

Main Gene Combinations and Genotype Identification of Hanwoo Quality with SNPHarvester

Jae-Young Bae^a, Jea-Young Lee^{1, a}

^aDepartment of Statistics, Yeungnam University

Abstract

It is known that human disease and the economic traits of livestock are significantly affected by a gene combination effect rather than a single gene effect. Existing methods to study this gene combination effect have disadvantages such as heavy computing, cost and time; therefore, to overcome those drawbacks, the SNPHarvester was developed to find the main gene combinations. In this paper, we looked for gene combinations using an adjusted linear regression model. This research finds that superior gene combinations which are related to the quality of the Korean beef cattle among sets of SNPs using SNPHarvester. We also identify the superior genotypes using a decision tree that can enhance the various qualities of Korean beef among selected a SNP combination.

Keywords: CART, Genotype, *k*-means, SNPHarvester.

1. 서론

인간의 질병과 가축의 경제적인 특성과 관련된 주요 유전자와 유전자 상호작용을 찾는 것은 유전학에서 주된 관심사 중의 하나이다. 이러한 주요 유전자와 조합을 찾기 위해서 수 많은 단일 염기 다형성(single nucleotide polymorphisms; SNPs)을 사용하였다. SNP들은 단일의 유전적 효과보다는 유전적 상호작용이 인간의 질병과 가축의 경제적인 특성에 더 많은 영향을 준다고 알려져 있다. 일반적인 통계적 모델에서 이런 상호작용효과를 보기 위한 모형으로 선형모형 등과 같은 방법이 널리 사용되고 있다. 하지만 유전자의 수가 많아 짐에 따라서 이런 통계적 모델로는 한계가 있는데, 모형이 복잡해지고 해석이 어려워지는 단점이 있다. 최근에는 이런 단점을 해결하고 많은 유전자들의 효과를 보기 위해서 다양한 방법들이 제시되었다.

유전자들의 조합을 찾는 방법으로 제시된 것에는 다중인자차원축소(multifactor dimensionality reduction; MDR; Ritchie 등, 2001)방법, CART(classification and regression tree)방법을 활용한 확장된 다중인자 차원축소(Expanded MDR; E-MDR; Lee 등, 2008)방법, SVM(support vector machine)을 이용한 다중인자 차원축소(support vector machine multifactor dimensionality reduction; SVM-MDR; Lee와 Lee, 2010)방법 등이 있다. 그러나 위의 방법들에도 수 많은 유전자들 중에서 우수 유전자를 찾을 때는 모든 경우의 수를 다 적용하여야 하기 때문에 계산이 복잡하며, 비용이 많이 들며, 시간이 많이 걸리는 단점이 있다. 이러한 단점을 해결하기 위하여 수 많은 유전자들 중에서 주요 SNP 조합을 찾는 방법으로 SNPHarvester (Yang 등, 2009)방법이 개발되었다.

SNPHarvester 방법은 인간의 질병과 가축의 경제적인 특성과 관련된 많은 유전자들 중에서 주요 유전자 조합을 찾는 방법으로 제시되었고, 사례-대조군과 같은 이분화된 데이터에 적용이 가능하다. 이

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

방법에 적용하기 위한 한우 데이터의 형질은 유전적인 효과와 환경적인 효과에 의해서 나타나고 있다. 하지만 본 연구에서는 유전적인 효과만으로 조합을 보기 위해서 환경적인 효과를 보정한 자료를 바탕으로 분석한다. 분석에 사용한 형질들은 한우의 품질(맛과 육질)에 영향을 주는 올레인산(oleic acid; C18:1), 포화지방산(saturated fatty acid; SFA), 불포화지방산(monounsaturated fatty acid; MUFA), 근내 지방도(marbling score; MS)이다 (Oh 등, 2011). 이 4개 형질에서 환경적인 효과를 보정한 품질에 관련된 형질 4개를 이용하여 K-평균 군집분석방법으로 이분화 시킨 뒤, 유전자 조합을 선별하기 위해서 32개의 SNP들을 SNPHarvester 방법에 적용한다. 그리고 선별된 유전자 조합을 각 경제형질에 적용하여서 한우의 품질에 영향을 주는 우수 유전자 조합으로 선별을 한다. 이 선별된 우수 유전자 조합을 CART방법을 활용하여 우수 유전자형(genotype)을 찾고, 이 유전자형의 효과를 보았다.

본 연구는 다음과 같이 구성되었다. 2절에서는 유전자 조합에 사용된 통계모형과 SNPHarvester 방법에 대해 소개를 하고, 3절에서는 유전자 조합과 우수 유전자형 선별결과에 대해서 나타내었다. 4절에서는 결론을 내었다.

2. 통계모형을 이용한 유전자 조합 선별 방법

2절에서는 유전자 조합을 선별하기 위해 사용된 통계모형과 방법에 대해 소개한다. 2.1절에서는 SNPHarvester 방법에 대해서 소개하고, 2.2절에서는 유전적인 효과만을 보기 위한 선형회귀모형 대해서 소개를 한다.

2.1. SNPHarvester 방법 및 절차

수 많은 SNP들을 이용해서 인간의 질병과 연관된 유전자들을 선별할 수 있다. 하지만 유전자의 조합을 찾기 위해서 사용되는 일반적인 통계적 방법들은 SNP들이 많을 때는 시간과 비용이 많이 들며, 계산이 복잡해지는 문제점을 가지고 있다. 이러한 문제점을 해결하고 많은 SNP들의 조합을 찾기 위해서 SNPHarvester 방법이 개발되었다. 즉, SNPHarvester 방법은 질병과 관련된 SNP들 가운데 주요 유전자 조합을 선별하는 방법이다. 이 방법은 특성치와 관련된 SNP들 중 초기에 몇 개의 SNP를 이용해서 조합으로 선정된 뒤, 그 조합에 있는 SNP 중 하나를 선택 안 된 SNP들과 바꿔가면서 스코어를 높이는 과정을 반복하는 것이다. 이 방법에서 사용되는 스코어 함수는 MDR방법에서의 정확도, χ^2 -통계량, B-통계량 등이 사용 될 수 있지만, 본 논문에서는 χ^2 -통계량을 스코어 함수로 사용한다. 이 방법을 이용하면 수 많은 SNP들 중 특성치와 관련된 주요한 SNP 그룹을 쉽게 찾을 수 있다. Figure 1과 아래의 Step은 SNPHarvester 방법의 절차를 나타낸 것이다 (Lee와 Kim, 2009).

Step 0. 전체 SNP 그룹을 정한다.

예) L 개의 SNP: $SNP_1, SNP_2, \dots, SNP_L$

Step 1. L 개의 SNP 그룹에서 랜덤으로 k 개의 SNP를 A라는 그룹으로 선택, 나머지는 SNP_i 로 둔다.

예) group A = $\{SNP_{s_1}, SNP_{s_2}, \dots, SNP_{s_k}\}$, $SNP_i (SNP_i \notin A)$, $i = 1, \dots, L$

Step 2. SNP_i 중 하나와 A그룹의 SNP와 1:1로 k 번 교체하여 각각 스코어를 낸다.

예) $A_1 = \{SNP_i, SNP_{s_2}, \dots, SNP_{s_k}\} \Rightarrow \text{Score}(A_1)$

$A_2 = \{SNP_{s_1}, SNP_i, \dots, SNP_{s_k}\} \Rightarrow \text{Score}(A_2)$

\vdots

$A_k = \{SNP_{s_1}, SNP_{s_2}, \dots, SNP_i\} \Rightarrow \text{Score}(A_k)$

Step 3. Step 2에서 계산된 스코어에서 가장 큰 값을 A^* 로 정한다.

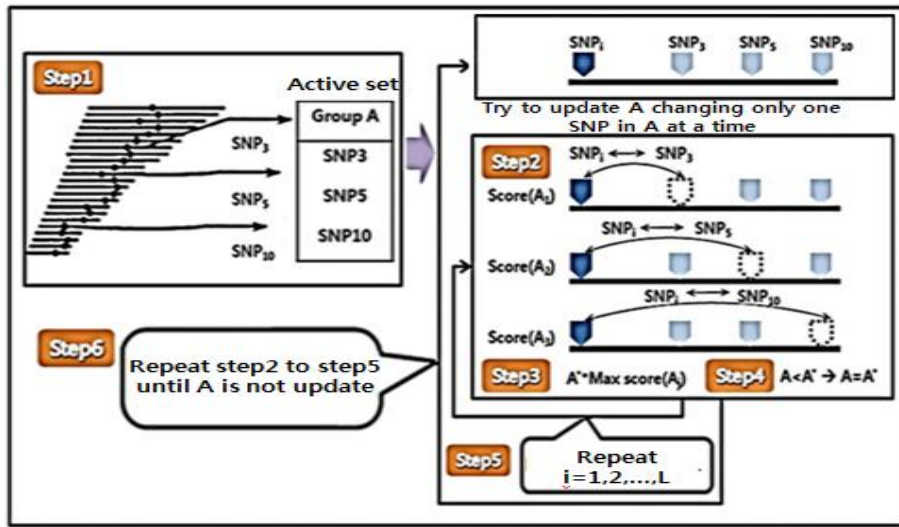


Figure 1: The procedure of SNPHarvester method

- Step 4. A^* 의 스코어와 A그룹의 스코어와 비교하여 $A^* > A$ 이면 A^* 그룹으로 교체, 아닐 경우 A그룹을 그대로 둔다.
- Step 5. 그룹이 선택되었다면, $SNP_i + 1$ 을 이용하여 Step 2~Step 4를 반복한다.
- Step 6. Step 2~Step 5 (q (반복수) = 2 or 3)를 반복하며, 최종으로 나온 A^* 스코어가 이전의 A 스코어보다 작을 경우 멈추게 된다.

위의 Figure 1은 Step 1~Step 6까지를 나타낸 그림이며, 위의 Step을 반복하여서 최종적으로 SNP 조합을 찾아내는 것이 SNPHarvester 방법이다. 그리고 여러번 반복을 해서 스코어를 높이더라도 그 SNP 조합이 통계적으로 유의성을 가지지 않는다면 의미가 없으므로, 통계적으로 유의한 조합을 선별하기 위해 $\alpha = 0.01$ 로 정한다. 처음에 선별된 조합수인 k 가 클수록 SNP 조합 내의 유전자형의 빈도가 없거나 작아질 수 있기 때문에 $k \leq \ln_3 N_d - 1$ 로 제한한다 (Yang 등, 2009). 여기서 N_d 는 case의 수이다. 본 논문에서는 SNPHarvester 방법을 적용하여 한우의 품질에 영향을 주는 우수 유전자 조합을 선별한다. 하지만 SNPHarvester 방법은 이분형 자료에만 적용가능 하기 때문에, 환경요인을 보정한 한우의 품질 관련 형질의 4개를 k -평균군집분석방법을 통해 이분화 하였다. 이러한 SNPHarvester 방법을 이용하여 환경요인이 보정된 형질과 관련된 주요 SNP 조합을 찾아보려고 한다.

2.2. 보정된 선형회귀모형을 통한 각 형질의 효과

실제 연구에서 연구된 한우의 각 형질은 유전분석에 있어 기본이 되는 개념으로 각 형질의 표현형이 그 형질의 유전적인 효과와 환경적인 효과에 의해서 결정된다는 것이다. 즉 형질의 표현형은 다음과 같이 두 부분으로 나눌 수 있다.

$$P = G + E, \tag{2.1}$$

여기서 P 는 각 형질의 표현형이고 G 는 각 형질의 유전적 효과, E 는 환경적 효과라 한다. 그래서 한우의 형질 연구에서는 환경적 요인과 유전적요인에 다음과 같은 선형 회귀모형을 사용한다 (Casas 등,

Table 1: Mean and standard deviation(SD) of each economic trait

trait	content	<i>N</i>	Mean ± SD
C18:1	Oleic acid	5130	44.30 ± 2.66
SFA	Saturated fatty acid	5130	40.60 ± 2.86
MUFA	Monounsaturated fatty acid	5130	53.50 ± 2.97
MS	Marbling score	5130	5.43 ± 1.94

2005).

$$y_{ijk} = \mu + \text{Farm}_i + \beta \text{Age} + \text{SNP}_j + e_{ijk} \quad (2.2)$$

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

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

y_{ijk} 는 한우 각각의 형질값이다. Farm_i 는 사육농장 (17군데)의 고정효과이며, β 는 사육일령에 대한 회귀계수이고, SNP_j 는 SNP의 고정효과, e_{ijk} 는 $N(0, \sigma^2)$ 인 확률변수이다. 사육장농장과 사육일령이 환경요인에 속하며, SNP들이 유전적 효과에 해당한다. 하지만 본 논문에서는 관심을 가지는 부분이 형질에 영향을 주는 유전적인 효과에 관심이 있기 때문에 환경적 요인을 보정한 선형회귀모형을 이용한다 (Lee와 Jin, 2012). 식 (2.3)에 환경적 요인을 보정한 선형회귀모형을 나타내었다.

$$y_{ijk} - (\text{Farm}_i + \beta \text{Age}) = \mu + \text{SNP}_j + e_{ijk}, \quad i = 1, \dots, f, j = 1, \dots, m, k = 1, \dots, n. \quad (2.3)$$

식 (2.3)의 선형회귀모형을 통해 환경적 요인을 보정하고, 한우의 품질에 영향을 주는 4개의 형질을 k -평균 군집분석에 적용해서 이분화시킨 다음, SNPHarvester 방법을 통해 유전자 조합을 찾았다.

3. 우수 SNP 조합 및 우수 유전자형 선별

3.1. 실험자료

이 데이터는 경북지역에서 18개의 씨수소를 통해 얻어진 513두의 한우로 구성되어 있다 (Oh 등, 2011). Oh 등은 한우 맛과 육질에 영향을 주는 경제형질 4개를 찾아내었다. 맛에 영향을 주는 형질에는 올레인산(Oleic acid; C18:1)과 포화지방산(Saturated fatty acid; SFA), 불포화지방산(Monounsaturated fatty acid; MUFA)이 있고, 육질에 영향을 주는 경제형질인 근내지방도(Marbling score; MS)를 분석에 사용했다. Oh 등에 의해 경제형질에 영향을 주는 단일염기다형성(Single nucleotide polymorphism; SNP) 중 32개의 SNP가 한우의 맛과 질에 연관된 4개의 형질에 영향을 주는 것으로 연구가 되었다. 연구된 32개의 SNP들은 g.6850+77A>G, g.8646+128A>G, g.10153A>G, g.10213T>C, g.10329C>T, g.13757C>A, g.14047C>T, g.14578A>G, g.1159-71208A>G, g.42555-29812G>A, c.1523G>T, g.12870T>C, g.13126T>C, g.15532C>A, g.16907T>C, g.17924G>A, sreb3T>C, sreb9T>C, g.2634+1018A>T, c.280A>G, c.388G>A, c.408G>C, g.3977-325T>C, c.456A>G, c.88G>A, c.194C>T, g.6960A>T, LPL7G>A, LPL67G>A, LPL76G>T, LPL80G>A, LPL99C>T이다.

SNPHarvester 방법은 데이터가 적으면, 초기 조합 생성시 조합수가 적어지기 때문에 데이터를 bootstrapping으로 10배 증폭시킨 5130두를 이용해 분석하였다. 아래의 Table 1은 5130두에 대한 평균과 표준편차를 나타낸 표이다.

3.2. SNPHarvester 적용 결과

SNPHarvester는 수많은 SNP들을 조합으로 만들어서 우수한 조합을 선별하는 것이다. 따라서 증폭된 데이터의 5130두에서 나온 32개의 SNP를 이 기법을 이용해서 한우의 품질에 연관성이 있는 주요

Table 2: Upper 10 SNP combinations associated with total economic trait

2 combinations			3 combinations		
SNP interaction	χ^2	<i>p</i>	SNP interaction	χ^2	<i>p</i>
(g.6350+77A>G, g.13126T>C)	407.99	0.000	(g.6350+77A>G, g.13126T>C, LPL7G>A)	578.09	0.000
(srebp3T>C, srebp9T>C)	366.50	0.000	(c.194C>T, srebp3T>C, g.10329C>T)	548.70	0.000
(g.42555-29812G>A, srebp9T>C)	341.72	0.000	(g.10213T>C, srebp3T>C, c.194C>T)	542.31	0.000
(g.17924G>A, c.194C>T)	340.61	0.000	(srebp3T>C, g.2634+1018A>T, c.194C>T)	527.03	0.000
(g.10213T>C, c.194C>T)	339.98	0.000	(g.42555-29812G>A, srebp9T>C, LPL99C>T)	490.21	0.000
(g.2634+1018A>T, c.194C>T)	333.97	0.000	(srebp3T>C, srebp9T>C, g.2634+1018A>T)	480.35	0.000
(g.6850+77A>G, g.42555-29812G>A)	314.47	0.000	(g.6960A>T, LPL67G>A, LPL76G>T)	476.51	0.000
(g.2634+1018A>T, g.42555-29812G>A)	290.88	0.000	(g.6960A>T, LPL67G>A, LPL80G>A)	452.22	0.000
(g.6850+77A>G, g.10329C>T)	266.37	0.000	(g.42555-29812G>A, c.1523G>T, LPL7G>A)	452.22	0.000
(g.6850+77A>G, LPL80G>A)	249.59	0.000	(srebp9T>C, g.6960A>T, LPL80G>A)	450.94	0.000

SNP 조합을 찾아보았다. 이 방법은 이분화된 데이터를 이용하기 때문에 데이터마이닝 기법 중 하나인 *K*-평균군집분석방법을 사용하여 4개의 형질을 모두 포함하여 이분화 하였다. 이분화된 데이터를 이용해서 SNPHarvester 방법을 사용하였고, 여기서 사용된 스코어 함수는 χ^2 -통계량이다. 그 결과 한우의 품질에 연관있는 상위 10개 주요 SNP조합을 다음 Table 2와 같이 나타냈다.

각 SNP 조합들 중 χ^2 값이 큰 조합들 중에서 상위 10개를 Table 2에 나타내었다. 한우의 품질에 영향을 많이 받는 SNP조합에서 χ^2 값이 큰 값을 3개의 조합에서 상위 4개를 선택하였다. 상위 4개의 조합은 (g.6350+77A>G, g.13126T>C, LPL7G>A), (c.194C>T, srebp3T>C, g.10329C>T), (g.10213T>C, srebp3T>C, c.194C>T), (srebp3T>C, g.2634+1018A>T, c.194C>T)를 선택하였으며, 3조합에서 선택한 각각의 SNP들 중에서 중복으로 선택이 된 2조합을 비교를 위해서 (g.6350+77A>G, g.13126T>C), (g.10213T>C, c.194C>T), (g.2634+1018A>T, c.194C>T)을 선별하였다. 이처럼 SNPHarvester 방법의 χ^2 -통계량을 이용하여 많은 SNP들 중에서 한우의 품질에 관련된 주요한 SNP 조합을 찾을 수 있었다. 하지만 SNPHarvester방법으로 찾은 SNP조합에 어떠한 유전자형으로 이루어져 있으며, 유전자형이 각 형질의 가치를 높일 수 있는지에 대해서는 알 수가 없다. 그래서 이를 알아보기 위해서 데이터마이닝 기법 중 하나인 CART 방법을 이용하였다. 한우의 품질에 영향을 주는 주요한 SNP조합내에서 어떤 유전자형이 많은 영향을 주는지 찾아보았다.

3.3. CART 적용 결과

SNPHarvester에 의해 무수히 많은 SNP들 중에서 우수한 SNP 조합을 찾을 수 있었다. 하지만 선택된 SNP조합들의 어떤 유전자형이 한우의 품질에 관련된 형질의 가치상승에 도움을 주는지 알 수가 없다. 따라서 3.2절에서 선택된 7개의 SNP 조합들을 각각 형질별로 데이터마이닝 기법 중 하나인 CART 방법에 적용하여 우수 유전자형과 다른 유전자형간의 차이를 살펴 보았다. 한우의 맛에 연관된 올레인산(C18:1)과 포화지방산(SFA), 불포화지방산(MUFA) 그리고 육질에 관련된 근내지방도(MS)에 대해 적용을 해서 우수 유전자형과 비교한 것을 Table 3~Table 6에 나타내었다.

Table 3~Table 6은 선별된 7개의 SNP조합을 이용해서 우수 유전자형과 비우수 유전자형간의 차이를 통계적 유의성을 통해 알아 보았다. 먼저 Table3은 선택된 7개의 SNP조합과 한우의 맛과 관련된 올레인산(C18:1)의 적용 결과이다. 올레인산에서 모든 조합이 통계적으로 유의하게 나타났으며(<0.001), 이 중에서 (g.6350+77A>G, g.13126T>C, LPL7G>A)조합이 가장 큰 차이를 보였다. 우수 유전자형은 GATTGA, AACCGA, GGTAA, AACTGA, AATTGA의 유전자형이 48.00으로 다른 유전자형에 비해서 가장 높았으며, 4.04의 평균차이를 보이고 있다. *t*값도 다른 조합과 비교 했을 때, 32.858로 가장 큰 차이를 보이고 있다.

Table 3: Genotype of superior SNP combinations selected by C18:1

SNP combinations	Genotype	N	Mean ± S.D	t (p-value)
g.6350+77A>G g.13126T>C LPL7G>A	GATTGA, AACCGA, GGTTAA, AACTGA, AATTGA Others	399 4731	48.00 ± 2.35 43.96 ± 2.40	32.858 (0.000)
c.194C>T srebp3T>C g.10329C>T	CCCTCT, TTTTCT, CTTTCC, CTCTCC, CCTTCT, CCCTTT, TTTTCC Others	1242 3888	45.68 ± 2.53 43.83 ± 2.51	22.497 (0.000)
g.10213T>C srebp3T>C c.194C>T	CTTTTT, CTCTTT, CCTTCT, CCCTCC, CCCTTT Others	699 4431	46.24 ± 2.47 43.97 ± 2.52	22.575 (0.000)
srebp3T>C g.2634+1018A>T c.194C>T	CCCTAA, TTTTAA, CTTTAA, CCTTAA Others	1603 3527	45.79 ± 2.57 43.59 ± 2.36	30.023 (0.000)
g.6350+77A>G g.13126T>C g.10213T>C	GATT, AATT, AACT Others TTCC	796 4334 202	46.44 ± 2.63 43.88 ± 2.43 47.10 ± 2.75	27.006 (0.000) 15.932
c.194C>T g.2634+1018A>T c.194C>T	Others AACC Others	4928 722 4408	44.16 ± 2.56 46.17 ± 2.71 43.96 ± 2.48	(0.000) 21.822 (0.000)

Table 4: Genotype of superior SNP combinations selected by SFA

SNP combinations	Genotype	N	Mean ± S.D	t (p-value)
g.6350+77A>G g.13126T>C LPL7G>A	GATTGA, AACCGA, GGTTAA, AACTGA, AATTGA Others	399 4731	37.62 ± 2.57 40.91 ± 2.71	-24.383 (0.000)
c.194C>T srebp3T>C g.10329C>T	CCCTCT, TTTTCT, CTTTCC, CTCTCC, CCTTCT, CCCTTT, TTTTCC Others	1242 3888	39.52 ± 2.62 41.02 ± 2.82	-16.619 (0.000)
g.10213T>C srebp3T>C c.194C>T	CTTTTT, CTCTTT, CCTTCT, CCCTCC, CCCTTT Others	699 4431	39.17 ± 2.56 40.89 ± 2.81	-15.222 (0.000)
srebp3T>C g.2634+1018A>T c.194C>T	CCCTAA, TTTTAA, CTTTAA, CCTTAA Others	1603 3527	39.42 ± 2.83 41.21 ± 2.67	-21.398 (0.000)
g.6350+77A>G g.13126T>C g.10213T>C	GATT, AATT, AACT Others TTCC	796 4334 202	38.94 ± 2.93 40.97 ± 2.71 37.88 ± 2.67	-19.111 (0.000) -15.031
c.194C>T g.2634+1018A>T c.194C>T	Others AACC Others	4928 722 4408	40.77 ± 2.79 39.30 ± 3.15 40.88 ± 2.73	(0.000) -14.082 (0.000)

Table 4는 선택된 7개의 조합과 한우의 맛과 관련된 포화지방산(SFA)의 적용 결과이다. 이 결과에서도 모든 SNP조합의 우수 유전자형과 차이가 있는 것을 볼 수 있으며(<0.001), 이 중에서도 (g.6350+77A>G, g.13126T>C, LPL7G>A)조합이 가장 큰 차이를 보이고 있다. 유전자형대는 C18:1과 같으며 37.62로 가장 좋았으며, 평균이 3.3차이를 보이고 있다. t값은 -24.383으로 다른 조합보다 큰 차이를 보이고 있다.

Table 5는 SNP조합과 맛에 관련된 불포화지방산(MUFA)의 적용 결과이다. 이 결과도 전체적으로 모든 SNP조합의 우수 유전자형과 차이가 있는 것을 볼 수 있으며(<0.001), 역시 (g.6350+77A>G,

Table 5: Genotype of superior SNP combinations selected by MUFA

SNP combinations	Genotype	N	Mean ± S.D	t (p-value)
g.6350+77A>G g.13126T>C LPL7G>A	GATTGA, AACCGA, GGTTAA, AACTGA, AATTGA Others	399 4731	57.38 ± 2.78 53.15 ± 2.71	29.245 (0.000)
c.194C>T srebp3T>C g.10329C>T	CCCTCT, TTTTCT, CTTTCC, CTCTCC, CCTTCT, CCCTTT, TTTTCC Others	1242 3888	55.00 ± 2.84 53.00 ± 2.80	21.760 (0.000)
g.10213T>C srebp3T>C c.194C>T	CTTTTT, CTCTTT, CCTTCT, CCCTCC, CCCTTT Others	699 4431	55.29 ± 2.81 53.20 ± 2.86	18.239 (0.000)
srebp3T>C g.2634+1018 A>T c.194C>T	CCCTAA, TTTTAA, CTTTAA, CCTTAA Others	1603 3527	55.05 ± 2.97 52.77 ± 2.63	27.582 (0.000)
g.6350+77A>G g.13126T>C g.10213T>C	GATT, AATT, AACT Others TTCC	796 4334 202	55.78 ± 3.13 53.06 ± 2.70 56.66 ± 2.80	25.524 (0.000) 16.434
c.194C>T g.2634+1018A>T c.194C>T	Others AACC Others	4928 722 4408	53.35 ± 2.87 55.26 ± 3.07 53.19 ± 2.81	(0.000) 18.104 (0.000)

Table 6: Genotype of superior SNP combinations selected by MS

SNP combinations	Genotype	N	Mean ± S.D	t (p-value)
g.6350+77A>G g.13126T>C LPL7G>A	GATTGA, AACCGA, GGTTAA, AACTGA, AATTGA Others	399 4731	6.44 ± 1.67 5.31 ± 1.92	12.777 (0.000)
c.194C>T srebp3T>C g.10329C>T	CCCTCT, TTTTCT, CTTTCC, CTCTCC, CCTTCT, CCCTTT, TTTTCC Others	1242 3888	5.99 ± 1.82 5.21 ± 1.92	12.700 (0.000)
g.10213T>C srebp3T>C c.194C>T	CTTTTT, CTCTTT, CCTTCT, CCCTCC, CCCTTT Others	699 4431	6.13 ± 1.82 5.28 ± 1.91	10.933 (0.000)
srebp3T>C g.2634+1018A>T c.194C>T	CCCTAA, TTTTAA, CTTTAA, CCTTAA Others	1603 3527	5.85 ± 1.91 5.19 ± 1.90	11.480 (0.000)
g.6350+77A>G g.13126T>C g.10213T>C	GATT, AATT, AACT Others TTCC	796 4334 202	5.93 ± 1.82 5.30 ± 1.93 6.02 ± 2.18	8.604 (0.000) 4.725
c.194C>T g.2634+1018A>T c.194C>T	Others AACC Others	4928 722 4408	5.37 ± 1.91 5.95 ± 2.01 5.31 ± 1.89	(0.000) 8.292 (0.000)

g.13126T>C, LPL7G>A)조합의 *t*값이 29.245로 다른 조합보다 가장 큰 것을 확인하였다. 평균이 57.38로 다른 유전자형 보다 가장 높았으며, 평균과 비교해서 4.13으로 가장 큰 차이를 나타내었다. Table 6은 근내지방도(MS)의 적용 결과인데, 이 역시 SNP조합이 차이를 보이고 있으며(<0.001), (g.6350+77A>G, g.13126T>C, LPL7G>A)조합이 7개의 조합 중 가장 큰 *t*값을 나타내었다. 평균이 6.44로 가장 좋은 유전자형이라 할 수 있다.

위의 표를 보았을 때, 선택된 7개의 SNP조합들 중에서 (g.6350+77A>G, g.13126T>C, LPL7G>A) 조합의 우수 유전자형이 다른 SNP조합에 비해서 가장 큰 차이를 보이는 것을 확인하였다. 이 조합들

Table 7: Genotype of superior SNP combinations and single selected by each traits

Traits	SNP	genotype	N	Mean \pm S.D	t (p-value)
C18:1	g.6350+77A>G	AA	357	46.42 \pm 3.12	16.374
		Others	4773	44.11 \pm 2.52	(0.000)
	g.13126T>C	TT	1459	45.65 \pm 2.45	25.250
		Others	3671	43.73 \pm 2.50	(0.000)
	LPL7G>A	AA	284	45.91 \pm 2.75	10.340
		Others	4846	44.18 \pm 2.59	(0.000)
	g.6350+77A>G g.13126T>C LPL7G>A	GATTGA, AACCGA, GGTAA, AACTGA, AATTGA Others	399 4731	48.00 \pm 2.35 43.93 \pm 2.40	32.858 (0.000)
SFA	g.6350+77A>G	AA	357	38.28 \pm 2.90	-16.090
		Others	4773	40.83 \pm 2.76	(0.000)
	g.13126T>C	TT	1459	39.55 \pm 2.76	-18.039
		Others	3671	41.09 \pm 2.76	(0.000)
	LPL7G>A	AA	284	39.41 \pm 2.82	-7.612
		Others	4846	40.76 \pm 2.83	(0.000)
	g.6350+77A>G g.13126T>C LPL7G>A	GATTGA, AACCGA, GGTAA, AACTGA, AATTGA Others	399 4731	37.62 \pm 2.57 40.91 \pm 2.71	-24.383 (0.000)
MUFA	g.6350+77A>G	AA	357	56.12 \pm 3.52	18.107
		Others	4773	53.29 \pm 2.79	(0.000)
	g.13126T>C	TT	1459	54.93 \pm 2.84	23.095
		Others	3671	52.91 \pm 2.78	(0.000)
	LPL7G>A	AA	284	54.99 \pm 3.15	8.361
		Others	4846	53.39 \pm 2.90	(0.000)
	g.6350+77A>G g.13126T>C LPL7G>A	GATTGA, AACCGA, GGTAA, AACTGA, AATTGA Others	399 4731	57.38 \pm 2.78 53.15 \pm 2.71	29.245 (0.000)
MS	g.6350+77A>G	AA	357	5.42 \pm 2.05	0.242
		Others	4773	5.40 \pm 1.91	(0.809)
	g.13126T>C	TT	1459	5.83 \pm 1.97	10.109
		Others	3671	5.23 \pm 1.88	(0.000)
	LPL7G>A	AA	284	6.25 \pm 2.05	7.246
		Others	4846	5.35 \pm 1.90	(0.000)
	g.6350+77A>G g.13126T>C LPL7G>A	GATTGA, AACCGA, GGTAA, AACTGA, AATTGA Others	399 4731	6.44 \pm 1.67 5.31 \pm 1.92	12.777 (0.000)

의 유전자형태를 살펴보면 GATTGA, AACCGA, GGTAA, AACTGA, AATTGA의 유전자형이 다른 유전자형에 비해서 가장 우수한 유전자형으로 나타났다. 이 조합을 우수 SNP조합으로 선택을 하고, 위의 유전자형을 우수 유전자형으로 선택을 하였다. 선택된 SNP조합과 우수 유전자형이 실제로 한우의 품질에 어느 정도의 효과와 가치를 주는지를 알기 위해서 조합과 조합하기 이전 각각의 단일 SNP와의 비교를 통해서 알아 보았다. 조합하기 이전 3개의 단일 SNP에서 각각 우수 유전자형을 선택하고, 단일 우수 유전자형을 각 품질에 관련된 형질의 효과를 구하고, 우수 SNP 조합의 효과와 비교를 하였다. Table 7은 각 형질별로 단일의 효과와 조합의 효과를 비교한 표이다.

Table 7은 (g.6350+77A>G, g.13126T>C, LPL7G>A)조합의 우수 유전자형인 GATTGA, AACCGA, GGTAA, AACTGA, AATTGA의 효과와 조합하기 이전 단일 SNP의 우수 유전자형의 효과를 나타낸 표이다. 먼저 맛에 연관된 올레인산(C18:1)에서 조합하기 이전의 단일 SNP는 (g.13126T>C)의

TT 유전자형이 다른 단일 SNP보다 큰 차이를 보이고 있다. t 값은 25.250이고 우수 유전자형의 평균은 45.65로 단일 효과중에서 높았으며, 평균이 1.92 차이를 보이고 있다. 하지만 SNP 조합과 비교 했을 때는 조합 이전의 단일보다는 조합의 우수 유전자형이 올레인산에 큰 영향을 주고 있다. t 값이 35.858이며, 우수 유전자형의 평균이 48.00으로 가장 뛰어 났으며, 평균차이는 4.04로 단일보다 우수하다고 할 수 있다. 즉, 단일효과보다는 조합에 의한 유전자형이 우수하다고 말할 수 있다. 맛과 연관된 포화지방산(SFA)에서 조합하기 이전의 단일 SNP는 (g.6350+77A>G)의 AA 유전자형이 다른 단일 SNP보다 큰 차이를 보이고 있다. 우수 유전자형의 평균이 38.28로 다른 단일 SNP보다 좋았으며, 2.55의 차이를 보이고 있다. 조합과 비교 했을 때는 t 값이 -24.383이며, 우수 유전자형의 평균이 37.62로 단일 보다 좋았으며, 3.29의 차이를 보였다. 포화지방산 역시 조합하기 이전의 단일효과 보다는 조합에 의한 유전자형이 우수 한것을 알 수 있었다. 불포화지방산(MUFA)에서도 (g.6350+77A>G)의 AA 유전자형이 다른 단일 SNP보다 좋았다. 우수 유전자형의 평균이 56.12로 단일 효과 중에서 우수 하였으며, 2.83의 평균차이를 보였다. 하지만 조합에 의한 우수 유전자형이 57.38로서 조합 전보다 우수한 것을 알 수 있다. 한우의 육질과 관련된 근내지방도(MS)에서는 (LPL7G>A)의 AA 유전자형이 다른 단일 유전자형보다 6.25로 우수하였지만 조합에 의한 우수 유전자형이 6.44로 더 우수 한 것을 확인할 수 있다. 한우의 품질과 관련된 4개의 형질과 우수 유전자형을 비교하였을 때, 조합하기 이전의 단일효과보다는 조합에 의한 우수 유전자형이 더 많은 영향과 우수한 효과를 나타내고 있는 것을 확인하였다. 즉, 한우의 품질에 영향을 주는 SNP조합 중에서 (g.6350+77A>G, g.13126T>C, LPL7G>A)가 가장 큰 영향을 주고 있으며, 이 중에서 GATTGA, AACCGA, GGTAA, AACTGA, AATTGA의 우수 유전자형태가 단일 유전자형태보다 한우의 품질 향상에 더 많은 영향을 주는것을 확인할 수 있다.

4. 결론 및 토의

본 연구는 소비자들에게 고품질의 한우를 제공하기 위해서 한우의 품질에 영향을 미치는 형질과 SNP를 이용해서 우수 유전자를 알아보고자 하였다. 특히, 각 형질의 효과는 유전적인 효과와 환경적인 효과에 의해 결정이 되는데, 본 연구에서는 환경적인 효과를 보정하고 유전적인 효과만을 토대로 그 형질의 효과를 알고자 하였다. 본 연구의 데이터는 Oh 등에 의해 연구된 경북지역에서 자란 18개의 씨수소로 통해 얻어진 513두를 환경적인 요인을 보정한 각 형질의 효과를 사용하고, bootstrapping으로 10배 증폭시킨 5130두를 이용하였다. 각 보정된 경제형질을 K 평균군집 방법으로 이분화 시킨 후, SNPHarvester 방법을 통해 32개의 SNP들 중 한우의 품질에 연관된 주요 SNP 조합을 선별하였다. 그 결과 한우의 맛의 형질인 올레인산(C18:1), 포화지방산(SFA), 불포화지방산(MUFA), 육질의 형질인 근내지방도(MS)와 연관된 우수 SNP 조합으로 (g.6350+77A>G, g.13126T>C, LPL7G>A), (c.194C>T, srebp3T>C, g.10329C>T), (g.10213T>C, srebp3T>C, c.194C>T), (srebp3T>C, g.2634+1018A>T, c.194C>T), (g.6350+77A>G, g.13126T>C), (g.10213T>C, c.194C>T), (g.2634+1018A>T, c.194C>T)이 선별 되었다. 이 유전자조합으로는 한우 품질의 효과상승을 알 수 없기 때문에, 각 조합에서 우수 유전자형을 통해서 실제적인 효과를 살펴보았다. 선별된 7개의 유전자 조합의 우수 유전자형을 알기 위해서, 데이터마이닝 기법 중 하나인 CART 방법을 이용하여서 우수 유전자형과 비우수 유전자형을 가려내었고, 이 두 그룹을 비교해보았다. 선별된 SNP 조합에서 (g.6350+77A>G, g.13126T>C, LPL7G>A)의 유전자형태가 다른 조합에 비해서 큰 차이를 보이고 있는 것을 Table 3~Table 6를 통해서 알 수 있었다. 즉, 연구 결과 한우 품질의 가치를 올릴 수 있는 유전자 조합은 (g.6350+77A>G, g.13126T>C, LPL7G>A)이었다. 이 조합의 우수 유전자형태로는 GATTGA, AACCGA, GGTAA, AACTGA, AATTGA 유전자형으로 밝혀졌으며, 이 우수 유전자 조합의 유전자형이 조합하기 이전 3개의 단일 SNP의 유전자형태와 비교를 하였을 때 단일효과보다는 상호작용에 의

한 효과가 한우의 품질에 더 많은 영향을 주는 것을 Table 7을 통해 확인하였다. 즉, 한우 품질을 향상 시키기 위해서 단일 유전자 보다는 조합에 의한 유전자 효과가 더 많은 영향을 주는 것을 확인 할 수 있었다.

References

- 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.
- Lee, J. Y. and Jin, M. H. (2012). Major gene interaction identification in Hanwoo by adjusted environmental effects, *Journal of the Korean Data & Information Society*, **23**, 467–474.
- Lee, J. Y. and Kim, D. C. (2009). Identify major gene-gene interaction effects using SNPHarvester, *Communications of the Korean Statistical Society*, **16**, 915–923.
- 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*, **6**, 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.
- Oh, D. Y., Lee, T. 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.
- Yang, C., He, Z., Wan, X., Yang, Q., Xue, H. and Yu, W. (2009). SNPHarvester: A filtering-based approach for detecting epistatic interactions in genome-wide association studies, *Bioinformatics*, **25**, 504–511.