Open Access
Asian-Australas J Anim Sci
Vol. 30, No. 3:309-319 March 2017
https://doi.org/10.5713/ajas.15.0608
pISSN 1011-2367 eISSN 1976-5517

AJAS
Asian-Australasian Journal of Animal Sciences

Genome-association analysis of Korean Holstein milk traits using genomic estimated breeding value

Donghyun Shin[1], Chul Lee[2], Kyoung-Do Park[3], Heebal Kim[1,2], and Kwang-hyeon Cho[4,*]

* **Corresponding Author:** Kwang-hyeon Cho
Tel: +82-41-580-3362, **Fax:** +82-41-580-3369,
E-mail: ckh1219@korea.kr

[1] Department of Agricultural Biotechnology, Animal Biotechnology, and Research Institute for Agriculture and Life Sciences, Seoul National University, Seoul 151-921, Korea
[2] Interdisciplinary Program in Bioinformatics, Seoul National University, Seoul 151-921, Korea
[3] The Animal Molecular Genetics & Breeding Center, Chonbuk National University, Jeonju 561-756, Korea
[4] Division of Animal Breeding and Genetics, National Institute of Animal Science, Rural Development Administration, Cheonan 331-801, Korea

Submitted Jul 19, 2015; Revised Aug 31, 2015; Accepted Oct 3, 2015

Objective: Holsteins are known as the world's highest-milk producing dairy cattle. The purpose of this study was to identify genetic regions strongly associated with milk traits (milk production, fat, and protein) using Korean Holstein data.
Methods: This study was performed using single nucleotide polymorphism (SNP) chip data (Illumina BovineSNP50 Beadchip) of 911 Korean Holstein individuals. We inferred each genomic estimated breeding values based on best linear unbiased prediction (BLUP) and ridge regression using BLUPF90 and R. We then performed a genome-wide association study and identified genetic regions related to milk traits.
Results: We identified 9, 6, and 17 significant genetic regions related to milk production, fat and protein, respectively. These genes are newly reported in the genetic association with milk traits of Holstein.
Conclusion: This study complements a recent Holstein genome-wide association studies that identified other SNPs and genes as the most significant variants. These results will help to expand the knowledge of the polygenic nature of milk production in Holsteins.

Keywords: Korean Holstein; Genome-wide Association Studies (GWAS); Milk Production; Milk Fat; Milk Protein

INTRODUCTION

Holsteins are the world's highest-milk producing dairy cattle. Approximately 2,000 years ago, the black Batavians and white Friesians cows were bred to produce better breed. These cattle have been continuously selected and genetically evolved into the efficient, high producing black-and-white dairy cattle, which we know as Holstein-Friesian. For the last several decades, intensive application of traditional animal breeding technologies has significantly improved milk performance throughout the world.

Technology of molecular biology has opened up the possibility of identifying genome regions and variants underlying complex traits such as milk production, fat and protein. Unlike the traditional animal breeding programs which rely on phenotype and pedigree information, genetic evaluated information provide a great potential to enhance selection accuracies and expedite the genetic improvement of animal productivity. Since the seminal work on quantitative trait locus (QTL) mapping by Georges et al [1], a large number of articles including detection of QTLs for milk production traits have been published. So far, approximate 1,345 QTLs for milk production traits had been reported via genome scans based on marker-QTL linkage analyses. The limitations of QTL mapping using linkage analysis and/or linkage disequilibrium (LD) based on the panels of low to moderate density markers have been well documented [2].

The advent of genome-wide panels including hundreds of thousands of single nucleotide polymorphisms (SNPs) has resulted in the development of commercial SNP chips and rapid,

Copyright © 2017 by Asian-Australasian Journal of Animal Sciences
This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

large-scale genotyping of common SNPs across large populations. These SNPs have been widely used for the detection and localization of QTL for complex traits in many species [3], and have proved powerful and useful in identification of casual mutations associated with economically important traits in livestock [4,5] as well as human diseases. At the same time, genome-wide association studies (GWAS) based on high throughput SNP genotyping technologies open a broad avenue for exploring genes associated with milk production traits in dairy cattle [6]. Most recently, along with maturing of genome sequencing and high throughput SNP genotyping technologies, GWAS is becoming practical for exploring genes associated with complex traits. Like this, GWAS has been widely accepted as a primary approach for gene finding, and it achieved huge success in identifying genes conferring modest disease risks in human.

Several studies focusing on identifying genes for milk production traits had been performed. Associations between milk traits and polymorphisms in candidate genes have produced a long list of potential markers with significant effects reported in regional Holstein cattle population [7]. Generally, most of the economic traits in dairy cattle are controlled by polymorphisms, genes of small or large effects. To find genetic variant related to milk production traits beyond previous studies, we performed GWAS with genomic estimated breeding values (gEBVs) using 1,941 Korean Holsteins data. Estimated breeding value (EBV) was used as the phenotype as it only considers the genetic component of phenotypic variance. We then used p-value integration method to detect significant genetic regions with reduction of false positive error. Using this approach, we identified 9, 6, and 17 significant genetic regions associated with milk production, fat and protein, respectively and most of these genetic regions were not reported, previously. The identified genetic regions and their genes could be considered as a preliminary foundation for further studies in Holstein milk production traits. Furthermore, the identified genetic regions may be used as potential candidate markers for selection in Korean dairy cattle breeding programs and provide unprecedented insight into the structure of Holstein cattle populations.

MATERIALS AND METHODS

Animals and data

We used pedigree data containing 1,941 individuals (from present to 3 generations ago) in Korean Holstein population to infer EBV. There were milk traits record data (milk production, milk fat, milk protein) of 1,169 individuals of total 1,941. We took 911 individuals samples to perform SNP chip analysis and 488 individuals of 911 were overlapped with individuals of pedigree data. DNA was extracted from nasal discharge samples or semen of 911 Holstein individuals. DNA was quantified and genotyped using the Illumina BovineSNP50 BeadChip containing 54,609 SNPs. Features of the Illumina BovineSNP50 BeadChip have been detailed previously. All samples were genotyped using BEADSTUDIO (Illumina lnc, San Diego, CA, USA).

Genotype quality control and imputation

The chip includes 54,609 SNPs that are distributed on the 29 bovine autosomes, X and Y chromosomes with an average density of one SNP per approximately 49 kb) from the cow genome, UMD 3.1. We used three criteria to perform quality control to reduce false positive results. So, we excluded SNPs with Hardy–Weinberg equilibrium test p-value of <0.001, a missing rate of >0.05 and minor allele frequency of <0.01. Additionally, because we used cows and bulls in association analysis, SNPs on the X and Y chromosome were also excluded retaining finally 41,099 autosomal SNPs. The remaining 41,099 SNPs were distributed evenly on the autosomes (Supplementary Figure 1). These quality control processes were performed using the software PLINK [8]. The 41,099 autosomal SNP data of Holsteins after quality control was imputed without panels using BEAGLE [9].

Inferring estimated breeding value of milk traits

We inferred EBVs of parity 1 records of three milk traits (milk production, milk fat, milk protein), respectively. To improve the accuracy of the EBV, we consolidated the number of environmental factors by reducing the factors deemed unnecessary. Considering size of pedigree data and phenotype records data, we used season and year in inferring EBVs. A single-trait animal model was used to estimate the genetic parameters as EBVs. The animal model used in this study was as followed:

$$y = Xb+Za+e$$

Where, y (n×1) was the vector of each milk traits, X (n×p) was the matrix of fixed effects (season and year in this study), Z (n×n) was the matrix of random effects (relationship matrix in this study), b and a were coefficients vector for X and Z, respectively and e (n×1) is the vector of residual error (meaning inexplicable factors). EBV is the coefficients vector of Z matrix. All parameters were estimated using the BLUPF90 program. We used parity 1 records of 1,941 individual in Korean Holstein population in this process and inferred EBVs of 1,941 individuals per each milk traits between 1990 and 2014. The fixed effects in this analysis were year and season (12 to 2: winter, 3 to 5: spring, 6 to 8: summer, 9 to 11: fall).

After inferring EBVs, we inferred gEBVs of 911 individuals which used to perform SNP analysis (milk production, milk fat and milk protein respectively). First, we estimated SNP effect of each trait using 488 individuals. The model of estimation SNP effect was as followed.

$$y = Za+e$$

Where, y (n×1) was the vector of each milk traits EBVs, Z (n×p)

was the matrix of SNP genotypes and e is the vector of the i.i.d. residual random error with e ~ N(0, Iσ_e^2), where σ_e^2 denotes a constant variance. a in this model is the coefficients vector for Z and marker effects of milk traits, simultaneously. We applied ridge regression to solve this model and we assigned ridge parameter (based on heritability of previous Korean Holstein population) to each model of milk traits (ridge parameter $\lambda = \sigma_e^2/\sigma_u^2$) [10].

$$\hat{a}^{ridge} = argmin_a\{\sum(Y - \sum z * a)^2 + \lambda * \sum a^2\}$$

In the ridge regression model, argmin (the argument of the minimum) meant that the set of points of the given argument for which the given function attained its minimum value in mathematics. And Z (n×p) is the matrix of SNP genotypes and a is the coefficient vector for Z. To check the accuracy of the solution of ridge regression, we used the 10-fold cross validation. Secondly, we estimated gEBV based on SNP effect, as follows:

$$\text{gEBV} = \sum_i^p \text{z} \times a_i$$

All calculations in estimating gEBV were performed using R ("MASS" packages).

Genome-wide association analysis

We performed single association analysis using PLINK, as follows:

$$y = xb+e$$

Where, y is a vector of each gEBVs of 911 genotyped individuals, x is each SNP information and b is coefficient value for x vector. After SNP association test, we used genomic control p-value instead of normal p-value. And we assigned integrated p-value to non-overlapped regions containing 5 SNPs to identify a significant genetic region instead of SNP. We performed p-value integration using R ("MADAM" packages). Genome-wide significance was defined based on genomic control p-value integration of 5 SNPs and Bonferroni method to correct p-value thresholds of significance after p-value integration: significant association of 0.05 false positives was used as a genome-wide significance. An overview of the results of test using Manhattan plots was produced by R. Because the SNPs were mapped on the UMD3.1 assembly, we used UMD3.1 gene information in Ensemble Genome Browser to investigate function of significant genetic region. We searched Ensemble gene ID and gene symbol which was overlapped with each regions. And then we assigned gene information to each regions. In this way, we identified relationship between each regions and animal trait through cow data of Animal QTL database.

RESULTS

Phenotypes used in this study were three traits related to milk of 1,169 Korean Holstein individuals (milk production, fat and protein) of parity 1. Milk production records were in range 3,473 kg to 13,734 kg. Mean and standard deviation were 8,730.68 kg and 1,481.864 kg, respectively. Milk fat records were in range 135 kg to 532 kg. Mean and standard deviation were 329.62 kg and 59.95 kg. In case of milk protein, records were in range 113 kg to 428 kg. Mean and standard deviation were 275.42 kg and 45.38 kg, respectively. All traits followed approximately normal distribution and their distributions are shown in Supplementary Figure 2. And all pairwise correlation of three traits were higher than 0.69 (milk production-fat: 0.69, milk production-protein: 0.925, milk fat-protein: 0.726) and their plots are shown in Supplementary Figure 3.

After quality control and imputation of 911 Korean Holstein individuals, we estimated gEBVs by two-step method. In first step, we estimated EBVs of each parity 1 milk traits using a single-trait animal model. We considered two fixed effect (season and year) in estimation EBVs and their relationships between traits and fixed effect are shown in Supplementary Figure 4 (season) and Supplementary Figure 5 (year). Through these relationship figures, we could identify that effect of year was more than season. We show EBVs distribution of each milk trait in Supplementary Figure 6. After estimation EBVs of 1,169 individuals (containing 488 SNP genotyped samples), we estimated the 41,099 SNP effect of each milk traits using 438 (training set of 10 fold cross validation strategy) of 488 individuals through ridge regression. SNP effect of each milk trait followed normal distribution and is shown in Supplementary Figure 7. We could estimate gEBVs through combining SNP genotyped information and estimated SNP effect. To test gEBVs accuracy, we compared EBVs with gEBVs using 49 individuals (test set of 10 fold cross validation strategy). The correlation coefficient of EBVs and gEBVs of each milk traits (milk production, fat and protein) of 49 individuals were 0.58, 0.70, and 0.68, respectively. Their correlation plots are shown in Supplementary Figure 8. In this way, we could estimate gEBVs of each three milk traits of 911 genotyped individuals and their distribution is shown in Supplementary Figure 9.

We compared the genotypes of 911 individuals with EBV as a phenotype, respectively. After performing single association analysis to these comparison, we identified that p-values from this analysis were the results of overestimation through an inflation factor much more than 1. The inflation factor of milk production, fat and protein were 2.19, 2.29, and 2.35, respectively. This was much higher than 1 and meant that using these p-values to identify significant genetic variants was not appropriate. This phenomenon is common in animal GWAS, because domesticated animals as Holsteins contain a massively structured population from small number of bulls and high linkage disequilibrium.

To reduce false positive errors, we applied two methods to detecting significant genetic variants. One was that we used genomic control p-values instead of normal p-values. These genomic control p-values of each milk trait were calculated in PLINK. We identified that there were no inflation in genomic control p-values of all milk traits through Quantile-Quantile plots (shown in Supplementary Figure 10). Additionally, the inflation factor of milk production, fat and protein were 1.003471, 1.013479, and 1.011284, respectively. But none of the 41,099 SNPs exceeded the threshold of Bonferroni multiple test based on genomic control p-values in milk association test (genomic p-value< 1.21E-06, equivalent to p-value = 0.05 after Bonferroni multiple correction). Milk fat and protein association test results were same to Milk production (Supplementary Figure 11). And then we integrated genomic control p-values of five SNPs into one p-value through Fisher's Method for combining p-values [11] and assigned a p-value to each region. We identified significant genetic regions which exceeded the threshold of Bonferroni multiple test based on integrated p-values (integrated p-values< 6.09E-06, equivalent to p-value = 0.05 after Bonferroni multiple correction). The results of region estimation in this GWAS study after chromosome sorting are in the Manhattan plots in Figure 1.

In association test of milk production, nine regions (containing forty five SNPs) were significant and were distributed into six chromosomes (Table 1). Seven of nine regions had overlapped sixteen Ensemble genes ID and fourteen of sixteen Ensemble genes were related to the protein coding genes. All of the cow QTLs were related to nine significant regions in the association test of milk production. In association test of milk fat, six regions (containing thirty SNPs) were significant and were distributed into three chromosomes (Table 2). Four of six regions had overlapped thirteen Ensemble genes ID and twelve of thirteen Ensemble genes were related to 11 protein coding genes. Ten cow QTL were related to two of six significant regions in the association test of milk fat. Additionally, six of ten QTL were Holstein breed specific. In association test of milk protein, seventeen regions (containing eighty five SNPs) were significant and were distributed into ten chromosomes (Supplementary Table 1). Fifteen of seventeen regions had overlapped sixty two Ensemble genes ID and fifty seven of sixty two Ensemble genes were related to the protein coding genes. Twenty three cow QTLs were related to six of seventeen significant regions in the association test of milk fat. Additionally, seven of twenty three QTLs were Holstein breed specific. Diacylglycerol O-Acyltransferase 1 (*DGAT1*) was known as a major gene for milk traits in cow. DGAT1 is located in 1,795,351 bp to 1,804,562 bp of chromosome 14. Flanking markers of *DGAT1* gene were ARS-BFGL-NGS-94706 (CHR14:1696470) and Hapmap52798-ss46526455 (CHR14: 1923292). Normal p-values of ARS-BFGL-NGS-94706 and Hapmap52798-ss46526455 were 0.04434 and 0.7297, respectively and did not passed criteria of Bonferroni multiple test. After p-value integration, the nearest genetic region of DGAT1 was located in 1,463,676 bp to 1,696,470 bp of chromosome 14. That region of which p-value was 0.136 and did not passed criteria of Bonferroni multiple test.

After three association test, we compared each result with other results. The most interesting comparison result was that CHR2:80605588-81002535 was significant in all three association tests. In this region, there were 4 Ensemble genes and three were protein coding genes (*TMEFF2*, transmembrane protein with EGF-like and two follistatin-like domains 2; *NABP1*, nucleic acid binding protein 1; and *SDPR*, serum deprivation response). There were two significant regions in both milk production and milk fat. Seven regions were significant in both milk production and milk protein. And we found two significant overlapped regions in between milk fat and milk protein.

DISCUSSION

To identify genetic regions underlying milk traits of Korean Holstein, we performed GWAS with p-value integration in this study. These results were based on 911 genotyped Holsteins in Korea. Before the association study, we estimated 1,941 Korean Holsteins EBVs and we inferred 41,099 SNP effects of each milk traits. Using these SNP effects, we estimated gEBVs of 911 genotyped Holsteins. EBVs contains only the genetic effect of phenotype and we could predict genetic capacity of each individual based on the record of those individuals and their relatives. EBV was used to rank breeding stock for selection in animal breeding and we decided that EBV were appropriate dependent variables in this study. In gEBV estimation, we assumed that heritability of the three milk traits (milk production, milk fat, and milk proteins) were 0.23, 0.20, and 0.19, respectively. These heritabilities were reported in a previous study using Korean Holstein population [12].

We decided that gEBV was more proper than phenotype. The reason for this was bulls were more important than cows in animal breeding. But bulls do not have milk production trait data. However, if we used gEBVs as phenotype instead of traits data, we can performed GWAS using cows and bulls. Additionally, milk production traits are strongly affected by environmental factors (herd, season, and so forth). Also, sample size is very important factor in GWAS. If we used phenotypes in this study, our sample size was only 488 individuals. This sample size is insufficient, especially considering recent GWAS trends. If we used gEBV as phenotypes, we can use 911 individuals in this analysis and perform GWAS with alarger sample size. Before real animal capacity tests, gEBV was directly used as selection indicator. Superior dairy cattle were selected based on gEBV as result of genomic selection.

GWAS is a promising method to discover common genetic variants that could explain disease or interesting economic traits of animals and plants. But inflation always is a problem in do-

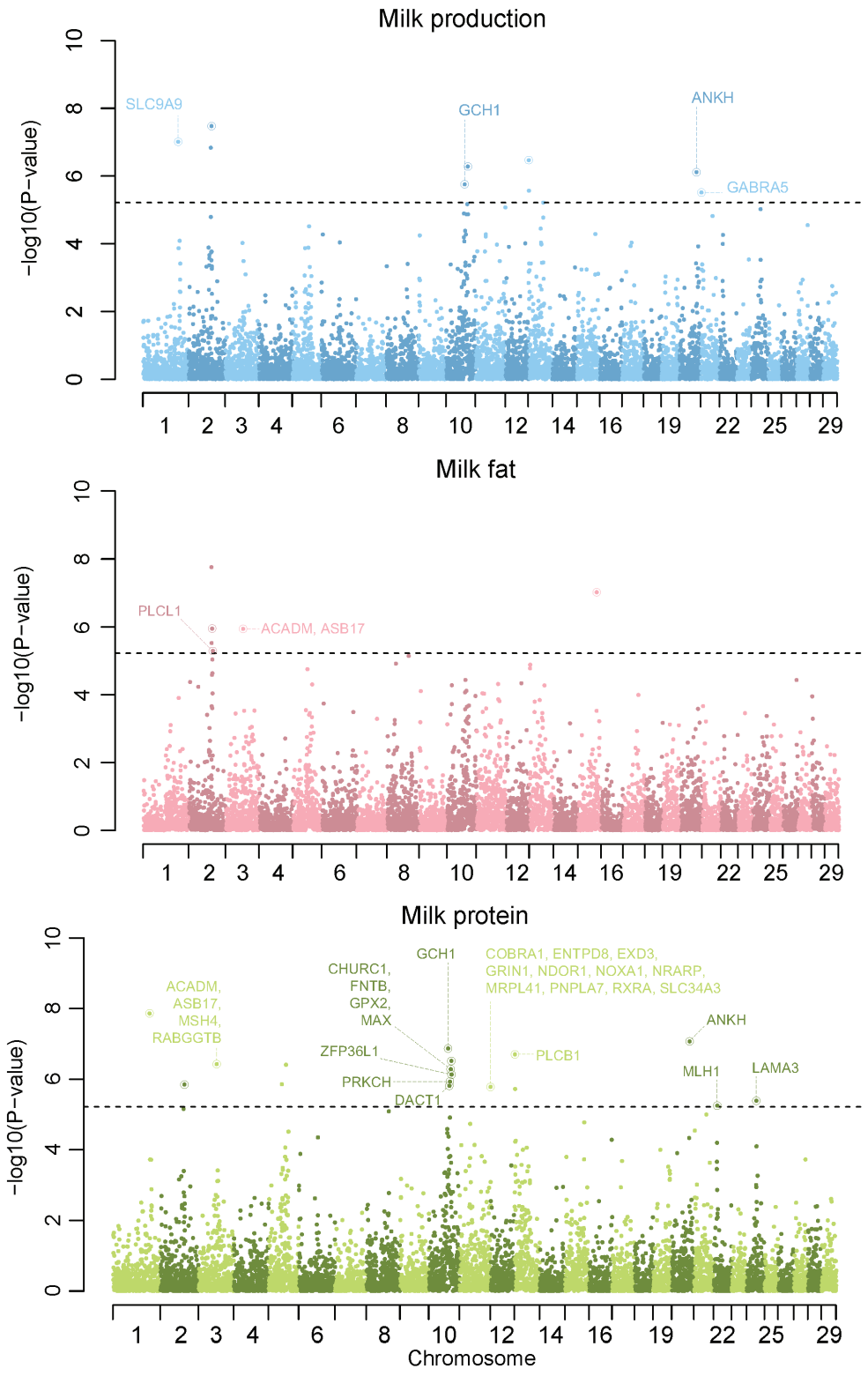

Figure 1. Manhattan plot of genome-wide association studies (GWAS) result of each milk traits after p-value integration. Each circle indicated each region containing 5 single nucleotide polymorphism (SNP) and circle with ring meant that region contained nearby gene. We provide the gene names which were significant in both GWAS and gene ontology analysis. a) milk production, b) milk fat, and c) milk protein. Grey dot line indicates the threshold of Bonferroni multiple test based on integrated p-values in each milk traits association test (genomic p-value<6.09E-06, equivalent to p-value = 0.05 after Bonferroni multiple correction).

mesticated animal GWAS. Lambda values of milk production, fat and protein were 2.190281, 2.299665, and 2.350455, which were much higher than expected. We guessed that a strong reason for this inflation was very large LD in the Korean Holstein

Table 1. Significant genetic regions in GWAS results of milk production

CHR	Region ID	regDIS	p-value	Bonferroni	SNP ID	SNP Pos	SNP p-value	Ensembl gene ID (Gene symbol)	QTL trait
1	CHR1:126549908 -126680933	131025	9.76E-08	8.01.E-04	BTB-00060751	126549908	6.12.E-03	ENSBTAG00000031178 (SLC9A9)	
					ARS-BFGL-NGS-98257	126567481	8.52.E-03		
					ARS-BFGL-NGS-113021	126606801	4.13.E-03		
					ARS-BFGL-NGS-25639	126651501	8.75.E-03		
					ARS-BFGL-NGS-100109	126680933	2.26.E-03		
2	CHR2:78302162 -78433731	131569	1.45E-07	1.19.E-03	Hapmap26185-BTA-157573	78302162	5.34.E-03		
					Hapmap40841-BTA-94957	78332368	4.66.E-03		
					BTB-01945480	78366932	2.95.E-03		
					BTB-01767882	78405537	1.20.E-02		
					BTB-01767855	78433731	7.74.E-03		
2	CHR2:80605588 -81002535	396947	3.35E-08	2.75.E-04	ARS-BFGL-NGS-41490	80605588	5.47.E-02	ENSBTAG00000014832 (TMEFF2)	
					ARS-BFGL-NGS-5680	80666057	9.90.E-02	ENSBTAG00000046400	
					Hapmap50262-BTA-122131	80687709	4.01.E-02	ENSBTAG00000018653 (NABP1)	
					ARS-BFGL-NGS-94696	80970515	1.18.E-04	ENSBTAG00000018497 (SDPR)	
					ARS-BFGL-NGS-102243	81002535	4.76.E-05		
10	CHR10:67538919 -67694539	155620	1.76E-06	1.44.E-02	ARS-BFGL-NGS-247	67538919	8.03.E-03	ENSBTAG00000040151 (GCH1)	
					ARS-BFGL-NGS-44563	67586540	1.72.E-03	ENSBTAG00000019120 (WDHD1)	
					BTA-74241-no-rs	67626007	2.88.E-02		
					ARS-BFGL-NGS-32233	67648206	8.47.E-03		
					ARS-BFGL-BAC-15431	67694539	3.89.E-02		
10	CHR10:80181163 -80567844	386681	5.25E-07	4.31.E-03	ARS-BFGL-NGS-117202	80181163	9.79.E-01	ENSBTAG00000045041 (7SK)	
					ARS-BFGL-BAC-11003	80410977	2.46.E-05	ENSBTAG00000018971 (RAD51B)	
					ARS-BFGL-NGS-41880	80525247	1.84.E-03	ENSBTAG00000014334 (ZFYVE26)	
					BTB-00437473	80546262	2.73.E-01		
					ARS-BFGL-NGS-3980	80567844	2.56.E-03		
13	CHR13:1278678 -1559165	280487	3.40E-07	2.79.E-03	ARS-BFGL-BAC-12483	1278678	8.63.E-03	ENSBTAG00000008338 (PLCB1)	
					Hapmap47208-BTA-15912	1299992	4.19.E-01		
					Hapmap60144-rs29013559	1397454	4.45.E-04		
					Hapmap45253-BTA-15908	1477972	3.68.E-03		
					ARS-BFGL-NGS-115902	1559165	3.13.E-03		
13	CHR13:1867669 -1982591	114922	2.71E-06	2.23.E-02	BTB-01324240	1867669	1.63.E-03		
					Hapmap39731-BTA-23124	1912749	6.18.E-01		
					BTB-01324017	1966648	4.02.E-01		
					ARS-USMARC-Parent-EF026087-rs29011643	1982209	7.22.E-04		
					UA-IFASA-5150	1982591	7.56.E-04		
20	CHR20:58292591 -58592622	300031	7.71E-07	6.33.E-03	Hapmap38462-BTA-110556	58292591	8.18.E-02	ENSBTAG00000013391 (ANKH)	
					ARS-BFGL-NGS-110091	58362004	4.83.E-04	ENSBTAG00000003186 (OTULIN)	
					ARS-BFGL-NGS-111931	58405641	4.22.E-03	ENSBTAG00000045869	
					ARS-BFGL-NGS-96125	58449212	6.31.E-03	ENSBTAG00000045215 (U6)	
					Hapmap41960-BTA-74781	58592622	4.64.E-02		
21	CHR21:4441252 -4671448	230196	3.07E-06	2.52.E-02	BTA-105737-no-rs	4441252	3.36.E-01	ENSBTAG00000003392 (GABRA5)	
					ARS-BFGL-NGS-36921	4482429	3.85.E-03		
					ARS-BFGL-BAC-30337	4558974	1.32.E-03		
					ARS-BFGL-NGS-18711	4638691	6.35.E-02		
					ARS-BFGL-NGS-12690	4671448	2.36.E-03		

GWAS, genome-wide association studies; CHR, chromosome; SNP, single nucleotide polymorphism; QTL, quantitative trait locus.

population. Korean Holsteins have been under intense directional artificial selection to increase milk quantity and quality. This selection could reduce genetic diversity of Korean Holstein population and increase LD. A previous study reported a reduction genetic diversity of Korean Holstein population through the effective population size [13]. We thought that inflation of p-values in animal GWAS was a general phenomenon and applied appropriate methods to detect significant genetic variants. First, we used genomic control p-values which did not have an inflation problem. But we could not detect significant genetic variants after multiple test through Bonferroni correction (Supplementary Figure 11). The reason no SNP was significant in these association tests was because milk traits are complex phenotypes affected by several or many genetic factors instead of a few strong genetic factors. So we identified significant genetic regions associated with milk traits instead of SNPs. Also, we assumed that SNP was representative of a certain region and that a region test with the trait were repeated by SNPs in that region, because the Holstein LD block was very large. We defined that each region of the Holstein genome consisted of 5 SNPs and did not overlap. This meant that a continuous five SNPs on the physical map was one region in this study. We could define 8,209 regions on whole genome of Korean Holsteins and the mean and standard deviation of region size were 243,570.4 bp and 142,286.2 bp.

Table 2. Significant genetic regions in GWAS results of milk fat

CHR	Region ID	regDIS	p-value	Bonferroni	SNP ID	SNP Pos	SNP p-value	Ensembl gene ID (Gene symbol)	QTL trait
2	CHR2:78302162 -78433731	131569	1.78E-08	1.46.E-04	Hapmap26185-BTA-157573	78302162	3.41.E-03		
					Hapmap40841-BTA-94957	78332368	2.18.E-03		
					BTB-01945480	78366932	2.06.E-03		
					BTB-01767882	78405537	9.17.E-03		
					BTB-01767855	78433731	4.18.E-03		
2	CHR2:78691609 -78893301	201692	3.05E-06	2.50.E-02	BTB-01860738	78691609	5.29.E-04		
					BTB-01860839	78726622	2.13.E-01		
					BTB-00103543	78749162	1.08.E-01		
					BTB-01374180	78872254	3.80.E-03		
					BTB-01374162	78893301	5.49.E-03		
2	CHR2:80605588 -81002535	396947	1.15E-06	9.45.E-03	ARS-BFGL-NGS-41490	80605588	6.49.E-02	ENSBTAG00000014832 (TMEFF2)	
					ARS-BFGL-NGS-5680	80666057	1.92.E-02	ENSBTAG00000046400	
					Hapmap50262-BTA-122131	80687709	4.03.E-02	ENSBTAG00000018653 (NABP1)	
					ARS-BFGL-NGS-94696	80970515	1.89.E-03	ENSBTAG00000018497 (SDPR)	
					ARS-BFGL-NGS-102243	81002535	8.33.E-04		
2	CHR2:86831095 -87004473	173378	5.21E-06	4.27.E-02	BTA-90292-no-rs	86831095	6.92.E-01	ENSBTAG00000007635 (PLCL1)	Meat_and_Carcass_Association (Intermuscular fat percentage: QTLID25126)
					BTA-90298-no-rs	86868918	2.31.E-01		
					Hapmap38934-BTA-117244	86909355	2.24.E-03		Meat_and_Carcass_Association (Subcutaneous fat: QTLID25127)
					Hapmap41106-BTA-90288	86959111	2.79.E-03		
					ARS-BFGL-NGS-55270	87004473	4.87.E-04		Meat_and_Carcass_Association (Intermuscular fat percentage: QTLID25128)
									Exterior_Association (Udder structure: QTLID25015)
									Milk_Association (Milk fat percentage: QTLID25003)
									Exterior_Association (Udder structure: QTLID25017)
3	CHR3:69182813 -69391345	208532	1.17E-06	9.61.E-03	BTB-01635025	69182813	3.76.E-01	ENSBTAG00000042950 (SNORD45)	
					BTB-00133734	69292666	6.16.E-03	ENSBTAG00000024240 (ACADM)	
					BTB-00133701	69315385	1.54.E-03	ENSBTAG00000005864 (ASB17)	
					BTB-00133671	69342335	6.32.E-03	ENSBTAG00000043447 (SNORD45)	
					BTB-01711286	69391345	3.58.E-03	ENSBTAG00000018447 (RABGGTB)	
								ENSBTAG00000018448 (MSH4)	
								ENSBTAG00000042353 (SNORD45)	
15	CHR15:67269656 -67429625	159969	9.74E-08	8.00.E-04	ARS-BFGL-NGS-36801	67269656	5.86.E-02	ENSBTAG00000044158	Reproduction_QTL (LDLRAD3)
					ARS-BFGL-NGS-111525	67323237	1.94.E-03		(Calving ease (direct) : QTLID15193)
					Hapmap41218-BTA-28547	67358867	3.53.E-03		Reproduction_QTL (Calving index: QTLID15194)
					ARS-BFGL-NGS-92508	67386416	3.68.E-05		
					BTB-00611649	67429625	2.88.E-01		Reproduction_QTL (Calving ease (maternal) : QTLID15192)
									Production_QTL (Average Daily Gain: QTLID22798)

GWAS, genome-wide association studies; CHR, chromosome; SNP, single nucleotide polymorphism; QTL, quantitative trait locus.

We integrated 5 SNP p-values into 1 region p-values to assign significant level to each 8,209 regions. We thought that this approach could detect significant genetic regions and exclude false positive error. Using this approach, we identified several genetic regions and genes related to milk traits which have not been reported, previously.

CHR2:80605588-81002535 was significant in all three association tests and contained three protein coding genes (*TMEFF2*, *NABP1*, and *SDPR*). TMEFF2 encodes transmembrane protein with epidermal growth factor (EGF)-like and two follistatin-like domains 2. EGF was reported to affect various milk production traits [14]. NABP1 encodes Single-stranded DNA ssDNA-binding protein that is ubiquitous and essential for a variety of DNA metabolic processes, including replication, recombination, and detection and repair of damage. SDPR encodes a calcium-independent phospholipid-binding protein whose expression increases in serum-starved cells. Serum related to density of several substances in milk and these affected milk production. So, we guessed that these genes in CHR2:80605588-81002535 were strongly related to diverse mechanisms of milk production.

There were 14 significant protein coding genes in the milk production association test, and we performed gene ontology analysis using them. Four terms were significant and three of total four terms were related to ion transport (Figure 2). Solute carrier family 9, subfamily a member 9 (*SLC9A9*), ankylosis protein homolog (*ANKH*), and gamma-aminobutyric acid A receptor, alpha 5 (*GABRA5*) genes were in these ion transport terms. Kramer et al reported in a previous GWAS study that SLC9A9 was in a region with possible high influence on the observed milk production trait [15]. ANKH encodes a multi-pass

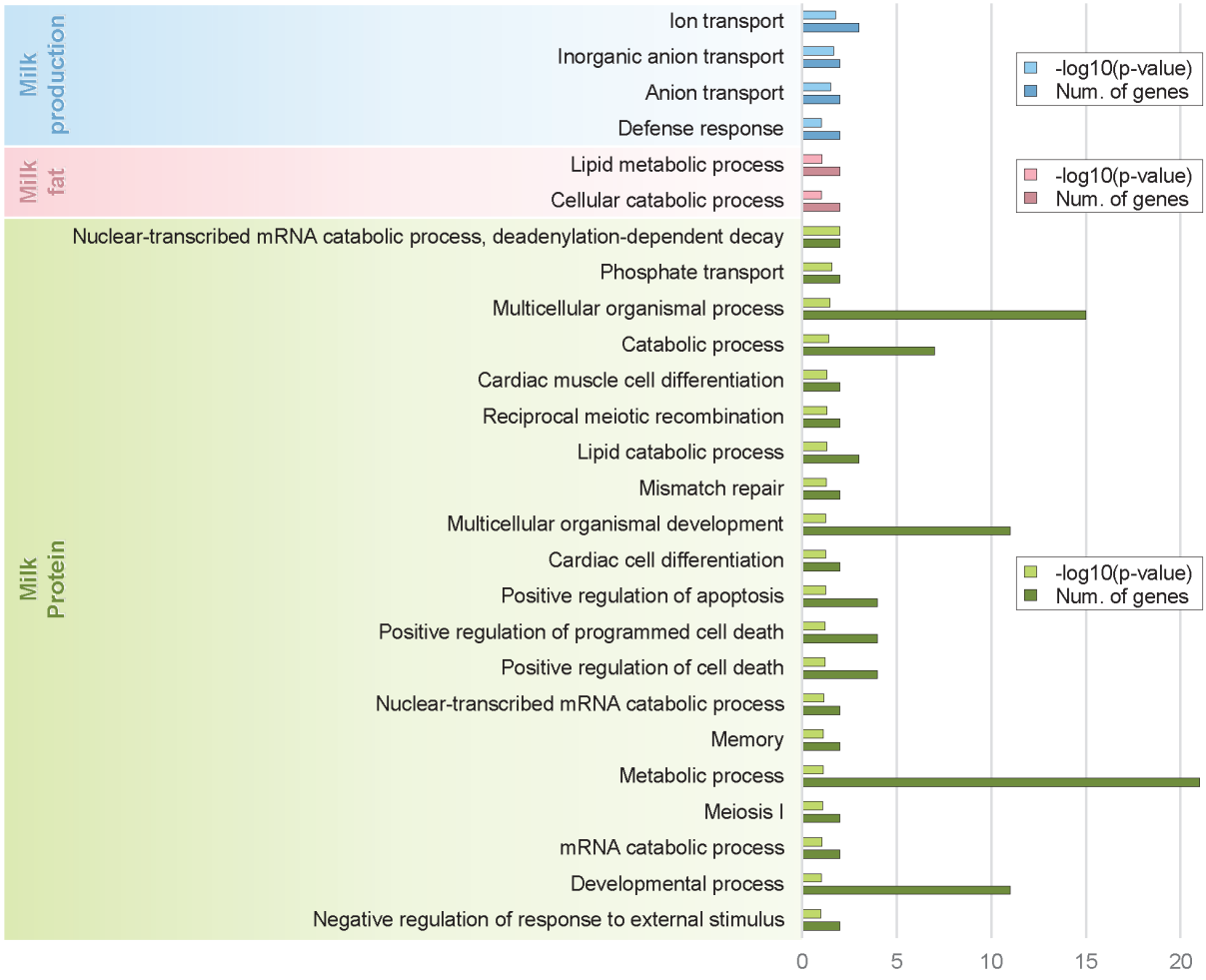

Figure 2. Result of gene ontology analysis using gene sets of significant genetic regions as each genome-wide association studies (GWAS) result of milk production (skyblue), fat (pink), and protein (green).

transmembrane protein that controls pyrophosphate level and GABRA5 is one of GABA subunit which are ligand-gated chloride channels. Previous studies reported that ion balance was very important to Holstein lactation. For example, maintenance of calcium homeostasis is critical for many functions as hormone secretion and cation–anion difference affects health status and lactation performance [16]. Phospholipase C, Beta 1 (*PLCB1*) encoded by this gene plays an important role in the intracellular transduction of many extracellular signals. PLCB1 with GABRA5 were reported as significant mammary gland genes affected by level of nutrient intake in pre-weaned Holstein heifers [17].

There were 11 protein coding genes in milk fat association test, and we identified their biological meaning in Holsteins through gene ontology analysis. Two terms were significant and one of them was related to the lipid metabolic process (Figure 2). *PLCL1* and *ACADM* (acyl-coenzyme A dehydrogenase, C-4 to C-12 straight chain) genes were in cluster of lipid metabolic processes. PLCL1 encodes a protein which is involved in an inositol phospholipid-based intracellular signaling cascade and a component in the phospho-dependent endocytosis process of GABA-A receptor. Also six QTLs (related to meat and carcass association: 3 QTLs, exterior association: 2 QTLs, milk association: 1 QTL) belonged to CHR2:86831095-87004473 region containing PLCL1 genes and three QTLs were Holstein specific [18]. Especially, a specific trait of milk association QTL (cattle QTL ID: 25003) was milk fat percentage and Holstein specific. ACADM is associated with lipid metabolism in fat depot and is the most important enzyme in the ACAD family [19]. Inside the mammary epithelial cell, the triglycerides synthesized at the outer surface of the smooth endoplasmic reticulum start coalescing and forming micro lipid droplets. Schlegel et al reported the relative mRNA abundances of *ACADM* genes involved in fatty acid oxidation in the liver of dairy cows in the transition period and at different stages of lactation [20]. Ran Zhang reported that low density lipoprotein receptor class A domain containing 3 (LDLRAD3) plays a central role in mammalian cholesterol metabolism through Next-Generation Sequencing in Transgenic Cattle [21]. Also four QTLs (related to reproduc-

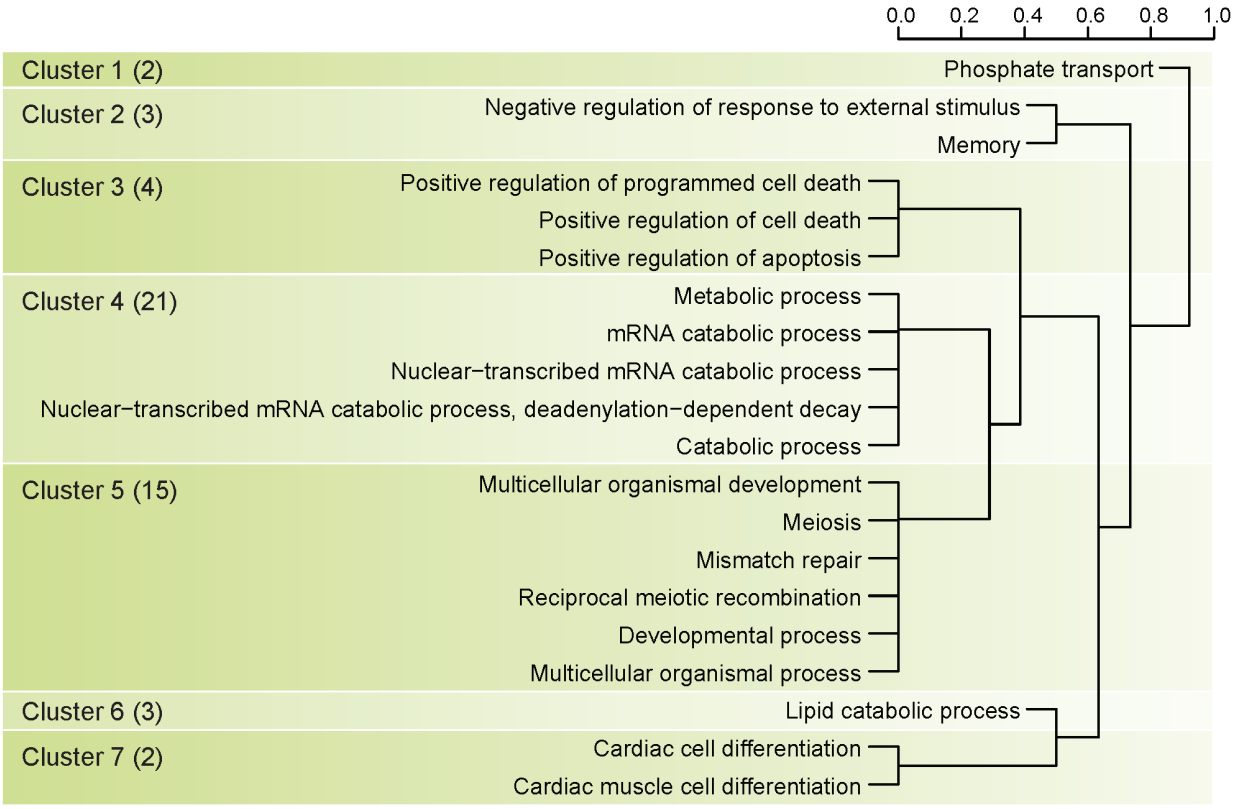

Figure 3. Clustering based on result of gene ontology analysis using gene set of significant genetic regions as genome-wide association studies (GWAS) result of milk protein. Number in parenthesis means number of genes in each cluster group.

tion: 3 QTLs, production: 1 QTL) belonged to CHR15:67269656-67429625 region containing *LDLRAD3* genes and three QTLs related to reproduction were Holstein specific [22]. Additionally, specific traits of three reproduction QTLs (cattle QTL ID: 15,193, 15,194, and 15,192) were calving ease (direct), calving index and calving ease (maternal), respectively.

There were 54 significant protein coding genes in the milk protein association test, and we identified twenty gene ontology terms to detect biological meaning of 54 genes related to milk protein (Figure 2). We clustered twenty gene ontology terms into 7 main terms using hierarchical clustering method (Figure 3). Cluster 4 in Figure 3 had the most number of genes related to milk protein (21 genes) and biological meaning of this cluster was mRNA metabolic process. The need of energy and protein during lactation increases dramatically. In dairy cows there is more than a 5-fold increase in energy and protein requirements from late gestation to lactation [23]. Another study using more precise measurements of daily tissue protein synthesis reported that there is a 4-fold increase in mRNA translation in lactating compared to non-lactating mammary tissue in the cow [24]. Because the efficiency to transform dietary nitrogen into milk proteins is low (25% to 30%), protein synthesis is a highly active and energetically costly process, with only a minor part of the synthetic machinery apparently being used for production of milk proteins. Also previous study reported the abundance of the milk proteins (with the exception of albumin, as discussed below) is highly-dependent on the transcription level [25]. Cluster 5 in Figure 3 had 15 genes related to milk protein and biological meaning in this clustering was multicellular organism process. Multicellular organisms are composed of many specialized cells which differ in structure and function. So we guessed that these 15 genes have a special relationship with the milk protein ingredients or mechanisms. Interestingly, Cluster 3 in Figure 3 had 4 genes and their biological meaning was programmed cell death. Programmed cell death was not directly associated with milk protein. But programmed cell death has substantial meaning in Holstein mammary biological system. The regulation of cell death initiation coupled to the removal of cell corpses is an integral part of the mammary gland life cycle [26]. During pregnancy, epithelial cells of the mammary gland expand to form branched and lobuloalveolar structures to allow milk production after birth of the offspring. Then on weaning of the progeny, the mammary gland undergoes an important remodeling step, termed involution, during which the unessential mammary epithelial cells die and are largely removed [27]. Additionally, Baik reported that protein kinase C eta (PRKCH) were differential expressed in mammary tissues of lactating dairy cows [28].

Our results strongly support a major involvement of milk production in the genetic predisposition for increasing capacity of Holstein milk production and suggest several novel genes as genetic factors in milk production. Our results are not overlapped by other some previous GWAS of Holstein production traits. But several previous studies reported that some of our results were related to milk production traits of Holstein. Also, we identified that some of our results overlapped QTLs of cattle milk production. Although we will have to collect more samples and further research will be needed, we thought that our investigated genetic regions were biologically related to milk production traits.

CONCLUSION

These candidate regions and genes in our results may provide insight into the genetic makeup underlying milk production of Korean Holsteins.

CONFLICT OF INTEREST

We certify that there is no conflict of interest with any financial organization regarding the material discussed in the manuscript.

ACKNOWLEDGMENTS

The work was supported by Project (PJ0092602015) of the National Livestock Research Institute of Rural Development Administration, Republic of Korea.

REFERENCES

1. Georges M, Nielsen D, Mackinnon M, et al. Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. Genetics 1995;139:907-20.
2. Meuwissen T, Goddard ME. Prediction of identity by descent probabilities from marker-haplotypes. Genet Sel Evol 2001;33:605-34.
3. Daw EW, Heath SC, Lu Y. Single-nucleotide polymorphism versus microsatellite markers in a combined linkage and segregation analysis of a quantitative trait. BMC Genetics 2005;6:S32.
4. Georges M. Mapping, fine mapping, and molecular dissection of quantitative trait loci in domestic animals. Annu Rev Genomics Hum Genet 2007;8:131-62.
5. Brym P, Kamiński S, Wójcik E. Nucleotide sequence polymorphism within exon 4 of the bovine prolactin gene and its associations with milk performance traits. J Appl Genet 2004;46:179-85.
6. Jiang L, Liu J, Sun D, et al. Genome wide association studies for milk production traits in Chinese Holstein population. PloS one 2010;5:e13661.
7. Fontanesi L, Calò D, Galimberti G, et al. A candidate gene association study for nine economically important traits in Italian Holstein cattle. Anim Genet 2014;45:576-80.
8. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81:559-75.
9. Browning BL, Browning SR. A fast, powerful method for detecting identity by descent. Am J Hum Genet 2011;88:173-82.
10. Endelman JB. Ridge regression and other kernels for genomic selection with R package rrBLUP. Plant Genome 2011;4:250-5.
11. Moreau Y, Aerts S, De Moor B, De Strooper B, Dabrowski M. Comparison and meta-analysis of microarray data: from the bench to the computer desk. TRENDS Genet 2003;19:570-7.
12. Cho C, Cho K, Choy Y, et al. Estimation of genetic parameters for milk production traits in Holstein dairy cattle. J Anim Sci Technol 2013;55:7-11.
13. Shin DH, Cho KH, Park KD, Lee HJ, Kim H. Accurate estimation of effective population size in the Korean dairy cattle based on linkage disequilibrium corrected by genomic relationship matrix. Asian-Australas J Anim Sci 2013;26:1672-9.
14. Ogorevc J, Kunej T, Razpet A, Dovc P. Database of cattle candidate genes and genetic markers for milk production and mastitis. Anim Genet 2009;40:832-51.
15. Kramer M, Erbe M, Seefried F, et al. Accuracy of direct genomic values for functional traits in Brown Swiss cattle. J Dairy Sci 2014;97:1774-81.
16. Wu W, Liu J, Xu G, Ye J. Calcium homeostasis, acid–base balance, and health status in periparturient Holstein cows fed diets with low cation–anion difference. Livest Sci 2008;117:7-14.
17. Piantoni P, Daniels K, Everts R, et al. Level of nutrient intake affects mammary gland gene expression profiles in preweaned Holstein heifers. J Dairy Sci 2012;95:2550-61.
18. Raven L-A, Cocks BG, Hayes BJ. Multibreed genome wide association can improve precision of mapping causative variants underlying milk production in dairy cattle. BMC Genomics 2014;15:62.
19. Illig T, Gieger C, Zhai G, et al. A genome-wide perspective of genetic variation in human metabolism. Nat Genet 2010;42:137-41.
20. Schlegel G, Keller J, Hirche F, et al. Expression of genes involved in hepatic carnitine synthesis and uptake in dairy cows in the transition period and at different stages of lactation. BMC Vet Res 2012;8:28.
21. Zhang R, Yin Y, Zhang Y, et al. Molecular characterization of transgene integration by next-generation sequencing in transgenic cattle. PLOS One 2012;7:e50348.
22. Sahana G, Guldbrandtsen B, Lund MS. Genome-wide association study for calving traits in Danish and Swedish Holstein cattle. J Dairy Sci 2011;94:479-86.
23. Church DC. The ruminant animal. Digestive physiology and nutrition. Englewood Cliffs, NJ: Prentice Hall; 1988.
24. Kühn C, Freyer G, Weikard R, Goldammer T, Schwerin M. Detection of QTL for milk production traits in cattle by application of a specifically developed marker map of BTA6. Anim Genet 1999;30:333-9.
25. Rhoads RE, Grudzien-Nogalska E. Translational regulation of milk protein synthesis at secretory activation. J Mammary Gland Biol

Neoplasia 2007;12:283-92.
26. Watson CJ, Kreuzaler PA. Remodeling mechanisms of the mammary gland during involution. Int J Dev Biol 2011;55:757.
27. Monks J, Smith-Steinhart C, Kruk ER, Fadok VA, Henson PM. Epithelial cells remove apoptotic epithelial cells during post-lactation involution of the mouse mammary gland. Biol Reprod 2008;78:586-94.
28. Baik M, Etchebarne B, Bong J, VandeHaar M. Gene expression profiling of liver and mammary tissues of lactating dairy cows. Asian-Australas J Animal Sci 2009;22:871-84.