Open Access
Asian Australas. J. Anim. Sci.
Vol. 29, No. 3 : 343-351 March 2016
http://dx.doi.org/10.5713/ajas.15.0525
www.ajas.info
pISSN 1011-2367 eISSN 1976-5517

Characterizing Milk Production Related Genes in Holstein Using RNA-seq

Minseok Seo[1,2,a], Hyun-Jeong Lee[1,3,a], Kwondo Kim[4], Kelsey Caetano-Anolles[5], Jin Young Jeong[6], Sungkwon Park[3,7], Young Kyun Oh[3], Seoae Cho[2], and Heebal Kim[1,2,5,*]

[1] Interdisciplinary Program in Bioinformatics, Seoul National University, Seoul 151-741, Korea

ABSTRACT: Although the chemical, physical, and nutritional properties of bovine milk have been extensively studied, only a few studies have attempted to characterize milk-synthesizing genes using RNA-seq data. RNA-seq data was collected from 21 Holstein samples, along with group information about milk production ability; milk yield; and protein, fat, and solid contents. Meta-analysis was employed in order to generally characterize genes related to milk production. In addition, we attempted to investigate the relationship between milk related traits, parity, and lactation period. We observed that milk fat is highly correlated with lactation period; this result indicates that this effect should be considered in the model in order to accurately detect milk production related genes. By employing our developed model, 271 genes were significantly (false discovery rate [FDR] adjusted p-value<0.1) detected as milk production related differentially expressed genes. Of these genes, five (albumin, nitric oxide synthase 3, RNA-binding region (RNP1, RRM) containing 3, secreted and transmembrane 1, and serine palmitoyltransferase, small subunit B) were technically validated using quantitative real-time polymerase chain reaction (qRT-PCR) in order to check the accuracy of RNA-seq analysis. Finally, 83 gene ontology biological processes including several blood vessel and mammary gland development related terms, were significantly detected using DAVID gene-set enrichment analysis. From these results, we observed that detected milk production related genes are highly enriched in the circulation system process and mammary gland related biological functions. In addition, we observed that detected genes including caveolin 1, mammary serum amyloid A3.2, lingual antimicrobial peptide, cathelicidin 4 (*CATHL4*), cathelicidin 6 (*CATHL6*) have been reported in other species as milk production related gene. For this reason, we concluded that our detected 271 genes would be strong candidates for determining milk production. (**Key Words:** RNA-seq, Holstein, Milk Production, Meta-analysis, Milk Yield, Differentially Expressed Gene)

INTRODUCTION

The milk composition and production rates of Holstein cattle make them extremely valuable from an economic perspective. Milk cows have been and continue to be genetically selected for improvement of their milk production. Milk production involves several factors, including production of lactose, fat, protein, vitamins, enzymes, and water. Although milk composition varies depending on species (cow, goat, and sheep) and breed (Holstein and Jersey), as well as the stage of lactation (Wickramasinghe et al., 2012) and other environmental factors, milk composition contents are generally reported as water (87.7%), lactose (4.9%), fat (3.4%), protein (3.3%), and others (0.7%) (Constantin and Csatlos, 2010). Of these components, the amount of solid (amount of milk excluding water), fat, protein, and total milk yield are representative economic traits (Vargas et al., 2002). In order to identify the increasing factors for milk production of these traits, many studies have been performed to characterize biomarkers in

* Corresponding Author: Heebal Kim. Tel: +82-2-880-4803, Fax: +82-2-883-8812, E-mail: heebal@snu.ac.kr
[2] CHO&KIM genomics, Seoul 151-919, Korea.
[3] Animal Nutritional & Physiology Team, National Institute of Animal Science, Jeonju 565-851, Korea.
[4] Department of Agricultural Biotechnology, Animal Biotechnology Major, and Research Institute of Agriculture and Life Sciences, Seoul National University, Seoul 151-921, Korea.
[5] Department of Animal Sciences, University of Illinois, Urbana, IL 61801, USA.
[6] Division of Animal Products R&D, National Institute of Animal Science, Jeonju 565-851, Korea.
[7] Department of Food Science and Technology, Sejong University, Seoul 143-747, Korea.
[a] These authors contributed equally to this work.
Submitted Jun. 18, 2015; Revised Sept. 17, 2015; Accepted Oct. 25, 2015

diverse omics-data on cattle species. Many studies take a transcriptome data analysis based approach in order to characterize genes. In one study (Cánovas et al., 2010), single nucleotide polymorphism (SNP) discovery analysis was performed on bovine RNA-seq data; many SNPs were detected in transcribed regions, which provides insight into gene expression. Another transcriptional study (Wickramasinghe et al., 2012) identified expressed genes related to transition lactation. Results of this study revealed a large number of genes expressed during the lactation period. Additionally, a small number of genes were only observed during specific lactation stages. While many transcriptional analyses have been performed related to the milk cow, only a few studies have been attempted to detect milk production related genes. One RNA-seq study (Cui et al., 2014), conducted analysis in order to detect milk yield related genes using a two-group comparison between high yielding (HY) and low yielding (LY) cattle. This study focused primarily on the identification of differentially expressed genes (DEGs) related to bovine milk production. However, lactation period and parity were not considered in these studies when conducting statistical tests. This is primarily due to a lack of biological replicates. As results of recent studies imply that lactation period and parity have a significant impact on milk yield (Yoon et al., 2004; Wickramasinghe et al., 2012), these factors should be considered in the identification of genes related to milk production. Development of recent RNA-seq techniques allow for more accurate detection of DEGs than microarray platforms (Wang et al., 2009). Using RNA-seq, transcriptome sequences as well as mRNA expression can be measured in a highly reproducible manner, meaning that between technical-replicates are highly correlated to each other. In this respect, an RNA-seq based approach is more suitable for identification of DEGs than use of a microarray platform. However, most recent RNA-seq based studies have used only a small number of biological replicates due to its high cost. Given the high accuracy, importance of technical replicates is less in RNA-seq than other platforms. However, the importance of biological replicates cannot be overemphasized for estimating variability. There are several advantages of using many biological replicates. Primarily, more complex experimental designs can be considered in statistical analysis; when considering multiple factors (i.e. lactation period and parity in milk yield related analysis), many biological replicates are needed. In this study, we performed RNA-seq analysis with several biological replicates in order to identify milk production related genes. The statistical analysis was performed with 21 Holstein RNA-samples with their four milk production ability traits: total milk-yield and three milk compositions such milk fat, protein, and solid.

MATERIALS AND METHODS

Animal information and procedures of RNA-seq experiment

RNA-seq data was collected on the same day from 21 randomly selected Holstein cows in high (n = 9) and low (n = 12) milk yielding ability groups, raised at the Kang Sung Won farm, Korea, in their 2nd to 4th lactation (n = 14, 4, and 3 samples in each parity group). Group information was coded as two groups: high (breeding value ≥0) and low (breeding value <0) based on the estimated breeding values derived from Korea Type-Production Index (KTPI) at the National Institute of Animal Science (NIAS). More detailed information can be found in Supplementary Table S1. Cows were kept in free stall housing, fed with total mixed ration and supplied with water *ad libitum*. They were milked twice a day, at 4 am and 4 pm, in a designated milking parlor and all Korea Hazard Analysis and Critical Control Point (HACCP) guidelines were followed. Milk samples were collected by hand-milking 2 to 3 hours after the evening milking at 60, 100 to 160, 180 to 210, and 240 to 270 days of lactation. More detailed animal information is included in (Supplementary Table S1). Samples were assessed for cell viability using the typan blue method and total RNA was extracted. Somatic cells were collected from fresh milk treated with 50 μL of 0.5 M ethylenediaminetetraacetic acid (EDTA), centrifuged at 1,800 rpm at 4℃ for 15 min, and washed with 10 mL of phosphate-buffered saline (PBS) (pH 7.2, diluted with 0.1% diethylpyrocarbonate) and 10 μL of 0.5 M EDTA. Cells were centrifuged at 1,800 rpm, 4℃ for 15 min, and re-suspended in PBS after supernatant removal. Total RNA isolation was performed according to the manufacturer instructions using the TRIzol reagent (Molecular Research Center, Cincinnati, OH, USA). Total RNA levels were quantified by absorbance at 260 nm using ND-1000 spectrophotometer (Fisher Thermo, Wilmington, MA, USA), and RNA integrity was assessed by 1% (w/v) agarose gel electrophoresis followed by ethidium bromide staining of the 28S and 18S bands. Total RNA (1 μg amounts) was reverse-transcribed into cDNA using an iScript cDNA Synthesis kit (Bio-Rad, Hercules, CA, USA), following the manual. Illumina Hiseq 2000 was used for RNA-seq based on the manufacturer instruction.

Read mapping on the genome reference and normalization

For removal of adapters, we used Trimmomatic with the following parameters: PE -phred33 ILLUMINACLIP: TruSeq3-PE.fa:2:30:10 MINLEN:75 2. These clean-reads were mapped to the reference genome (BosTau7) from the UCSC database using Bowtie2, included in Tophat2. This is one of the most commonly used tools for mapping to the genome reference. The aligned result from Tophat2 was

converted from BAM to SAM format using SAMtools. Next, gene expression levels were estimated using the HTseq package, implemented in python and cross-referencing the Bos taurus7 gene transfer format (.GTF) file. From the read mapping result, annotated gene expression levels were estimated.

DEG analysis using negative-binomial distribution based generalized linear model

In order to detect DEGs corresponding to milk production ability (milk yield, fat, protein, and solid), we employed the edgeR package implemented within R. EdgeR allows for two group comparison using a multi factorial design by including explanatory variables on the linear predictor. Lactation period and parity also can be considered on the model as follows:

$$\log(\theta)_i = \beta_0 + \beta_1 \cdot \mathrm{Group}_{1i} + \beta_2 \cdot \mathrm{LP}_{2i} + \beta_3 \cdot \mathrm{PR}_{3i} \quad (1)$$

Samples indexed by *i*, LP and PR, are lactation period and parity respectively. Using group information, statistical tests were performed for each trait. One complication of carrying out a study on milk-production related genes using OMICS data is that many salient traits exist which should be considered in statistical analysis, including milk yield, and fat, protein, and solid content. In order to consider multiple variables simultaneously, there are two representative approaches: multivariate-model based analysis (Chauhan and Hayes, 1991; Hill et al., 1999) and meta-analysis (Onetti and Grummer, 2004; Glasser et al., 2008). While multivariate analysis has strong statistical power compared to simply combining results of univariate analysis, this approach can create many false-positive results and interpretation of results can be difficult (Thompson and Higgins, 2002). On the other hand, meta-analysis is typically used to combine probability values (p-values) from each univariate analysis. Although this approach is simple and less powerful, interpretation of results is easier than that of multivariate-model based analysis. Given this advantage, the GRACOMICS tool was recently developed for comparing multiple results using OMICS data (Seo et al., 2015). In the present study, we used a meta-analysis approach implemented in GRACOMICS to identify milk production related DEGs through comparison and/or combination of the four milk production ability information (total milk yield and fat, protein, and solid) using RNA-seq analysis. We adjusted for multiple testing errors at a significance level of false discovery rate (FDR) adjusted $p<0.1$. After DEG detection, DAVID gene-set enrichment analysis was performed in order to characterize functional differences between high and low milk producing groups.

Technical validation of detected significant milk production related genes

All primers were designed using reference sequences published by the National Center for Biotechnology Information. Real-time polymerase chain reaction (PCR) was performed using QuantiTect SYBR Green RT-PCR Master Mix (Qiagen, Valencia, CA, USA) and 7500 Fast Sequence Detection System (Applied Biosystems, Foster City, CA, UDA). Briefly, PCR was performed in a final reaction volume of 25 μL containing 200 ng cDNA, 12.5 μL SYBR Green RT-PCR Master Mix, and 1.25 μL of each of two primer solutions (10 μM). Parameters for thermal cycling were as follows: 50℃ for 2 m and 95℃ for 15 m followed by 40 cycles at 94℃ for 15 s, 60℃ for 30 s, and 72℃ for 30 s. ΔCt values were used to determine relative fold changes in mRNA levels. A total of 23 biological replicates including RNA-sequenced samples and an additional five samples (Supplementary Table S1) were employed for the experiment. All data was normalized with the housekeeping Ribosomal protein S9 (*RPS9*) gene. A student t-test was used for significance testing between high and low yielding groups.

RESULTS

Investigating the relationship between milk production related traits and influencing factors such as parity and lactation period

In order to identify relationships between the milk related traits, correlation coefficients were calculated among four milk-production related traits including two influencing factors; parity and lactation period (Figure 1A). In this figure, strong positive correlations can be observed between the amount of milk yield, fat, protein, and solid in phenotypic traits (0.58 to 0.98) and milk fat, protein, and solid content (0.47 to 0.52). Particularly, milk yield, solid, and protein were highly correlated with each other in phenotypic traits (0.96 to 0.98). While parity and lactation period were not highly correlated with these traits (Absolute correlation: 0.06 to 0.58), milk fat content and lactation period were highly correlated (–0.58). In order to statistically test the relationship between traits and two factors, a linear regression was used. Results revealed that milk fat content was highly associated with lactation period (p-value: 0.005) as shown in Table 1 and Figure 1B. Although parity and lactation period were not highly correlated with entire traits, relatively higher correlations were observed in the amount of 305 milk-fat and milk-fat content. While we expected that milk yield would be highly associated with parity and lactation period due to the fact that many milk related studies imply that milk yield is highly influenced by these factors, no significant association (p-value: 0.543) was observed in our analysis

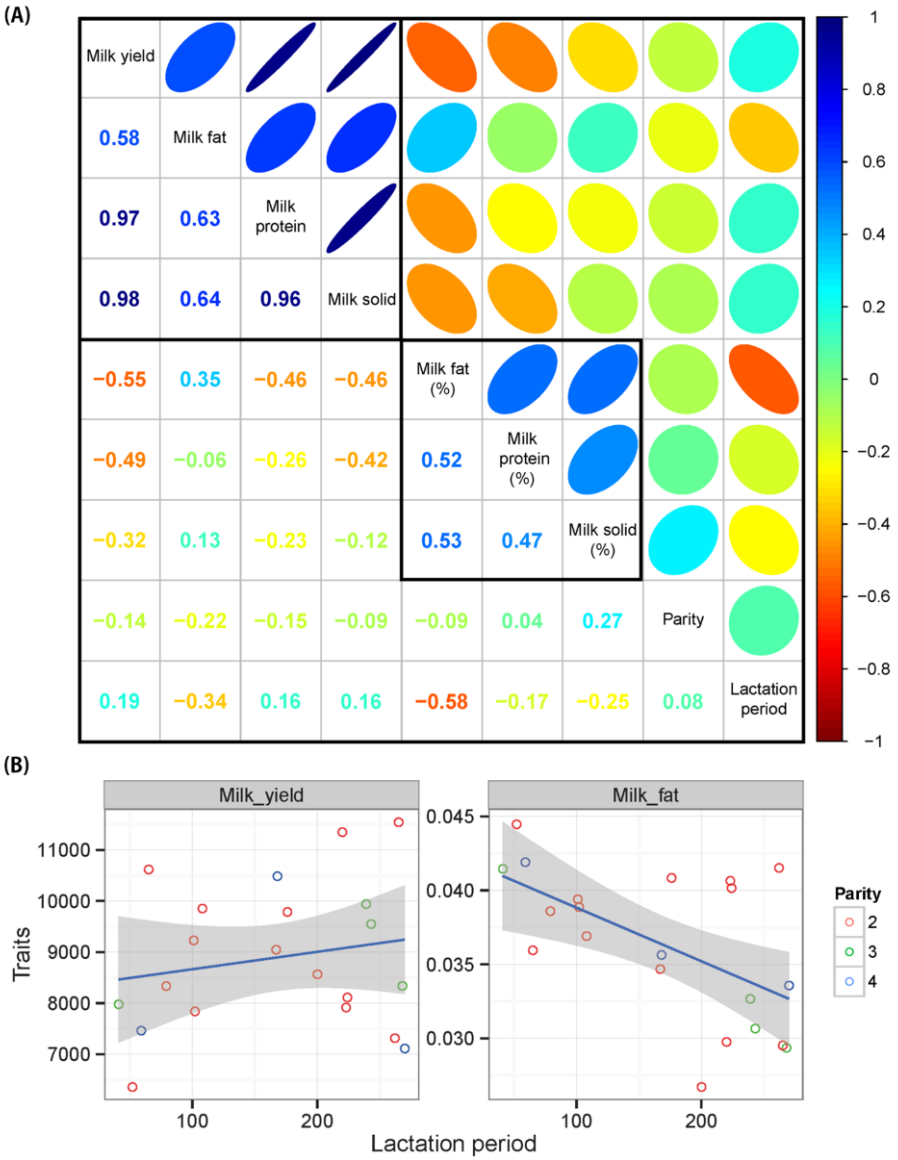

Figure 1. Relationships among milk related traits. (A) Pearson correlation coefficients among the traits including parity and lactation period using 21 Holstein samples. Absolute correlation, 0.58 to 0.98 were observed in phenotypic traits. The milk yield, fat, protein, and solid content presented is the result of 305 days of yield (kg). Percentages of the milk fat, protein, and solid represent the content of the each component in samples. (B) Linear regression fitted plots between milk related traits and lactation period. The red, green, and blue colors represent 2, 3, and 4 times parity, respectively. Only composition of milk-fat was significantly observed in association test (p-value <0.05).

(Figure 1B and Table 1). Given these results, we confirm that milk related traits and component contents were highly correlated each other. In addition, parity and lactation period should be considered in the model when detecting DEGs, particularly in studies investigating milk-fat.

Table 1. Association test results between milk related trait and major influence factors such as parity and lactation period

Traits	Parity	Lactation period
Milk yield	0.543	0.397
Milk fat	0.333	0.127
Milk protein	0.509	0.475
Milk solid	0.693	0.499
Milk fat (%)	0.686	0.005*
Milk protein (%)	0.870	0.452
Milk solid (%)	0.243	0.272

Identification of milk production related DEGs

We performed two group tests between high and low ability groups related to the four milk related traits (milk yield, fat, protein, and solid). As a result, 536, 332, 282, and 431 genes were identified as significant (under the 5% significance level) DEGs relating to milk yield, fat, protein, and solid contents, respectively. In order to compare

detected DEGs across the each trait, we created a Venn diagram (Figure 2A). As expected, many DEGs overlap with each other between traits. Only 200 (37.3%), 103 (31%), 28 (9.9%), and 230 (53.3%) DEGs were identified as unique to each specific trait; a total of 83 DEGs were commonly identified in all traits. Based on the detected DEGs, hierarchical clustering analysis was performed as shown in Figure 2B along with information of milk yielding ability. In this figure, we observed that two groups, HY and LY, were clearly separated except for a single LY sample. Although one sample was included in the HY cluster, our clustering result revealed smaller within group distances and without group distance between HY and LY. This confirms successful identification of detected DEGs. Finally, in order to investigate milk production related genes overall, four univariate analysis results were combined by using meta-analysis methods implemented in the GRACOMICS. As a result, 271 genes were significantly detected as milk production related genes (FDR adjusted p-value<0.1). As shown in Figure 2A, significantly detected genes showed high expression in the LY group, similar to the pattern previously shown. The lingual antimicrobial peptide (*LAP*) was observed as the most significant gene (Combined p-value: 0.0 and log2 fold-change: 2.286) in the comparison between HY and LY as well as meta-analysis. The serine palmitoyltransferase, small subunit B (*SPTSSB*; 1.21E-07), FXYD domain containing ion transport regulator 3 (*FXYD3*; 8.47E-08), and palmdelphin (*PALMD*; 7.09E-09) were found to be differentially expressed in the HY and LY groups.

Technical validation of the significantly detected DEGs using qRT-PCR

For technical validation of RNA-seq analysis results, we randomly selected 5 DEGs and qRT-PCR was performed on these genes using 23 biological replicated samples. All genes were significantly detected based on the meta-analysis results of qRT-PCR analyses at 10% significance level (Table 2). Furthermore, we observed that these genes were also significantly detected in the comparison between HY and LY (milk yielding ability groups). Of the five validated genes, albumin (*ALB*) was significantly detected as an overall milk production related traits in RNA-seq analysis, but only significantly detected as a milk yield

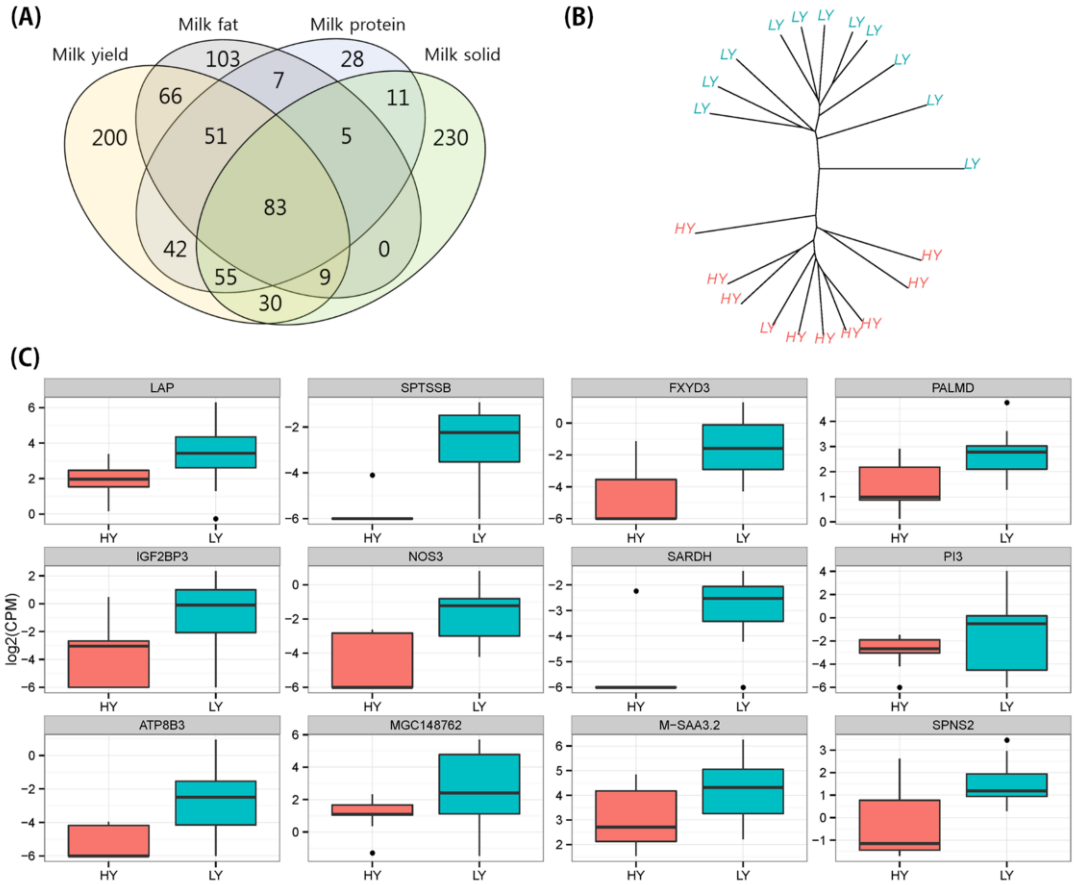

Figure 2. Identification of the differentially expressed genes between high yielding (HY) and low yielding (LY). (A) Venn-diagram for the comparing significant genes across the four types of milk production abilities: total milk yield and milk fat, protein, and solid content. (B) Hierarchical clustering analysis using normalized gene expression of the significantly detected genes between HY and LY (p-value<0.05) with number of group (k = 2). (C) The box-plots of most significant 12 genes comparing HY and LY.

Table 2. Technical validation results of the significantly detected DEGs in RNA-seq analysis[1]

Gene_symbol	Milk_yield	Milk_fat	Milk_protein	Milk_solid	Combined p-value
	---------- Statistical test results in RNA-seq data ----------				
ALB	3.67E-04*	5.49E-04*	0.002774*	0.057027*	8.42E-08*
NOS3	1.12E-04*	0.002912*	1.40E-05*	0.011061*	0*
RNPC3	0.016529*	0.182063	0.25322	0.899593	0.068049*
SECTM1	0.012424*	0.332108*	0.181233	0.914573	0.067949*
SPTSSB	2.99E-05*	0.003118*	0.010716*	0.048322*	1.21E-07*
	---------- Statistical test results in qRT-PCR data ----------				
ALB	0.022979*	0.169538	0.224582	0.408038	0.044196*
NOS3	0.013338*	0.145807	0.052182*	0.004669*	3.01E-04*
RNPC3	0.050623*	0.095703*	0.02621	0.036443*	0.00184*
SECTM1	0.003012*	0.010871*	0.044447	0.143359	1.55E-04*
SPTSSB	0.032168*	0.074156*	0.02278*	0.266572	0.004415*

DEGs, differentially expressed genes; *ALB*, albumin; *NOS3*, nitric oxide synthase 3; *RNPC3*, RNA-binding region (RNP1, RRM) containing 3; *SECTM1*, secreted and transmembrane 1; *SPTSSB*, serine palmitoyltransferase, small subunit B; qRT-PCR, quantitative real-time polymerase chain reaction.
[1] Randomly selected 5 DEGs; *ALB*, *NOS3*, *RNPC3*, *SECTM1*, and *SPTSSB*, were technically validated using qRT-PCR.
* Represents significant result (p-value<0.1).

related trait in qRT-PCR analysis. For nitric oxide synthase 3 (*NOS3*), the only non-significant results was observed in the comparison of milk fat production ability using qRT-PCR analysis. Contrastively, RNA-binding region (RNP1, RRM) containing 3 (*RNPC3*) was only observed as a DEG in the comparison of milk yielding ability groups using RNA-seq, however, qRT-PCR analysis showed that *RNPC3* is a significant DEG in the comparison of milk yielding ability as well as milk fat and solid production ability. Finally, secreted and transmembrane 1 (*SECTM1*) and *SPTSSB* were significantly observed in most comparison results in both platforms. Technical validation, confirmed that our RNA-seq analysis was successfully and accurately performed.

Gene-set enrichment analysis using significantly detected DEGs

In order assign biological meaning to our results, DAVID gene-set enrichment analysis was performed using 271 detected milk production related DEGs. As a result, 83 gene ontology (GO) biological processes were significantly observed (Supplementary Table S2). We observed several significantly enriched categories such as metabolic process related terms; *icosanoid metabolic process* (p-value: 0.0818) and *unsaturated fatty acid metabolic process* (0.0818), and defense response related terms; *defense response* (0.049), *response to bacterium* (0.0019), *defense response to bacterium* (0.0013). Of many significant terms, we observed two representative GO term clusters- mammary gland and blood circulation related terms. Among significantly detected GO terms, 12 processes were associated with the blood vessel system and 4 terms were linked closely to mammary gland development. For further investigation of these relationships, we visualized an ancestor chart using these terms (Figure 3). As shown in Figure 3, circulatory system process and gland development terms were highly enriched. We observed that most circulation related terms contain caveolin 1 (*CAV1*), myosin, light chain 3, alkali (*MYL3*), endothelin 2 (*EDN2*), and *NOS3* genes. Additionally, gland development related terms included *CAV1*, homeobox A3 (*HOXA3*), E74-like factor 3 (*ELF3*), forkhead box A1 (*FOXA1*), homeobox A9 (*HOXA9*) genes as shown in Supplementary Table S2. In light of these results, *CAV1* gene may be a strong candidate to serve as a key marker for characterization of milk production as it appears to act as a hub gene engaging in two biological processes, blood circulation and gland development. In addition, *CAV1* gene is involved in diverse regulation processes such as regulation of protein amino acid phosphorylation (0.098), regulation of cell proliferation (0.021), regulation of peptidyl-tyrosine phosphorylation (0.0775), and regulation of tyrosine phosphorylation of Stat5 protein (0.0998).

DISCUSSION

From RNA-seq analysis on 21 Holstein samples, 271 milk production related genes and 83 related biological terms were identified. Of detected genes, mammary serum amyloid A3.2 (*M-SAA3.2*), sarcosine dehydrogenase (*SARDH*), *LAP*, *NOS3*, claudin 6 (*CLDN6*), and polypeptide N-acetylgalactosaminyltransferase 16 (*GALNTL1*), were found to be the most significantly detected in meta-analysis (Combined p-value: 0.0). *M-SAA3.2* has been widely researched in the milk production (McDonald et al., 2001; Molenaar et al., 2009; Cui et al., 2014). Results of these studies suggest *M-SAA3.2* as a promising candidate gene related to milk production. Particularly, genetic studies using quantitative trait loci (QTL) analysis have implied that significant SNPs between HP and LP are located in the

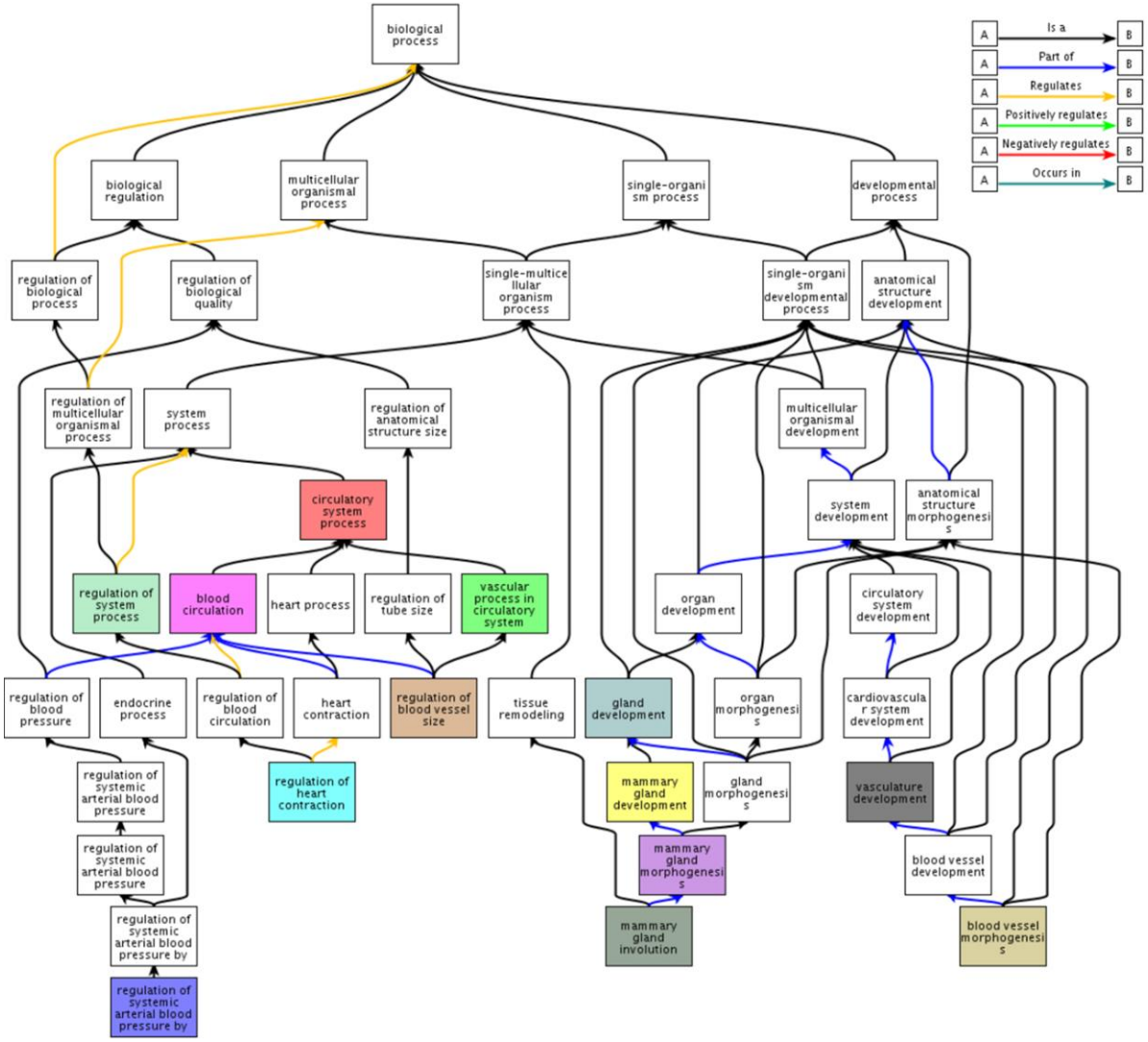

Figure 3. The ancestor chart of the significantly detected biological process from DAVID analysis. Of significantly observed gene ontology (GO) terms, blood circulation and gland development related terms were visualized as ancestor chart using QuickGO. Each box represents GO term and colored boxes are significantly observed terms in the DAVID gene-set analysis.

M-SAA3.2 gene. Expression of *CLDN6* has been uncovered as a cue to induction of epithelial differentiation in mouse stem cells (Sugimoto et al., 2013). Epithelial differentiation is a known necessary process in mammary gland development (Hennighausen and Robinson, 2001). From the results of the present DEG analysis and previous studies, it appears that *CLDN6* is associated to epithelial differentiation in cattle, which result in the difference of milk production related traits. In previous studies, *GALNTL1* has been found to contain QTL region with bovine marbling trait (Takasuga et al., 2007) and to be down-regulated during bovine intramuscular adipogenesis (Mizoguchi et al., 2010). We also found that *GALNTL1* was significantly expressed in univariate analysis of the milk fat trait (log2fold change: 5.296), which was the most significant difference of all genes analyzed (FDR adjust p-value: 0.034). Additionally, the differentially expression of *GALNTL1* in all univariate analyses as well as in meta-analysis implies that this gene may be associated to overall milk yield process. Another DEG, lingual antimicrobial peptide (*LAP*), is related with the defense mechanism to bacterium (Yarus et al., 1996). This result corresponds with results of our gene-set analysis, which contained many defense mechanism related terms including killing of cells of another organism (p-value: 6.27E-05), cell killing (6.46E-04), defense response to bacterium (0.0013), response to bacterium (0.0019), defense response to fungus (0.0058), and defense response (0.049). The commonality of these terms included *CATHL4* (Combined p-value: 9.75E-05), *LAP* (0.0), *ALB* (8.42E-08), and *CATHL6* (2.51E-05) genes. When considering that our experiment was performed using somatic milk cells, it is reasonable to detect several defense mechanism terms.

We performed technical validation experiment to

confirm our RNA-seq results. 5 randomly selected genes were confirmed as milk production related genes using qRT-PCR (Table 2). Of those genes, *ALB* and *NOS3* have been previously reported as milk-related. *ALB* is responsible for synthesizing albumin; bovine serum albumin (BSA) is a component of bovine milk (Pepe et al., 2013). *NOS3* synthesizes nitric oxide synthase 3; it has been shown to associate with nipple erection in humans (Tezer et al., 2012). Dairy cattle may have an identical mechanism of nipple erection. Because nipple erection can induce milk ejection reflex and accompany the ejection itself, the relationship between *NOS3* gene and milk yield is clearly deducible. The other three genes, RNA-binding region (RNP1, RRM) containing 3 (*RNPC3*), secreted and transmembrane 1 (*SECTM1*), and serine palmitoyltransferase, small subunit B (*SPTSSB*), have not been previously reported in milk production studies. Discovery of these genes is a novel finding in relation to milk production, and may potential serve as future candidate genes for selective breeding.

From gene-set enrichment analysis, two representative biological processes were observed as shown in (Figure 3 and Supplementary Table S2): blood vessel and mammary gland development. Generally, blood vessel development is required for cell proliferation, which leads to tissue development, including the mammary gland. Furthermore, the development of mammary glands is directly and indirectly connected with milk production as a tissue responsible for milk yielding. For example, during the dry period, mammary tissue experiences extensive cell proliferation and turnover to compensate for cell loss during the preceding lactation period and to replace senescent secretory cells (Tao et al., 2011). This process, called mammary gland involution is an important prerequisite of the subsequent lactation as a part of the reproductive cycle in dairy cows (Accorsi et al., 2002). A previous study showed that cell turnover is the major determinant for the lactation curve after lactation peak (Sorensen et al., 2006). Therefore, the identification of gene-sets related to blood vessel and mammary gland is unsurprising. *CAV1* was found significant in results of RNA-seq analysis (8.66E-04) as well as gene-set analysis as shown in (Supplementary Table S2). *CAV1* is well known as representative milk production related gene (Park et al., 2002) in rodent species; *CAV1* null mice show accelerated development and premature milk production as well as lobuloalveolar compartment. Given this result, we suggest that *CAV1* gene may perform a similar role in the cattle species. In addition, 8 terms related to cell proliferation or anti-apoptosis were also identified through gene-set enrichment analysis. In general, tissue development requires cell proliferation, which is also applied to mammary tissue as explained above. On the contrary, apoptosis is in opposition to cell proliferation in terms of tissue development. Therefore, these terms are indirectly associated to milk production along with blood vessel and mammary gland development. A majority of biological processes also identified belonged to tissue development or tissue development-related category, which imply that tissue development, not milk component, might be a major factor for enhancing milk production.

The present study performed meta-analysis by combining four RNA-seq analysis results. As far as we know, this is the first attempt to use this approach to study milk-production at the transcriptomic level. Generally, this approach is widely used in genome level data analysis for higher statistical power (Onetti and Grummer, 2004; Glasser et al., 2008; Neale et al., 2010; Minozzi et al., 2012). The main reason for using meta-analysis is to identify correlation between multiple traits as shown in Figure 1. Because of high correlation, a multivariate regression model cannot be used, therefore many studies have used a meta-analysis approach in this situation (Mensink et al., 2003; Renehan et al., 2008). As phenotypes can often show strong correlation, it can be difficult to simultaneously consider then in a statistical model. In this situation, overall importance can be calculated using meta-analysis with results from the independently performing statistical analysis with each trait. In this paper, four milk-production related traits (ilk yield, fat, protein, and solid, content) were also highly correlated with each other. Therefore, meta-analysis was employed to characterize milk production related genes via four milk-production related traits. As a result of our research, we present 271 milk production related genes and their enriched biological processes, which can be used as valuable resource in future bovine transcriptome analysis.

CONFLICT OF INTEREST

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

ACKNOWLEDGMENTS

This work was carried out with the support of "Cooperative Research Program for Agriculture Science & Technology Development (Project No. PJ01040603, PJ01203101)" Rural Development Administration, Republic of Korea.

REFERENCES

Accorsi, P. A., B. Pacioni, C. Pezzi, M. Forni, D. J. Flint, and E.

Seren. 2002. Role of prolactin, growth hormone and insulin-like growth factor 1 in mammary gland involution in the dairy cow. J. Dairy Sci 85:507-513.

Cánovas, A., G. Rincon, A. Islas-Trejo, S. Wickramasinghe, and J. F. Medrano. 2010. SNP discovery in the bovine milk transcriptome using RNA-Seq technology. Mamm. Genome 21:592-598.

Chauhan, V. and J. Hayes. 1991. Genetic parameters for first lactation milk production and composition traits for Holsteins using multivariate restricted maximum likelihood. J Dairy Sci 74:603-610.

Constantin, A. and C. Csatlos. 2010. Research on the influence of microwave treatment on milk composition. Bulletin of the Transilvania University of Braşov 3:52.

Cui, X. G., Y. L. Hou, S. H. Yang, Y. Xie, S. L. Zhang, Y. Zhang, Q. Zhang, X. M. Lu, G. E. Liu, and D. X. Sun. 2014. Transcriptional profiling of mammary gland in Holstein cows with extremely different milk protein and fat percentage using RNA sequencing. BMC Genomics 15:226.

Glasser, F., A. Ferlay, and Y. Chilliard. 2008. Oilseed lipid supplements and fatty acid composition of cow milk: A meta-analysis. J. Dairy Sci. 91:4687-4703.

Hennighausen, L. and G. W. Robinson. 2001. Signaling pathways in mammary gland development. Dev. Cell 1:467-475.

Hill, P. D., J. C. Aldag, and R. T. Chatterton. 1999. Effects of pumping style on milk production in mothers of non-nursing preterm infants. J. Hum. Lact. 15:209-216.

McDonald, T. L., M. A. Larson, D. R. Mack, and A. Weber. 2001. Elevated extrahepatic expression and secretion of mammary-associated serum amyloid A 3 (M-SAA3) into colostrum. Vet. Immunol. Immunopathol. 83:203-211.

Mensink, R. P., P. L. Zock, A. D. Kester, and M. B. Katan. 2003. Effects of dietary fatty acids and carbohydrates on the ratio of serum total to HDL cholesterol and on serum lipids and apolipoproteins: a meta-analysis of 60 controlled trials. Am. J. Clin. Nutr. 77:1146-1155.

Minozzi, G., J. L. Williams, A. Stella, F. Strozzi, M. Luini, M. L. Settles, J. F. Taylor, R. H. Whitlock, R. Zanella, and H. L. Neibergs. 2012. Meta-analysis of two genome-wide association studies of bovine paratuberculosis. Plos One 7:e32578.

Mizoguchi, Y., T. Hirano, T. Itoh, H. Aso, A. Takasuga, Y. Sugimoto, and T. Watanabe. 2010. Differentially expressed genes during bovine intramuscular adipocyte differentiation profiled by serial analysis of gene expression. Anim. Genet. 41:436-441.

Molenaar, A. J., D. P. Harris, G. H. Rajan, M. L. Pearson, M. R. Callaghan, L. Sommer, V. C. Farr, K. E. Oden, M. C. Miles, and R. S. Petrova et al. 2009. The acute-phase protein serum amyloid A3 is expressed in the bovine mammary gland and plays a role in host defence. Biomarkers 14:26-37.

Neale, B. M., S. E. Medland, S. Ripke, P. Asherson, B. Franke, K.-P. Lesch, S. V. Faraone, T. T. Nguyen, H. Schäfer, and P. Holmans et al. 2010. Meta-analysis of genome-wide association studies of attention-deficit/hyperactivity disorder. J. Am. Acad. Child Adolesc. Psychiatry 49:884-897.

Onetti, S. G. and R. R. Grummer. 2004. Response of lactating cows to three supplemental fat sources as affected by forage in the diet and stage of lactation: A meta-analysis of literature. Anim. Feed Sci. Technol. 115:65-82.

Park, D. S., H. Lee, P. G. Frank, B. Razani, A. V. Nguyen, A. F. Parlow, R. G. Russell, J. Hulit, R. G. Pestell, and M. P. Lisanti. 2002. Caveolin-1-deficient mice show accelerated mammary gland development during pregnancy, premature lactation, and hyperactivation of the Jak-2/STAT5a signaling cascade. Mol. Biol. Cell 13:3416-3430.

Pepe, G., G. C. Tenore, R. Mastrocinque, P. Stusio, and P. Campiglia. 2013. Potential anticarcinogenic peptides from bovine milk. J. Amino Acids Article ID 939804.

Renehan, A. G., M. Tyson, M. Egger, R. F. Heller, and M. Zwahlen. 2008. Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies. Lancet 371:569-578.

Seo, M., J. Yoon, and T. Park. 2015. GRACOMICS: software for graphical comparison of multiple results with omics data. BMC Genomics 16:256.

Sorensen, M., J. V. Nørgaard, P. K. Theil, M. Vestergaard, and K. Sejrsen. 2006. Cell turnover and activity in mammary tissue during lactation and the dry period in dairy cows. J. Dairy Sci. 89:4632-4639.

Sugimoto, K., N. Ichikawa-Tomikawa, S. Satohisa, Y. Akashi, R. Kanai, T. Saito, N. Sawada, and H. Chiba. 2013. The tight-junction protein claudin-6 induces epithelial differentiation from mouse F9 and embryonic stem cells. PloS one 8:e75106.

Takasuga, A., T. Watanabe, Y. Mizoguchi, T. Hirano, N. Ihara, A. Takano, K. Yokouchi, A. Fujikawa, K. Chiba, and N. Kobayashi et al. 2007. Identification of bovine QTL for growth and carcass traits in Japanese Black cattle by replication and identical-by-descent mapping. Mamm. Genome 18:125-136.

Tao, S., J. W. Bubolz, B. C. Do Amaral, I. M. Thompson, M. J. Hayen, S. E. Johnson, and G. E. Dahl. 2011. Effect of heat stress during the dry period on mammary gland development. J. Dairy Sci. 94:5976-5986.

Tezer, M., Y. Ozluk, O. Sanli, O. Asoglu, and A. Kadioglu. 2012. Nitric oxide may mediate nipple erection. J. Androl. 33:805-810.

Thompson, S. G. and J. Higgins. 2002. How should meta-regression analyses be undertaken and interpreted? Stat. Med. 21:1559-1573.

Vargas, B., A. F. Groen, M. Herrero, and J. A. Van Arendonk. 2002. Economic values for production and functional traits in Holstein cattle of Costa Rica. Livest. Prod. Sci. 75:101-116.

Wang, Z., M. Gerstein, and M. Snyder. 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet. 10:57-63.

Wickramasinghe, S., G. Rincon, A. Islas-Trejo, and J. F. Medrano. 2012. Transcriptional profiling of bovine milk using RNA sequencing. BMC Genomics 13:45.

Yarus, S., J. M. Rosen, A. M. Cole, and G. Diamond. 1996. Production of active bovine tracheal antimicrobial peptide in milk of transgenic mice. Proc. Nat. Acad. Sci. 93:14118-14121.

Yoon, J. T., J. H. Lee, C. K. Kim, Y. C. Chung, and C.-H. Kim. 2004. Effects of milk production, season, parity and lactation period on variations of milk urea nitrogen concentration and milk components of Holstein dairy cows. Asian Australas. J. Anim. Sci. 17:479-484.