REVIEW ARTICLE

Molecular insights into gene expression in dairy cattle: Techniques and trends

Marko Oroz[1], Mihaela Oroz[1], Mario Shihabi[2], Ino Čurik[2], Marija Spehar[3], Nikola Raguž[1], Boris Lukić[1]

[1]Faculty of Agrobiotechnical Sciences Osijek, Josip Juraj StrossmayerUniversity of Osijek, Vladimira Preloga 1, Osijek, Croatia (moroz@fazos.hr)
[2]University of Zagreb, Faculty of Agriculture, Svetošimunska cesta 25, Zagreb, Croatia
[3]Centre for Livestock Breeding,Svetošimunska cesta 25,Zagreb, Croatia

Summary
Gene expression is a complex biological process influenced by both genetic and environmental factors, involving a series of biochemical reactions. Production traits are significantly affected by gene expression, and understanding these mechanisms allows for production optimization, reduced disease occurrence, improved reproduction, and overall enhancement of production quality. Whole-genome sequencing enables the identification of individual gene functions within the genome and their role in the production performance of an individual. By regulating gene expression, cells can precisely control the production of functional proteins, ensuring optimal physiological conditions. The main methods used to study gene expression in livestock include microarray hybridization, RNA sequencing (RNA-Seq), and real-time PCR (qPCR). Microarrays allow many genes to be analyzed at once using special cDNA chips. RNA-Seq measures the activity of all genes in a sample and gives a detailed picture of gene expression. Because of its accuracy and wide range of uses, RNA-Seq is now considered the most powerful tool for gene expression analysis.

Key words: gene expression, RNA-Seq, cattle, milk production

Introduction

In intensive dairy cattle systems, the primary objective is to achieve adequate milk fat and protein quality along with a proportionally high daily milk yield. Milk production is a technologically complex process involving the interaction between cattle, the production environment, and humans, with its outcomes being influenced by numerous genetic and environmental factors that determine the expression of an individual's genetic potential. Gene expression is the process by which genetic information is translated into functional molecules, such as RNA and proteins, enabling cells to perform specific functions, and it is influenced by both genetic and non-genetic factors (Cassar-Malek et al., 2008). This involves the transcription of DNA into RNA and the subsequent translation of RNA into proteins. Gene expression regulation is crucial for cell differentiation and organism development, and plays a significant role in regulating physiological and production traits, including the nutritional composition of milk and lactation parameters (Gao et al., 2013). Although gene expression data is not routinely incorporated into genomic evaluations, it is highly valuable in scientific research for identifying and classifying genomic sequences associated with production and reproductive traits, thereby providing more accurate information on specific genes and traits to enhance breeding programmes (Fang et al., 2025).

The connection between gene expression and milk traits of cows

Towards the end of the last century and the beginning of this one, molecular biology emerged as a key research area in livestock breeding (Cassar-Malek et al., 2008). Moving beyond mere identification of DNA sequences has enabled a more precise definition of gene functions at the whole-genome level. Organizing sequences into functional groups is based on their expression, which lays the foundation for further experiments aimed at the detailed characterization of the function of the final gene product (Moody, 2001). Gene expression is determined by both genetic and environmental factors. The efficiency of gene expression research is achieved by analyzing phenotypic variability within and between breeds, as well as by comparing animals with significant deviations in traits such as growth dynamics, body condition, and production quality parameters. Genetic research has significantly contributed to the understanding of numerous physiological processes and organ functions (Cassar-Malek et al., 2008). A large number of gene expression studies have been conducted in the field of dairy cattle production, leading to the key insights. According to Gao et al. (2013), the mammary gland is the only organ that, after the maturation process, undergoes cycles of proliferation and involution, characterized by phases of increased cell growth followed by eventual cell regression. Epithelial cells of the mammary gland are characterized by their ability to produce and secrete milk (Canovas et al., 2014). Understanding the molecular processes in these cells can significantly improve management and breeding technologies for dairy cattle.

Breeding programs have significantly influenced the metabolic balance of dairy cows, with the aim of achieving high milk production driving nutritional demands far beyond the animals' natural intake capacity (McCabe et al., 2012). This process has resulted in a disturbance of the physiological status known as negative energy balance (NEB), a physiological condition with distinctly adverse effect on animal health and fertility, further highlighting its economic importance in milk production (McCabe et al., 2012). Like most mammals, dairy cows enter a period of negative energy balance shortly after calving, as their nutrient requirements during lactation exceed their intake from food (Loor et al., 2006). After calving, the transition from colostrum to milk production involves major changes in gene expression in the mammary gland. These changes play an important role in starting and supporting lactation. During this period, called secretory activation, genes responsible for lactose production become more active. Notably, the increased expression of LALBA and B4GALT1 genes is positively correlated with higher milk and lactose production (Shangraw and McFadden, 2024). Increased expression of several genes is commonly observed during the periparturient period, although this pattern may not always be consistent.The transcriptomic profile of mammary secretory cells provides direct insights into key genes that may act as limiting factors for high milk production (Shangraw and McFadden, 2024).

Thus, in the studies by Bionaz and Loor (2011), increased expression of the LALBA, CSN3, FASN, and LPIN1 genes was observed in the mammary gland 15 days after calving, compared to the prepartum period. It is known that increased milking frequency during the postpartum period positively affects milk yield, particulary during the secretory activation phase (Wall and McFadden, 2007). By measuring gene expression during the period of intensive milking, key genes responsible for lactose biosynthesis and increased milk production could be identified (Shangraw and McFadden, 2024). The discovery of transcriptomic profiles of mammary gland epithelial cells (MEC) during lactation is crucial for identifying candidate genes associated with milk production traits (Yang et al., 2015). In addition to MECs, mammary gland tissue also includes myoepithelial and mesenchymal cells. As a result, RNA isolated from mammary epithelial cells (MECs) may not accurately reflect the complete gene expression profile of these cells. Given that milk fat globules (MFGs) are derived exclusively from MECs, RNA obtained from MFGs offers a more reliable representation of gene expression in the mammary epithelium (Maningat et al., 2009). Research on gene expression during lactation using MFGs represents

a simpler alternative to MEC analysis. MFGs can be easily collected at any time during lactation, whereas mammary gland tissue biopsy is invasive, costly, and disrupts the natural course of lactation (Yang et al., 2015). A significant number of differentially expressed genes have been identified using RNA-seq sequencing. In cows with higher milk, fat, and protein yields at peak production, a large proportion of these genes exhibited lower expression. Specifically, 91.01% of the differentially expressed genes in a specific interval showed lower expression during peak production (Yang et al., 2015). The findings of Cui et al. (2014), studying the differential gene expression in the mammary gland tissue of cows with extremely high and low fat and protein content, align with the findings of Yang et al. (2015), further supporting the hypothesis of a connection between gene expression dynamics and milk containing higher fat and protein levels. Using RNA-seq sequencing technology, Yang et al. (2015) investigated differential gene expression in MFGs of lactating cows with significantly different production outcomes.

Experimental methods for gene expression analysis

The regulation of gene expression enables cells to adapt to different conditions and ensures that the synthesized functional product, usually a protein, is produced at the right time and in appropriate amounts. Numerous methodologies are available for gene expression analysis, with microarray hybridization, RNA sequencing (RNA-Seq), and real-time quantitative PCR (qPCR) being the most commonly used as highly reliable (Sun et al, 2012).

The microarray hybridization method allows the simultaneous analysis of gene expression for a large number of genes and is divided into three main types: 1) oligonucleotide chips, which are created by synthesizing oligonucleotides directly onto a glass slide; 2) oligonucleotide chips, which are made by placing previously synthesized oligonucleotides onto glass slides or nylon membranes; and 3) cDNA chips, which are created by inserting PCR-amplified inserts from cDNA library copies onto glass slides or nylon membranes (Moody, 2001). Hybridization-based approaches usually involve the incubation of fluorescently labeled cDNA with microarrays which can be either custom-made or commercially available high-density oligonucleotide microarrays (Clark et al., 2002.).

Moody (2001) conducted an experiment using two mRNA samples to demonstrate the fundamental principles of the microarray hybridization technique. In this procedure, messenger RNA (mRNA) from both samples is reverse transcribed into complementary DNA (cDNA) and labeled with fluorescent markers.The newly formed cDNA is hybridized onto a microarray containing sequences from thousands of genes. The hybridization signal, once detected is proportional to the amount of mRNA transcript in the original sample. By comparing the hybridization signals of individual genes across different samples, differences in gene expression can be identified (Moody, 2001). The analysis of microarray data typically involves three main steps: 1) the identification and quantification of hybridization intensity, 2) data visualization for pattern recognition, and 3) the application of clustering techniques to classify gene expression profiles. However, hybridization-based methods face several limitations, such as the requirement for prior knowledge of the target genome sequences, high background noise which arises from cross-hybridization, and a dynamic range which is restricted to some extent for detecting gene expression across various transcript levels (Ciaramella and Staiano, 2019). Finally, comparing expression levels between different experiments may require complex normalization processes for results and data (Wang et al., 2009).

*Real-time quantitative PCR (*qPCR*)* is a molecular techinque that enables the simultaneous amplification and quantification of target DNA in real time (Pryor and Wittwer, 2006). This method quantifies the initial amount of target DNA in a sample by monitoring fluorescence during the PCR reaction, enabling for both absolute and relative quantification (Singh and Roy-

Chowdhuri, 2016). Owing to its high sensitivity and specificity, qPCR is extensively employed in genetic analyses of specific genes, genotyping, gene expression profiling, and the identification of polymorphisms and mutations. The primary limitation of qPCR lies in its dependence on prior knowledge of target sequences, which prevents its use in the discovery of novel genes (Nola et al., 2006). One of the main limitations of qPCR is its high sensitivity to sample contamination and primer-dimer formation, which can result in inaccurate outcomes. Consequently, its dynamic range is narrower than that of sequencing-based methods, making it less effective for detecting extremely low or high expression levels across a wide array of genes (Smith and Osborn, 2009). *RNA-Seq* is a powerful high-throughput sequencing approach that allows comprehensive and quantitative profiling of the entire transcriptome (Wang et al., 2009). It relies on next-generation sequencing technologies to convert RNA molecules into complementary DNA (cDNA) fragments, which are then sequenced and mapped to a reference genome. One of the key advantages of RNA-Seq is its high sensitivity and low background noise, as reads can be uniquely assigned to specific genomic regions. Unlike hybridization-based methods, RNA-Seq offers a broad dynamic range of detection, with expression levels directly reflecting the number of sequence reads per transcript. This enables both the detection of highly expressed genes and the identification of transcripts with low abundance. As the technology advances, future developments are expected to improve the resolution of complex transcriptomes, particularly in distinguishing alternative splicing events and rare RNA isoforms (Wang et al., 2009).

Conclusions

Gene expression is a complex mechanism influenced by both genetic and environmental factors. Understanding the expression of individual genes can lead to improvements in production efficiency by optimizing genetic pathways that enhance desirable traits. In addition to production outcomes, understanding and controlling gene expression can improve the health condition of animals, their fitness, and reduce the frequency of certain diseases. Over the past two decades, numerous experimental methods for gene expression analysis have been developed, significantly influencing breeding programs. These methods enable more precise quantification and comparison of gene expression with high sensitivity. They vary in their analytical approaches, sample types and the quantity of genes studied. Among all methods, RNA-Seq stands out due to its comprehensive approach and precise data processing. As technological advancements and bioinformatics tools continue to evolve, transcriptome analysis will become even more refined, potentially leading to the integration of new gene expression insights into breeding programs for livestock species.

References

Bionaz, M., Loor J.J. (2008). Gene networks driving bovine milk fat synthesis during the lactation cycle. BMC Genomics, 31(9): 366.

Bionaz, M., Loor, J.J. (2011.) Gene networks driving bovine mammary protein synthesis during the lactation cycyle. Bioinformatic and Biology Insights, 5: 83-98.

Canovas, A., Rincon, G., Bevilacqua, C., Islas-Trejo, A., Brenaut, P., Hovey, R.C., Boutinad, M., Morgenthaler, C., VanKlompenberg, M.K., Martin, P., Medrano, J.F. (2014). Comparasion of five different RNA sources to exmaine the lactating bovine mammary gland transcriptome using RNA-Sequencing. Scientific Reports, 4, 5297.

Cassar-Malek, I., Picard, B., Bernard, C., Hocquette, J.-F. (2008). Application of gene expression studies in livestock production systems: a European perspective. Australian Journal of Experimental Agriculture, 48: 701-710.

Ciaramella, A., Staiano, A. (2019). The Role of Clustering and Visualization Techniques in Gene Microarray Dana. Algorithms, 12(6): 123.

Clark, T.A., Sugnet, C.W., Ares, M Jr. (2002). Genomewide analysis of mRNA processing in yeast using splicing-specific microarrays. Science, 3, 296(5569): 907-10.

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

Fang, L., Teng, J., Lin, Q. *et al.* (2025). The Farm Animal Genotype–Tissue Expression (FarmGTEx) Project. Nature Genetics, 57: 786–796.

Gao, Y., Lin, X., Shi. K., Yan, Z., Wang, Z. (2013). Bovine mammary gene expression profiling during the onset of lactation. PLoS One 8.

Loor, J.J., Dann, H.M., Guretzky, N.A., Everts, R.E., Oliveira, R., Green, C.A., Litherland, N.B., Rodriguez-Zas, S.L., Lewin, H.A., Drackley, J.K. (2006). Plane of nutrition prepartum alters hepatic gene expression and function in dairy cows as assessed by longitudinal transcript and metabolic profiling. Physiol Genomics, 3, 27(1): 29-41.

Maningat, P.D., Sen, P., Rijnkels, M., Sunehag, A.L., Hadsell, D.L., Bray, M., Haymong, M.W. (2009). Gene expression in the human mammary epithelium during lactation: the milk fat globule transcriptome. Physiological Genomics, 3. 3(1): 12-22.

McCabe, M., Waters, S., Morris, D., Kenny, D., Lynn, D., Creevey, C. (2012). RNA-seq analysis of differential gene expression in liver from lactating dairy cows divergent in negative energy balance. BMC Genomics, 13: 193.

Moody, D.E. (2001). Genomics techniques: An overview of methods for the study of gene expression. Journal of Animal Science, 79: 128–135.

Nolan, T., Hands, R.E., Bustin S.A. (2006). Quantification of mRNA usin real-time RT-PCR. Nature Protocols, 1(3): 1559-82.

Pryor, R.J, Wittwer, C.T. (2006). Real-time polymerase chain reaction and melting curve analysis. Methods in Molecular Biology, 336: 19-32.

Shangraw, E.M., McFadden, T.B. (2024). Regulation of gene expression in milk secretory pathways during secretory activation in Holstein cows. Journal of Dairy Science, 107(12): 11761-11773.

Singh, C., Roy-Chowdhuri, S. (2016). Quantitative Real-Time PCR: Recent Advances. Methods in Molecular Biology, 1392: 161-76.

Smith, C.J., Osborn, A.M. (2009). Advantages and limitations of quantitative PCR (Q-PCR)-based approaches in microbial ecology. FEMS Microbiology Ecology, 67(1): 6-20.

Sun, Z., Kuczek, T., Zhu, Y. (2012). Statistical Calibration of qRT-PCR, Microarray and RNA-Seq Gene Expression Data with Measurement Error Models. The Annals of Applied Statistics, 8(2).

Wall, E.H., McFadden, T.B. (2007). The milk yield response to frequent milking in early lactation of dairy cows is locally regulated. Journal of Dairy Science, 90(2): 716-20.

Wall, E.H., McFadden, T.B. (2012). Triennial Lactation Symposium: A local affair: How the mammary gland adapts to changes in milking frequency. Journal of Animal Science, 90(5): 1695-707.

Wang, Z., Gerstein, M., Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. Nature Review Genetics, 10(1): 57-63.

Yang, J., Jiang, J., Liu, X., Wang, H., Guo, G., Zhang, Q., Jang, L. (2015). Differential expression of genes in milk of dairy cattle during lactation. Animal Genetics, 47(2): 174-80.