 큰 발전을 위한 매우 중요한 과제라 할 수 있다. 이를 위한 많은 연구가 이루어지고 있으며, 특히 광범위 유전자 관련 (Genome-Wide Association; GWA) 연구에서는 무수히 많은 유전자들을 이용하여 인간의 질병에 관련된 유전자를 찾아왔다.

일반적으로 인간의 질병과 가축의 경제적인 특성은 단일 유전자의 효과로 설명하기 보다는 이들 유전자의 조합으로 인한 상호작용으로 해석하고 있다. 이러한 상호작용을 분석하기 위해 선형모형과 같은 표준 통계모형을 이용한 통계학적인 방법은 유전적인 정보와 관련된 특성들과 QTL (Quantitative Trait Loci)의 관계를 보다 상세히 분석하고 식별할 수 있게 되었다. 하지만, 유전자의 수가 많아지게 되고, 그에 따라 유전자 조합이 많아지므로 모형이 복잡해지게 되어 이들 상호작용에 대한 해석이 어려워지는 경우도 발생하였다. 그래서 다중인자 차원 축소방법 (Multifactor Dimensionality Reduction: MDR)이 개발되었고 (Ritchie 등, 2001; Chung 등, 2005), 이분형 자료에만 활용 가능한 MDR 방법을 연속형 데이터에도 적용 가능하게 보완한 CART 알고리즘을 활용한 Expanded multifactor dimensionality reduction (E-MDR, Lee 등, 2008)과 더미변수를 활용한 Dummy multifactor dimensionality

<sup>1</sup> 교신저자: 경북 경산시 대동 214-1, 영남대학교 통계학과, 교수. E-mail: jlee@yu.ac.kr

<sup>2</sup> 경북 경산시 대동 214-1, 영남대학교 통계학과, 대학원, 석사과정.

<sup>3</sup> 경북 경산시 대동 214-1, 영남대학교 통계학과, 대학원, 박사.

reduction (D-MDR; Lee 등, 2009)을 제안하였다. 하지만 이 방법들은 상호작용에 영향을 준 QTL들의 하위 집단의 유전자형에 대해서 고려하지 않은 문제점이 있다. 복잡한 특성들에 영향을 주는 최종 값은 단 하나의 상호작용에서 찾기보다 복잡한 상호작용에서 결정되어야 한다. 따라서 많은 QTL들을 한 번에 비교하여 특성치에 영향을 주는 QTL의 상호작용과 그 하위집단인 유전자형을 규명하는 방법으로 유전자형 행렬 맵핑 (Genotype Matrix Mapping: GMM)이 개발되었다 (Sachiko 등, 2007). 본 연구에서의 데이터는 농협중앙회 한우개량사업소의 후대검정집단인 30차에서 35차 국가 후대검정우 집단 476두로 구성된다. 한우의 여러 경제형질인 일당증체량 (ADG:Average Daily Gain)과 도체중 (CWT: Carcass cold WeighT), 등심단면적 (LMA: Longissimus Muscle dorsi Area), 근내지방도 (MS: Marbling Score)는 모든 F1 자손으로부터 수집되었고, 한국 축산물등급판정소의 규격에 따라 측정되었다.

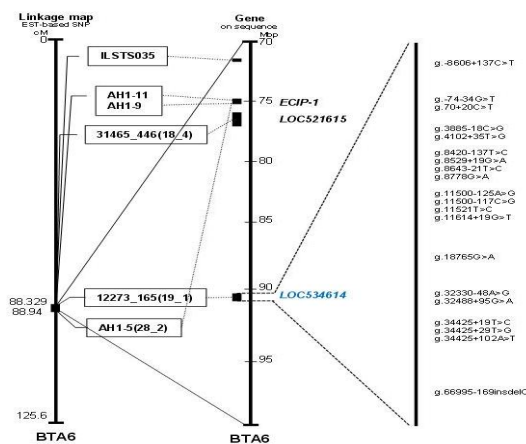


그림 1.1 Linkage map과 Gene on sequence에 기반한 SNPs의 발굴과정 (Lee, Y., 2009)

그림 1.1의 EST-based SNP 연관지도 (Snelling 등, 2005)에서 총 33개의 SNPs를 확인할 수 있다 (Lee, 2009). 후보 유전자로 판단되는 LOC534614 유전자내 SNPs 20개 중 대립유전자의 빈도가 0.1 미만이거나 유전자형의 빈도가 치우친 3개의 SNP (g.-8606+137C>T, g.-74346>T, g.70x200>T)를 제외한 17개의 SNPs를 한우의 경제형질에 영향을 주는 후보 SNPs로 판단한다.

본 논문에서는 Lee (2009)가 선정한 후보 SNPs 17개를 대상으로 Genotype matrix napping (GMM) 방법을 이용하여 한우의 경제형질인 일당증체량 (ADG)과 도체중 (CWT), 등심단면적 (LMA), 근내지방도 (MS)등에 영향을 우수 유전자형 조합을 선별하고, GMM 방법에서 적용하지 않은 순열검정을 우수유전자형조합에 적용하여 통계적 유의성을 검증받게 한다. 즉, 먼저 2장에서 한우의 경제형질에 영향을 주는 유전자 조합을 찾는 방법인 GMM방법과 순열검정에 대해 소개한다. 그리고 3장에서는 한우데이터에 적용하여 종합적인 경제형질에 관련된 주요 유전자를 GMM 방법을 통해서 선별하고 순열 검정을 통해서 통계적 유의성을 검증하여 최종 우수 유전자형 조합을 규명한다.

## 2. 유전자 행렬 맵핑 소개

### 2.1. 유전자 행렬 맵핑 (Genotype matrix mapping)

유전자 상호작용의 중요성이 강조되고 있지만, 전통적인 유전자 분석에서는 상호작용을 무시하거나

유전자 상호작용의 효과를 추가적인 편차로 규정하였다. 유전자 상호작용을 분석하기 위해서 로지스틱 회귀분석, 신경망 분석 그리고 비모수적 분석 방법을 비롯해서 다른 많은 방법들이 연구 되었다. 하지만 이러한 방법들은 많은 QTL 사이의 상호작용 중 가장 뛰어난 QTL을 발견하기 위해 발전되었지만, 상호작용에 참여한 QTL들의 하위집단인 유전자형은 고려하지 않았다는 문제점이 있다. 그 문제점을 해결하기 위해서 Sachiko 등 (2007)은 GMM 방법을 제시하였으며, GMM방법을 통해서 많은 QTL을 한 번에 비교하여 특성치에 영향을 주는 QTL의 상호작용 및 QTL의 하위집단인 유전자형을 발견할 수 있었다.

GMM 방법은 다양한 체계를 가진 복잡한 특성의 유전적 변이에서 QTL 상호작용의 선별에 유용하며, 많은 QTL을 한 번에 비교하여 복잡한 특성과 유전자의 상호작용 및 연관성을 찾기 위해 개발되었다 (Sachiko 등, 2007). 그림 2.1과 같이 각 마커들은 Genotype Matrix (GM)으로 나타낼 수 있으며 모 집단에서 유전자형의 총 개수에 기준하여 배분된다.

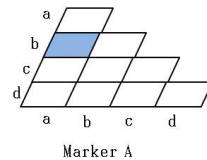


그림 2.1 Marker A의 GM 도식화 (Sachiko 등, 2007)

그리고 유전자형과 특성치 사이에 유의한 연관성은 일원배치 분산분석에 기초하며, 마커 A에서 유전자형 'a'와 'b' 사이의 상호작용은 마커 A의 GM에서 'ab'를 가지는 셀과 그렇지 않은 셀에 해당하는 특성치와 비교하여 판단하게 된다. 최종적으로 그림 2.2와 같이 다중 GM을 통하여 Genotype matrix network (GMN)이라 불리는 가상 네트워크로 각 셀들을 연결하여 유전자형 조합을 생성한다. 부분적인 네트워크가 특성치와 유의한 연관성을 나타내면, 네트워크 내 특정 유전자형 조합과 특성치와의 상호작용을 찾아낸다.

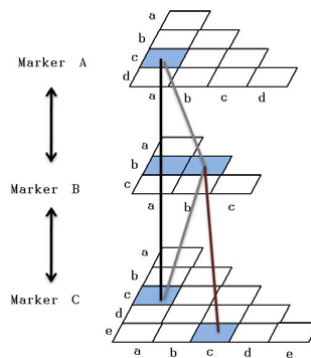


그림 2.2 GMN에 의한 마커들 사이의 상호작용 선별 (Sachiko 등, 2007)

이러한 GMM 방법의 절차는 다음과 같다.

- **절차 1.** 각 마커들을 GM으로 나타내고, GMN을 통하여 각 셀들의 조합을 생성한다.

- 절차 2. 조합의 유의성을 평가하기 위해 F-측정치를 계산하며, 그 방법은 다음과 같다.
1. N개의 개별치들로 구성된 전체 집단을 S라고 하고, 전체 집단을 중복 없이 두 집단  $S_0$ 와  $S_1$ 으로 나눈다. 여기에서  $S_1$ 은 특정한 유전자형을 지닌 그룹으로, 이외의 집단을  $S_0$ 로 한다.
  2. 집단 간 평균제곱을 계산한다.

$$MSA = \sum_{j=0}^1 |S_j|(\mu_j - \mu)^2 \quad (2.1)$$

단,  $\mu_0, \mu_1$  :  $S, S_0, S_1$  집단의 특성치 평균,  
 $|X|$  : X 집단의 개체수

3. 집단내 평균제곱을 계산한다.

$$MSW = \frac{\sum_{j=0}^1 \sum_{s_i \in S_j} (P_i - \mu_j)^2}{N - 2} \quad (2.2)$$

단,  $P_i$  :  $S_i$  집단에서 i 번째 개체의 특성값

4. 집단간 평균제곱값을 집단내 평균제곱값으로 나누어 F-측정치를 계산한다.

$$F = \frac{MSA}{MSW} = \frac{(N - 2)|S_0|(\mu_0 - \mu)^2 + |S_1|(\mu_1 - \mu)^2}{\sum_{s_i \in S_0} (P_i - \mu_0)^2 + \sum_{s_i \in S_1} (P_i - \mu_1)^2} \quad (2.3)$$

위의 방법을 통하여 F-측정치가 산출되며, F-측정치가 클수록 유전자형과 특성들 사이의 상호작용이 강한 것으로 나타난다. 산출된 F-측정치의 최대값을  $F_{opt}$ 라고 하고 현재 유전자형 조합 이외 다른 조합으로 더 큰 F-측정치가 산출되면, 현재  $F_{opt}$ 를 새로 산출된 F-측정치로 수정한다. 모든 유전자형 조합에 대한 계산이 끝난 후 가장 큰 F-측정치를 가지는 유전자형 조합이 특성치와 연관성이 가장 큰 유전자형 조합으로 판단하게 된다.

이러한 GMM 방법을 활용하기 위해 Kasusa DNA 연구소에서 제공된 프로그램을 사용한다. GMM 방법은 특성치에 영향을 주는 유전자 상호작용 우수 유전자 선별 시 직접 유전자형 조합을 찾아주기 때문에 다른 방법들보다 절차 및 계산이 간편하며, 하나의 개별적인 유전자형 조합을 선별할 수 있다는 장점이 있다. 하지만, 개체수가 매우 작은 경우 특성치와 연관된 유전자형 조합의 정확한 영향력을 파악하기 힘들다는 점이 있다.

일반적인 GMM 방법에서는 F-측정치가 가장 큰 조합을 선택하게 된다. 하지만 본 논문에서는 각 경제형질에 대해 상위 15개의 조합을 우선 선정하고, 이를 근거로 하여 공통적으로 포함된 유전자형 조합을 종합적인 경제형질에 영향을 주는 최종 우수 유전자형 조합으로 선택한다. 하지만, GMM 방법을 통해서 구해진 F-측정치는 단순히 F-측정치가 큰 값을 가지는 유전자형 조합이 특성치에 높은 영향을 준다는 내용만 있을 뿐 통계적인 유의성을 검증하지 않았다. 그렇기 때문에 실제로 F-측정치에 의해서 선별된 우수 유전자형 조합이 특성치에 유의한 영향을 주는가에 대해서 통계적인 검정이 필요하다. 따라서 선별된 유전자형 조합에 대해 순열 검정을 실시하여 통계적 유의성을 검증한다.

## 2.2. GMM 방법의 순열검정 적용

Sachiko 등 (2007)이 제시한 F-측정치는 통계적 유의성을 검증하지 않았기 때문에 통계적인 유의성 검정이 반드시 필요하다. 본 논문에서는 F-측정치의 통계적 유의성을 검증해주기 위해 순열검정을 활

용하여 10,000번의 시뮬레이션을 실시한 후 P-값을 계산한다. 순열검정 절차를 살펴보면 다음과 같다 (Good, 2000).

- **절차 1. 가설 설정**
  - GMM 방법을 통해 선별된 우수 유전자형 조합이 특성치에 영향력이 있다.
- **절차 2. 통계량과 기각역 설정**
  - 분석에 사용할 통계량으로 F-측정치를 선택한다.
  - 특정 유전자형 조합을 가진 그룹과 그 외의 그룹으로 나누어 그룹간 데이터를 서로 바꾸었을 때 특정 유전자 조합을 가진 그룹의 통계량이 높다면 특성치에 영향력이 있다고 판단한다.
- **절차 3. 선별된 유전자형 조합에 대한 통계량 계산**
  - GMM 방법을 통해 선별된 우수 유전자형 조합의 F-측정치를 계산한다.
- **절차 4. 관측치의 재배열과 재배열 후의 통계량 계산**
  - 두 그룹의 데이터를 n개 만큼 랜덤 추출하여 서로 교환한 후 그룹의 F-측정치를 구한다. 이 과정을 10,000번 반복한다.
- **절차 5. 결론 (P-값 계산)**
  - 각 F-측정치를 내림차순으로 정렬한 후 기존의 F-측정치와 비교하여 P-값을 구한다.

즉, GMM 방법을 통해서 선별된 우수 유전자형 조합을 가진 그룹과 가지지 않은 그룹으로 분류하고 유전자형을 가진 그룹에 대해 F-측정치를 구한다. 그리고 각 그룹의 데이터를 n개 교환하여 F-측정치를 구하고, 이를 10,000번 반복하여 구해진 F-측정치를 내림차순으로 정렬하여 P-값을 구하게 된다.

### 3. 한우 경제형질에 대한 우수 유전자형 조합의 선별 결과

#### 3.1. 개요

한우의 경제형질에 영향을 주는 유전자 상호작용 우수 유전자를 선별하기 위해 GMM 방법을 활용하여 한우의 경제형질에 영향을 주는 유전자 상호작용 우수 유전자 조합 및 유전자형을 선별하였다. 선별된 우수 유전자형 조합에 대한 유의성 검정 및 경제형질과의 유의성 분석을 실시하였다. 본 논문에서 활용한 데이터는 농협중앙회 한우개량사업소의 후대검정집단인 30차에서 35차 국가 후대검정우 집단 한우 476두에서 결측치를 제외한 후 붓스트랩 샘플링 방법 (Efron 등, 1993)을 적용시킨 3830두를 대상으로, Lee (2009)의 연구에서 최종적으로 선택된 17개의 SNPs를 분석에 사용하였다.

#### 3.2. 한우의 종합 경제형질에 대한 우수 유전자형 선별

한우의 주요 경제형질인 일당증체량, 도체증량, 근내지방도, 등심단면적에 GMM 방법을 적용하여 각 경제형질에서 상위 15개의 유전자형 조합을 찾았다. 표 3.1은 GMM 방법을 활용하여 근내지방도에 대한 우수 유전자 조합을 나열한 것이다. 근내지방도에 대한 우수 유전자형 조합에서는 g.34425+102A>T(AA)의 SNP Marker와 g.66995-169insdelC를 가진 조합이 우수한 것으로 선별되었다. 이 밖에 나머지 일당증체량, 도체증량, 등심단면적에서도 동일하게 GMM 방법을 적용시켜 3개의 유전자형 조합에서 우수한 조합 상위 15개의 조합을 살펴보았다. 그 중에서 각 경제형질 모두에서 공통적으로 나타나는 우수 유전자 조합으로 g.34425+102A>T(AA) g.11521T>C(TT) g.4102+36T>G(GT) 포함한 6개의 유전자형 조합 g.34425+102A>T(AA) g.11500-125A>G(AA) g.4102+36T>G(GT) 과 g.34425+102A>T(AA) g.8778G>A(GG) g.4102+36T>G(GT), g.34425+

표 3.1 GMM방법을 활용한 한우 경제형질 (근내지방도)에 대한 상위 15개 우수 유전자 조합

경제형질	유전자 조합			F-측정치
근내지방도 (MS)	g.11614+19GT(GT)	g.11500-117C>g(CG)	g.3885-18c>G(CC)	100.432
	g.34425+102A>T(AA)	g.11521T>C(TT)	g.4102+36T>G(GT)	72.1975
	g.34425+102A>T(AA)	g.11500-125A>G(AA)	g.4102+36T>G(GT)	72.1975
	g.34425+102A>T(AA)	g.8778G>A(GG)	g.4102+36T>G(GT)	72.1975
	g.34425+102A>T(AA)	g.8529+19G>A(GG)	g.4102+36T>G(GT)	72.1975
	g.34425+102A>T(AA)	g.8420-137T>C(TT)	g.4102+36T>G(GT)	72.1975
	g.34425+102A>T(AA)	g.4102+36T>G(GT)	g.-74-34G>T(GG)	72.1975
	g.66995-169insdelC(AA)	g.34425+102A>T(TT)	g.11614+19G>T(GT)	71.4085
	g.66995-169insdelC(AA)	g.11614+19G>T(GT)	g.11521T>C(TT)	71.4085
	g.66995-169insdelC(AA)	g.11614+19G>T(GT)	g.11500-125A>G(AA)	71.4085
	g.66995-169insdelC(AA)	g.11614+19G>T(GT)	g.8778G>A(GG)	71.4085
	g.66995-169insdelC(AA)	g.11614+19G>T(GT)	g.8529+19G>A(GG)	71.4085
	g.66995-169insdelC(AA)	g.11614+19G>T(GT)	g.8420-137T>C(TT)	71.4085
	g.34425+102A>T(TT)	g.32488+95G>A(GG)	g.11614+19G>T(GT)	71.4085
	g.34425+102A>T(TT)	g.32330-48A>G(AA)	g.11614+19G>T(GT)	71.4085

\*All P-value < 0.0001

102A>T(AA) g.8529+19G>A(GG) g.4102+36T>G(GT), g.34425+102 A>T(AA) g.8420-137T>C(TT) g.4102+36T>G(GT), 그리고 g.34425+102A>T(AA) g.4102+36T>G(GT) g.-74-34G>T(G) G)이 한우의 종합경제형질에 유의한 연관이 있었으며, 이를 우수 유전자형 조합으로 최종 선별 되었다 (표 3.2).

표 3.2 GMM에 의해서 최종 선별된 우수 유전자형 조합

	유전자 조합		유전자형
g.34425+102A>T	g.11521T>C	g.4102+36T>G	AATTGT
g.34425+102A>T	g.11500-125A>G	g.4102+36T>G	AAAAGT
g.34425+102A>T	g.8778G>A	g.4102+36T>G	AAGGGT
g.34425+102A>T	g.8529+19G>A	g.4102+36T>G	AAGGGT
g.34425+102A>T	g.8420-137T>C	g.4102+36T>G	AATTGT
g.34425+102A>T	g.4102+36T>G	g.-74-34G>T	AAGTGG

#### 4. GMM 방법을 활용하여 선별된 우수 유전자형 분석

3.2절에서는 한우의 종합적인 특성인 경제형질에 대한 우수 유전자형 조합을 GMM 방법을 이용하여 찾을 수가 있었다. 이렇게 선별된 우수 유전자형 조합들이 각각의 경제형질에서 통계적으로 유의한지를 알아보기 위해서 각각의 경제형질에 대해서 t-검정과 순열검정을 실시하였다.

표 4.1 선별된 우수 유전자형에 대한 t-test 및 순열검정 적용 결과

경제형질	유전자 조합	유전자형	t-test (P-value)	순열검정 (P-value)
일당증체량 (ADG)	g.34425+102A>T g.11521T>C	AATTGT	8.94(<0.0001)	0.0001
	g.34425+102A>T g.11521T>C	AATTGT	8.94(<0.0001)	0.0001
	g.4102+36T>G			
	g.34425+102A>T g.11500-125A>G	AAAAGT	8.94(<0.0001)	0.0001
	g.4102+36T>G			
	g.34425+102A>T g.8778G>A	AAGGGT	8.94(<0.0001)	0.0001
	g.4102+36T>G			
	g.34425+102A>T g.8529+19G>A	AAGGGT	8.94(<0.0001)	0.0001
	g.4102+36T>G			
	g.34425+102A>T g.8420-137T>C	AATTGT	8.94(<0.0001)	0.0001
g.4102+36T>G				
g.34425+102A>T g.4102+36T>G	AAGTGG	8.94(<0.0001)	0.0001	
g.-74-34G>T				
도체중량 (CWT)	g.34425+102A>T g.11521T>C	AATTGT	12.74(<0.0001)	0.0001
	g.4102+36T>G			
	g.34425+102A>T g.11500-125A>G	AAAAGT	12.74(<0.0001)	0.0001
	g.4102+36T>G			
	g.34425+102A>T g.8778G>A	AAGGGT	12.74(<0.0001)	0.0001
	g.4102+36T>G			
	g.34425+102A>T g.8529+19G>A	AAGGGT	12.74(<0.0001)	0.0001
	g.4102+36T>G			
	g.34425+102A>T g.8420-137T>C	AATTGT	12.74(<0.0001)	0.0001
	g.4102+36T>G			
g.34425+102A>T g.4102+36T>G	AAGTGG	12.74(<0.0001)	0.0001	
g.-74-34G>T				
등심단면적 (LMA)	g.34425+102A>T g.11521T>C	AATTGT	6.67(<0.0001)	0.0001
	g.4102+36T>G			
	g.34425+102A>T g.11500-125A>G	AAAAGT	6.67(<0.0001)	0.0001
	g.4102+36T>G			
	g.34425+102A>T g.8778G>A	AAGGGT	6.67(<0.0001)	0.0001
	g.4102+36T>G			
	g.34425+102A>T g.8529+19G>A	AAGGGT	6.67(<0.0001)	0.0001
	g.4102+36T>G			
	g.34425+102A>T g.8420-137T>C	AATTGT	6.67(<0.0001)	0.0001
	g.4102+36T>G			
g.34425+102A>T g.4102+36T>G	AAGTGG	6.67(<0.0001)	0.0001	
g.-74-34G>T				
근내지방도 (MS)	g.34425+102A>T g.11521T>C	AATTGT	6.56(<0.0001)	0.0001
	g.4102+36T>G			
	g.34425+102A>T g.11500-125A>G	AAAAGT	6.56(<0.0001)	0.0001
	g.4102+36T>G			
	g.34425+102A>T g.8778G>A	AAGGGT	6.56(<0.0001)	0.0001
	g.4102+36T>G			
	g.34425+102A>T g.8529+19G>A	AAGGGT	6.56(<0.0001)	0.0001
	g.4102+36T>G			
	g.34425+102A>T g.8420-137T>C	AATTGT	6.56(<0.0001)	0.0001
	g.4102+36T>G			
g.34425+102A>T g.4102+36T>G	AAGTGG	6.56(<0.0001)	0.0001	
g.-74-34G>T				

표 3.3을 살펴보면 선별된 6개의 유전자형 조합이 각각의 경제형질 모두에서 순열검정 결과 P-값이 0.0001로써 종합경제형질에 모두 유의한 영향을 미치는 것으로 나타났다. 또한 t-검정에서도 모든 경제형질에서도 유의한 영향을 나타났다. 그 결과를 바탕으로 표 3.2에서 선별된 6개의 유전자형 조합을 한

우의 종합경제형질에 영향을 주는 우수 유전자형 조합으로 최종 선택 할 수 있었다. 마지막으로 최종 우수 유전자형 조합으로 선별된 6개의 조합을 살펴보면 6개의 조합이 모두 각 경제형질에서 동일한 t-값과 P-값을 가지는 것을 확인 할 수 있다. 이는 선별된 6개의 조합이 실질적으로는 서로 동질하다는 것을 추측할 수 있겠다.

## 5. 결론 및 토의

본 논문은 농협중앙회 후대검정우 집단 476두 중 결측치를 제외한 후 붓스트랩 샘플링 방법 (Efron 등, 1993)을 적용시킨 3830두에 대해서 GMM 방법을 통해 17개의 SNPs 중 한우의 경제형질에 연관된 우수 유전자형 조합을 선별하였다.

GMM 방법을 통해서 선별된 유전자 조합 및 유전자형은 최종 6개로써 g.34425+102A>T(AA) g.11521T>C(TT) g.4102+36T>G(GT), g.34425+102A>T(AA) g.11500-125A>G(AA) g.4102+36T>G(GT), g.34425+102A>T(AA) g.8778G>A(GG) g.4102+36T>G(GT), g.34425+102A>T(AA) g.8529+19G>A(GG) g.4102+36T>G(GT), g.34425+102A>T(AA) g.8420-137T>C(TT) g.4102+36T>G(GT), g.34425+102A>T(AA) g.4102+36T>G(GT) g.-74-34G>T(GG)가 선별되었다. 이 유전자형 조합이 한우의 경제형질에 유의한 영향을 검증하기위해 순열검정과 t-검정을 적용하였다. 그 결과 우수 유전자 조합에서 선별된 우수 유전자형 조합이 통계적으로 유의한 것으로 밝혀졌다. 따라서 우리는 GMM 방법을 통해서 선별한 위의 우수 유전자형 조합이 한우의 종합적인 경제형질에 유의한 영향을 주는 것을 규명할 수 있었다. 그리고 최종적으로 선별된 6개의 조합을 살펴보니 t-검정 결과와 P-값이 각 경제형질에서는 동일한 값을 보였다. 이는 6개의 조합이 어느 정도 동질하다는 것을 추측할 수 있으며, 차후 연구를 통해서 이들 조합의 동질성에 대해 확인할 필요가 있다고 여겨진다.

## 참고문헌

- Chung, Y. J., Lee, S. Y. and Park, T. S. (2005). Multifactor dimensionality reduction in the presence of missing observations. *Journal of Korea Statistical Society, Proceedings of the Autumn Conference*, **1**, 31-36.
- Efron, B. and Tibshirani, R. (1993). An introduction to the bootstrap, *Chapman & Hall/CRC*, New York.
- Good, P. (2000). *Permutation test: A ractical guide to resampling method for testing hypotheses*, Springer-Verlag Berlin and Heidelberg GmbH & Co., New York.
- 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-Australian Journal of Animal Sciences* **6**, 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(2)**, 435-442
- 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.
- 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.
- Sachiko, I., Akihiro, N. and Satoshi, T. (2007). Genotype matrix mapping: Searching for quantitative trait loci interactions in genetic variation in complex trait. *DNA Research*, **14**, 217-225.
- 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.



## Detection of major genotypes combination by genotype matrix mapping

Jea Young Lee<sup>1</sup> · Jong Hyeong Lee<sup>2</sup> · Yong Won Lee<sup>3</sup>

<sup>123</sup>Department of statistics, Yeungnam

Received 28 February 2010, revised 29 April 2010, accepted 6 May 2010

### Abstract

It is important to identify the interaction of genes about human disease and characteristic value. Many studies as like logistic analysis, have associated being pursued, but, previous methods did not consider the sub-group of the genotypes. So, QTL interactions and the GMM (genotype matrix mapping) have been developed. In this study, we detect the superior genotype combination to have an impact on economic traits of Korean cattle based on the study over GMM method. Thus, we identified interaction effects of single nucleotide polymorphisms (SNPs) responsible for average daily gain (ADG), marbling score (MS), carcass cold weight (CWT), longissimus muscle dorsi area (LMA) using GMM method. In addition, we examine significance of the major genotype combination selected by implementing permutation test of the F-measure which was not obtained by Sachiko et al.

*Keywords:* GMM (genotype matrix mapping), hanwoo Economic traits, SNP (single nucleotide polymorphisms).

---

<sup>1</sup> Corresponding author: Professor, Department of Statistics, Yeungnam University, Kyungsan, Korea.  
E-mail: jlee@yu.ac.kr

<sup>2</sup> Master student, Department of Statistics, Yeungnam University, Kyungsan, Korea.

<sup>3</sup> Doctor of philosophy, Department of Statistics, Yeungnam University, Kyungsan, Korea.