

방사형그래프를 활용한 한우의 품질관련 주요 유전자 상호작용 효과 규명

이제영¹ · 배재영² · 이진목³ · 오동엽⁴ · 이성원⁵

¹²³⁵영남대학교 통계학과 · ⁴영남대학교 생명공학부

접수 2012년 12월 21일, 수정 2013년 1월 7일, 게재확정 2013년 1월 17일

요약

최근 인간의 질병과 관련된 유전자 연구가 많이 진행되고 있다. 인간의 질병에 관련된 유전자는 단일효과보다 상호작용에 의한 효과가 중요하다고 한다. 무수히 많은 유전자들 중에서 상호작용을 찾기 위해서 선형모형 등의 통계적 방법을 이용할 시에는 계산의 복잡성 및 비용등 여러 단점이 생긴다. 따라서 이런 수많은 유전자들 중 주요 유전자 상호작용을 찾는 방법으로 SNPHarvester가 개발되었다. 본 연구에서는 최근에 밝혀진 한우의 지방산과 경제형질에 관련된 32개의 우수 단일 염기 다형성(SNP)를 이용하여 주요 유전자 조합을 SNPHarvester방법에 적용하여 분석한 결과를 시각적으로 쉽게 확인하기위해 방사형 그래프로 나타냄으로써 한우의 맛과 육질을 높일 수 있는 단일 염기 다형성 조합을 시각적으로 찾아내고, 각 형질별 단일효과와 상호작용 효과의 차이를 비교하여 단일효과보다 상호작용에 의한 효과가 더 큰 차이를 시각적 효과로 규명하였다.

주요용어: 단일 염기 다형성, 방사형 그래프, SNPHarvester.

1. 서론

최근 유전학에서는 인간의 질병이나 가축의 경제적인 특성에 관련된 주요 유전자와 유전자의 상호작용을 규명하는 것이 중요한 연구 과제중 하나이다. 일반적으로 인간의 질병이나 가축의 경제적인 특성은 많은 유전자 중에서 유전자 하나의 단일효과가 아니라 여러 유전자의 상호작용으로 일어난다. 과거에는 여러 유전자의 상호작용을 고려한 모형으로 선형모형과 같은 단순한 통계적 모형을 사용했다. 하지만 유전자의 수가 많아지면서 모형이 복잡해지고 해석하는데 어려움이 많이 존재한다. 또한 유전자들을 바탕으로 모형화된 경우에는 상호작용을 하였을 때, 테이블 셀에 관측 값이 없을 수도 있다. 이 경우는 추정 값과 오차 값이 더 크게 존재할 수 있다 (Hosmer와 Lemeshow, 2000). 이런 어려움을 해결하고 유전자들 간의 상호작용 효과를 보기 위해서 다양한 통계적 방법들이 제시되고 있다.

유전자들의 상호작용을 찾는 통계적인 방법으로 다중인자차원축소 (multifactor dimensionality reduction; MDR) (Ritchie 등, 2001; Chung 등, 2005), CART (classification and regression tree)방법을 활용한 확장된 다중인자 차원축소 (expanded MDR; E-MDR) (Lee 등, 2008)방법, SVM (support

¹ (712-749) 경상북도 경산시 대학로 280 영남대학교 이과대학 통계학과, 교수.

² (712-749) 경상북도 경산시 대학로 280 영남대학교 이과대학 통계학과, 석사과정.

³ (712-749) 경상북도 경산시 대학로 280 영남대학교 이과대학 통계학과, 석사과정.

⁴ (750-600) 경상북도 영주시 축산기술연구소, 영남대학교 생명공학부, 박사과정.

⁵ (712-749) 교신저자: 경상북도 경산시 대학로 280 영남대학교 이과대학 통계학과, 강사.

E-mail: rswsky@yumail.ac.kr

vector machine)을 이용한 다중인자 차원축소 (support vector machine multifactor dimensionality reduction; SVM-M DR) (Lee와 Lee, 2010)방법 등이 있다. 그러나 위의 방법들에는 적은 유전자의 수에서 유전자 조합을 찾는 데 유용한 방법으로 수많은 유전자들 중에서 유전자 조합을 찾을 경우에는 계산이 복잡하고, 비용이 많이 들고, 시간이 오래 걸린다는 단점이 존재한다. 위의 방법들의 단점을 해결하기 위해서 수많은 유전자들 중에서 빠르고 쉽게 주요 유전자 조합을 찾는 통계적 방법으로 SNPHarvester (Yang 등, 2009)방법이 개발되었다.

SNPHarvester 방법은 무수히 많은 유전자들 중 인간의 질병 유무에 연관 있는 주요 유전자 그룹을 찾는 방법으로 사례-대조와 같은 이분형 데이터에 적용이 가능하다. 이 방법을 통해서 질병과 관련된 주요 유전자들의 조합을 찾아 낼 수 있다. 그리고 무수히 많은 유전자들 중 주요 유전자 그룹을 찾는 방법으로 유전자 그룹 내부의 유전자형 (genotype)이 어떤 경우에 인간의 질병을 줄이거나 가축의 경제형질을 높일 수 있는지 알 수 없다. 그래서 다양한 통계적 방법을 사용하여 우수한 유전자형을 찾아본다. 즉, SNPHarvester 방법을 이용하여 주요 유전자 그룹을 찾아낸 후 다양한 통계적 방법을 이용하여 우수 유전자형과 비우수 유전자형을 선별한다.

한우의 경제형질은 유전적인 효과와 환경적인 효과로 나타난다고 한다. 본 연구에서는 환경적인 효과를 제외하고 유전적인 효과만을 이용하여 우수한 유전자그룹을 찾기 위해 환경적인 효과를 보정한 자료를 이용하여 분석 한다. 분석에 사용된 경제형질로는 한우의 맛과 육질에 영향을 미치는 4개의 경제형질로 올레인산 (oleic acid; C18:1)과, 포화지방산 (saturated fatty acid; SFA), 불포화지방산 (monounsaturated fatty acid; MUFA), 그리고 근내지방도 (marbling score; MS) 이다 (Oh 등, 2011). 앞에서 소개한 바와 같이 SNPHarvester는 이분형 데이터에 적용이 가능하기 때문에 4개의 경제형질을 데이터마이닝 기법을 사용하여 주요 유전자 그룹을 찾아낸 뒤, 한우의 경제형질을 높일 수 있는 유전자형을 얻을 수 있다. (Park과 Cho, 2012). 이러한 주요 유전자 조합과 유전자형을 방사형 그래프에 적용함으로써 우수 유전자조합과 우수 유전자형을 시각적으로 쉽게 볼 수 있도록 나타내었다.

본 논문의 2절에서는 통계모형과 SNPHarvester 방법에 대하여 소개를 하고, 3절에서는 SNPHarvester방법을 소개하고, SNPHarvester기법을 통해 얻은 결과를 방사형 그래프에 적용하여 시각적 효과를 규명하였다. 4절에서 그 결과를 요약 하였다.

2. 회귀모형과 SNPHarvester 소개

2장에서는 본 연구에서 사용한 통계적 방법에 대해서 소개 한다. 2.1절에서는 유전적 효과만을 보기 위한 환경요인을 보정한 모형, 2.2절에서는 주요 유전자 조합을 찾기 위해 사용한 SNPHarvester 방법에 대해서 소개를 한다.

2.1. 유전적 효과만을 보기 위한 모형

실제로 수집된 한우의 각 형질들은 유전적 분석에 있어서 기본이 되는 개념으로 각 형질들은 표현형으로 되어 있으며, 이 표현형은 환경적요인과 유전적 요인에 의해서 효과가 결정이 된다. 즉 형질의 표현형은 다음과 같이 두 부분으로 나눌 수 있다.

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

여기서 P는 각 형질의 표현형으로써 그 효과의 값이 되며, 환경적 요인인 E와 유전적 요인인 G에 의해서 그 효과의 값이 결정이 된다고 연구되었다. 그래서 형질은 환경적 요인과 유전적 요인에 의해서 다음과 같은 선형 회귀모형이 사용이 된다 (Casas 등 2005).

$$y_{ijk} = \mu + Farm_i + \beta Age + SNP_j + e_{ijk}, \quad i = 1, \dots, f; j = 1, \dots, m; k = 1, \dots, n \quad (2.2)$$

식 (2.2)에서 y_{ijk} 는 형질 각각의 효과가 되며, 환경적 요인의 Farm, Age는 한우의 사육 농장과 사육 일령이며, β 는 사육일령에 대한 회귀계수가 된다. 유전적 효과는 SNP이며, e_{ijk} 는 $N(0, \sigma^2)$ 인 확률변수이다. 하지만 본 논문에서는 유전적인 효과에만 관심이 있기 때문에 환경적 요인을 보정한 선형회귀모형을 이용한다 (Matsuhashi 등, 2011). 식 (2.3)에 환경적 요인을 보정한 선형회귀모형을 나타내었다 (Lee와 Jin, 2012).

$$y_{ijk} - (Farm_i + \beta Age) = \mu + SNP_j + e_{ijk}, \quad i = 1, \dots, f; \quad j = 1, \dots, m; \quad k = 1, \dots, n \quad (2.3)$$

식 (2.3)에서 환경적 요인을 보정한 선형회귀모형을 분석하기 위해 Minitab 버전 16을 활용하였다.

2.2. SNPHarvester 소개 및 절차

수많은 SNP들에서 주요 유전자 조합을 찾기 위해서, 일반적인 통계방법들을 이용 할 때에는 시간과 비용이 많이 들며, 계산이 복잡해지는 한계점을 가지고 있다. 이런 한계점을 보완하고 쉽고 빠르게 주요 유전자 조합을 찾기 위해서 SNPHarvester방법 이 개발이 되었다 (Yang 등, 2009). 즉 이 방법은 질병과 관련된 수많은 SNP들 중에서 영향을 많이 미치는 조합으로 선별이 가능하다. 어떠한 특성 값과 관련된 SNP들 중에서 초기에 k 개의 SNP를 이용해서 조합으로 선택 한 뒤, 그 조합에서 선택되지 않은 SNP들 중 하나를 초기에 선택된 SNP들과 바꾸면서, 스코어를 계산한다. 그 스코어의 비교와 그 과정의 반복을 통해 주요 SNP조합을 선별하는 방법이다. 이 방법에서 스코어는 B-통계량, χ^2 -통계량 등을 이용하여서 선별할 수 있지만, 본 연구에서는 χ^2 -통계량을 스코어 함수로 이용한다. 이 방법을 이용하면 많은 SNP들 중 주요 SNP 그룹을 쉽고 빠르게 찾을 수 있다는 장점을 가지고 있다. 아래의 단계는 SNPHarvester방법의 절차를 나타낸 것이다 (Lee와 Bae, 2012).

단계 0. 전체 SNP그룹을 정한다.

예) L개의 SNP : $SNP_1, SNP_2, \dots, SNP_L$

단계 1. 전체 L개의 SNP그룹에서 랜덤으로 k 개의 SNP를 선택한 후, A라는 그룹을 선택한다. 그리고 나머지 SNP를 SNP_i 로 둔다.

예) 그룹 A = $SNP_{s1}, SNP_{s2}, \dots, SNP_{sk}$

SNP_i ($SNP_i \notin A$), $i = 1, \dots, L$.

단계 2. 포함되지 않은 SNP들 중 하나와 A그룹의 SNP를 1:1로 바꿔가면서 스코어를 낸다.

예) $A_1 = SNP_i, SNP_{s2}, \dots, SNP_{sk} \Rightarrow \text{Score}(A_1)$

$A_2 = SNP_{s1}, SNP_i, \dots, SNP_{sk} \Rightarrow \text{Score}(A_2)$

⋮

$A_k = SNP_{s1}, SNP_{s2}, \dots, SNP_i \Rightarrow \text{Score}(A_k)$

단계 3. 단계2에서 나온 각각의 스코어 중 가장 큰 값을 A^* 로 정한다.

단계 4. 단계3에서 나온 A^* 그룹의 스코어와 초기 A그룹의 스코어를 비교하여

$A^* > A \Rightarrow A^*$ 그룹으로 교체

$A^* < A \Rightarrow A$ 그룹으로 유지

단계 5. 단계4에서 그룹이 결정이 되었다면, SNP_{i+1} 를 이용하여 단계2~단계4를 반복한다.

단계 6. 단계2~단계5 (반복수 $q=2$ 혹은 3)를 반복하며, 최종으로 나온 A^* 가 이전의 A보다 작을 경우 멈추게 된다.

위의 단계1~단계6까지 반복을 통해 주요 SNP조합을 선별하는 것이 SNPHarvester가 된다. 그리고 반복을 통하여 스코어 값을 높이더라도, SNP조합이 통계적으로 유의하지 않다면 아무런 의미가 없어지기 때문에 유의성 검정은 본페로니 수정 (Bonferroni corrections)에 의해 $\alpha=0.01$ 로 검정을 한다. 그리고 A의 그룹수인 k 가 클수록 SNP그룹 내의 유전자형에서의 빈도가 없거나 작아질 수 있으므로 k 를 $k \leq \ln_3 N_d - 1$ 로 제한한다. 여기서 N_d 는 사례의 수이다. 이러한 SNPHarvester 방법을 이용하여 한우의 경제형질에 관련된 주요 SNP 조합을 찾을 수 있지만, 그 SNP 조합의 효과와 그 유전자형은 알 수가 없기 때문에 데이터마이닝기법을 이용해서 그 효과와 유전자형을 확인할 수 있다.

3. 통계적 기법 적용결과

3.1. 실험자료

본 연구에서 사용된 데이터는 대구경북한우 클러스트 사업단으로부터 Oh 등 (2011)에 의해서 수집된 자료이다. 경북지역에서 18개의 씨수소를 통해 얻어진 513두와 32개의 SNP를 이용하였으며, 환경적 요인인 사육농장 17곳과 사육일령 744~1240일인 한우와 SNP의 유전적인 요인을 이용하였다. Oh 등 (2011)은 한우의 맛과 육질에 영향을 주는 형질 4개를 찾아내었다. 맛에 영향을 주는 것으로 알려진 형질로 올레인산 (oleic acid; C18:1)과 포화지방산 (saturated fatty acid; SFA), 불포화지방산 (monounsaturated fatty acid; MUFA)이 있으며, 육질에 영향을 주는 근내지방도 (marbling score; MS)를 활용하였다. Oh 등 (2011)에 의해 한우의 맛과 육질에 영향을 주는 단일염기다형성(single nucleotide polymorphism; SNP)중 단일 요인으로써 주요한 영향을 주는 32개의 SNP들을 이용하였다. 본 연구에서 사용된 SNP들은 2012년도 Lee와 Bae의 연구에서 사용한 32개의 SNP들이다 (Lee와 Bae, 2012).

SNPHarvester 방법은 데이터가 적을수록, 초기에 생성되는 조합의 수가 적어지기 때문에 데이터를 붓스트랩으로 10배 증폭시킨 5130두를 분석에 이용하였다.

3.2. SNPHarvester에 의한 주요 SNP 그룹선별

SNPHarvester는 수많은 SNP들을 조합으로 만들어서 주요한 SNP조합을 선별하는 것이다. 따라서 증폭된 데이터의 5130두와 32개의 SNP를 이용해서 한우의 맛과 육질에 관련된 주요 SNP조합을 찾아보았다. 이 기법은 이분화된 데이터를 이용하기 때문에 수치형으로 이루어진 형질들을 이분화 하기 위해서 데이터마이닝 기법중 하나인 K-평균군집분석방법을 4개의 형질을 모두 포함하여 이분화 하였다. 이분화 시킨 데이터를 SNPHarvester에 적용하였으며, 여기서 사용된 스코어 함수는 χ^2 -통계량을 이용하였다. 그 결과 χ^2 -통계량 순으로 한우의 품질에 관련된 상위 10개 주요 SNP 조합이 Lee와 Bae의 연구를 통해서 선별이 되었다 (Lee와 Bae, 2012).

Lee와 Bae (2011)는 7개의 우수 SNP 조합을 선별하였으며, 이 조합들은 통계적으로 유의성을 만족하였다 ($p < 0.001$). 아울러 이들 유전자 조합에서 우수 유전자형을 밝힐 수 있었다. 이 논문에서는 이들 규명된 효과를 시각적으로 쉽게 확인하기 위해서 방사형 그래프를 활용하여 분석하였다.

3.3. 주요 SNP 조합의 방사형 그래프 적용결과

Lee와 Bae (2012)의 연구에서 밝혀진 우수 SNP 조합들을 시각적으로 쉽게 비교하기 위해서 방사형 그래프를 통해서 그 효과를 확인해 보았다. 4개의 형질별로 SNP 조합들의 유전자형 간의 효과는 통계적으로 유의성을 만족하였다 ($p < 0.001$). 형질 중에서 먼저 한우의 맛과 관련된 올레인산 (C18:1)의 효과를 방사형 그래프를 통해 나타내었다. 선택된 SNP의 모든 조합에서 유전자형 간에 차이가 있었으며 ($p < 0.001$), Figure 3.1의 C18:1을 통해서 이 조합들 중에서 (g.6350+77A>G, g.13126T>C, LPL7G >A)조합의 우수 유전자형인 GATTGA, AACCGA, GGTAA, AACTGA, AATTGA가 ‘평

‘평균±표준오차’가 48.00 ± 2.35 로 가장 뛰어난 효과를 보여주고 있으며, 우수 유전자형과 비우수 유전자형의 ‘평균±표준오차’가 43.96 ± 2.40 로 평균차가 4.04로 가장 큰 차이를 보이고 있는 것을 방사형 그래프를 통해서 매우 쉽게 확인 할 수 있었다.

Figure 3.1의 SFA는 한우의 맛에 영향을 주는 포화지방산에 대한 방사형 그래프이다. 이 형질은 값이 작을수록 한우에 좋은 영향을 주는 것으로, 이 형질에서도 SNP의 조합에서 유전자형 간에 차이가 있었다 ($p < 0.001$). 포화지방산에서도 (g.6350+77A>G, g.13126T>C, LPL7G>A)조합 우수 유전자형의 ‘평균±표준오차’가 37.62 ± 2.57 로 다른 조합에 비해서 뛰어난 효과를 보여주고 있으며, 우수 유전자형과 비우수 유전자형의 ‘평균±표준오차’인 40.91 ± 2.71 로 평균차 -3.29로 가장 큰 차이를 보이고 있는 것을 방사형 그래프를 통해서 쉽게 확인 할 수 있었다.



Figure 3.1 Radial graph on genotype

Figure 3.1의 MUFA는 불포화지방산에 대한 효과를 방사형 그래프로 나타낸 것이다. 이 형질에서도 모든 SNP 조합에서 차이를 보이고 있으며 ($p < 0.001$), (g.6350+77A>G, g.13126T>C, LPL7G>A) 조합의 우수 유전자형의 ‘평균±표준오차’가 57.38 ± 2.78 로 다른 조합에 비해서 뛰어난 효과를 보여주고 있으며, 우수 유전자형과 비우수 유전자형의 ‘평균±표준오차’인 53.15 ± 2.71 로 평균차 4.23으로 가장 큰 차이를 보이는 것을 Figure 3.1을 통해서 아주 쉽게 확인 할 수 있었다. 선택된 조합의 우수 유전자형이 그 효과가 가장 크게 나타났으며, 비 우수 유전자형과의 차이도 가장 크게 나타나는 것을 확인하였다. 불포화지방산에서도 이 조합이 가장 우수한 SNP조합이며, 우수 유전자형임을 확인 할 수 있었다.

Figure 3.1의 MS는 한우의 육질과 관련된 근내지방도의 방사형 그래프로써, (g.6350 +77A>G, g.13126T>C, LPL7G>A) 조합이 다른 조합에 비해서 평균이 6.44로 가장 높은 효과를 나타내었으며, 유전자형간의 차이에서도 4.23으로 가장 큰 차이를 보이는 것을 쉽게 확인 할 수 있다. 근내지방도 역시 이 조합을 우수 SNP조합과 우수 유전자형임을 확인 하였다.

앞의 방사형 그래프를 통해서 주요 SNP 조합 중에서 한우의 맛과 육질에 가장 우수한 영향을 미치는 조합으로 (g.6350+77A>G, g.13126T>C, LPL7G>A)조합, 그 유전자형으로 GATTGA, AACCGA, GGTAA, AACTGA, AATTGA가 우수함을 확인 할 수 있었다. 이 우수 SNP 조합과 유전자형의 가

치와 효과를 확인하기 위해서 단일과 비교를 통해서 알아보았다 (Lee와 Bae, 2012). 이 논문에서는 조합과 단일의 효과를 시각적으로 쉽게 확인하기 위해서 방사형 그래프를 활용하여 분석하였다.

3.4. 우수 SNP 조합과 단일 SNP의 방사형 그래프 적용 결과

각각의 형질들과 7개의 주요 SNP 조합들을 비교하였을 때, (g.6350+77A>G, g.13126T>C, LPL7G >A) 조합이 가장 우수한 유전자 조합으로 선별이 되었으며, GATTGA, AACCGA, GGTTAA, AACTGA, AATTGA가 우수 유전자형으로 선별이 되었다 (Lee와 Bae, 2012). 이 조합과 유전자형의 효과나 가치가 단일의 우수 유전자형의 효과와 비교한 것을 Figure 3.2인 방사형 그래프를 통해 확인해 보았다. Lee와 Bae (2012)의 연구에서 각 형질별로 단일 SNP의 유전자형간의 차이가 통계적으로 유의한 것을 확인 하였다 ($p < 0.001$). 먼저, 올레인산 (C18:1)에서 단일 모두 차이를 보이고 있으며($p < 0.001$), 이 중에서 (g.6350+77A>G)의 SNP가 다른 단일보다는 우수 유전자형의 효과와 유전자형 간의 차이가 가장 좋은 것을 확인 할 수 있었다. 하지만 단일 보다는 조합에 의한 효과 (우수유전자 평균±표준오차 = 48.00±2.35, 비우수 평균±표준오차 = 43.24±2.40)가 유전자형 간에도 큰 평균차 = 4.07를 보이는 것을 방사형 그래프를 통해 쉽게 확인 할 수 있었다.

마찬가지로 포화지방산 (SFA)의 경우에서도 단일 모두 통계적으로 유의한 차이를 보이고 있었으며 ($p < 0.001$), 이 중에서 그 값이 가장 작은 (g.6350+77A>G)가 우수 하였다. 하지만 조합과 함께 비교하였을 때는 단일에 비해서 그 값도 작으며, 유전자형간의 차이도 크다는 것을 쉽게 확인 할 수 있었다. 불포화지방산 (MUFA)과 근내지방도 (MS)역시 단일 보다는 조합에 의한 효과가 우수하다는 것을 방사형 그래프를 통해서 확인 하였다. 즉, 한우의 맛과 육질에 영향을 주는 SNP 조합 중에 (g.6350+77A>G, g.13126T>C, LPL7G>A)조합이 가장 큰 영향을 주고 있으며, GATTGA, AACCGA, GGTTAA, AACTGA, AATTGA의 우수 유전자 형태가 단일 유전자형태보다 한우의 품질 향상에 더 많은 영향을 주는 것을 방사형 그래프를 통해서 확인 할 수 있다.

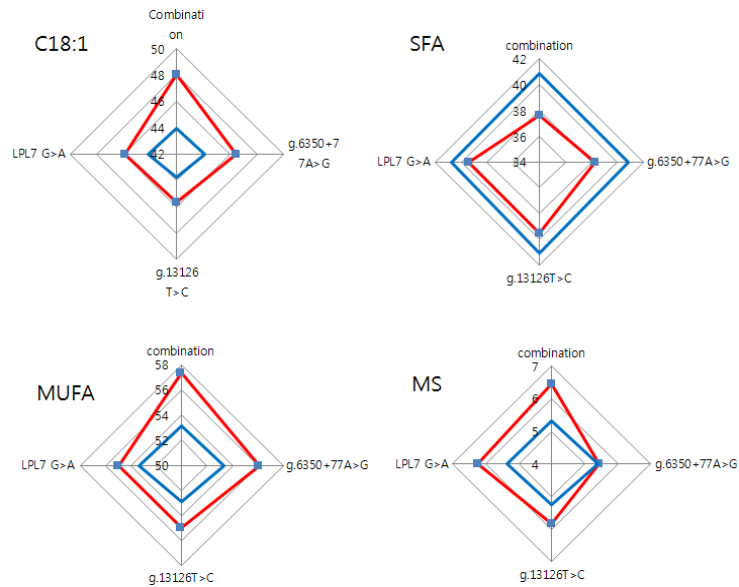


Figure 3.2 Radial graph for each character of interaction effect and single effect

4. 결론 및 토의

본 연구는 Lee와 Bae (2012)를 통해 통계적 기법인 SNPHarvester 방법을 이용하여 밝혀진 우수 SNP 조합과 유전자형의 효과를 시각적으로 쉽게 확인하기 위해서 방사형 그래프를 활용하였다. Lee와 Bae (2012)의 연구에서 우수 SNP 조합과 유전자형을 찾기 위해서 연구의 데이터는 대구경북한우 클러스터 사업단에서 Oh 등 (2011)에 의해 연구된 최근 513두의 한우 지방산과 경제형질 관련 32개의 우수 SNP들을 이용하였다. SNP만의 유전적 효과만을 분석하기 위해서 수집된 자료에 환경적 요인인 사육 농장과 사육일령을 보정하였으며, 그리고 충분한 표본자료를 생성하기 위해 붓스트랩으로 10배 증폭시킨 5130두를 이용하였다. 우수 SNP들의 조합 결합 시 MDR 등의 방법에서는 표본이 커지면 상호작용에서 빈 셀이 많아지는 단점을 가지고 있다. 이런 경우 SNP들을 쉽고 빠르게 조합을 찾을 수 있는 통계적 방법인 SNPHarvester 방법을 활용하였다. 각 보정된 경제형질을 SNPHarvester 방법을 통해 32개의 SNP들 중 한우의 품질에 연관된 우수 SNP 조합이 선별되었다. 그 결과 한우의 맛의 형질인 올레인산 (C18:1), 포화지방산 (SFA), 불포화지방산 (MUFA), 육질의 형질인 근내지방도 (MS)와 연관된 7개의 SNP조합을 선별되었다. 7개의 SNP조합의 효과를 시각적으로 쉽게 나타내어 비교하기 위해 본 논문에서는 방사형 그래프를 활용하여 분석하였다. 그 결과 SNP 조합은 (g.6350+77A>G, g.13126T>C, LPL7G>A)조합의 우수 유전자 형태가 다른 조합의 우수 유전자에 비해서 큰 값을 가지면서, 유전자들 간에는 가장 큰 차이를 보이고 있는 것을 Figure 3.1의 방사형 그래프를 통해서 알 수 있었다. 즉, 연구 결과 SNPHarvester 기법을 통해 얻은 결과를 시각적으로 쉽게 나타내기 위해 방사형 그래프에 적용하여 한우 품질의 가치를 올릴 수 있는 유전자 조합은 (g.6350+77A>G, g.13126T>C, LPL7G>A)이며, 이 조합의 우수 유전자 형태로는 GATTGA, AACCGA, GGTAA, AACTGA, AATTGA 유전자 형이 가장 우수함을 확인하였다. 그리고 이 우수 유전자 조합의 유전자형이 조합하기 이전 3개의 단일 SNP의 유전자 형태와 비교를 하였을 때, 단일효과보다는 상호작용에 의한 효과가 한우의 품질에 더 많은 영향을 주는 것을 Figure 3.1의 방사형 그래프를 통해서 규명하였다. 즉, 한우 품질을 향상시키기 위해서 단일 유전자 보다는 조합에 의한 상호작용 효과가 한우의 품질에 더 많은 영향을 주는 것을 방사형 그래프를 통해 쉽게 규명 할 수 있었다.

참고문헌

- Casas, E., White, S. N., Riley, D. G., Smith, T. P. I., Brennehan, 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.
- 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 of the Korean Statistical Society*, 31-36.
- Hosmer, D. W. and Lemeshow, S. (2000). *Applied logistic regression*, John Wiley & Sons, New York.
- Lee, J. Y. and Bae, J. Y., (2011). The effect investigation of the delirium by Bayesian network and radial graph. *Journal of the Korean Data & Information Science Society*, **22**, 911-919.
- Lee, J. Y. and Bae, J. Y. (2012). Main gene combinations and genotype identification of Hanwoo quality with SNPHarvester. *Communications of the Korean Statistical Society*, **19**, 799-808.
- Lee, J. Y. and Jin, M. H. (2012). Major gene interaction identification in Hanwoo by adjusted environmental effect. *Journal of the Korean Data & Information Science Society*, **23**, 464-474.
- Lee, J. Y., Kwon, J. C. and Kim, J. J. (2008) Multifactor dimensionality reduction(MDR) analysis to detect single nucleotide poly-morphisms 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.

- Matsuhashi, T., Maruyama, S., Uemoto, Y., Kobayashi, N., Mannen, H., Abe, T., Sakaguchi, S. and Kobayashi, E. (2011). Effects of bovine fatty acid synthase, stearoyl-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, 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.
- Park, H. C. and Cho, K. H. (2012). A study on decision tree creation using marginally conditional variables. *Journal of the Korean Data & Information Science Society*, **23**, 299-307.
- 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.

Major gene interactions effect identification on the quality of Hanwoo by radial graph

Jea-Young Lee¹ · Jae-Young Bae² · Jin-Mok Lee³ · Dong-yep Oh⁴ · Seong-Won Lee⁵

¹²³⁵Department of Statistics, Yeungnam University

⁴School of Biotechnology, Yeungnam University

Received 21 December 2012, revised 7 January 2013, accepted 17 January 2013

Abstract

It is well known that disease of human and economic traits of livestock are affected a lot by gene combination effect rather than a single gene effect. But existing methods have disadvantages such as heavy computing, many expenses and long time. In order to overcome those drawbacks, SNPHarvester was developed to find the main gene combinations among the many genes. In this paper, we used the superior gene combination which are related to the quality of the Korean beef cattle among sets of SNPs by SNPHarvester, and identified the superior genotypes using radial graph which can enhance various qualities of Korean beef among selected SNP combinations.

Keywords: Radial graph, SNP, SNPHarvester.

¹ Professor, Department of Statistics, Yeungnam University, Kyongsan 712-749, Korea.

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

³ Graduate student, Department of Statistics, Yeungnam University, Kyongsan 712-749, Korea.

⁴ Graduate student, School of Biotechnology, Yeungnam University, Livestock Research Institute, Yeungju 750-600, Korea.

⁵ Corresponding author: Instructor, Department of Statistics, Yeungnam University, Kyongsan 712-749, Korea. E-mail: rswsky@yumail.ac.kr