

베이지안 회귀를 이용한 국내 홀스타인 젖소의 유량형질 관련 DGAT1유전자 효과 검증[†]

조광현¹ · 조충일² · 박경도³ · 이준호⁴

¹²농촌진흥청 국립축산과학원

³⁴한경대학교 유전정보연구소

접수 2015년 7월 20일, 수정 2015년 10월 29일, 게재확정 2015년 11월 24일

요약

젖소의 유생산 형질에 가장 큰 영향을 미치는 유전자들 중 하나로 알려진 DGAT1 유전자의 효과를 국내 젖소 종축의 고밀도 유전체 정보를 이용하여 검증하기 위하여 본 연구를 수행하였다. 국내 젖소 씨수소로 구성된 353두의 고밀도 유전체 정보, 혈통, 추정 육종가 및 신뢰도 정보를 수집하였으며, 단일염기다형성 효과를 추정하기 위한 종속변량으로 가장 정확한 유전체 육종가를 예측할 수 있는 DeRegressed EBV를 산출하여 분석에 이용하였다. BovineSNP50 v2 패널을 이용하여 구명한 고밀도 유전자형 정보 중 유효성검증 과정을 통하여 41,051개 SNP을 선정하였으며, 각 단일염기다형성의 실제적 유전체 육종가 기여도를 확인하기 위하여 유전체 선발방법 중 하나인 베이스B ($\pi=0.99$) 방법을 이용하여 SNP 효과를 추정하였다. 1메가 베이스페어의 구간으로 구성된 유전체 전장의 2,516개 윈도우 별 유전분산 설명력을 계산한 결과 상위 1, 3 윈도우가 DGAT1유전자 주변에서 발견되었으며, 이 두 윈도우의 유전분산 설명력은 각각 0.51% 및 0.48%인 것으로 나타났다. DGAT1유전자는 유전체 선발에 상업적으로 이용되는 50k SNP chip에 포함되어있지 않기 때문에 직접적인 유전자의 효과가 명확하게 드러나지는 않지만 DGAT1 유전자에 인접한 단일염기다형성들 간의 연관불평형에 의하여 주변 윈도우에서 가장 높은 유전분산 설명력을 보이는 것으로 사료된다.

주요용어: 단일염기다형성, 베이지안 회귀, 유량, 젖소, DGAT1 유전자.

1. 서론

최근 고밀도 SNP (Single Nucleotide Polymorphism) chip이 상용화되면서 가축의 종축선발을 위한 유전능력평가에 유전체 정보를 활용하는 유전체 선발 기법이 광범위하게 이용되고 있다. 기존의 표현형과 혈통정보를 이용하여 종축의 유전능력을 평가하는 방법보다 더 높은 정확도를 보이고 있으며, 특히 조기선발에서도 높은 정확도를 보여 많은 유전적 개량량을 유발하는 새로운 육종기술로 자리매김하고 있다 (Meuwissen 등, 2001). 특히 젖소의 개량에 이러한 유전체 선발 방법이 가장 선구적으로 연구되어 안정적 적용단계에 이르고 있으며 (Hayes 등, 2009; Boichard 등, 2015), 더 나아가 국가 간 유전능력평가에도 유전체 정보를 활용하는 방법이 제안되고 있다 (Mrode 등, 2010).

[†] 본 논문은 농촌진흥청 연구사업 (세부과제번호: PJ009260)의 지원에 의해 이루어진 것임

¹ (331-801) 충남 천안시 서북구 성환읍 신방1길 114, 농촌진흥청 국립축산과학원, 연구사.

² (331-801) 충남 천안시 서북구 성환읍 신방1길 114, 농촌진흥청 국립축산과학원, 박사후연구원.

³ (456-749) 경기도 안성시 중앙로 327, 국립한경대학교 유전정보연구소, 연구교수.

⁴ 교신저자: (456-749) 경기도 안성시 중앙로 327, 국립한경대학교 유전정보연구소, 전임연구원.

E-mail: zoonolee@gmail.com

이러한 유전체 선발 기술은 양적형질좌위를 (Quantitative Trait Loci; QTL) 탐색하거나 검증하는 기술에도 활용되고 있다 (Saatchi 등, 2014; Kemper 등, 2015). 기존의 QTL 탐색에는 하나의 SNP 유전자형과 표현형과의 연관성을 분석하는 단일표지인자회귀 또는 단일표지인자분석 (Single Marker Regression; SMR or Single Marker Analysis; SMA)을 이용하였으나, SNP 마커 효과를 고정효과 (fixed effect)로 간주하여 유전력이 고려되지 않고, 모든 효과들을 동시에 모델링하는 것이 아니기 때문에 과대추정이 발생할 수 있어 최근의 QTL 탐색은 유전체 선발기법을 이용하는 추세이다. 또한 SMR이나 SMA에서 다수의 SNP 마커를 이용할 경우 다중검정에 의한 보정이 요구된다 (Oh 등, 2014; Seo 등, 2014). 유전체 선발방법은 실제적인 유전자의 효과를 구명하기보다는 이러한 QTL유전자와 연관불평형상태로 존재하는 주변 마커에 의하여 그 효과를 구명한다. 따라서 분석에 이용되는 모든 SNP 마커들이 형질발현에 관여하는 QTL과 연관불평형이 존재하도록 다수의 SNP를 이용해야 하며, 유전체 전장에 촘촘히 분포해야 한다 (Meuwissen 등, 2001). 이러한 고밀도 SNP유전정보를 이용할 경우 유전체 내 특정 부위에서 형질에 연관된 SNP효과가 크게 나타나며, 이러한 SNP의 물리적 위치를 추적하여 QTL을 탐색하거나 연관불평형을 이루는 SNP들의 효과를 합하여 그 효과의 크기를 가능하게 된다.

젓소의 14번 염색체 도입부에 존재하는 DGAT1 (Diacylglycerol O-acyltransferase1)유전자는 유생 산 형질에 깊이 관여하고 있다는 연구 결과가 과거부터 다수 보고되었으며 (Grisart 등, 2002; Schennink 등, 2007), 국내에서도 위에서 언급한 SMA를 이용하여 DGAT1유전자의 직접적인 유전자형에 따른 유의성을 분석한 결과가 보고되었다 (Son 등, 2009년). 그러나 가장 일반적으로 젓소 개량에 이용되는 50k SNP chip에는 이 유전자의 변이가 직접적으로 포함되어 있지 않다. 젓소의 개량에 DGAT1유전자의 효과가 적용되기 위해서는 50k SNP chip을 이용한 경우에도 DGAT1유전자 주변 SNP들에서 연관 불평형을 통하여 그 효과가 나타나는지 확인해야 하며, DGAT1유전자의 효과가 나타나지 않을 경우 직접적인 DGAT1유전자의 유전자형을 구명하여 분석에 포함시켜야 할 것이다. 또한 SMR 또는 SMA 방법을 이용할 경우 SNP효과를 고정효과로 간주하여 유전력이 고려되지 못하고 나머지 유전체정보가 형질발현에 관여하는 것을 보정하지 못하여 과대추정되기 때문에 유의성은 확인할 수 있으나 실제적인 육종가의 기여 정도는 확인하기 어렵다. 따라서 본 연구는 DGAT1유전자의 효과가 유전체 선발기법을 적용한 고밀도 유전체 전장분석에서 어떻게 나타나는지 확인하고 그 효과의 유전분산 기여도를 가능하고자 본 연구를 수행하였다.

2. 자료와 연구방법

2.1. 연구자료

국내 젓소 씨수소로 구성된 353두의 고밀도 유전체 정보, 혈통, 유량에 대한 추정 육종가 (Estimated Breeding Value; EBV) 및 신뢰도 정보를 수집하였다. 대상 개체들의 고밀도 SNP 유전자형 구명은 Illumina BovineSNP50 v2 패널을 이용하였으며, Minor allele frequency 0.01 미만, 결측률 10% 이상, Hardy-Weinberg법칙에 위배 (χ^2 test, $p < 0.001$), 위치정보 누락 및 성염색체상의 SNP들은 제거하여 41,051개 SNP정보를 분석에 이용하였다.

2.2. Deregressed breeding value 계산

국내 젓소의 유전능력평가에 따라 추정된 육종가를 이용하여 SNP효과를 추정하기 위한 종속변량으로 이용되는 Deregressed EBV를 계산하였다. SNP효과 추정에 표현형을 이용할 경우 씨수소는 우유를 생산하지 못하기 때문에 표현형이 없으며, 딸소들의 기록들과 혈연관계를 이용하여 추정된 씨수소의 육종가를 이용할 경우 육종가 추정모형을 통하여 유전 부분 이외의 임의 잔차값이 제거되어 표현형에 비

하여 유전력만큼 분산이 감소된 상태이기 때문에 SNP 효과를 추정하기 위한 유전분산과 잔차분산을 가정하기 어렵다. 또한 씨수소별 추정 육종가는 딸소의 기록 수와 혈연관계에 따라 신뢰도가 각각 다르다. 따라서 육종가의 분산을 표현형 수준까지 증가시키고 여기에 추정 육종가 신뢰도에 따라 가중치를 부여하는 Deregressed EBV(DEBV)를 계산하여 분석에 이용하였다. 유전체선발 모형에서의 종속변량으로는 표현형, 육종가, 자손의 표현형 평균 및 Deregressed EBV등이 이용될 수 있으며, 이중 DEBV를 종속변량으로 이용하였을 경우에 가장 높은 정확도를 나타내는 것으로 보고되고 있어 (Ostersen 등, 2011) 젖소의 유전체 육종가 추정 및 QTL 탐색에 많이 이용되고 있다 (Tiezzi 와 Maltecca, 2015; Ding 등, 2013). DEBV의 계산을 위하여 유전체정보를 생성한 모든 개체들과 부모개체들의 육종가 및 신뢰도를 수집하였으며, Garrick 등 (2009)이 제시한 방법을 이용하여 DEBV를 계산하였다. 우선 각 씨수소의 육종가에서 부모의 능력 (Parent average of breeding value)을 보정하고 여기에 각 개체별 가중치를 곱하여 DEBV를 계산하였다. 육종가에 곱해지는 가중치는 아래와 같다.

$$w_i = \frac{(1 - h^2)}{[c + (1 - r_i^2)/r_i^2]h^2} \quad (2.1)$$

여기서, w_i 는 i 번째 개체의 육종가에 곱해지는 가중치, c 는 분석에 이용된 약 4만 개의 SNP 유전자형이 모든 유전체 정보를 포함하고 있는 것은 아니기 때문에 전체 유전분산 중 SNP 마커로 설명되지 못하는 비율로 0.1을 적용하였으며, h^2 는 유량 형질의 유전력으로 국내 젖소의 육종가 추정에 이용되는 0.23을 이용하였고, r_i^2 는 i 번째 개체의 추정육종가 신뢰도이다. Figure 2.1에서와 같이 DEBV의 경우 가중치에 의하여 분산이 육종가에서 표현형 수준까지 증가하였으므로 EBV보다 더 큰 범위를 나타내고 있으며, 추정 육종가의 신뢰도가 대부분의 씨수소들에서 유사하여 중심축에 모여 있지만 일부 개체들의 경우 낮거나 높은 신뢰도에 따라 중심축에서 다소 벗어난 경우도 확인할 수 있다.

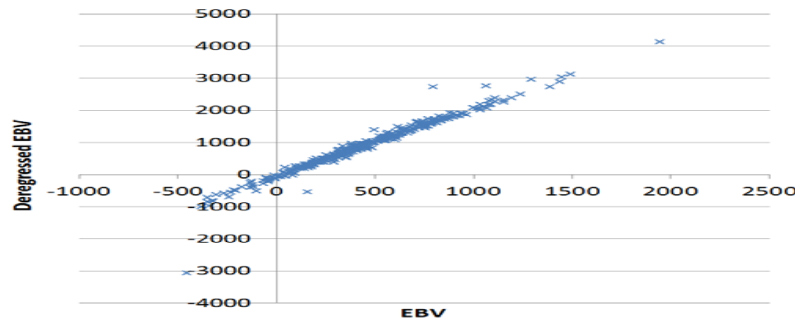


Figure 2.1 Plots of DEBV on EBV for all genotyped animals

2.3. 베이저안을 이용한 유전체전장의 SNP 효과 추정

Meuwissen 등 (2001)이 제안한 베이저안 회귀 방법인 BayesB 방법을 이용하여 각 41,051개 SNP 효과를 추정하였으며 분석 모형은 아래와 같다.

$$y_i = \mu + \sum_{j=1}^N z_{ij}\alpha_j + e_i \quad (2.2)$$

여기서, y_i 는 i 번째 개체의 DEBV, z_{ij} 는 i 번째 개체 j 번째 SNP의 -0, 1로 코딩된 유전자형, α_j 는 j 번째 SNP 효과, N 은 SNP수 (41,051), e_i 는 i 번째 개체의 잔차이다. 모든 SNP들에서 각각 서로 다른 효

과 분산을 가정하는 BayesA의 사전분포 가정은 다음과 같다.

$$(e_i|\sigma_e^2) \sim (iid)N(0, \sigma_e^2), \sigma_e^2 \sim \nu_e S_e^2 \chi_{\nu_e}^{-2} \quad (2.3)$$

분산이 σ_e^2 인 조건에서의 각 관측치별 잔차 e_i 는 평균이 0이고 분산이 σ_e^2 인 독립이며 동일한 (Independent and identically distributed; iid) 정규분포들을 따른다. 또한 σ_e^2 는 자유도가 ν_e 이고 척도 매개변수 (scale parameter)가 S_e^2 인 척도화된 역카이제곱 (scaled inverse chi-square) 분포를 따른다.

$$(\alpha_j|\sigma_j^2) \sim (iid)N(0, \sigma_j^2), \sigma_j^2 \sim \nu_\alpha S_{\nu_\alpha}^2 \chi_{\nu_\alpha}^{-2} \quad (2.4)$$

j 번째 SNP의 분산 σ_j^2 의 조건에서 j 번째 SNP의 효과 α_j 는 평균이 0이고 분산이 σ_j^2 인 정규분포를 따르며, σ_j^2 는 자유도가 ν_α 이고 척도 매개변수가 $S_{\nu_\alpha}^2$ 인 척도화된 역카이제곱 분포를 따른다. 비조건부 분포 (unconditional distribution) α_j 는 평균이 0, 척도 매개변수가 $S_{\nu_\alpha}^2$, 자유도가 ν_α 인 독립이며 동일한 univariate-t 분포들을 따른다.

$$\alpha_j \sim (iid)t(0, S_{\nu_\alpha}^2, \nu_\alpha) \quad (2.5)$$

표지인자 효과의 자유도 ν_α 는 4를, S_α^2 에는 유량 유전분산을 모든 SNP의 이형빈도의 합으로 나누어 적용하였다. BayesA 방법 적용을 위한 깃샘플링에는 아래와 같은 분포들을 이용하였다 (Gianola 등, 2009).

σ_e^2 의 완전 조건부 분포 (full conditional) 확률밀도함수는 다음과 같은 척도화된 역카이제곱 분포를 따른다

$$f(\sigma_e^2|y, \mu, \alpha) \sim \tilde{\nu}_e \tilde{S}_e^2 \chi_{\tilde{\nu}_e}^{-2} \quad (2.6)$$

여기서, 자유도는 $\tilde{\nu}_e = n + \nu_e$ 이고 n 은 표현형 수, ν_e 에는 4를 적용하였다. 척도 매개변수는 아래와 같다.

$$\tilde{S}_e^2 = \frac{SSE + \nu_e S_e^2}{\tilde{\nu}_e} \quad (2.7)$$

여기서, SSE는 $e'e$ 이며, e 는 잔차벡터이다 ($e = \mathbf{y} - \mathbf{1}\mu - \Sigma\mathbf{z}'\alpha$, 벡터는 소문자 볼드체, 행렬은 대문자 볼드체로 표기).

평균 μ 는 아래와 같은 정규분포를 따른다.

$$f(\mu|\mathbf{y}, \alpha, \sigma_e^2) \sim N\left(\frac{\mathbf{1}'(\mathbf{y} - \mathbf{Z}\alpha)}{n}, \frac{\sigma_e^2}{n}\right) \quad (2.8)$$

$\boldsymbol{\xi} = [\sigma_1^2, \sigma_2^2, \dots, \sigma_k^2]$, $\boldsymbol{\xi}_{j-}$ 를 j 번째 표지인자 분산을 제외한 나머지 분산이라 할 경우 j 번째 SNP 효과 분산 σ_j^2 의 완전 조건부 분포 확률밀도함수는 다음과 같은 척도화된 역카이제곱분포를 따른다.

$$f(\sigma_j^2|\mathbf{y}, \mu, \alpha, \boldsymbol{\xi}_{j-}, \sigma_e^2) \sim \tilde{\nu}_\alpha \tilde{S}_\alpha^2 \chi_{\tilde{\nu}_\alpha}^{-2} \quad (2.9)$$

여기서, 자유도는 $\tilde{\nu}_\alpha = 1 + \nu_\alpha$ 이고 척도 매개변수는 아래와 같다

$$\tilde{S}_\alpha^2 = \frac{\alpha_j^2 + \nu_\alpha S_\alpha^2}{\tilde{\nu}_\alpha} \quad (2.10)$$

j 번째 SNP효과 α_j 의 완전 조건부 분포 확률밀도함수는 다음과 같은 정규분포를 따른다.

$$f(\alpha_j | \mathbf{y}, \mu, \alpha, \sigma_e^2) \sim N\left(\hat{\alpha}_j, \frac{\sigma_e^2}{c_j}\right) \quad (2.11)$$

여기서, $\hat{\alpha}_j = \frac{z_j' \mathbf{w}}{c_j}$, $\mathbf{w} = \mathbf{y} - 1\mu - \sum_{j' \neq j} \mathbf{x}_{j'} \alpha_{j'}$, $c_j = (\mathbf{x}_j \mathbf{x}_j + \frac{\sigma_e^2}{\sigma_j^2})$ 이다.

BayesB 방법은 BayesA 방법과 마찬가지로 각 SNP별로 서로 다른 분산을 가정하지만 BayesB 방법에서는 $\pi(\pi)$ 확률에 따라 모든 SNP가 효과를 가지지는 않는 것을 가정하기 때문에 Metropolis-Hastings 샘플링을 별도로 적용한다. 따라서 아래와 같이 π 조건부 SNP효과 ($\alpha|\pi$)는 π 가 적용된 확률에 따라 다르게 샘플링된다 (Habier 등, 2011).

$$\alpha_i | \pi, \sigma_j^2 = \begin{cases} N(0, \sigma_j^2) & \text{with probability } (1 - \pi) \\ 0 & \text{with probability } \pi \end{cases} \quad (2.12)$$

SNP효과추정을 위하여 GelSel 프로그램을 이용하였으며, 모든 모수에 대하여 4만 번의 샘플링을 수행하고 초기 4천 번의 샘플링은 Burn-in으로 간주하여 제외하였다.

효과를 가지는 유전자 영역의 주변 SNP들이 높은 연관불평형을 가지는 경우 SNP들의 유전자형 빈도가 서로 독립적이지 않기 때문에 효과가 비슷하게 추정되며, 그 효과의 크기가 작게 나타나는 경우가 발생할 수 있다. 예를 들어 한 유전자와 완전 연관불평형상태 (complete linkage)인 4개의 SNP마커가 인접하여 존재하고 고밀도 유전체 전장 SNP chip에 (5만~70만개 SNP로 구성) 이 4개의 마커들이 포함되어 있을 경우 유전체선발기법을 적용한 효과 추정치는 SMR방법과는 다르게 유전자의 효과가 각 4개의 SNP에서 동등하게 4등분되어 추정된다. 따라서 보다 정확한 QTL영역을 검증하고 그 QTL의 효과 크기를 가능하기 위하여 서로 겹치지 않은 1.0 Mbp 구간의 window를 생성하였고 각 window내 존재하는 모든 SNP들의 효과 분산을 모두 합하여 이 window가 얼마나 많은 유전분산을 설명하는지 검증하였다 (Saatchi 등, 2014).

3. 유전체 전장 분석 및 DGAT1 유전자의 효과 검증

3.1. SNP 효과 추정 결과

BayesB 방법에 π 값을 0.99를 적용하여 유전체 전장의 41,051개 SNP 효과를 추정하였다. π 값을 0.99로 적용하는 것의 의미는 효과를 구명하는 모든 SNP들 중 약 1%의 SNP만이 효과를 가지고 나머지 99%SNP는 효과를 가지지 않는다는 가정을 적용한 것이다 (Meuwissen 등, 2001; Habier 등, 2011). 즉 41,051개의 SNP중 약 400여개의 SNP가 형질발현에 관여하고 나머지 SNP들은 영향하지 않는다는 사전 가정을 적용하였다. Ding 등 (2013)의 연구결과에서 젖소의 유량에 대한 유전체 육종가 추정의 정확도는 0.95의 π 가정에서 가장 높은 신뢰도를 보이고 있으나 본 연구에서는 유전체 육종가 추정 보다는 DGAT1 유전자의 효과를 검증하기 위하여 수행되었으므로 효과가 존재하는 SNP의 추정치가 확연하게 드러나는 높은 π 값을 가정하여 분석하였다.

BayesB 방법에 의한 유전체 전장의 각 SNP 효과 추정결과는 Figure 3.1과 같다. 유량형질에 관여하는 SNP들은 유전체 전장에 다양하게 존재하는 것을 확인할 수 있으며, 가장 큰 효과를 나타내는 단일 SNP은 4번 염색체에서 나타나지만 DGAT1유전자가 존재하는 14번 염색체에서 다수의 SNP들이 효과를 나타내고 있는 것을 확인할 수 있다. 유량형질에 관여하는 효과 크기에 대한 상위 10개 SNP들 중 2개 SNP가 DGAT1 유전자 영역을 포함하는 Window에 존재하였으며, 이외에 상위 4, 9 SNP도 DGAT1유전자 주변 윈도우에 나타나는 것을 확인할 수 있다 (Table 3.1). DGAT1유전자 주변 이 아닌

4, 8, 19, 21, 23번 염색체상의 특정 SNP들에서도 큰 효과들이 나타나고 있으나, 타품종 및 다른 관련 형질에 대하여 4번 (Daetwyler 등, 2008; Mai 등, 2014), 19번 (Raven 등, 2014), 21번 (Schnabel 등, 2005) 및 23번 (Maxa 등, 2012) 염색체 영역에 대한 SNP효과가 일부 보고되었을 뿐이기 때문에 현 단계에서 이 부분의 QTL을 명확하게 구명 할 수는 없다.

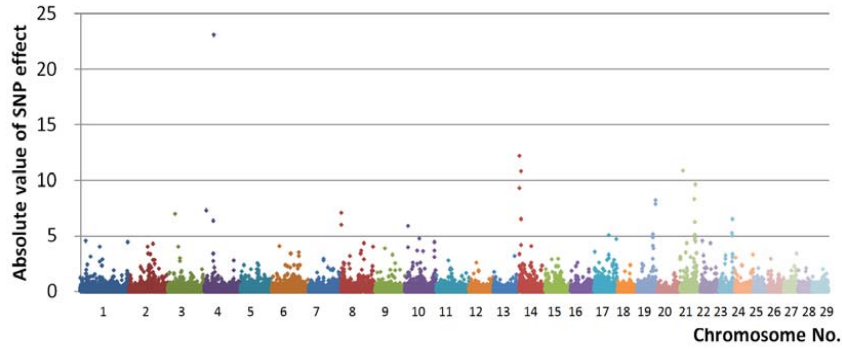


Figure 3.1 Plots of whole genome SNP effects on milk yield trait

Table 3.1 Top 10 SNPs associated to milk yield

Rank	map position (Chromosome_Position, bp)	Effect	Minor allele frequency	Window No.
1	BTA04_32831585	23.070	0.305	450
2	BTA14_1801116	12.230	0.200	1498**
3	BTA21_12095049	10.840	0.331	2039
4	BTA14_5428037	10.830	0.428	1502*
5	BTA14_1651311	9.327	0.205	1498**
6	BTA19_59300925	8.211	0.489	1950
7	BTA19_59276357	7.846	0.466	1950
8	BTA08_1154683	7.073	0.456	894
9	BTA14_5225467	6.525	0.360	1502*
10	BTA23_46110329	6.488	0.424	2207

** : Window including DGAT1 region, * : Window nearby DGAT1 region (less than 4.0 Mbp apart)

DGAT1 유전자 주변 특정 영역의 SNP효과들을 나타낸 Figure 3.2에서 DGAT1 유전자와 물리적 거리가 가장 가까운 1,801,116bp에 위치한 SNP의 효과가 크게 나타나고 있으며, DGAT1 유전자 주변 5.0Mb 내에서도 높은 효과를 보이는 SNP들이 다수 존재하는 것을 확인할 수 있다 (Figure 3.2). DGAT1 유전자 주변에서 높은 효과가 드러나는 결과는 Raven 등 (2014), Meredith 등 (2012) 및 Hayes 등 (2010)의 유전체 전장분석 연구결과와 유사하게 나타나고 있으나 효과의 크기는 다소 낮게 나타났다.

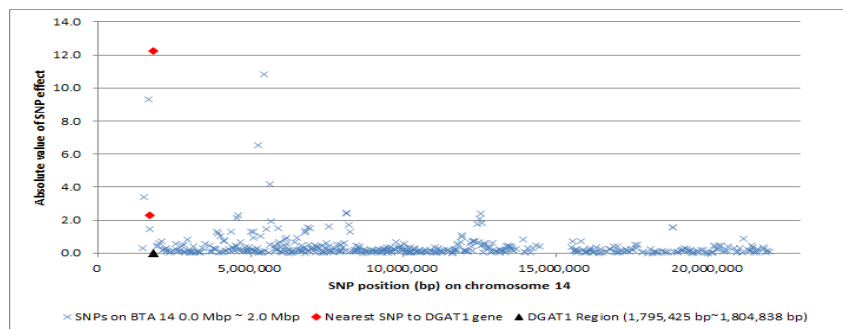


Figure 3.2 Zoomed plots of SNP effects nearby DGAT1 region

3.2. 각 Window별 유전분산 설명력 검증

1 Mbp 단위로 SNP들을 군집화한 Window를 이용하여 얼마나 많은 유전분산을 설명하는지 구명한 Figure 3.3에서 DGAT1유전자의 효과는 단일 SNP를 이용하여 나타낸 Figure 3.1 보다 확연하게 드러나고 있다. 이는 DGAT1유전자 주변의 연관불평형을 이루는 SNP들이 효과를 나누어 나타내면서 단일 SNP효과는 상대적으로 작게 나타나지만 1 Mbp window에서는 효과를 나타내는 SNP들이 합쳐져 DGAT1유전자의 효과가 더 명확하게 드러나는 것으로 사료된다. 유전체 전장의 2,516개 window별 유전분산 설명력을 계산한 결과 상위 1, 3 window가 DGAT1유전자 주변에서 발견되었으며, 이 두 window의 유전분산 설명력은 각각 0.51% 및 0.48%인 것으로 나타났다 (Figure 3.3, Figure 3.4). DGAT1유전자 인접한 window 이외에 약 4.0Mbp 떨어진 다른 window에서도 높은 효과를 나타내나 현재까지 이 부위의 알려진 유전자가 아직 없는 것을 고려할 때에 이 window의 효과는 DGAT1유전자와의 연관불평형을 이루는 SNP들에 의하여 나타나는 것으로 사료된다 (Figure 3.5). 이러한 현상은 Zhang (2015)의 연구에서 참조집단의 규모가 작으면 SNP효과가 명확하게 추정되지 않지만 500두 이상으로 규모를 증가시켰을 경우 주요 유전자 효과가 명확하게 드러나는 것을 고려할 때에 앞으로 보강될 국내 참조집단에 의하여 해결할 수 있을 것으로 예상된다. DGAT1유전자에 가장 인접한 window와 DGAT1유전자의 주변에서 연관불평형이 존재하는 것으로 예상되는 window들에서 가장 높은 유전분산 설명력이 나타남에 따라 DGAT1유전자가 유생산량 발현에 깊이 관여하고 있음을 확인할 수 있으며, DGAT1유전자의 유전변이가 일반적으로 젖소 개량에 이용되는 50k SNP chip에는 직접적으로 포함되어 있지 않지만 연관불평형에 의하여 유전체 육종가 추정 시 고려될 수 있음을 확인하였다.

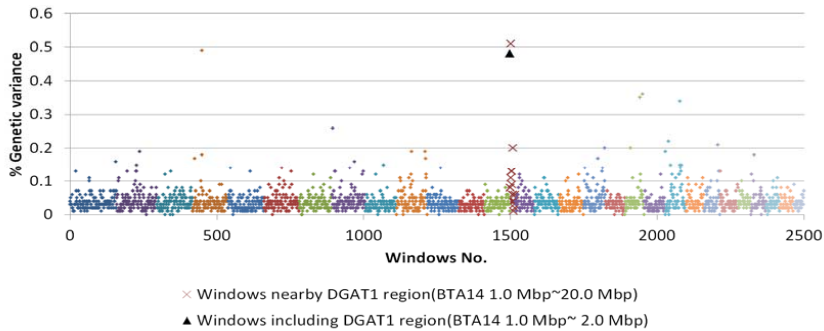


Figure 3.3 Plots of genetic variance for milk yield explained by 1.0 Mbp windows

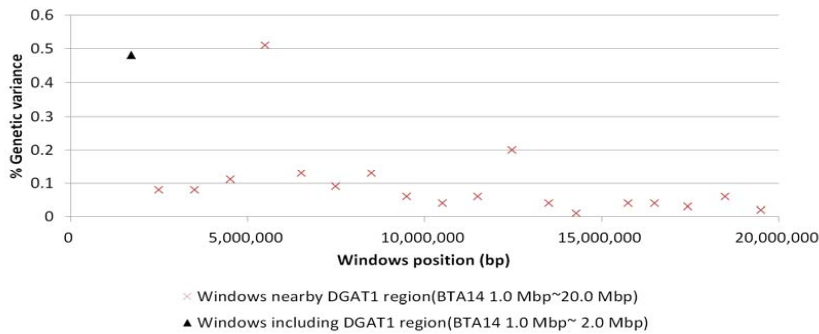


Figure 3.4 Zoomed plots of genetic variance explained by windows nearby DGAT1 gene region

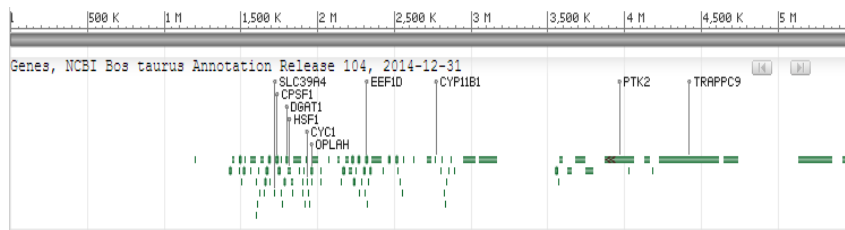


Figure 3.5 Gene information near by DGAT1 gene region

4. 결론

국내 젖소집단을 이용하여 분석한 본 연구에서 DGAT1 유전자가 유량에 대하여 매우 큰 영향을 미치고 있음을 검증하였으며, 현재 상용화된 50k 고밀도 유전체 정보를 이용하여 유전체 육종가를 추정할 경우 이러한 DGAT1 유전자의 효과는 DGAT1 유전자 영역 주변에서 DGAT1 유전자와 높은 연관불평형을 이루는 다른 SNP들에 반영되어 나타남을 확인하였다. 보다 더 정확한 유전체 육종가 추정을 위해서는 주요 종축에 대하여 별도로 DGAT1 유전자 영역의 변이를 구명한 후 기존 50k SNP 유전자형정보에 추가하는 것이 필요하며, 효과가 명확한 DGAT1 유전자 이외의 다른 유전자 효과를 구명하기 위해서는 참조집단의 규모를 증가시키는 것이 바람직할 것으로 사료된다.

References

- Boichard, D., Ducrocq V. and Fritz, S. (2015). Sustainable dairy cattle selection in the genomic era. *Journal of Animal Breeding and Genetics*, **132**, 135-143.
- Daetwyler, H. D., Schenkel, F. S., Sargolzaei, M. and Robinson J. A. B. (2008). A Genome scan to select quantitative trait loci for economically important traits in holstein cattle using two methods and a dense single nucleotide polymorphism map. *Journal of Dairy Science*, **91**, 3225-3236.
- Ding, X., Zhang, Z., Li, X., Wang, S., Wu, X., Sun, D., Yu, Y., Liu, J., Wang, Y., Zhang, Y., Zhang, S., Zhang, Y. and Zhang, Q. (2013). Accuracy of genomic prediction for milk production traits in the Chinese Holstein population using a reference population consisting of cows. *Journal of Dairy Science*, **96**, 5315-5323.
- Garrick, D. J., Taylor, J. F. and Fernando, R. L. (2009) Deregressing estimated breeding values and weighting information for genomic regression analyses. *Genetics Selection Evolution*, **41**, 55-63.
- Gianola, D., de los Campos, G., Hill, W. G. Manfredi, E. and Fernando, R. (2009). Additive genetic variability and the bayesian alphabet. *Genetics*, **183**, 347-363.
- Grisart, B., Coppieters, W., Farnir, F., Karim, L., Ford, C., Berzi, P., Cambisano, N., Mni, M., Reid, S., Simon, P., Spelman, R., Georges, M. and Snell, R.. (2002) Positional candidate cloning of a QTL in dairy cattle: identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition. *Genome Research*, **12**, 222-231.
- Habier, D., Fernando, R. L., Kizilkaya, K. and Garrick, D. J. (2011). Extension of the bayesian alphabet for genomic selection. *BMC Bioinformatics*, **12**, 186-198.
- Hayes, B. J., Bowman, P. J., Chamberlain, A. J. and Goddard, M. E. (2009). Invited review: Genomic selection in dairy cattle: progress and challenges. *Journal of Dairy Science*, **92**, 433-443.
- Hayes, B. J., Pryce, J., Chamberlain, A. J., Bowman, P. J. and Goddard, M. E. (2010). Genetic architecture of complex traits and accuracy of genomic prediction: Coat colour, milk-fat percentage, and type in Holstein cattle as contrasting model traits. *PLoS Genetics*, **6**, e1001139.
- Kemper, K. E., Reich, C. M., Bowman, P. J., Jagt, C. J., Chamberlain, A. J., Mason, B. A., Hayes, B. J. and Goddard, M. E. (2015). Improved precision of QTL mapping using a nonlinear Bayesian method in a multi-breed population leads to greater accuracy of across-breed genomic predictions. *Genetics Selection Evolution*, **47**, 29-46.

- Mai, M. D., Sahana, G., Christiansen, F. B. and Guldbandsen, B. (2014). A genome-wide association study for milk production traits in Danish Jersey cattle using a 50K single nucleotide polymorphism chip. *Journal of Animal Science*, **88**, 3522-3528.
- Maxa, J., Neuditschko, M., Russ, I., Forster, M. and Medugorac, I. (2012). Genome-wide association mapping of milk production traits in Braunvieh cattle. *Journal of Dairy Science*, **88**, 3522-3528.
- Meredith, B. K., Kearney, F. J., Finlay, E. K., Bradley, D. G., Fahey, A. G., Berry, D. P. and Lynn, D. J. (2012). Genome-wide associations for milk production and somatic cell score in Holstein-Friesian cattle in Ireland. *BMC Genomics*, **13**, 21-32.
- Mrode, R., Coffey, M. and Berry, D. P. (2010). Understanding genomic evaluations from various evaluation methods and GMACE. *Interbull Bulletin*, **42**, 52-55.
- Oh, D. Y., Yeo, J. S. and Lee, J. Y. (2014). Major SNP identification for oleic acid and marbling score which are associated with Korean cattle. *Journal of the Korean Data & Information Science Society*, **25**, 1011-1024.
- Ostensen, T., Christensen, O. F., Henryon, M., Nielsen, B., Su, G. and Madsen, P. (2011). Deregressed EBV as the response variable yield more reliable genomic predictions than traditional EBV in pure-bred pigs. *Genetics Selection Evolution*, **43**, 38-44.
- Raven, L., Cocks, B. G. and Hayes, B. J. (2014). Multibreed genome wide association can improve precision of mapping causative variants underlying milk production in dairy cattle. *BMC Genomics*, **15**, 62-76.
- Saatchi M., Schnabel, R. D., Taylor, J. F. and Garrick, D. J. (2014). Large-effect pleiotropic or closely linked QTL segregate within and across ten US cattle breeds. *BMC Genomics*, **15**, 442-458.
- Schnabel, R. D., Sonstegard, T. S., Taylor, J. F. and Ashwell M. S. (2005). Whole-genome scan to detect QTL for milk production, conformation, fertility and functional traits in two US Holstein families. *Animal Genetics*, **36**, 408-416.
- Schennink, A., Stoop, W. M., Visker, M. H. P. W., Heck, J. M. L., Bovenhuis, H., van der Poel, J. J., van Valenberg, H. J. F. and van Arendonk, J. A. M. (2007). DGAT1 underlies large genetic variation in milk-fat composition of dairy cows. *Animal Genetics*, **38**, 467-473.
- Seo, S. M., Kim, T. H. and Kim, J. H. (2014). Comparison and analysis of multiple testing methods for microarray gene expression data. *Journal of the Korean Data & Information Science Society*, **25**, 971-986.
- Son, J. Y., Jeong H. J., Yu, S. L., Lee, J. H. Do, C. H., Ryoo, S. H. and Sand, B. C. (2009). Characterization of the DGAT1 gene in the Korean Holstein dairy cattle population. *Journal of Agricultural Science*, **36**, 167-177.
- Tiezzi, F. and Maltecca, C. (2015). Accounting for trait architecture in genomic predictions of US Holstein cattle using a weighted realized relationship matrix. *Genetics Selection Evolution*, **47**, 24-37.
- Zhang, Z., Erbe, M., He, J., Ober, U., Gao, N., Zhang, H., Simianer, H. and Li, J. (2015). Accuracy of whole-genome prediction using genetic architecture-enhanced variance-covariance matrix. *Genes Genomes Genetics*, **5**, 615-627.

Validation of diacylglycerol O-acyltransferase1 gene effect on milk yield using Bayesian regression[†]

Kwang-Hyun Cho¹ · Chung-Il Cho² · Kyong-Do Park³ · Joon-Ho Lee⁴

¹²National Institute of Animal Science, RDA

³⁴Animal Genomics & Breeding Center, Hankyong National University

Received 20 July 2015, revised 29 October 2015, accepted 24 November 2015

Abstract

DGAT1(diacylglycerol O-acyltransferase1) gene is well known as a major gene of milk production in dairy cattle. This study was conducted to investigate how the DGAT1 gene effect on milk yield was appeared from the genome wide association (GWA) using high density whole genome SNP chip. The data set used in this study consisted of 353 Korean Holstein sires with 50k SNP genotypes and deregressed estimated breeding values of milk yield. After quality control 41,051 SNPs were selected and locations on chromosome were mapped using UMD 3.1. Bayesian regression of BayesB method ($\pi=0.99$) was used to estimate the SNP effects and genomic breeding values. Percentages of variance explained by 1 Mb non-overlapping windows were calculated to detect the QTL region. As the result of this study, top 1 and 3 of 2,516 windows were seen around DGAT1 gene region and 0.51% and 0.48% of genetic variance were explained by these two windows. Although SNPs on the DGAT1 gene region are excluded in commercial 50k SNP chip, the effect of DGAT1 gene seem to be reflected on GWA by the SNPs which are in linkage disequilibrium with DGAT1 gene.

Keywords: Bayesian Regression, DGAT1 gene, Holstein, milk yield, SNP.

[†] This work was carried out with the support of “Cooperative Research Program for Agriculture Science and Technology Development (Project No. PJ009260)” Rural Development Administration, Republic of Korea

¹ Junior researcher, National Institute of Animal Science, RDA, Chungnam 330-801, Korea.

² Post-Doc, National Institute of Animal Science, RDA, Chungnam 330-801, Korea.

³ Research professor, Animal Genomics & Breeding Center, Hankyong National University, Gyeonggi-do 456-749, Korea.

⁴ Corresponding author: Researcher, Animal Genomics & Breeding Center, Hankyong National University, Gyeonggi-do 456-749, Korea. E-mail: zoonolee@gmail.com