



# Identification of Key Genes and Potential Pathways Associated with Mastitis Induced by *E. coli*

Elham Darang<sup>1</sup> · Zahra Pezeshkian<sup>1</sup> · Seyed Ziaeddin Mirhoseini<sup>1</sup> · Shahrokh Ghovvati<sup>1</sup>

Received: 9 August 2021 / Accepted: 22 June 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

## Abstract

Mastitis is one of the most important diseases of dairy cattle. It is an infectious disease leading to an inflammatory reaction in the cow's mammary gland. *Escherichia coli* is one of the common bacteria which induce mastitis in cows. The aim of this study was to identify key genes and potential pathways associated with mastitis induced by *E. coli* in dairy cattle using bioinformatics analysis. The gene expression profile of ten samples including five adjacent tissues from a quarter infected with *Escherichia coli* and five tissues from a healthy quarter of dairy cattle was assessed using GEO2R. Gene ontology and pathway analysis were performed using bioinformatics tools. A total of 156 differentially expressed genes were detected which 95 genes were upregulated and 61 genes were downregulated in adjacent tissue of quarter infected compared with healthy tissue. Cellular oxidant detoxification and oxidation–reduction process were the most significant biological process terms in gene ontology analysis. The most important pathways of DEGs were the biosynthesis of amino acids, p53 signaling pathway, and Metabolic pathways. Three important modules were identified and their path enrichment analysis was performed. There are 10 core genes, among which *SOD2*, *COL1A2*, *COL3A1*, *POSTN*, *ALDH18A1*, and *CBS* may be the main genes associated with mastitis, which can be considered as candidate genes in the prevention and early diagnosis program of mastitis.

**Keywords** Mastitis · *Escherichia coli* · Hub genes · Protein–protein interaction · Bioinformatics analysis

---

✉ Shahrokh Ghovvati  
Ghovvati@guilan.ac.ir; Ghovvati@yahoo.co.uk

<sup>1</sup> Department of Animal Sciences, Faculty of Agriculture, University of Guilan, Rasht 41635-1314, Guilan, Iran

## Introduction

Due to the increasing demand of consumers for healthy products, attention to animal health has become more important (Palii et al. 2020). Mastitis is one of the most common diseases that can affect dairy cattle during dry periods and lactation. It is considered as the most expensive dairy cattle disease in the world, which causes a lot of damage to the dairy industry every year (Halasa et al. 2007; Hamed et al. 2020). In addition to the cost of treatment and veterinary medicine and death from the disease, the influence of mastitis in the dairy cattle economy can include reduced milk production and quality, the indirect impact of the disease on lower reproductive rates, and increased risk of culling in the herd. The use of antibiotics in the treatment of mastitis can also cause side effects of its residues on human health (Barlow 2011; Schukken et al. 2009).

In general, the causative agent of mastitis is divided into two categories: infectious and environmental factors, which are epidemiologically different. The main bacteria that cause infectious mastitis are *Staphylococcus aureus*, *Streptococcus agalactiae*, *Streptococcus dysagalactiae*, and *Mycoplasma bovis*. In addition, the bacteria that cause peripheral mastitis include *Escherichia coli* and *Streptococcus dysagalactiae*. Due to the control of the causative agents of infectious mastitis, the incidence of environmental mastitis in the herd has increased. *E. coli* is the most important environmental pathogen in dairy cow mastitis, which causes great economic damage to livestock (Phuektes et al. 2001; Zhao and Lacasse 2008).

*E. coli* infection has two different types of responses. The first response, which is localized, involves inflammation and the cellular immune response and occurs in a quarter infected with a bacterial infection, but the other response, which is systemic, may occur in a quarter around infected tissue and distant organs such as the liver. The systemic response involves the processing and antigen presentation, cytokines, protein degradation, and apoptosis. The interaction effects of these responses indicate disease and infection in the animal (Mitterhuemer et al. 2010).

The goal of the breeding program is to increase resistance to mastitis in dairy herds. To achieve this goal, genes that affect mastitis should be identified. Based on analog biological roles and comparison of transcripts of infected and healthy animals and determinants of gene and network hubs, this project supports proteins and significant modules effective in disease that helps to understand the host response to the pathogen (Gorji et al. 2019).

It will be possible to study the expression of thousands of genes normally in a short time using microarray technology (Blohm and Guiseppe-Elie 2001). In this regard, bioinformatics and computational biology have overcome the challenges associated with the processing of biological information, and by recognizing the difference in the expression of genes and their interactions, a better understanding of cellular mechanisms and disease-causing processes, including mastitis, is obtained. This study aimed to identify key genes and potential pathways associated with mastitis induced by *E. coli* in dairy cattle by bioinformatics analysis.

## Materials and Methods

### Data Collection

In this study, the gene expression profile (GSE15022 (which was obtained with the help of platform GPL2112 Affymetrix Bovine Genome Array) was used. The data contain ten samples of mammary quarter tissue, including five tissue samples around quarter infected with *E. coli*, compared to five quarter tissue samples of healthy dairy cows. RNA probes for hybridization of sample arrays were developed by Ambion Kit in 2010 by Mitterhuemer et al. The gene expression data were downloaded from the public database the Gene Expression Omnibus (GEO) database of NCBI (<http://www.ncbi.nlm.nih.gov/geo/>) (Barrett et al. 2013).

### Identification of DEGs

The online statistical tool, GEO2R (<http://www.ncbi.nlm.nih.gov/geo/geo2r>) (Davis and Meltzer 2007) is an R-based web program that helps users analyze GEO data and identify and visualize significant differences in gene expression. This tool uses a *t* test to compare two groups (control and treatment) or ANOVA to compare more groups. In this study, GEO2R analysis was used to compare tissue samples around quarter infected with *E. coli* and quarter tissue of healthy dairy cows. The adjusted *P* value was applied to correct for the occurrence of false-positive results. The GEO2R inbuilt method, Benjamini and Hochberg (false discovery rate), were used. The cut-off criteria were  $|LogFC| \geq 1$  and adjusted *p* value  $< 0.05$ . Finally, DEGs including upregulated and down-regulated genes were obtained.

### PPI Network and Module Analysis

To further investigate the association of DEGs at the protein level, the associated PPI networks were plotted using the Search Tool for the Retrieval of Interacting Genes (STRING, <http://string-db.org/>) (Szklarczyk et al. 2019). A maximum number of interactors = 0 and a confidence score  $\geq 0.4$  were chosen as the cutoff criteria. The STRING database is an online database available for evaluating protein networks.

The Molecular Complex Detection (MCODE) v1.6.1 plugin of Cytoscape v3.8.0 software was applied to clarify the biological significance of modules of the PPI network with max. depth = 100, k-core = 2, node score cutoff = 0.2 and degree cutoff = 2. Then, the DAVID tool was used for functional enrichment analysis of each module.

### Gene ontology and KEGG Pathway Analysis

For functional annotation of GO and analysis of KEGG pathway enrichment, David (<https://david.ncifcrf.gov/>) was used (Sherman and Lempicki 2009). This web tool is an online bioinformatics resource for any functional evaluation of high-throughput gene expression profiles. The DEGs were exposed to ClueGO v2.5.7 to obtain

complete Gene Ontological terms (GO) (Bindea et al. 2009). The  $p$ -values  $< 0.05$  were considered to be significant.

## Finding Hub Genes

In the PPI network, the genes that are most closely related to other genes are called gene hubs. Hub genes are conserved in the evolutionary pathway and are known as molecular biomarkers in various diseases. CytoHubba is a simple Cytoscape plugin that ranks the importance of nodes in the PPI network with different algorithms for identifying key biological elements (Chin et al. 2014). In this study, using the degree algorithm, ten genes with a rank higher than 8 were selected as hub genes.

## Results

### Identification of DEGs

The gene expression profile of GSE15022 which contains 10 samples including five adjacent tissue of quarter infected with *E.coli* and five tissue of healthy quarter of dairy cattle was assessed using GEO2R. A total of 156 differentially expressed genes were detected which 95 genes were up-regulated and 61 genes were downregulated in adjacent tissue of quarter infected compared with healthy tissue (Table 1). Volcano plot display differentially expressed genes between adjacent tissue of quarter infected and healthy tissue of quarters of dairy cattle (Fig. 1).

### PPI Network and Sub-modules Analysis

Based on the data in the STRING database, the PPI network consisting of 153 nodes and 168 edges was plotted using Cytoscape software (Fig. 2).

The Cytoscape plugin MCODE displayed the closely interlinked areas from the PPI network in the shape of clusters. The top three clusters that were significant with MCODE scores of 6.5, 3.3, and 3.1 were selected (Table 2).

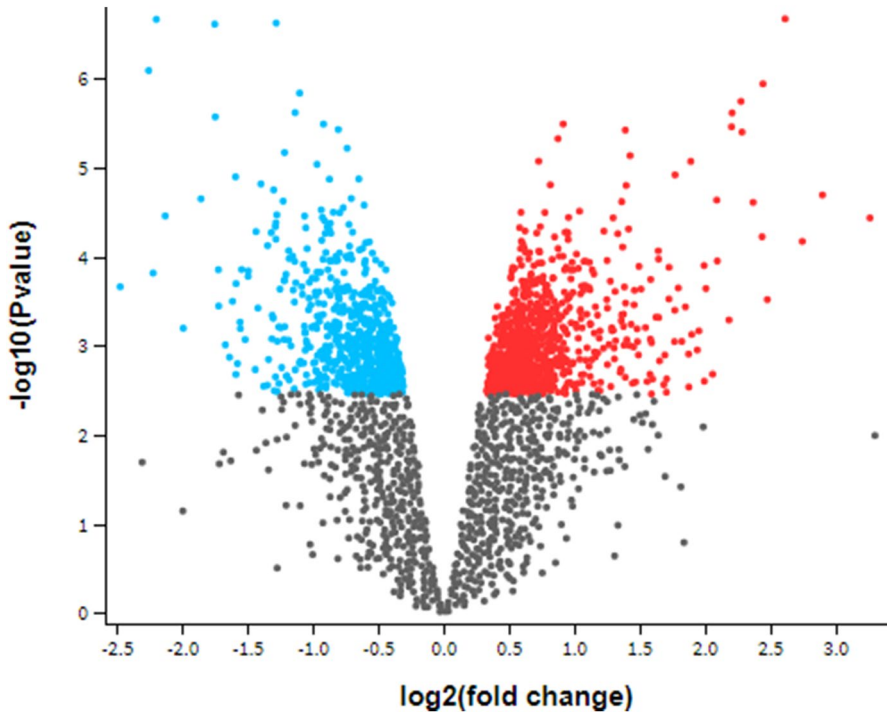
### DAVID Enrichment Analysis

To obtain a better understanding of important differentially expressed genes, the DAVID v6.8 online server was used for GO analysis and KEGG enriched pathways. In gene ontology, three parts of functional groups including biological process (BP), cellular performance (CC), and molecular function (MF) were analyzed. The modified Fisher exact  $p$ -value (EASE score)  $\leq 0.05$  was considered strongly enriched.

It was found that the DEGs are enriched in 20 terms of the biological process which cellular oxidant detoxification (GO: 0098869), oxidation–reduction process (GO: 0055114), and positive regulation of intrinsic apoptotic signaling pathway (GO: 2001244) are the most significant biological process (Table 3). About the cellular component analysis, the DEGs were mostly enriched in extracellular space

**Table 1** One hundred and fifty six differentially expressed genes including 95 up-regulated genes and 61 down-regulated genes which were identified between adjacent tissue of quarter infected of infected tissue with *E. coli* and healthy quarter of dairy cattle

DEGs	Gene names
Upregulated DEGs	S100A8, ATF5, S100A9, TNFRSF6B, CH13L1, WARS, S100A12, CFI, ZMYND12, CYP2B6, GPX3, CCDC39, DTX3L, SDS, HMGCL, MPZL3, C6, BAX, CYBRD1, ATXN3, GARS, GKAP1, MTHFD2, CBS, CSNK2A2, METTL12, PHGDH, SARS, H2B, SOD2, NSL1, LOXL4, PCK2, UBE2L6, GABARAPL1, GNLY, BOLA-N, HSP90AB1, GSS, CCDC53, CXCL14, CLU, GTPBP2, RPS27L, SEC22C, FAM159B, PIR, SVOPL, RGN, PHB, SLC3A2, CARHSP1, SAT1, MAP1LC3B, MID1IP1, ADORA2B, MT2A, SESN2, NPC1, SLC18A2, PABPC4, SESN1, CDO1, BLOC1S2, CYP4F2, TRIB3, SAA3, MED10, BOLA, THRSP, CDKN1A, HSPD1, NXPE3, NCF4, GLRX2, PSPH, HSPE1, BOLA-A, MTPP, NT5C2, NCAPH, ANGPT2, HIST2H2AC, PLLP, RND1, SNHG12, GRO1, C28H10orf10, HSPH1, TSPAN2, KCNJ15, MT1E, AKR1C4, SHPK, FKBP5
Downregulated DEGs	NREP, OGN, IGFBP5, COL21A1, CDH11, LUM, CCDC80, LOXL1, NDFIP1, CYP1A1, MYH1, BEND5, ISLR, FXYD6, EHF, PRELP, FBLN5, RGS5, SFN, STMN1, CCDC184, RRM2, TMOD4, CNN1, PMPA1, PDGFRL, COL1A2, AHCYL2, MSR1, IGFBP6, ALOX15, KIAA0101, IDH1, NTN4, PTHLH, GEM, SSR3, PI16, MMP23B, SLC25A21, ART3, FMOD, FABP5, ABCG2, TM4SF18, AGR2, RASD1, COL3A1, ACSM1, CNGA1, ALDH18A1, BDH1, RETSAT, SFRP4, G0S2, POSTN, PIK3C2G, PON3, RGS1, CCL26, FKBP14



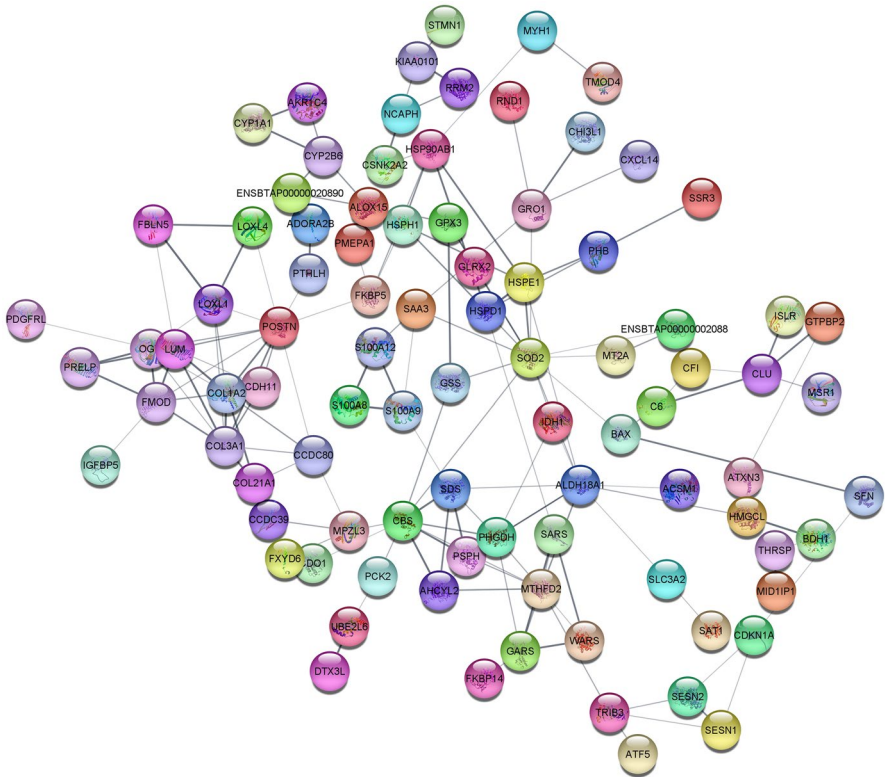
**Fig. 1** Volcano plot displaying differentially expressed genes between adjacent tissue of quarter infected with *E. coli* and healthy quarters of dairy cattle. Significant up-regulated genes were showed in red dot (Up) and significant down-regulated genes were showed in blue dot (Down) (adj.*p*value < 0.05)

(GO: 0005615), mitochondrial matrix (GO: 0005759), and extracellular exosome (GO: 0070062) (Table 4). The gene ontology molecular function analysis revealed the involvement of DEGs in oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen (GO: 0016705), monooxygenase activity (GO: 0004497), and iron ion binding (GO: 0005506) (Table 5).

The KEGG pathway enrichment analysis revealed the association of the DEGs in ten KEGG pathways. The most significant pathways were Biosynthesis of amino acids (bta01230), p53 signaling pathway (bta04115), Cysteine and methionine metabolism (bta00270), and Metabolic pathways (bta01100) (Table 6).

### ClueGO Enrichment Analysis

The Cytoscape plugin ClueGO was applied to investigate the functional enrichment of the DEGs of the dataset. ClueGo assisted to visualize the GO terms of the detected PPI complex network. The BP and MF terms of the GO functional enrichment analysis of the complex PPI network are described in Fig. 3. The statistical options for ClueGO enrichment analysis were set according to a hypergeometric test which is two-sided with  $p \leq 0.05$ , Benjamini–Hochberg correction, and kappa score  $\geq 0.4$  as



**Fig. 2** Protein–protein interaction network constructed with the differentially expressed genes

**Table 2** The most interlinked regions are clustered from the DEGs of GSE1502 dataset using MCODE

Cluster	Node IDs	Edges	Nodes	Score
1	COL3A1, COL1A2, LUM, CDH11, LOXL1, OGN, POSTN, FMOD	23	8	6.571
2	SESN1, SESN2, TRIB3, CDKN1A	5	4	3.333
3	HSPE1, HSP90AB1, HSPD1, WARS, PHGDH, SARS, CBS, HSPH1, GARS, PSPH	14	10	3.111

a primary criterion. The BP terms for DEGs were mostly enriched in response to a toxic substance (GO:0009636), L-serine metabolic process (GO:0006563), chaperone-mediated protein folding (GO:0061077), cellular response to xenobiotic stimulus (GO:0071466), xenobiotic metabolic process (GO:0006805), and establishment of the skin barrier (GO:0061436), and the MF term for DEGs were mostly enriched in RAGE receptor binding (GO:0050786), sulfur compound biosynthetic process (GO:0044272), oxidoreductase activity, acting on single donors with incorporation of molecular oxygen, incorporation of two atoms of oxygen (GO:0016702) and extracellular matrix organization (GO:0030198) (Fig. 4).

**Table 3** Significantly enriched GO terms in biological process of differentially expressed genes ( $p < 0.05$ )

Term	Description	Count	PValue	Genes
GO:0098869	Cellular oxidant detoxification	5	2.23E-04	GPX3, SESN1, SESN2, S100A9, S100A8
GO:0055114	Oxidation–reduction process	12	2.70E-04	BDHI, MTHFD2, GPX3, ALOX15, LOXL4, CYBRD1, PHGDH, GLRX2, AKR1C4, CDO1, SOD2, LOXL1
GO:2001244	Positive regulation of intrinsic apoptotic signaling pathway	4	0.0015	BAX, S100A9, CLU, S100A8
GO:0030199	Collagen fibril organization	4	0.0027	COL3A1, COL1A2, LUM, FMOD
GO:0009636	Response to toxic substance	4	0.0047	PON3, CYP1A1, BAX, SLC18A2
GO:0002474	Antigen processing and presentation of peptide antigen via MHC class I	3	0.0068	BOLA-A, BOLA, BOLA-N
GO:0.014,002	Astrocyte development	3	0.0079	TSPAN2, S100A9, S100A8
GO:0.006,801	Superoxide metabolic process	2	0.0091	CBS, SOD2
GO:0006919	Activation of cysteine-type endopeptidase activity involved in apoptotic process	4	0.0095	BAX, S100A9, S100A8, HSPD1
GO:0006914	Autophagy	4	0.0108	NPC1, SESN2, S100A9, S100A8
GO:0006955	Immune response	7	0.0189	C6, TNFRSF6B, BOLA-A, GRO1, BOLA, CXCL14, BOLA-N
GO:0035963	Cellular response to interleukin-13	2	0.0194	SHPK, ALOX15
GO:0002793	Positive regulation of peptide secretion	2	0.0194	S100A9, S100A8
GO:0070488	Neutrophil aggregation	2	0.019	S100A9, S100A8
GO:0007409	Axonogenesis	4	0.026	LUM, OGN, PRELP, FMOD
GO:0006564	L-serine biosynthetic process	2	0.029	PHGDH, PSPH
GO:1901031	Regulation of response to reactive oxygen species	2	0.029	SESN1, SESN2
GO:0010811	Positive regulation of cell–substrate adhesion	3	0.0303	CCDC80, ALOX15, AGR2
GO:0061077	Chaperone-mediated protein folding	3	0.0323	FKBP14, CLU, FKBP5
GO:0045606	Positive regulation of epidermal cell differentiation	2	0.0479	SFRP4, SFN



**Table 4** Significantly enriched GO terms in cellular component of differentially expressed genes ( $p < 0.05$ )

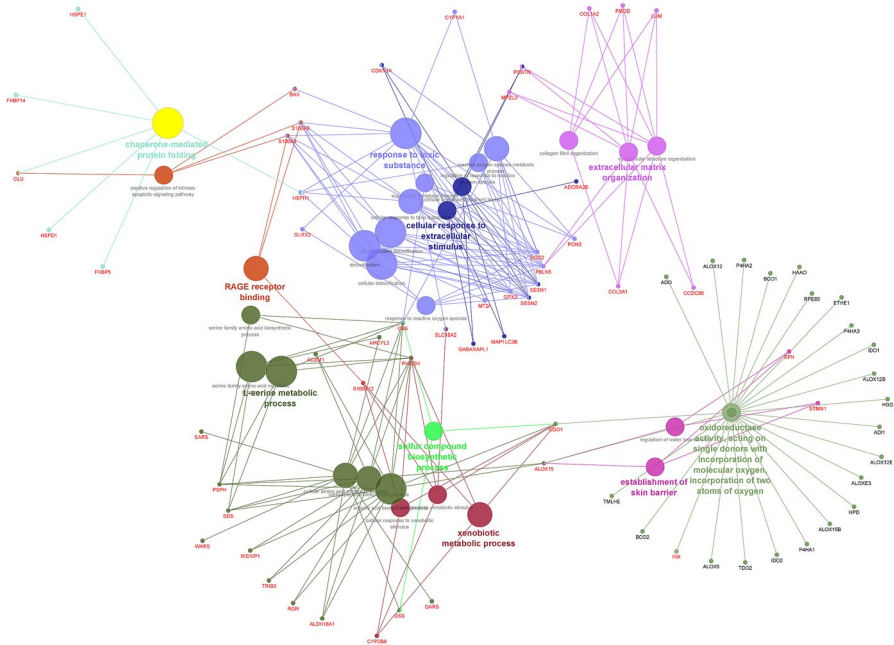
Term	Description	Count	P Value	Genes
GO:0005615	Extracellular space	27	1.07E-06	PON3, SAA3, CFI, LOXL4, PRELP, GRO1, CLU, FBLN5, PTHLH, LOXL1, HSPD1, AGR2, H2B, IGFBP6, ANGPT2, IGFBP5, LUM, GPX3, SFRP4, COL3A1, COL1A2, MTHFD2, OGN, CHI3L1, FMOD, S100A9, S100A8
GO:0005759	Mitochondrial matrix	8	3.80E-04	ATXN3, HMGCL, BDH1, ACSM1, SOD2, HSPH1, HSPD1
GO:0070062	Extracellular exosome	37	5.15E-04	PON3, ACSM1, CFI, LOXL4, PRELP, PHB, FBLN5, ART3, HSPD1, ISLR, HSPH1, C6, STMN1, H2B, SFN, IGFBP6, PCK2, WARS, LUM, PLLP, GPX3, SARS, CYBRD1, HSPH1, CARHSP1, COL1A2, FABP5, NPC1, CDH11, OGN, CHI3L1, BAX, GARS, PI16, S100A9, S100A8, FKBP5
GO:0005829	Cytosol	20	0.004	AHCYL2, GABARAPL1, CDKN1A, GSS, IDH1, MID1I1, NCF4, ALOX15, CLU, SAT1, HSPD1, MT2A, MAP1LC3B, CARHSP1, HSPH1, MTHFD2, RGS1, BAX, THRSF, CDO1
GO:0042612	MHC class I protein complex	3	0.0066	BOLA-A, BOLA, BOLA-N
GO:0005739	Mitochondrion	16	0.0228	GABARAPL1, HSP90AB1, SDS, ACSM1, GOS2, GLRX2, HSPH1, MTHFD2, BLOC1S2, SESN2, AGR2, BAX, GARS, ALDH18A1, C28H100RF10, PCK2
GO:0031012	Extracellular matrix	5	0.02539	MMP23B, OGN, CLU, FBLN5, LOXL1
GO:0005578	Proteinaceous extracellular matrix	6	0.0289	COL3A1, LUM, OGN, PRELP, FMOD, FBLN5

**Table 5** Significantly enriched GO terms in molecular function of differentially expressed genes ( $p < 0.05$ )

Term	Description	Count	P Value	Genes
GO:0016705	Oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen	3	0.003	CYP2B6, CYP4F2, CYP1A1
GO:0004497	Monooxygenase activity	3	0.003	CYP2B6, CYP4F2, CYP1A1
GO:0005506	Iron ion binding	6	0.012	CYP2B6, CYP4F2, ALOX15, CYP1A1, CDO1
GO:0035662	Toll-like receptor 4 binding	2	0.019	S100A9, S100A8
GO:0000287	Magnesium ion binding	6	0.023	HMGCL, MTHFD2, IDH1, GSS, PSPH, GEM
GO:0020037	Heme binding	5	0.029	CYP2B6, CBS, CYP4F2, CYP1A1
GO:0050544	Arachidonic acid binding	2	0.038	S100A9, S100A8
GO:0070728	Leucine binding	2	0.038	SESN1, SESN2
GO:0016641	Oxidoreductase activity, acting on the CH-NH2 group of donors, oxygen as acceptor	2	0.038	LOXL4, LOXL1
GO:0051287	NAD binding	3	0.042	AHCYL2, IDH1, PHGDH
GO:0030145	Manganese ion binding	2	0.042	HMGCL, SOD2
GO:0005201	Extracellular matrix structural constituent	3	0.049	COL3A1, COL1A2, PRELP

**Table 6** KEGG pathway analysis of differentially expressed genes ( $p < 0.05$ )

Term	Count	P-Value	Genes
bia01230:biosynthesis of amino acids	6	0.0012	SDS, CBS, IDH1, PHGDH, ALDH18A1, PSPH
bia04115:p53 signaling pathway	6	0.0012	CDKN1A, RRM2, SESN1, SESN2, BAX, SFN
bia00590:Arachidonic acid metabolism	5	0.0089	CYP2B6, GPX3, CYP4F2, ALOX15
bia00270:cysteine and methionine metabolism	4	0.009	AHCYL2, SDS, CBS, CDO1
bia01100:metabolic pathways	24	0.0095	AHCYL2, RRM2, SDS, GSS, CYP4F2, ACSMI, IDH1, ALOX15, PIK3C2G, NT5C2, SAT1, HMGCL, CYP2B6, BDH1, MTHFD2, CBS, RGN, CYP11A1, PHGDH, ALDH18A1, CDO1, PSPH, PCK2
bia00260:glycine, serine and threonine metabolism	4	0.0111	SDS, CBS, PHGDH, PSPH
bia00480:glutathione metabolism	4	0.0233	RRM2, GPX3, IDH1, GSS
bia01130:biosynthesis of antibiotics	7	0.0296	SDS, CBS, IDH1, RGN, PHGDH, PSPH, PCK2
bia01200:carbon metabolism	5	0.0353	SDS, IDH1, RGN, PHGDH, PSPH
bia00650:butanoate metabolism	3	0.04	HMGCL, BDH1, ACSMI



**Fig. 3** Gene Ontology enrichment terms was visualized by the ClueGO plugin of Cytoscape. Important BP and MF of the DEGs are presented with the particular gene interactions. Functional nodes and edges are described the connectivity of the GO terms network with a kappa score of 0.4 and  $p$ -value  $\leq 0.05$ . This  $p$ -value indicates the size of the node. The color code of each node represents the special functional categorizes that they are involved in. The most important functional GO terms are showed by bold fonts and the names of the DEGs involved in each class are presented in red font (Color figure online)

The KEGG pathway analysis from ClueGO showed that many DEGs were significantly enriched in Glycine, serine and threonine metabolism (KEGG: 00260), Cysteine and methionine metabolism (KEGG: 00270), Glutathione metabolism (KEGG: 00480), Butanoate metabolism (KEGG: 00650), p53 signaling pathway (KEGG: 04115), Ferroptosis (KEGG: 04216) and Arachidonic acid metabolism (KEGG: 00590) (Fig. 4).

### Finding Hub Genes

PPI network using cytoHubba plugin showed that 10 genes have the highest interaction compared to other genes identified as hub genes (Table 7). Among them, *SOD2*, *ALDH18A1*, and *MTHFD2* genes did not belong to any of the three modules. *CBS* and *HGDH* were present in Module 3 and the remaining five genes belonged to Module 1.



**Fig. 4** Pathway term enrichment that is visualized by the ClueGo plugin of Cytoscape. The plugin delivers a comprehensive enrichment analysis for DEGs, including KEGG and REACTOME pathways. Functional nodes and edges describe the connectivity of the pathway terms network with a kappa score of 0.4 and  $p$ -value  $\leq 0.05$ . This  $p$ -value indicates the size of the node. The color code of each node represents the special functional categories that they are involved in. The most important functional pathway terms are showed by bold fonts and the names of the DEGs involved in each class are presented in red font (Color figure online)

## Discussion

Mastitis in dairy cows is of great economic importance because of the damage it does to the dairy industry, its quality, and composition, as well as dairy products. The most significant gene ontologies of DEGs in biological process enrichment were cellular oxidant detoxification and oxidation–reduction process. In addition, seven genes were enriched in immune response, which indicates the development of an immune response against *E. coli*. This was similar to the results reported by Buitenhuis et al. (2011) that indicated mastitis in *E. coli* infection significantly increased the expression of upregulated genes in the immune response biological

**Table 7** Top 10 in network string-cytoscape input.txt ranked by Degree method

Rank	Name	Score	Cluster
1	SOD2	13	–
2	CBS	10	3
3	POSTN	10	1
4	ALDH18A1	10	–
5	MTHFD2	9	–
6	COL1A2	8	1
7	COL3A1	8	1
8	LUM	8	1
9	OGN	8	1
10	PHGDH	8	3

process. Mitterhuemer et al. (2010) reported that the DEGs between udder quarters infected with *E. coli*, adjacent to infected tissue quarters, and control animals were mainly involved in immune response.

The common biological process terms using ClueGO and DAVID were response to toxic substance, L-serine metabolic process and chaperone-mediated protein folding. Cellular oxidant detoxification was subgroup of response to toxic substance GO.

Abdelmegid et al. (2017) investigated the host defense-related proteins of cows with *S. aureus* subclinical mastitis. One of the significant GO in BP term was cellular oxidant detoxification. So, this GO might be important in the response to any bacterial infection leading to mastitis (Abdelmegid et al. 2017). Ahlawat et al. (2021) compared gene expression profiling of milk somatic cells of cattle and buffaloes. Cellular oxidant detoxification was a BP term that was significant in both species. As somatic cells are important in mastitis, this GO can be identified as a significant GO in mastitis. Therefore, the DEGs involved in this GO are important too (Ahlawat et al. 2021).

Dai et al. (2018) used RNA-seq transcriptomics analysis of the mammary gland of bovine during the dry period and lactation. One of the significant GO was the oxidation–reduction process which was showed the importance of this GO in lactation (Dai et al. 2018). The oxidation–reduction process is a way to degrade organic contaminants which can be applied for the inactivation of pathogenic contaminants (Ali et al. 2020).

The KEGG pathway analysis of DEGs showed that the most significant pathway was enriched in the biosynthesis of amino acids and the p53 signaling pathway. Most of the genes (24 genes) were involved in the metabolic pathways.

The gut microbiome related to milk protein in buffaloes and Holstein cattle were investigated in a study. The results showed that the biosynthesis of amino acids is a significant pathway that also positively correlated to the genera of *Parasutterella*, *Sutterella*, *Dorea*, and *Parabacteroides* (Zhang et al. 2017). Another study surveyed the key proteins in Aspirin inhibiting effect on *Staphylococcus xylosus*. *S. xylosus* is a pathogen that causes infection in cow mastitis. The biosynthesis of amino acids

was one of the most significant pathways in the study. Therefore, it can be concluded that the pathway of the biosynthesis of the amino acids is an important pathway in response to pathogen and bacterial infection (Xu et al. 2017).

FU et al. (2019) reported that p53 suppresses the progression of mammary epithelial tumors and their metastasis through the ZEB1/p53 signaling pathway. Other significant pathways are Arachidonic acid metabolism, Cysteine and methionine metabolism, Glycine, serine, and threonine metabolism, Glutathione metabolism, Biosynthesis of antibiotics, Carbon metabolism, and Butanoate metabolism (Fu et al. 2019).

In addition, the results of KEGG pathway analysis using ClueGO showed that many DEGs are involved in the mentioned pathways such as Glycine, serine and threonine metabolism, Cysteine and methionine metabolism, Glutathione metabolism, Butanoate metabolism, p53 signaling pathway, Ferroptosis, Arachidonic acid metabolism.

Among DEGs, 10 hub genes had the largest degree. Superoxide dismutase 2 (*SOD2*) is a manganese-dependent superoxide dismutase. Becuwe et al. (2014) reported that *SOD2* could act as a biomarker in the metastatic progression of breast tumors. In a study, Jensen et al. (2013) investigated mastitis caused by *E. coli* and *S. aureus* adjacent to non-infected quarters. The results showed that *SOD2* was up-regulated during *S. aureus* infection and acted as a gene involved in mammary gland response to the infection (Jensen et al. 2013). A bioinformatics analysis investigated the microRNA and putative target genes in bovine mammary tissue infected with *Streptococcus uberis*. The results showed that *SOD2* was among differentially expressed genes due to infection (Naeem et al. 2012). Another study investigated the distinct local and systemic transcriptome responses which were induced by *E. coli* infection in the mammary gland of cows. Enhanced expression of *SOD2* was observed as an indicator of oxidative stress and an active defense reaction in infected and neighboring healthy quarters (Mitterhuemer et al. 2010). A study surveyed differentiating *S. aureus* from *E. coli* mastitis after udder infection. *SOD2* was one of the DEGs included in the *E. coli* infected samples (Günther et al. 2017).

Cystathionine beta-synthase (*CBS*) was involved in the metabolic pathway of glycine, serine, and threonine metabolism, biosynthesis of amino acids, biosynthesis of antibiotics, and cysteine and methionine metabolism and was one of the genes involved in Module 3. The *CBS* gene encodes the enzyme cystathionine beta-synthase. This enzyme is responsible for using vitamin B6 as a cofactor to make cystatoin from serine and homocysteine. Cystathionine is used to make taurine. Taurine plays a variety of roles in the cell, including regulating electrolyte balance, strengthening the immune system, and antioxidant function. Therefore, increased expression of this gene can be one of the signs of the cell's immune response against *E. coli* infection (Zuhra et al. 2020).

Aldehyde dehydrogenase 18 family, member A1 (*ALDH18A1*) is a member of the aldehyde dehydrogenase family and encodes a bifunctional NADPH- and ATP-dependent mitochondrial enzyme. *ALDH18A1* plays an important role in the biosynthesis of the amino acids of proline, ornithine, and arginine (Holmes 2017). *ALDH18A1* gene encodes the enzyme synthetase of P5C, which catalyzes glutamate to P5C and then P5C catalyzed by the enzyme P5C reductase (PYCR) to proline

(D’Aniello et al. 2020). Also, Proline is one of the main amino acids of collagen and is needed to stabilize the triple helix of collagen (D’Aniello et al. 2020). The results of a study on the identification of candidate genes affecting mastitis in cows showed that *ALDH18A1* was among DEGs in mammary tissue after intramammary infection with *E. coli* or *Streptococcus uberis* (Chen et al. 2015). *ALDH18A1* is one of the genes in the collagen synthesis pathway (Kocher et al. 2021). Therefore, *ALDH18A1* can be considered an important gene in mastitis induced by *E. coli*.

Collagen, as the most abundant protein in the body, is essential for maintaining the natural structure and strength of connective tissue (Chen et al. 2020). Glycine, proline, and hydroxyproline are the main amino acids that make up the collagen structure. Threonine, serine, and choline are involved, in the synthesis of glycine. Also synthesis proline from arginine and hydroxyproline from the rest of proline in collagen (Li and Wu 2018). Therefore, the amino acid biosynthesis pathway enriched in Kegg is important for collagen synthesis. Collagen, type III, alpha 1 (*COL3A1*), and Collagen, type I, alpha 2 (*COL1A2*) genes were identified as other hub genes that encode type 3 and type 1 collagen, respectively. In a study on sheep, Transcriptional Profiles of the mammary gland were investigated during the lactation period. *COL3A1* was found to be DEG in this study. This evidence suggests that *COL3A1* may play an important role in the sheep mammary epithelial cells cycle and can be considered as a candidate hub gene for future research in cows (Chen et al. 2020). A study evaluating various miRNAs expressed in dairy heifers and dairy cows showed that *COL1A2* is one of the genes associated with the cytoskeleton and extracellular matrix and is associated with mammary gland functional activity (Wicik et al. 2016). In a study by Tsutsui et al. (2020) On the pattern of expression of different types of fibril collagen in different periods of pregnancy, lactation, and after breastfeeding, it was stated that type I collagen produces a suitable fibrillar structure during pregnancy, while this structure after pregnancy broken. during calving temporary dense fibrillar network is formed by type III fibrillar collagen. it has been generally reported that these collagens are involved in the structural reconstruction of the mammary glands and their function (Tsutsui et al. 2020). In a study, miRNA and mRNA Expression Profiles in Mammary Glands of Holstein Cows infected by *S. aureus* were investigated and *COL1A2* was one of the most significant DEGs involved in many pathways (Wang et al. 2021).

Younis et al. (2016) reported that Osteoglycin (*OGN*) which induces bone formation, Lumican (*LUM*) which is important in the development of tissue-engineered cartilage, and periostin osteoblast-specific factor (*POSTN*) which is important in extracellular matrix mineralization along with *COL1A2* was involved in *E. coli*-induced defense response in the breast during mastitis induced by *E. coli* (Younis et al. 2016). Gopinath et al (2022) reported that *OGN* not only induces the formation of bovine endochondral bone but also plays an important role in collagen fibrillogenesis. *POSTN* has also been reported to form a network structure in interaction with several proteins, including collagen V (Gopinath et al. 2022). Lumican participates in the structural organization of tissues (Giatagana et al. 2021). Barreto et al. (2020) reported that *LUM*, through its pathogen-related molecular pattern (PAMP), as a mediator of TLR4 activation, causes inflammation, cartilage destruction, and polarization of macrophages in osteoarthritis. Choi et al. (2022) reported that serine



is involved in the proliferation of breast cancer cells and that the *PHGDH* gene, which encodes the enzyme phosphoglycerate dehydrogenase, catalyzes the synthesis of this amino acid.

As a result, this bioinformatics analysis identified ten hub genes that might play an important role in mastitis induced by *E. coli*. *SOD2*, *COLIA2*, *COL3A1*, *POSTN*, *ALDH18A1*, and *CBS* might be the core genes related to mastitis. They can be considered as candidate genes in a breeding program of mastitis prevention and early detection. Further experimental studies are required to confirm the findings of the present analysis.

**Author Contributions** ED and ZP: designing, analyzing, and writing the manuscript. SZM and SG: project administration, Supervision, Validation, review & editing. All authors have read and agreed to the published version of the manuscript.

**Funding** Not applicable.

**Data Availability** The gene expression profile (GSE1502 (was downloaded from the public database of the Gene Expression Omnibus (GEO) database of NCBI (<http://www.ncbi.nlm.nih.gov/geo/>).

**Code Availability** Not applicable.

## Declarations

**Conflict of interest** The authors have no conflict of interest in this article.

**Ethical Approval** The manuscript does not contain clinical studies or animal use.

**Consent to Participate** Not applicable.

**Consent for Publication** Not applicable.

## References

- Abdelmegid S, Murugaiyan J, Abo-Ismael M, Caswell JL, Kelton D, Kirby GM (2017) Identification of host defense-related proteins using label-free quantitative proteomic analysis of milk whey from cows with *Staphylococcus aureus* subclinical mastitis. *Int J Mol Sci* 19:78
- Ahlawat S, Arora R, Sharma U, Sharma A, Girdhar A, Sharma R et al (2021) Comparative gene expression profiling of milk somatic cells of Sahiwal cattle and Murrah buffaloes. *Gene* 764:145101
- Ali A, Liang Y, Ahmed S, Yang B, Guo B, Yang Y (2020) Mutual contaminants relational realization and photocatalytic treatment using Cu<sub>2</sub>MgSnS<sub>4</sub> decorated BaTiO<sub>3</sub>. *Appl Mater Today* 18:100534
- Barlow J (2011) Mastitis therapy and antimicrobial susceptibility: a multispecies review with a focus on antibiotic treatment of mastitis in dairy cattle. *J Mammary Gland Biol Neoplasia* 16:383–407. <https://doi.org/10.1007/s10911-011-9235-z>
- Barreto G, Senturk B, Colombo L, Brück O, Neidenbach P et al (2020) Lumican is upregulated in osteoarthritis and contributes to TLR4-induced pro-inflammatory activation of cartilage degradation and macrophage polarization. *Osteoarthr Cartil* 28(1):92–101. <https://doi.org/10.1016/j.joca.2019.10.011>

- Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M et al (2013) NCBI GEO: Archive for functional genomics data sets - update. *Nucleic Acids Res* 41:991–995. <https://doi.org/10.1093/nar/gks1193>
- Becuwe P, Ennen M, Klotz R, Barbieux C, Grandemange S (2014) Manganese superoxide dismutase in breast cancer: from molecular mechanisms of gene regulation to biological and clinical significance. *Free Radic Biol Med* 77:139–151. <https://doi.org/10.1016/j.freeradbiomed.2014.08.026>
- Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A et al (2009) ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* 25:1091–1093
- Blohm DH, Guiseppi-Elie A (2001) New developments in microarray technology. *Curr Opin Biotechnol* 12:41–47. [https://doi.org/10.1016/S0958-1669\(00\)00175-0](https://doi.org/10.1016/S0958-1669(00)00175-0)
- Buitenhuys B, Røntved CM, Edwards SM, Ingvarsen KL, Sørensen P (2011) In depth analysis of genes and pathways of the mammary gland involved in the pathogenesis of bovine *Escherichia coli*-mastitis. *BMC Genom*. <https://doi.org/10.1186/1471-2164-12-130>
- Chen X, Cheng Z, Zhang S, Werling D, Wathes DC (2015) Combining genome wide association studies and differential gene expression data analyses identifies candidate genes affecting mastitis caused by two different pathogens in the dairy cow. *Open J Anim Sci* 5:358–393
- Chen W, Lv X, Wang Y, Zhang X, Wang S, Hussain Z et al (2020) Transcriptional profiles of long non-coding RNA and mRNA in sheep mammary gland during lactation period. *Front Genet* 11:946
- Chin CH, Chen SH, Wu HH, Ho CW, Ko MT, Lin CY (2014) Cytohubba: identifying hub objects and sub-networks from complex interactome. *BMC Syst Biol* 8:1–7. <https://doi.org/10.1186/1752-0509-8-S4-S11>
- Choi B-H, Rawat V, Högström J, Burns PA, Conger KO et al (2022) Lineage-specific silencing of PSAT1 induces serine autotrophy and sensitivity to dietary serine starvation in luminal breast tumors. *Cell Rep* 38(3):110278
- Dai W, Zou Y, White RR, Liu J, Liu H (2018) Transcriptomic profiles of the bovine mammary gland during lactation and the dry period. *Funct Integr Genomics* 18:125–140
- D’Aniello C, Patriarca EJ, Phang JM, Minchiotti G (2020) Proline metabolism in tumor growth and metastatic progression. *Front Oncol*. <https://doi.org/10.3389/fonc.2020.00776>
- Davis S, Meltzer PS (2007) GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. *Bioinformatics* 23:1846–1847
- Fu R, Han CF, Ni T, Di L, Liu LJ, Lv WC et al (2019) A ZEB1/p53 signaling axis in stromal fibroblasts promotes mammary epithelial tumours. *Nat Commun*. <https://doi.org/10.1038/s41467-019-11278-7>
- Giatragana EM, Berdiaki A, Tsatsakis A, Tzanakakis GN, Nikitovic D (2021) Lumican in Carcinogenesis—Revisited. *Biomolecules* 11(9):1319
- Gopinath P, Natarajan A, Sathyanarayanan A, Sridev V, Gopal G (2022) The multifaceted role of Matrix-cellular Proteins in health and cancer, as biomarkers and therapeutic targets. *Gene* 812:146137
- Gorji AE, Roudbari Z, Sadeghi B, Sadeghi J (2019) Transcriptomic analysis on the promoter regions discover gene networks involving mastitis in cattle. *Microb Pathog* 137:103801
- Günther J, Petzl W, Bauer I, Ponsuksili S, Zerbe H, Schuberth HJ et al (2017) Differentiating *Staphylococcus aureus* from *Escherichia coli* mastitis: *S. aureus* triggers unbalanced immune-dampening and host cell invasion immediately after udder infection. *Sci Rep* 7:1–14
- Halasa T, Huijps K, Østerås O, Hogeveen H (2007) Economic effects of bovine mastitis and mastitis management. *A Review Vet Q* 29:18–31. <https://doi.org/10.1080/01652176.2007.9695224>
- Hamed AE, Mohammed A, Kamel E (2020) Incidence and economics of clinical mastitis of Holstein friesian dairy cows under Egyptian condition. *BVMJ* 39:119–124
- Holmes RS (2017) Comparative and evolutionary studies of *ALDH18A1* genes and proteins. *Chem Biol Interact* 276:2–8. <https://doi.org/10.1016/j.cbi.2016.12.012>
- Jensen K, Günther J, Talbot R, Ponsuksili S, Zerbe H, Schuberth HJ et al (2013) *Escherichia coli*- and *Staphylococcus aureus*-induced mastitis differentially modulate transcriptional responses in neighbouring uninfected bovine mammary gland quarters. *BMC Genomics* 14:1–19
- Kocher F, Tymozzuk P, Amann A, Sprung S, Salcher S et al (2021) Deregulated glutamate to pro-collagen conversion is associated with adverse outcome in lung cancer and may be targeted by renin-angiotensin-aldosterone system (RAS) inhibition. *Lung Cancer* 159:84–95. <https://doi.org/10.1016/j.lungcan.2021.06.020>
- Li P, Wu G (2018) Roles of dietary glycine, proline, and hydroxyproline in collagen synthesis and animal growth. *Amino Acids* 50(1):29–38

- Mitterhuemer S, Petzl W, Krebs S, Mehne D, Klanner A, Wolf E et al (2010) *Escherichia coli* infection induces distinct local and systemic transcriptome responses in the mammary gland. BMC Genomics. <https://doi.org/10.1186/1471-2164-11-138>
- Naeem A, Zhong K, Moisés SJ, Drackley JK, Moyes KM, Loor JJ (2012) Bioinformatics analysis of microRNA and putative target genes in bovine mammary tissue infected with *Streptococcus uberis*. J Dairy Sci 95:6397–6408
- Palii AP, Kovalchuk YO, Boyko YA, Bondaruk YV, Diachuk PV, Duka TM et al (2020) Impact of various milking equipment on incidence of mastitis in dairy herd. Vinnytsia Nat Med Uni 10:160–165
- Phuektes P, Mansell PD, Browning GF (2001) Multiplex polymerase chain reaction assay for simultaneous detection of *Staphylococcus aureus* and Streptococcal causes of bovine mastitis. J Dairy Sci 84:1140–1148. [https://doi.org/10.3168/jds.S0022-0302\(01\)74574-2](https://doi.org/10.3168/jds.S0022-0302(01)74574-2)
- Schukken YH, Hertl J, Bar D, Bennett GJ, González RN, Rauch B et al (2009) Effects of repeated gram-positive and gram-negative clinical mastitis episodes on milk yield loss in Holstein dairy cows. J Dairy Sci 92:3091–3105. <https://doi.org/10.3168/jds.2008-1557>
- Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4:44
- Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J et al (2019) STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res 47:D607–D613
- Tsutsui S, Wakasa H, Tsugami Y, Suzuki T, Nishimura, (2020) Distinct expression patterns of fibrillar collagen Types I, III, and V in association with mammary gland remodeling during pregnancy, lactation and weaning. J Mammary Gland Biol Neoplasia 25(3):219–232
- Wang X, Fan Y, Han Z, Gong Z, Peng Y, Mao Y et al (2021) Integrative analysis of miRNA and mRNA expression profiles in mammary glands of holstein cows artificially infected with *Staphylococcus aureus*. Pathogens 10:506
- Wicik Z, Gajewska M, Majewska A, Walkiewicz D, Osińska E et al (2016) Characterization of micro-RNA profile in mammary tissue of dairy and beef breed heifers. J Anim Breed Genet 133(1):31–42. <https://doi.org/10.1111/jbg.12172>
- Xu CG, Yang YB, Zhou YH, Hao MQ, Ren VZ, Wang XT et al (2017) Comparative proteomic analysis provides insight into the key proteins as possible targets involved in aspirin inhibiting biofilm formation of *Staphylococcus xylosum*. Front Pharmacol 8:543
- Younis S, Javed Q, Blumenberg M (2016) Meta-analysis of transcriptional responses to mastitis-causing *Escherichia coli*. PLoS ONE 11(3):1–18
- Zhang J, Xu C, Huo D, Hu Q, Peng Q (2017) Comparative study of the gut microbiome potentially related to milk protein in Murrah buffaloes (*Bubalus bubalis*) and Chinese Holstein cattle. Sci Rep 7:1–11
- Zhao X, Lacasse P (2008) Mammary tissue damage during bovine mastitis: causes and control. J Anim Sci 86:57–65. <https://doi.org/10.2527/jas.2007-0302>
- Zuhra K, Augsburg F, Majtan T, Szabo C (2020) Cystathionine- $\beta$ -synthase: molecular regulation and pharmacological inhibition. Biomolecules 10(5):697