BioTechnologia vol. 99(2) • pp. 153–163 • 2018
Journal of Biotechnology, Computational Biology and Bionanotechnology RESEARCH PAPERS
http://doi.org/10.5114/bta.2018.75659

Transcriptome signature of the lactation process, identified by meta-analysis of microarray and RNA-Seq data

MOHAMMAD FARHADIAN[1], SEYED ABBAS RAFAT[1]*, KARIM HASANPUR[1], ESMAEIL EBRAHIMIE[2, 3, 4, 5]

[1] Department of Animal Science, Faculty of Agriculture University of Tabriz, Tabriz, Iran

[2] The University of Adelaide, Adelaide Medical, South Australia, Adelaide, Australia

[3] Institute of Biotechnology, Shiraz University, Shiraz, Fars, Iran

[4] School of Information Technology and Mathematical Sciences, Division of Information Technology, Engineering and the Environment, The University of South Australia, Adelaide, South Australia, Australia

[5] School of Biological Sciences, Faculty of Science and Engineering, Flinders University, Adelaide, South Australia, Australia

Abstract

Lactation plays the crucial role in mammals' life. Uncovering the transcriptome signature of lactation process helps to understand the molecular basis of milk production. To identify the genes that express differentially between early and late lactation, publicly available microarray transcriptomic datasets of dairy cattle were investigated and the array results were validated by a next-generation sequencing dataset (RNA-Seq data from sheep). Limma and edgeR packages were used for the analysis of the microarray and RNA-Seq datasets, respectively. Five common differentially expressed genes (DEGs), namely glutathione s-transferase mu 3 (*GSTM3*), EGF containing fibulin-like extracellular matrix protein 1(*EFEMP1*), fibulin 1(*FBLN1*), gelsolin (*GSN*), and fibrinogen-like 2 (*FGL2*), were identified. The involvement of *EFEMP1* in the lactation process has been reported for the first time. The identified DEGs are involved in the development of the immune system and cell differentiation of the mammary gland. A gene ontology network analysis revealed the key role of the *GSN* gene in the regulation of two important functions of actin nucleation and barbed-end actin filament capping. The gene ontology enrichment analysis showed that the function of calcium ion binding is statistically ($P < 0.05$) enriched by the identified transcriptomic signature. The approach presented in this study provides an integrative framework for finding the signature of the lactation process by utilizing global transcriptome data, gene ontology (GO) regulatory network, and enrichment analysis.

Key words: co-expression network, microarray, milk production, RNA-Seq

Introduction

Lactation, the secretion of milk from mammary glands, is a complex function in mammals, crucial for the feeding of newborn animals. To execute the shifts from the early to the peak and from the peak to the late lactation stages, mammary glands undergo dramatic physiological changes. Several genes play vital roles in many of the mammary glands functions, and the gene expression changes during different lactation stages reflect the biology of lactation and the functions of mammary glands in the mammals. The milk yield is dependent on the number and metabolic activity of secretory cells in the milking stage of livestock's lives. Accordingly, the lactation performance may be improved by increasing the mammary cell proliferation or decreasing the apoptosis of secretory cells (Lin et al., 2015). Furthermore, factors involved in the regulation of these processes in the various stages of lactation can directly affect the mammary functions and milk yield. Studying the expression profiles of genes involved in the milk production over different stages of lactation may provide insights into the molecular properties of mammary gland biology, morphogenesis, and metabolic activity. It also improves our understanding of how milk composition undergoes considerable modifications during lactation (Suárez-Vega et al., 2015). During the last two decades, high-throughput

* Corresponding author: Department of Animal Science, Faculty of Agriculture University of Tabriz, Tabriz, Iran; e-mail: Rafata @tabrizu.ac.ir

technologies such as microarray and next-generation sequencing have enabled researchers to investigate more thoroughly than before the transcriptomic biosignatures of a given biological sample. There are several databanks that collect raw data for further analyses, such as the European Bioinformatics Institute (EBI) and the National Center for Biotechnology Information (NCBI). The microarray technology allows the measurements of the mRNA levels of tens of thousands of genes simultaneously in a given sample. The microarray has been widely applied in various fields of molecular genetics and functional genomics to understand the biological mechanisms underlying various processes (Rays et al., 1996), to discover the molecular causes of novel subgroups of diseases (Golub et al., 1999; Alizadeh et al., 2000), and to examine drug responses (Dan et al., 2002). RNA-Seq, a more recent sequence-based method of gene expression analysis, uses next-generation sequencing technologies to obtain sequences from fragments (reads) of cDNA that have adaptors ligated to one or both ends (Wang et al., 2009). Unlike microarrays, RNA-Seq does not require prior knowledge about the sequences of genes, and therefore, is more suitable to whole transcriptome studies even in species without a genome draft. RNA-Seq enables the generation of extensive transcriptome information providing the capability to characterize transcripts, to quantify expression, and to identify differential regulation in a single experiment (Wang et al., 2009).

Information about gene co-expression is useful for understanding gene functions (Obayashi and Kinoshita, 2010; Obayashi et al., 2012). Several databases on gene co-expression analyses based on gene expression data measurements have been developed thus far (Ebrahimie et al., 2014). These public databases provide information about the relative expression levels of many genes simultaneously. The change in the gene expression levels between two genes can lead to the determination of the similarity of expression called gene co-expression at which two or more genes show a high correlation among different whole transcriptome experiments or conditions. It is expected that the co-expressed genes correlate with each other in a specific phenomenon (Mansouri et al., 2014).

High-throughput analyses have been extensively used in animal genetics in the recent years (Stothard et al., 2011; Wiggans et al., 2011). These technologies can be used for selection, i.e., the use of genetic markers to select the best animal/plant for large-scale production or to develop animal/plant resistance to diseases.

In dairy animals, there are two important stages in the lactation curve: early lactation until the peak production (hereinafter called pre-peak), and late lactation (hereinafter called post-peak) where the persistency of lactation is an important feature of more-efficient lactating animals (Vijayakumar et al., 2017). The most distinguished feature of the milk production curve, however, is the peak production at which the lactating animal experiences a negative energy balance because of the high energy exit (with the milk) and a low energy intake (from food). Several high-throughput gene expression studies on milk production using microarray data in different animals have been performed to identify the effective genes in different stages of lactation and non-lactation of different species (Bongiorni et al., 2009; Bionaz et al., 2012; Izumi et al., 2014; Vander Jagt et al., 2015; van de Moosdijk & van Amerongen, 2016). Moreover, in another study, milk production was investigated using RNA-Seq in sheep (Suárez-Vega et al., 2015).

The principal aim of the current study was to identify the differential expression genes (DEGs) between early lactation and late lactation, as the transcriptomic signatures of lactation using high-throughput gene expression data. Furthermore, a variety of computational biology analyses, including gene ontology (GO) regulatory network, hypergeometric test, and co-expression network analysis, were carried out.

Materials and methods

Data collection and sources

To detect the genes that undergo significant changes in expression and contribute to the differences in milk production in various stages of lactation, we re-analyzed the publicly available datasets. The flowchart of an analytical scheme for meta-analysis is presented in Figure 1.

According to the availability of the datasets, we classified the data into two stages of lactation, namely the pre-peak and the post-peak stages. The data of a microarray study, with the accession number of GSE19055 of mammary gland biopsy specimens of dairy cattle, were obtained from NCBI (https://www.ncbi.nlm.nih.gov/). These data were sampled at −30, −15, 1, 15, 30, 60, 120, 240, and 300 days (d) relative to parturition (Bionaz

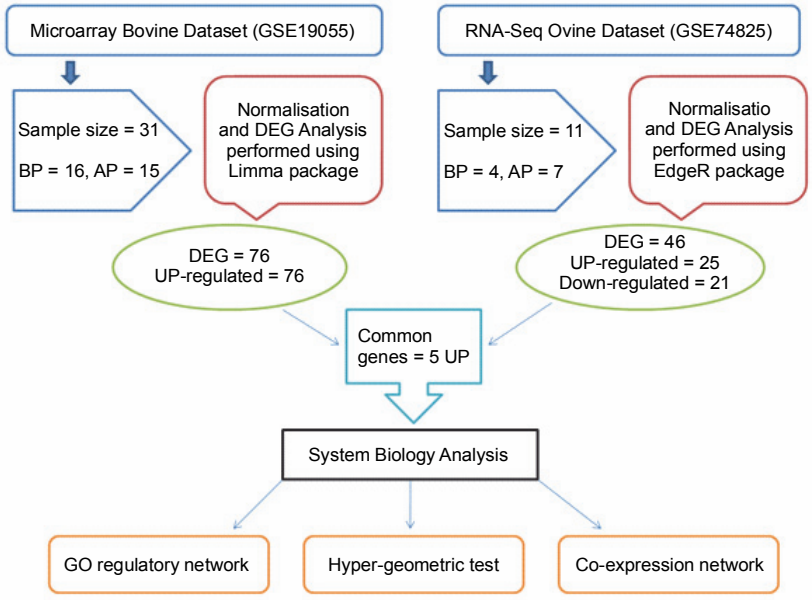

Fig. 1. Flowchart of the meta-analysis of milk production in microarray and RNA-Seq dataset; BP – before peak; AP – after peak

et al., 2012). The objective of this previous study was to investigate the mammary gland transcriptome differences between late pregnancy and different time-points of subsequent lactation. Samples related to 30 days and 15 days before parturition as well as those related to 1 day and 60 days after parturition were excluded from the analyses. The Limma package was used for the normalization and differential expression analysis of this dataset (Ritchie et al., 2015). Subtraction, loess, and quantile methods were used for the background correction and within and between array normalization, respectively (Bionaz et al., 2012).

A sheep milk RNA-Seq dataset with the accession number of GSE74825 was re-analyzed to validate the results of the microarray dataset. The samples of this dataset were collected on 10, 50, 120, and 150 days after parturition (Suárez-Vega et al., 2015). Samples related to day 10 were considered pre-peak, while samples related to days 120 and 150 after parturition were considered post-peak. The EdgeR package was used for the analysis of this dataset (Law et al., 2016; McCarthy et al., 2012; Robinson et al., 2010). Finally, the common DEGs were selected as the ultimate DEGs.

GO regulatory network construction

To predict a novel type of regulatory networks based on GO interaction, a comparative GO was applied on common statistically significant regulated genes (Fruzangohar et al., 2015; Fruzangohar et al., 2017). Instead of gene relationships, the GO terms of biological processes were investigated for the construction of this type of network. A regulatory relationship was established between the biological process (BP) terms based on interactions extracted from the GO database (Botstein et al., 2000; Consortium, 2004). For the given gene sample of the common regulated genes, the comparative GO built a GO network based on the regulatory relationship.

In this work, the GO enrichment of up-regulated genes was compared with the genome of *Bos taurus* by using a hypergeometric distribution with a comparative GO web application (Fruzangohar et al., 2013). The hypergeometric test based on Fisher's exact test was used to find the differential distribution of GO enrichment between our samples and the *B. taurus* genome.

Co-expression analysis and co-expression-based network prediction analysis for DEGs

To perform the co-expression analysis, COXPRESdb (http://coxpresdb.jp/) was used (Okamura et al., 2014). We analyzed the co-expressed genes with the *GSTM3*, *EFEMP1*, *FBLN1*, *GSN*, and *FGL2* genes to find the potentially co-expressed genes. Based on the mutual ranking (MR) index calculated as a measure of the co-expression/relation, a co-expression-based network of the

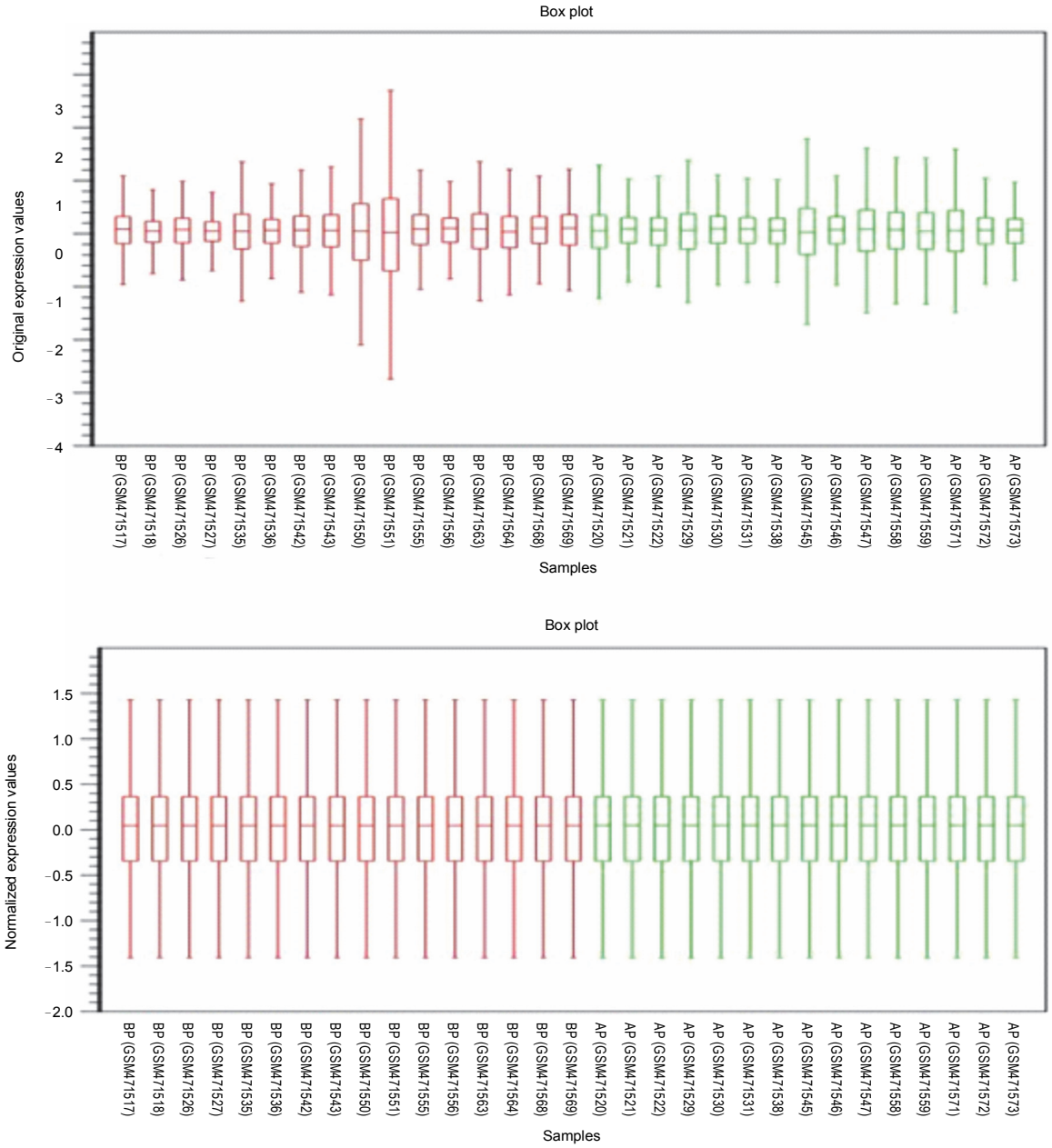

Fig. 2. Quality control of microarray transcriptomic data; box plot of microarray data: above – before normalization and low – after normalization

GSTM3, *EFEMP1*, *FBLN1*, *GSN*, and *FGL2* genes was constructed.

Results and discussion

Data pre-processing and differential expression analysis

Raw microarray data are usually biased because of some known effects such as mRNA preparation, cDNA synthesis, and nonspecific hybridization. Therefore, to ensure the acquisition of accurate results, the pre-processing and normalization of the microarray data is necessary prior to the main analysis. Pre-processing removes the non-biological variation of the data and helps to improve a subsequent analysis by providing a suitable scale (Bolstad et al., 2003). Background correction is one of the pre-processing steps. Some of the non-biological biases are systematic and can be normalized both within and between the array levels. RNA-Seq data can also be biased by all of the steps of library preparation, PCR, and sequencing. To eliminate the effect of the

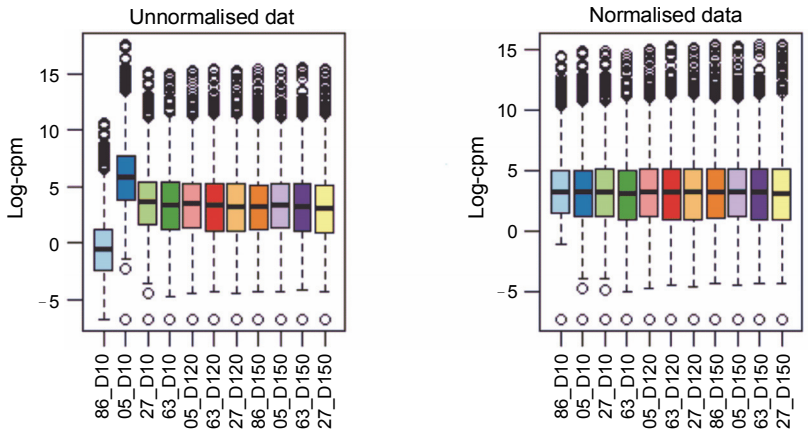

Fig. 3. Quality control of RNA-Seq transcriptomic data; box plot for un-normalized and normalized data

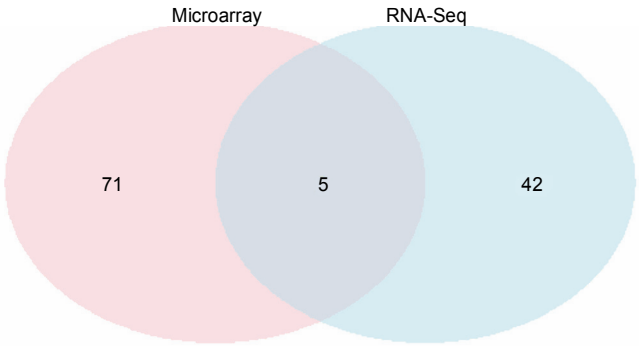

Fig. 4. Venn diagram of microarray and RNA-Seq dataset

known non-biological factors on the RNA-Seq data, the trimmed mean of m-values (TMM) method was used to correct the read-count differences that exist among the samples (Robinson and Oshlack, 2010). Box-plots of normalized *vs.* raw data are shown in Figure 2 and Figure 3 for the microarray and the RNA-Seq data, respectively.

In Figure 2, the samples indicated as "before" and "after" the peak are shown with red and green colors, respectively (BP – before peak; AP – after peak).

The results of the microarray data analysis revealed a total of 76 DEGs between the pre-peak and post-peak stages of lactation using FDR corrected *P* values ($P < 0.05$). All of these genes were up-regulated at pre-peak as compared to post-peak.

As mentioned above, the RNA-Seq data were analyzed to validate the results of the microarray data analysis. Because of the unavailability of the RNA-Seq data with both pre-peak and post-peak samples in dairy cattle, we utilized the Assaf sheep breed's RNA-Seq data, instead. In total, there were 46 DEGs between the pre-peak and the post-peak stages of lactation. Out of which, 25 were up- and 21 were down-regulated at pre-peak. The Venn diagram of common genes between the microarray and the RNA-Seq dataset is shown in Figure 4.

When we compared the results of the two analyses, we discovered five DEGs in common. The gene names and their description are given in Table 1. The *GSTM3*, *EFEMP1*, *FBLN1*, *GSN*, and *FGL2* genes were up-regulated in the pre-peak as compared to the post-peak stage of lactation. *GSN* is known as an effective gene of the lipid metabolism pathway (Mach et al., 2011). The *GSN* gene has been reported as a gene related to milk performance and mammary morphology (Zheng et al., 2017). In a previous study (Stute et al., 2012), the expression of *GSTM3* was assayed between the lactating and the non-lactating dairy cows. The results indicated that *GSTM3* was significantly reduced with advances in age. The results of the analyses of protein levels showed that *GSTM3* exhibited the highest fold change in the peak lactation against the late lactation phases (Zheng et al., 2017).

The RNase A family 5 (*RNASE5*) pathway, responsible for a large genetic variation in milk production among dairy cattle, consists of 11 genes, of which *FBLN1* is one (Raven et al., 2013). Moreover, *FBLN1* has been reported to play an important role in the development and cell differentiation of mammary glands (Menzies et al., 2009). *FGL2* is a part of the immune system, and its expression in cows with multiple milking times is significantly higher than in cows with milked once or twice a day (Connor et al., 2008). Therefore, the results of the current study are in partial accordance with the previous

Table 1. Common significant genes between microarray and RNA-Seq data in lactation process

Gene symbol	Gene description	Up/Down	Fold change in microarray	Fold change in RNA-Seq
GSTM3	Glutathione S-transferase mu 3 (brain)	Up	1.95	1.5
EFEMP1	EGF containing fibulin-like extracellular matrix protein 1	Up	1.55	1.82
FBLN1	Fibulin 1	Up	1.93	2.54
GSN	Gelsolin	Up	1.74	1.96
FGL2	Fibrinogen-like 2	Up	1.89	1.76

findings (Raven et al., 2013; Connor et al., 2008; Menzies et al., 2009; Zheng et al., 2017; Stute et al., 2012). According to the reports mentioned above, all of the DEGs identified in the current study, other than *EFEMP1*, are in direct or indirect relation to the milk production. By improving the functioning of the immune system, and stimulating cell differentiation during lactation, the identified DEGs significantly affect milk production. *EFEMP1* has been for the first time reported herein as a candidate gene that may contribute to the differences in milk production at different time-points of lactation.

This gene (*EFEMP1*) encodes a member of the fibulin family of extracellular-matrix glycoproteins. The encoded protein contains tandemly repeated epidermal growth factor-like repeats followed by a C-terminus fibulin-type domain. Mutations in this gene are associated with Doyne honeycomb retinal dystrophy (Timpl et al., 2003).

For the selection of studies, there were some restrictions including having samples for the entire lactation (BF and AF) and finding the microarray and RNA-Seq data from the same species.

GO regulatory network for common DEGs in both microarray and RNA-Seq

The GO regulatory network for DEGs is shown in Figure 5. Using a regulatory network, some researchers interpreted the relationship between the GO terms and their corresponding genes (Fruzangohar et al., 2013). The predicted GO network (Figure 5) showed that two functional groups, namely "*actin nucleation*" and "*barbed-end actin filament capping*," were located in the center of the network and had the highest number of interactions with the other GO terms. Interestingly, the "*actin nucleation*" and "*barbed-end actin filament capping*" functional groups have the *GSN* gene in common.

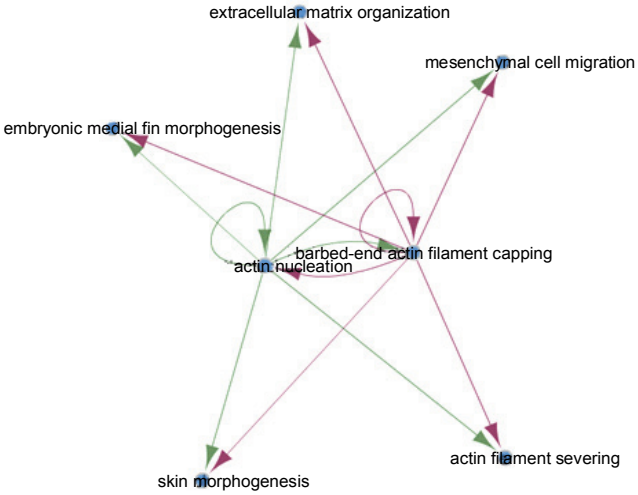

Fig. 5. Gene ontology (GO) regulatory network of DEGs between pre-peak and post-peak stages of lactation. The circles indicate their splicing variants in each gene ontology group. Green arrows indicate up-regulation and red arrows indicate down-regulation of gene expression

According to a study, *actin nucleation* and *barbed-end actin filament capping* have been indicated to be involved in the immune system (Obino et al., 2016). Therefore, we concluded that these two central GO functions, in the GO regulatory network, influence milk production through the activation of the immune system. GO harboring *GSN* has interactions with other GO terms including *extracellular matrix organization*, *actin filament severing*, *embryonic medial fin morphogenesis*, *skin morphogenesis*, and *mesenchymal cell migration*. The other GO terms, as indicated in Figure 5, have the *FBLN1* gene in common. Stromal – epithelial interactions play important roles in the development of the mammary ductal tree and are important in the embryonic and postnatal evolvement. The stroma consists of mesenchymal cells (fibroblasts, blood cells, and leuko-

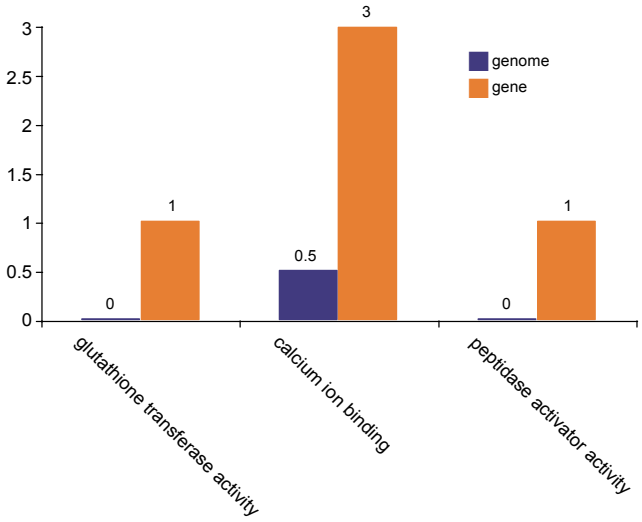

Fig. 6. Gene ontology (GO) distribution of DEGs between pre-peak and post-peak stages of lactation *vs.* GO distribution of genome during biological process. The protein levels produced by DEGs and genome relative to their genes are presented in blue and yellow boxes, respectively

cytes) and extracellular matrix (ECM) (laminin, fibronectin, collagen, proteoglycans, etc.), which influence the mammary gland development. Mammary gland development is one of the most important factors in lactation (Kass et al., 2007).

Comparison of GO distribution of common DEGs within the Bos taurus genome

A comparison of the GO enrichment of the common up-regulated genes versus the GO distribution of the *Bos taurus* genome performed using Fisher's exact test (hypergeometry) at the "biological process" level is presented in Figure 6. Among the three levels of the GO terms, only the biological effect indicated a significant term. Interestingly, during a biological process, the functions significantly up-regulated by DEGs were related to the *calcium ion binding* GO term (Fig. 6).

Based on the data presented in Figure 6, the calcium ion binding biological process is significantly higher than the genome contribution ($P < 0.05$). Calcium is a major chemical element present in milk. In the aqueous phase of milk, calcium concentrates in the forms of ionic calcium, calcium phosphate, and calcium citrate (Tanaka et al., 2011). The concentration of calcium ions in mammary glands is thought to be associated with the integrity of the mammary gland during lactation (Neville and Peaker, 1981). It can therefore be concluded that the calcium ion concentration in mammary glands plays an important role in regulating the physiological process of milk production.

Genes co-expressed with DEGs

Genes that co-express with the DEGs were identified using the COXPRESdb navigator (Obayashi et al., 2012). The 100 genes highly co-expressed with each DEG and their MR are presented in Tables S1–S5. Interestingly, during lactation in the different experiments, all of the identified DEGs showed high levels of co-expression (Tables S1–S5). The MR index is a co-expression measurement performed by taking the geometric average of Pearson's correlation coefficient rank from gene A to gene B and that of gene B to gene A. The MR values confirmed a highly significant co-expression and co-occurrence of DEGs during lactation. The MR index was used instead of Pearson's correlation coefficient because the geometric average (MR index) is more accurate (Obayashi et al., 2012). A smaller value of the MR index indicates a higher co-expression (Ebrahimie et al., 2014).

The *C1R* (complement component 1, r subcomponent) gene is among the genes showing the highest co-expression with the *EFEMP* gene (Tables S1 and S7) and is involved in the mammary gland function (Khalil et al., 2011). Therefore, it is interesting that with the co-expression analysis of genes involved in developing a specific trait, the other genes related to this trait can be detected. Another gene, decorin (*DCN*) is a mammary-gland extracellular-matrix gene (Suárez-Vega et al., 2015), which also co-expresses with the *FBLN* gene (Table S2 and Table S7).

The toll-like receptor 8 (*TLR8*) gene shows a high co-expression with the *FGL2* gene (Table S3 and Table S7). It enhances the mammary gland's defense system during mastitis induced by *Escherichia coli* infection in goats (Zhu et al., 2007). A study performed on Canadian Holstein cattle on the enrichment of lactation persistency resulted in the identification of the cytochrome b5 reductase 3 (*CYB5R3*) gene (Do et al., 2017). Interestingly, in our research, the CYB5R3 gene showed a high co-expression with the *GSN* gene (Table S4 and Table S7). *GTSM3* also showed co-expression with other genes belonging to its family (*GSTM2*, *GSTM1*, and *GSTM4*) (Table S5 and Table S7).

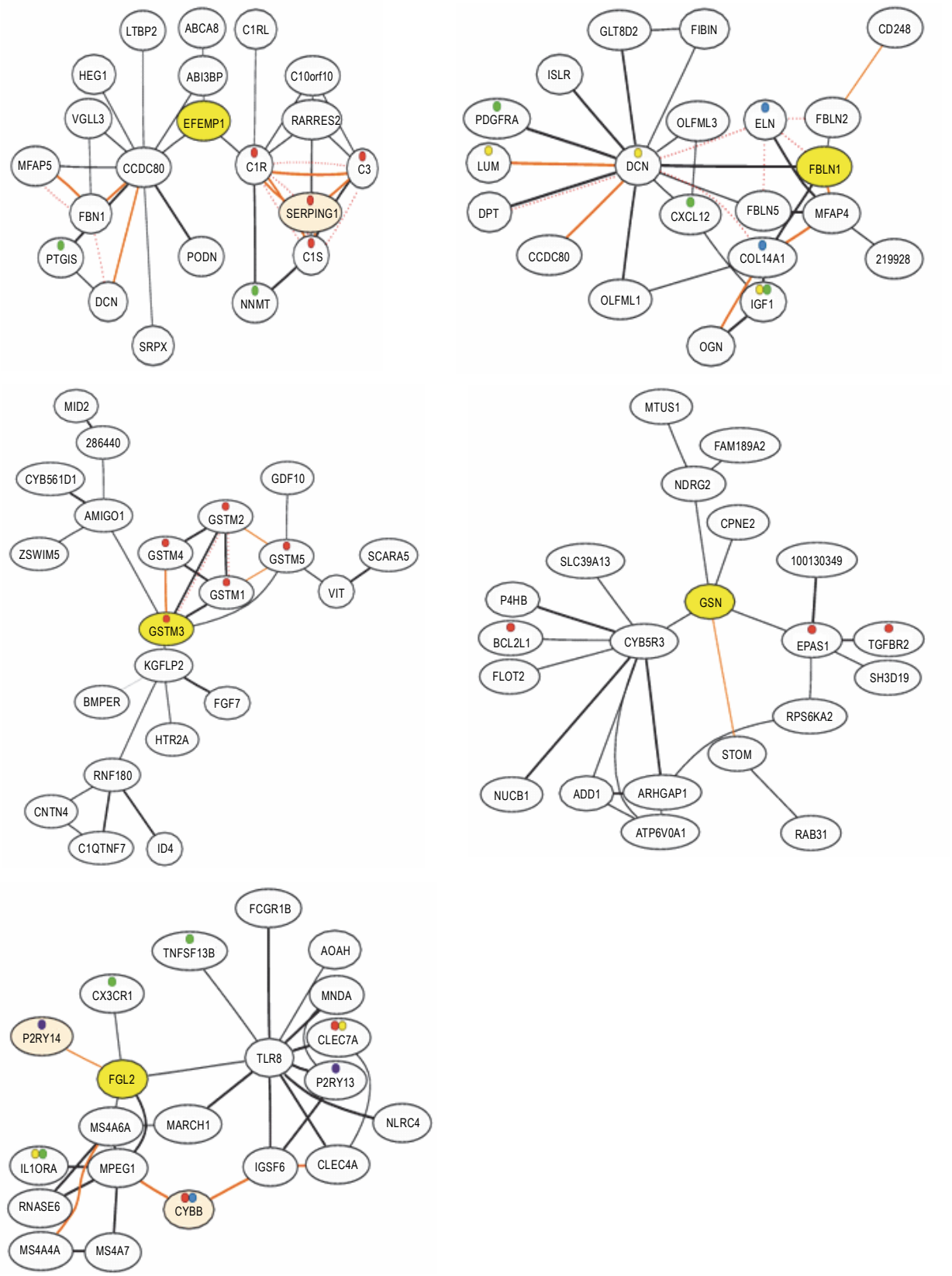

Fig. 7. Co-expression-based network of DEGs involved in pre- and post-peak lactation including *EFEMP1*, *FBLN1*, *GSTM3*, *GSN*, and *FGL2*; the yellow node is the guide gene; the width of the edges corresponds to the strength of the co-expression; marks on the nodes indicate the common KEGG annotations; the orange edge indicates the conserved co-expression, which is also observed in other species

Co-expression-based network of DEGs

On the basis of the co-expression analyses of the genes, which were selected from different experiments using MR, the expression networks were constructed and visualized using COXPRESdb (Fig. 7).

In addition, functional annotations of these DEGs were extracted on the basis of the biological process classification of Gene Ontology (Table S6). As presented in Table S6, a comparison of the functional annotation of four DEGs suggested that the *EFEMP1* and *FBN1* genes were involved in the organization of the ECM. The ECM is an intricate network composed of an array of macromolecules, whose importance is becoming increasingly apparent. The ECM is an integral part of the machinery that regulates cell function; its role in cell differentiation and tissue-specific gene expression has been described (Lin and Bissell, 1993).

Conclusions

The microarray and RNA-Seq datasets are the most frequent and important gene expression data, publicly available for finding reliable candidate genes in a trait of interest. In this research, using both microarray and RNA-Seq datasets related to milk lactation in cow and sheep, we identified five DEGs between the two studied stages of lactation, i.e., pre-peak versus post-peak. The identified five DEGs may contribute to the entire milk production of lactating dairy animals by influencing their productive potential at different stages of lactation. Therefore, although there are many factors that influence the milk production potential of dairy animals, the consideration of these candidate genes in the selection and breeding procedures is recommended. Further investigation, however, is required to elucidate the effect of these DEGs on milk production and to validate the results of the current work.

Conflict of interest

The authors declare that they have no conflict of interest to report with respect to this paper.

Acknowledgment

We would like to thank Department of Animal Science, Tabriz University, for support in writing this manuscript.

References

Alizadeh A.A., Eisen M.B., Davis R.E., Ma C., Lossos I.S., Rosenwald A., Boldrick J.C., Sabet H., Tran T., Yu X. (2000) *Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling*. Nature 403: 503–511.

Bionaz M., Periasamy K., Rodriguez-Zas S.L., Everts R.E., Lewin H.A., Hurley W.L., Loor J.J. (2012) *Old and new stories: revelations from functional analysis of the bovine mammary transcriptome during the lactation cycle*. PloS one 7: e33268.

Bolstad B.M., Irizarry R.A., Åstrand M., Speed T.P. (2003) *A comparison of normalization methods for high density oligonucleotide array data based on variance and bias*. Bioinformatics 19: 185–193.

Bongiorni S., Chillemi G., Prosperini G., Bueno S., Valentini A., Pariset L. (2009) *A tool for sheep product quality: custom microarrays from public databases*. Nutrients 1: 235–250.

Botstein D., Cherry J.M., Ashburner M., Ball C., Blake J., Butler H., Davis A., Dolinski K., Dwight S., Eppig J. (2000) *Gene Ontology: tool for the unification of biology*. Nat. Genet. 25: 25–29.

Connor E.E., Siferd S., Elsasser T.H., Evock-Clover C.M., Van Tassell C.P., Sonstegard T.S., Fernandes V.M., Capuco A.V. (2008) *Effects of increased milking frequency on gene expression in the bovine mammary gland*. BMC Genom. 9: 362.

Consortium GO (2004) *The Gene Ontology (GO) database and informatics resource*. Nucl. Acids Res. 32: D258–D261.

Dan S., Tsunoda T., Kitahara O., Yanagawa R., Zembutsu H., Katagiri T., Yamazaki K., Nakamura Y., Yamori T. (2002) *An integrated database of chemosensitivity to 55 anticancer drugs and gene expression profiles of 39 human cancer cell lines*. Cancer Res. 62: 1139–1147.

Do D., Bissonnette N., Lacasse P., Miglior F., Sargolzaei M., Zhao X., Ibeagha-Awemu E. (2017) *Genome-wide association analysis and pathways enrichment for lactation persistency in Canadian Holstein cattle*. J. Dairy Sci. 100: 1955–1970.

Ebrahimie M., Esmaeili F., Cheraghi S., Houshmand F., Shabani L., Ebrahimie E. (2014) *Efficient and simple production of insulin-producing cells from embryonal carcinoma stem cells using mouse neonate pancreas extract, as a natural inducer*. PloS One 9: e90885.

Fruzangohar M., Ebrahimie E., Adelson D.L. (2017) *A novel hypothesis-unbiased method for Gene Ontology enrichment based on transcriptome data*. PloS One 12: e0170486.

Fruzangohar M., Ebrahimie E., Ogunniyi A.D., Mahdi L.K., Paton J.C., Adelson D.L. (2013) *Comparative GO: a web application for comparative gene ontology and gene ontology-based gene selection in bacteria*. PloS One 8: e58759.

Fruzangohar M., Ebrahimie E., Ogunniyi A.D., Mahdi L.K., Paton J.C., Adelson D.L. (2015) *Correction: Comparative GO: a web application for comparative gene ontology and gene ontology-based gene selection in bacteria*. PloS One 10.

Golub T.R., Slonim D.K., Tamayo P., Huard C., Gaasenbeek M., Mesirov J.P., Coller H., Loh M.L., Downing J.R., Caligiuri M.A. (1999) *Molecular classification of cancer: class discovery and class prediction by gene expression monitoring*. Science 286: 531–537.

Izumi H., Kosaka N., Shimizu T., Sekine K., Ochiya T., Takase M. (2014) *Time-dependent expression profiles of microRNAs and mRNAs in rat milk whey*. PloS One 9: e88843.

Kass L., Erler J.T., Dembo M., Weaver V.M. (2007) *Mammary epithelial cell: influence of extracellular matrix composition and organization during development and tumorigenesis*. Int. J. Biochem. Cell Biol. 39: 1987–1994.

Khalil E., Digby M.R., Thomson P.C., Lefèvre C., Mailer S.L., Pooley C., Nicholas K.R. (2011) *Acute involution in the tammar wallaby: identification of genes and putative novel milk proteins implicated in mammary gland function*. Genomics 97: 372–378.

Law C.W., Alhamdoosh M., Su S., Smyth G.K., Ritchie M.E. (2016) *RNA-seq analysis is easy as 1-2-3 with limma, Glimma and edgeR*. F1000Res. 5: 1408.

Lin C.Q., Bissell M.J. (1993) *Multi-faceted regulation of cell differentiation by extracellular matrix*. FASEB J. 7: 737–743.

Lin J., Bao Z.K., Zhang Q., Hu W.W., Yu Q.H., Yang Q. (2015) *Transcriptome analysis of the mammary gland from GH transgenic goats during involution*. Gene 565: 228–234.

Mach N., Jacobs A., Kruijt L., Van Baal J., Smits M. (2011) *Alteration of gene expression in mammary gland tissue of dairy cows in response to dietary unsaturated fatty acids*. Animal 5: 1217–1230.

Mansouri A., Esmaeili F., Nejatpour A., Houshmand F., Shabani L., Ebrahimie E. (2014) *Differentiation of P19 embryonal carcinoma stem cells into insulin-producing cells promoted by pancreas-conditioned medium*. J. Tissue Eng. Regener. Med. 13.

McCarthy D.J., Chen Y., Smyth G.K. (2012) *Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation*. Nucl. Acids Res. 40: 4288–4297.

Menzies K.K., Lefèvre C., Macmillan K.L., Nicholas K.R. (2009) *Insulin regulates milk protein synthesis at multiple levels in the bovine mammary gland*. Funct. Integr. Genomics 9: 197–217.

Neville M., Peaker M. (1981) *Ionized calcium in milk and the integrity of the mammary epithelium in the goat*. J. Physiol. 313: 561–570.

Obayashi T., Kinoshita K. (2010) *COXPRESdb: a database to compare gene coexpression in seven model animals*. Nucl. Acids Res. 39: D1016–D1022.

Obayashi T., Okamura Y., Ito S., Tadaka S., Motoike I.N., Kinoshita K. (2012) *COXPRESdb: a database of comparative gene coexpression networks of eleven species for mammals*. Nucl. Acids Res. 41: D1014–D1020.

Obino D., Farina F., Malbec O., Sáez P.J., Maurin M., Gaillard J., Dingli F., Loew D., Gautreau A., Yuseff M.-I. (2016) *Actin nucleation at the centrosome controls lymphocyte polarity*. Nature Commun. 7.

Okamura Y., Aoki Y., Obayashi T., Tadaka S., Ito S., Narise T., Kinoshita K. (2014) *COXPRESdb in 2015: coexpression database for animal species by DNA-microarray and RNAseq-based expression data with multiple quality assessment systems*. Nucl. Acids Res. 43: D82–D86.

Raven L.-A., Cocks B.G., Pryce J.E., Cottrell J.J., Hayes B.J. (2013) *Genes of the RNASE5 pathway contain SNP associated with milk production traits in dairy cattle*. Genetics Select. Evol. 45: 25.

Rays M., Chen Y., Su Y.A. (1996) *Use of a cDNA microarray to analyse gene expression patterns in human cancer*. Nat. Genet. 14.

Ritchie M.E., Phipson B., Wu D., Hu Y., Law C.W., Shi W., Smyth G.K. (2015) *Limma powers differential expression analyses for RNA-sequencing and microarray studies*. Nucl. Acids Res. 43: e47.

Robinson M.D., McCarthy D.J., Smyth G.K. (2010) *EdgeR: a bioconductor package for differential expression analysis of digital gene expression data*. Bioinformatics 26: 139–140.

Robinson M.D., Oshlack A. (2010) *A scaling normalization method for differential expression analysis of RNA-seq data*. Genome Biol. 11: R25.

Stothard P., Choi J.-W., Basu U., Sumner-Thomson J.M., Meng Y., Liao X., Moore S.S. (2011) *Whole genome resequencing of black Angus and Holstein cattle for SNP and CNV discovery*. BMC Genom. 12: 559.

Stute P., Sielker S., Wood C.E., Register T.C., Lees C.J., Dewi F.N., Williams J.K., Wagner J.D., Stefenelli U., Cline J.M. (2012) *Life stage differences in mammary gland gene expression profile in non-human primates*. Breast Cancer Res. Treat. 133: 617–634.

Suárez-Vega A., Gutiérrez-Gil B., Klopp C., Robert-Granie C., Tosser-Klopp G., Arranz J.J. (2015) *Characterization and comparative analysis of the milk transcriptome in two dairy sheep breeds using RNA sequencing*. Sci. Rep. 5: 18399.

Tanaka M., Suzuki T., Kotb S., Kamiya Y. (2011) *Relation of calcium activity in milk and milk production of Holstein cows in hot season*. Asian-Austral. J. Anim. Sci. 24: 1372–1376.

Timpl R., Sasaki T., Kostka G., Chu M.-L. (2003) *Fibulins: a versatile family of extracellular matrix proteins*. Nature Rev. Mol. Cell Biol. 4: 479–489.

van de Moosdijk A.A.A., van Amerongen R. (2016) *Identification of reliable reference genes for qRT-PCR studies of the developing mouse mammary gland.* Sci. Rep. 6.

Vander Jagt C., Whitley J., Cocks B., Goddard M. (2015). *Gene expression in the mammary gland of the tammar wallaby during the lactation cycle reveals conserved mechanisms regulating mammalian lactation*. Send to
Reprod. Fertil. Dev. doi: 10.1071/RD14210.

Vijayakumar M., Park J.H., Ki K.S., Lim D.H., Kim S.B., Park S.M., Jeong H.Y., Park B.Y., Kim T.I. (2017) *The effect of lactation number, stage, length, and milking frequency on milk yield in Korean Holstein dairy cows using automatic milking system*. Asian-Australas. J. Anim. Sci. 30: 1093.

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

Wiggans G., VanRaden P., Cooper T. (2011) *The genomic evaluation system in the United States: Past, present, future*. J. Dairy Sci. 94: 3202–3211.

Zheng X., Ning C., Dong Y., Zhao P., Li J., Fan Z., Li J., Yu Y., Mrode R., Liu J.-F. (2017) *Quantitative proteome analysis of bovine mammary gland reveals protein dynamic changes involved in peak and late lactation stages.* Biochem. Biophys. Res. Commun. 494: 292–297.

Zhu Y.-M., Miao J.-F., Zhang Y.-S., Li Z., Zou S.-X., Deng Y.-E. (2007) *CpG-ODN enhances mammary gland defense during mastitis induced by Escherichia coli infection in goats*. Vet. Immunol. Immunopathol. 120: 168–176.