



Integrative Systems Biology Analysis Elucidates Mastitis Disease Underlying Functional Modules in Dairy Cattle

Nooshin Ghahramani^{1*}, Jalil Shodja¹, Seyed Abbas Rafat¹, Bahman Panahi² and Karim Hasanpur¹

¹ Department of Animal Science, Faculty of Agriculture, University of Tabriz, Tabriz, Iran, ² Department of Genomics, Branch for Northwest & West Region, Agricultural Biotechnology Research Institute of Iran (ABRII), Agricultural Research, Education and Extension Organization (AREEO), Tabriz, Iran

OPEN ACCESS

Edited by:

Aline Silva Mello Cesar,
University of São Paulo, Brazil

Reviewed by:

Claas Heuer,
ST Genetics, United States
Andressa Oliveira De Lima,
University of Washington,
United States

*Correspondence:

Nooshin Ghahramani
gh.nooshin69@gmail.com

Specialty section:

This article was submitted to
Livestock Genomics,
a section of the journal
Frontiers in Genetics

Received: 20 May 2021

Accepted: 30 August 2021

Published: 08 October 2021

Citation:

Ghahramani N, Shodja J, Rafat SA,
Panahi B and Hasanpur K (2021)
Integrative Systems Biology Analysis
Elucidates Mastitis Disease Underlying
Functional Modules in Dairy Cattle.
Front. Genet. 12:712306.
doi: 10.3389/fgene.2021.712306

Background: Mastitis is the most prevalent disease in dairy cattle and one of the most significant bovine pathologies affecting milk production, animal health, and reproduction. In addition, mastitis is the most common, expensive, and contagious infection in the dairy industry.

Methods: A meta-analysis of microarray and RNA-seq data was conducted to identify candidate genes and functional modules associated with mastitis disease. The results were then applied to systems biology analysis *via* weighted gene coexpression network analysis (WGCNA), Gene Ontology, enrichment analysis for the Kyoto Encyclopedia of Genes and Genomes (KEGG), and modeling using machine-learning algorithms.

Results: Microarray and RNA-seq datasets were generated for 2,089 and 2,794 meta-genes, respectively. Between microarray and RNA-seq datasets, a total of 360 meta-genes were found that were significantly enriched as “peroxisome,” “NOD-like receptor signaling pathway,” “IL-17 signaling pathway,” and “TNF signaling pathway” KEGG pathways. The turquoise module ($n = 214$ genes) and the brown module ($n = 57$ genes) were identified as critical functional modules associated with mastitis through WGCNA. *PRDX5*, *RAB5C*, *ACTN4*, *SLC25A16*, *MAPK6*, *CD53*, *NCKAP1L*, *ARHGEF2*, *COL9A1*, and *PTPRC* genes were detected as hub genes in identified functional modules. Finally, using attribute weighting and machine-learning methods, hub genes that are sufficiently informative in *Escherichia coli* mastitis were used to optimize predictive models. The constructed model proposed the optimal approach for the meta-genes and validated several high-ranked genes as biomarkers for *E. coli* mastitis using the decision tree (DT) method.

Conclusion: The candidate genes and pathways proposed in this study may shed new light on the underlying molecular mechanisms of mastitis disease and suggest new approaches for diagnosing and treating *E. coli* mastitis in dairy cattle.

Keywords: attribute weighting, *E. coli*, machine learning, mastitis, meta-analysis, WGCNA

BACKGROUND

Mastitis is a severe disease characterized as an inflammatory condition affecting the mammary glands (Gelasakis et al., 2015). *Escherichia coli*, *Staphylococcus aureus*, and *Streptococcus uberis* are the three major bacterial pathogens associated with mastitis disease (Vasudevan et al., 2003), with *E. coli* causing severe inflammation in dairy cattle (Vangroenweghe et al., 2005). The focus of current research has shifted to elucidating the underlying mechanisms and developing effective treatment strategies for mastitis disease (Takeshima et al., 2008; Compton et al., 2009). *E. coli* typically infects the mammary glands during the dry period, and inflammation progresses during the early stages of lactation (Burvenich et al., 2003). Recent research indicates that *E. coli*'s pathogenicity is entirely dependent on a protein called FecA (Blum et al., 2018).

Recent advancements in high-throughput transcriptome profiling technologies, such as microarray and RNA sequencing (RNA-seq), have enabled opportunities for precision critical care medicine to understand better the molecular mechanisms underlying diverse biological functions (Bansal and Di Bernardo, 2007; Farhadian et al., 2020; Panahi et al., 2020). On the other hand, identifying disease biomarkers can aid breeders in optimizing their genetic programs for dairy animals (Kulkarni and Kaliwal, 2013; Duarte et al., 2015; Lai et al., 2017). Previous research identified *TNF*- and *SAA3* (Swanson et al., 2009), *STAT3*, *MAPK14*, *TNF* (Gorji et al., 2019), *IL8RB*, *CXCL6*, *MMP9* (Li et al., 2019), *IRF9*, *CCL* (Buitenhuis et al., 2011), *S100A12*, *MT2A*, *SOD2* (Mitterhuemer et al., 2010), *CXCL8*, *CXCL2*, *S100A9* (Sharifi et al., 2018), *PSMA6*, *HCK*, and *STAT1* (Bakhtiarizadeh et al., 2020) as potential biomarkers for mastitis disease.

Meta-analysis is a systematic and quantitative study methodology used to evaluate prior research and reach a conclusion (Haidich, 2010). On the other hand, independent studies have limitations in sample size, statistical power, and the reliability of the results (Panahi and Hejazi, 2021). Meta-analysis has demonstrated that combining *p* values resolves several issues (Rhodes et al., 2002; Tseng et al., 2012; Panahi et al., 2019a). When combining *p* values using Fisher's technique, the null hypothesis is that the actual effect is zero in each of the combined datasets (Evangelou and Ioannidis, 2013), suggesting that the techniques should be sensitive even when only a subset of the combined datasets has an impact magnitude more significant than zero. Fisher's approach outperformed other methods for establishing associations. In addition, the *p* value combination method shows considerable promise for identifying novel markers or differentially expressed genes (DEGs) (Evangelou and Ioannidis, 2013). Moreover, connectivity analysis of known meta-genes has been presented as a promising approach for disentangling the complicated method (Panahi et al., 2020).

Weighted gene coexpression network analysis (WGCNA) has been proposed as a versatile tool for gene coexpression analysis, which provide valuable information about gene connectivity based on gene expression levels (Ebrahimie et al., 2014; Farhadian et al., 2021). A combination of machine-learning algorithms and microarray meta-analysis was used to identify mastitis genes in dairy cattle (Sharifi et al., 2018). However, they did not include

RNA-seq data in their analysis and instead focused on the expression patterns of meta-genes.

The present study is the first that the authors are aware of that integrates meta-analysis of microarray and RNA-seq datasets, connectivity analysis, and model optimization in mastitis disease. Thus, in this integrative study, we identified master genes associated with mastitis disease using a combination of meta-analysis, WGCNA, and machine-learning algorithms.

MATERIALS AND METHODS

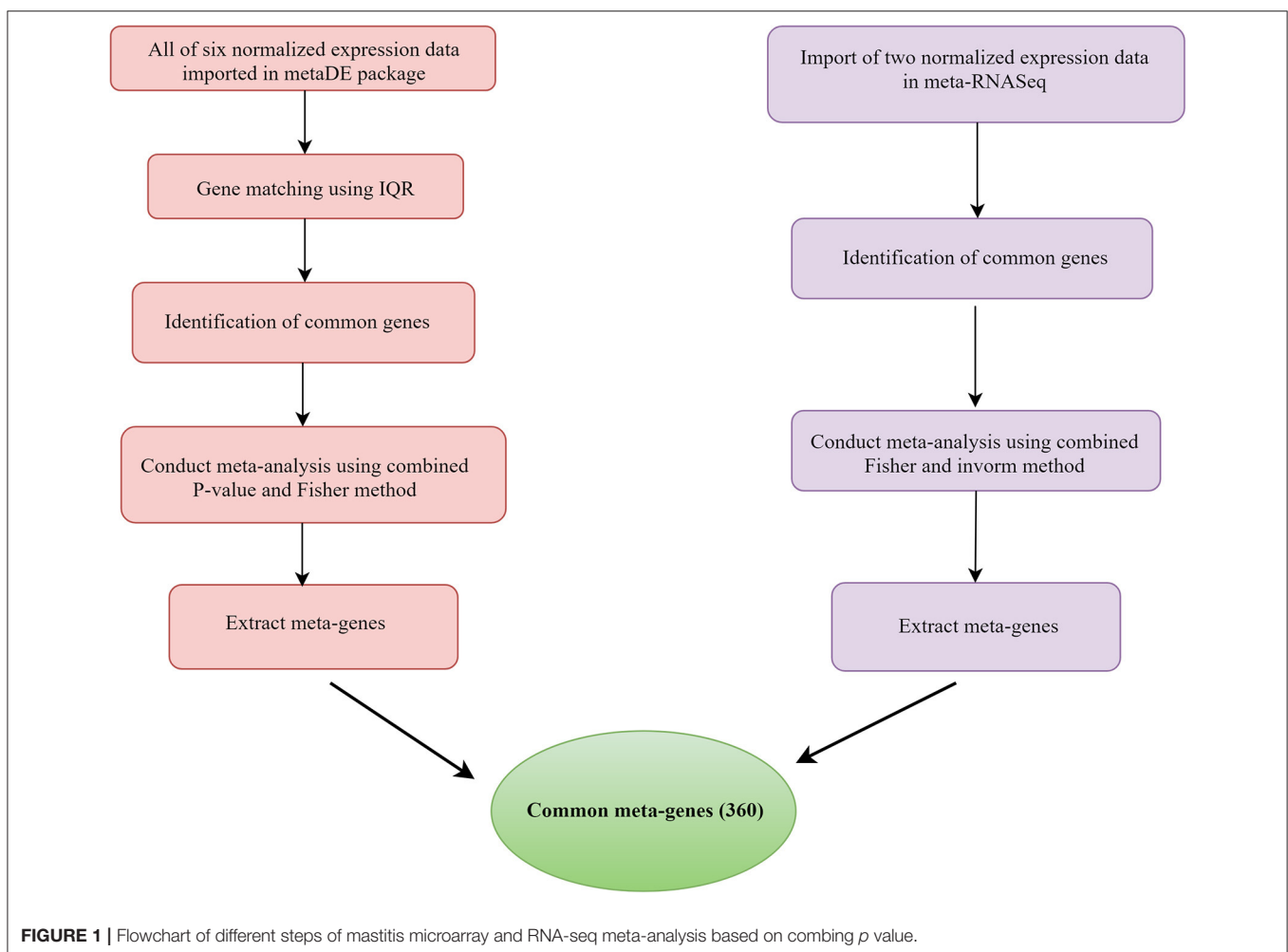
Data Collection

The National Center for Biotechnology Information's Gene Expression Omnibus (GEO) repository (<https://www.ncbi.nlm.nih.gov/gds/>) was explored for datasets related to dairy cattle mastitis. This database was searched for RNA-seq and microarray studies using the keywords "*Bos taurus*," "mastitis," and "*Escherichia coli*." For this research, six microarrays and two RNA-seq datasets were chosen. **Table 1** lists the platform, accession number, species, and references for each dataset. All healthy and mastitis animal samples were from the *Bos taurus* species, which have a high sensitivity to *E. coli*. Fifteen healthy German Holstein Frisian cows in midlactation (3–6 months postpartum) were included in dataset GSE15025. The animals were inoculated with *E. coli* in one-quarter and died after 6 h ($n = 5$) or 24 h ($n = 5$) in two distinct infection scenarios. Five heifers were used as controls; they were not given any medication and died after 24 h. At 4 to 6 weeks following parturition, 16 healthy primiparous Danish Holstein-Friesian cows were tested with *E. coli* for the GSE24217 dataset. The overall udder health of 24 dairy cows was assessed before the disease challenge. Control quarters were selected based on bacteriological tests performed before *E. coli* inoculation and the quarter foremilk SCC at 24 and 192 h. From the total German Holstein population, 11 heifers at day 42 postpartum, either with high or low sensitivity to mastitis, were chosen for the GSE24560 dataset. Heat-inactivated *E. coli* plus *S. aureus* was used to challenge the cells, as a control. The cells were collected after 1, 6, and 24 h, and mRNA expression was compared. Four first-lactation Holstein cows in the fourth month of lactation were also experimentally inoculated with the mastitis-causing *E. coli* pathogen strain 1303 for the GSE25413 dataset. The transcriptomes of the treated and untreated cells were examined at 1, 3, 6, and 24 h. In the GSE32186 dataset, four first-lactation Holstein cows were given primary MEC (pbMEC) cultures for 6 h, and some cultures were stimulated. *E. coli* particles were collected from the udders of three healthy, pregnant (day 130 of gestation) cows in the middle of their first lactation 12 or 42 h later. Six Holstein Friesian cows were challenged with *E. coli* mastitis for the GSE50685 investigations. Every 6 h after infection, blood and milk samples were taken. At successive milkings, the treatment was repeated (12, 24, and 36 h postchallenge). At 24 h ($n = 3$) and 48 h ($n = 3$) following infection, the cows were sacrificed for tissue collection. GSE75379 and GSE159286 were two datasets related to RNA-seq. Sixteen healthy primiparous Holstein cows at 4–6 weeks of lactation were included in the GSE75379 dataset. Biopsy specimens of healthy and diseased udder tissue were taken at

TABLE 1 | Selected microarray and RNA-seq datasets for systems biology analysis of mastitis disease.

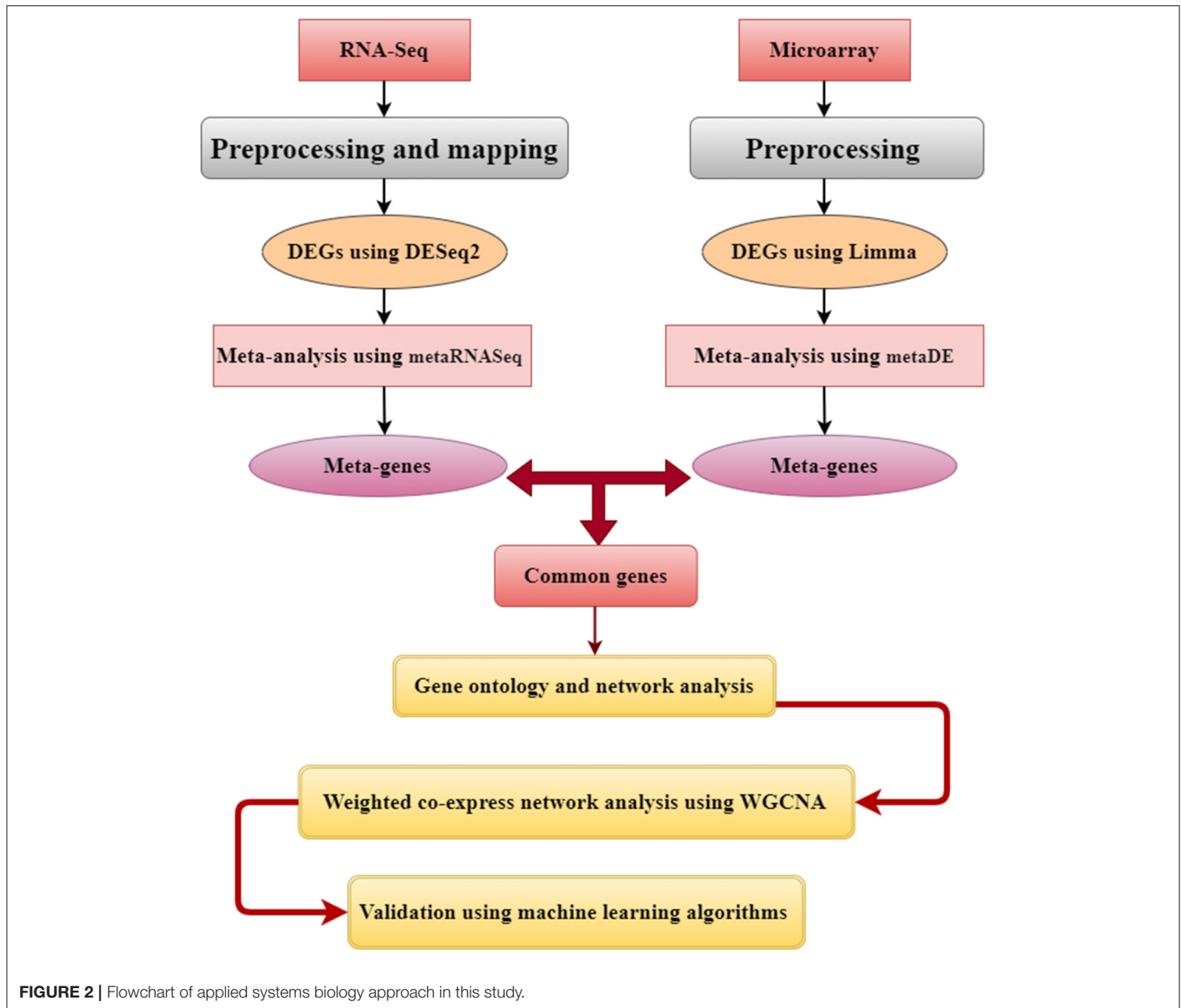
Accession no.	Species	Bacteria	Platform	Samples* (C:T)	References
Microarray datasets					
GSE15025	<i>Bos taurus</i>	<i>Escherichia coli</i>	Affymetrix	15:15	Mitterhuemer et al., 2010
GSE24217	<i>B. taurus</i>	<i>E. coli</i>	Affymetrix	23:26	Buitenhuis et al., 2011
GSE24560	<i>B. taurus</i>	<i>E. coli</i>	Affymetrix	27:31	Brand et al., 2011
GSE25413	<i>B. taurus</i>	<i>E. coli</i>	Affymetrix	6:24	Günther et al., 2011
GSE32186	<i>B. taurus</i>	<i>E. coli</i>	Affymetrix	12:12	Günther et al., 2012
GSE50685	<i>B. taurus</i>	<i>E. coli</i>	Affymetrix	5:15	Sipka et al., 2014
RNA-seq datasets					
GSE75379	<i>B. taurus</i>	<i>E. coli</i>	Illumina	6:12	Moyes et al., 2016
GSE159286	<i>B. taurus</i>	<i>E. coli</i>	Illumina	53:52	Cebon et al., 2020

*Number of healthy and infected samples.



$T = 24$ h and $T = 192$ h after infection. In total, 12 heifers were intramuscularly vaccinated with heat-killed *E. coli* for the GSE159286 investigations. Half of the heifers (IM group, $n = 6$) received a booster injection 2 months later. Others (IMM group, $n = 6$) received 50 g of protein concentrate produced from *E.*

coli in two quarters. Cows were then challenged with an *E. coli* P4 bacterial suspension infusion at 49 days in milk inside one healthy quarter (10^3 bacteria). Before the trial, blood was taken 7 days after the first and second injections (immunization phase) and then at 0, 12, and 40 h following infection (challenge phase).

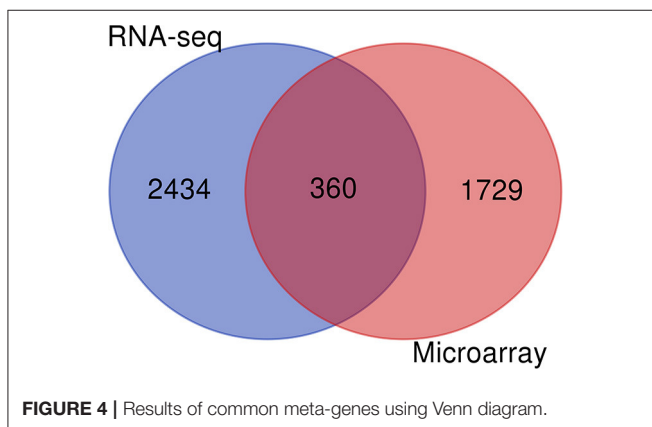
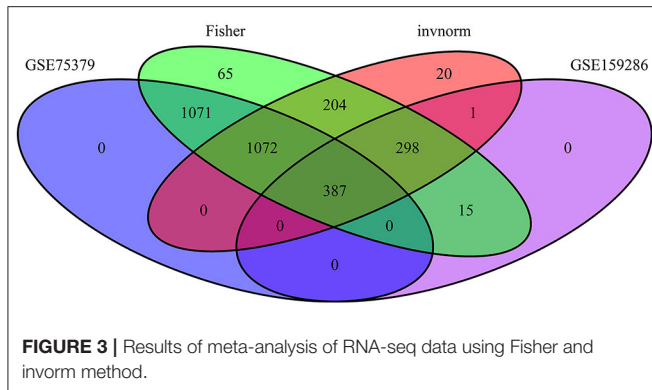


Preprocessing and Analysis of Microarray Datasets

The GEO database was used to collect all microarray expression raw data and associated annotations for each study, and microarray data were preprocessed to obtain reliable findings. Nonbiological data variances were then removed, and appropriate scales were used for further analysis (Bolstad et al., 2003). The quantile normalization method and batch effects reduction were used to conduct effective gene expression analysis and eliminate variability between studies. The Limma software (2.16.0) (Smyth et al., 2005) was used to calculate DEGs among each control and treatment group after preprocessing the raw data. DEGs were deemed significant when the false discovery rate (FDR) using the Benjamini–Hochberg method was $p < 0.05$ and the logarithm of fold change $> \pm 0.5$ (Benjamini and Hochberg, 1995).

Preprocessing and Analysis of Individual RNA-seq Datasets

The data generated by RNA-seq can be skewed due to biases introduced during library preparation, polymerase chain reaction, and sequencing. The technique of trimmed mean of m -values was used to eliminate the effect of known nonbiological features on the RNA-seq data (Robinson and Oshlack, 2010). Each sample was inspected for quality using the FastQC tool version 0.11.5 (Trapnell et al., 2012), and low-quality reads were trimmed using the Trimmomatic (v 0.32) software (Goldman et al., 2006). Bowtie2 (2.2.4) software was used to index reference genomes, and clean reads were then mapped to the *B. taurus* reference genome (ARS-UCD1.2 version) employing Tophat2 (2.0.10) software (Kim et al., 2013; Love et al., 2014) with default settings. The sample mapping rates are listed in **Supplementary Material 1, Table 1**. The htseq-count package



(2.7.3) (Anders et al., 2013) was used to calculate the expression count matrix. The Bioconductor DESeq2 software (1.10.1) was used to determine the differential gene expressions in each research (Love et al., 2014). In terms of normalization and batch effect correction, the methods outlined in the studies (Farhadian et al., 2021; Panahi and Hejazi, 2021) were followed. In summary, genes with a CV <10% were initially removed. The group was then used as a covariate, using DESeq2 library size, size factor normalization factors. The variance-stabilizing transformations (VSTs) function was used to reduce sample disparities. The VST function does not typically remove variations related to the batch or other variables. As a result, the “removeBatchEffect” function was used to remove batch variations. The blind = false option was selected as re-estimation of the dispersion values was not required. This process leveled the library size and other normalization variables. Each study’s samples were normalized jointly, implying that each dataset was normalized individually (Love et al., 2014).

Meta-Analysis of Microarray and RNA-seq Datasets

In microarray studies, the MetaDE package (1.0.5) was used to identify meta-genes (Wang et al., 2012). The meta-analysis included the following steps: after quantile normalization, labeling samples as “Infected” or “Healthy”; the “Gene Symbol” was matched to multiple probe IDs using the interquartile range

for probe selection (Wang et al., 2012). Merging genes is an approach used to determine which genes should be studied further (Wang et al., 2021). Because the number of genes in the research varied, the multiple gene expression datasets may not have been adequately matched by genes. In this study, the direct merging method was used to obtain common genes across different investigations. The Fisher technique (Marot et al., 2020) was used to identify meta-genes in RNA-seq data using the metaRNASeq software (1.0.5). Initially, the DEGs for each study were defined using the DESeq2 package (1.30.1), and the corresponding *p* value was extracted. Then, the fishcomb function included in the metaRNASeq package was used to combine the *p* values. For downstream analysis, the genes shared between meta-analyses of microarray and RNA-seq data were extracted (Figure 1).

Weighted Gene Coexpression Network Analysis

Common genes were used to construct coexpression networks using the WGCNA Bioconductor R package (version 3.5.1) to understand the correlation patterns among genes and identify significant modules associated with mastitis disease (Langfelder and Horvath, 2008). Initially, networks were constructed using Pearson correlations across all common genes (Botía et al., 2017). A soft threshold was used to evaluate the correlation and noise-filtering power to fulfill the scale-free topology requirement. The weighted gene coexpression network design promotes strong correlations at the cost of low correlations by increasing the absolute magnitude of the correlation to a soft threshold (Langfelder and Horvath, 2008). The topological overlap matrix (TOM) and its corresponding dissimilarity matrix (1-TOM) were used to visualize the network, which resulted in a network diagram for module detection. Module eigengenes and module membership were used to identify the hub genes for each significant coexpressed module (Langfelder and Horvath, 2008). The following parameters were used to construct the modules: cut height of 0.975, minimum module size of 30 genes, “hybrid” method, and deepSplit = 2.

Functional Enrichment Analysis

The STRING (Szklarczyk et al., 2015) database was used to conduct enrichment analysis on the Kyoto Encyclopedia of Genes and Genomes (KEGG) and Gene Ontology (GO) (Dennis et al., 2003). The FDR (<0.05) correction was used to determine the statistical significance of GO and KEGG terms.

Protein–Protein Interaction Networks of Common Genes

Gene network analysis of protein–protein interaction between common genes was performed using Cytoscape software to visualize gene networks and identify hub genes. Hub genes are defined as those with the highest degree of connectivity and those with a greater biological significance than other gene members (Shannon et al., 2003).

TABLE 2 | Significant common meta-genes in mastitis disease.

Gene symbol	Full name	Chromosome	Compartment
<i>ELMO2</i>	Engulfment and cell motility 2	20	Cytosol
<i>ORM1</i>	Orosomucoid 1 (alpha-1-acid glycoprotein)	9	Extracellular
<i>ABCB7</i>	ATP-binding cassette subfamily B member 7	X	Mitochondrion
<i>LRRC41</i>	Leucine-rich repeat containing 41	1	Nucleus
<i>CXCL3</i>	C-X-C motif chemokine ligand 3	4	Extracellular
<i>SOD2</i>	Superoxide dismutase 2	6	Extracellular
<i>ZFYVE1</i>	Zinc finger FYVE-type containing 1	14	Mitochondrion
<i>SORL1</i>	Sortilin-related receptor 1	11	Plasma membrane
<i>PTCD3</i>	Pentatricopeptide repeat domain 3	2	Mitochondrion
<i>RPS6KA5</i>	Ribosomal protein S6 kinase A5	14	Nucleus
<i>LHFPL2</i>	LHFPL tetraspan subfamily member 2	5	Plasma membrane
<i>TRIQK</i>	Triple QxxK/R motif containing	8	Endoplasmic reticulum
<i>MAOA</i>	Monoamine oxidase A	X	Mitochondrion
<i>CORO2A</i>	Coronin 2A	9	Cytosol
<i>TPM3</i>	Tropomyosin 3	1	Extracellular
<i>PTPRC</i>	Protein tyrosine phosphatase receptor type C	1	Plasma membrane
<i>LOC407171</i>	Fc gamma 2 receptor	18	Extracellular

Supervised Machine-Learning Models

The common meta-genes identified were utilized to select features using 10 different weighting algorithms, including information gain, information gain ratio, χ^2 , deviation, rule, support vector machine, Gini index, uncertainty, relief, and PCA to validate the hub genes' efficacy in distinguishing different genes involved in mastitis disease (Farhadian et al., 2018a, 2021). The Rapid Miner software (Rapid Miner 5.0.001, Dortmund, Germany) was used for attribute weighting (Ebrahimi et al., 2011; Farhadian et al., 2018a; Panahi et al., 2019a; Nami et al., 2021). The primary objective of attribute weighting algorithms was to extract a subset of input features (genes) by excluding those that contained little or no information (Panahi et al., 2019b). The decision trees (DTs) were constructed using features with weighting values greater than 0.5. The DTs were constructed using the following methods: information gain, information gain ratio, Gini index, and accuracy criteria. **Figure 2** depicts the flowchart of an analytical strategy for microarray and RNA-seq.

RESULTS

Meta-Analysis

We conducted a meta-analysis of DEGs using data from microarray and RNA-seq experiments. Six raw microarray datasets containing 211 samples and two RNA-seq datasets containing 123 independent dairy cattle experiments were chosen separately for the meta-analysis. Finally, a total of 2,089 and 2,794 meta-genes in response to *E. coli* mastitis in microarray and RNA-seq data, respectively, were observed using the Fisher method in the metaDE and metaRNASeq packages. **Supplementary Material 2** and **Figure 3** contain the results of the meta-analysis of RNA-seq data.

Identification of Common Genes by Meta-Gene Comparison

A total of 360 genes were identified as common meta-genes in meta-analysis of microarray and RNA-seq data (**Figure 4**). **Table 2** and **Supplementary Material 3** contain additional information about the significant common meta-genes.

Functional Enrichment Analysis of Common Genes

The STRING database was used to conduct GO analyses on 360 common meta-genes to ascertain their biological process (BP), molecular function (MF), and cellular component (CC) roles in mastitis disease. The results found 170, 33, and 36 GO terms for BPs, MFs, and CCs, respectively. The terms "cellular process," "response to stimulus," "biological regulation," "regulation of a biological process," and "regulation of a cellular process" were used to denote the most critical process in the BP category. "Binding," "ion binding," "actin binding," "cation binding," and "metal ion binding" were all significantly overrepresented in the MF category. In terms of CC, the terms "intracellular," "cell," "cytoplasm," "intracellular organelle," and "organelle" were significantly enriched. Additional information is available in **Table 3** and **Supplementary Material 4**.

This analysis identified a total of nine significant KEGG pathways. In addition, the results indicated that the "peroxisome," "NOD-like receptor signaling pathway," "IL-17 signaling pathway," and "TNF signaling pathway" were significantly overrepresented. **Table 4** contains additional information about KEGG pathways.

Cytoscape demonstrates the involvement of DEGs in protein-protein interaction. **Figure 5** illustrates the gene network

TABLE 3 | Significant GO terms of common genes.

Term ID	Description	GO terms	p value
GO:0009987	Cellular process	BP	1.52E-10
GO:0050896	Response to stimulus	BP	1.96E-10
GO:0065007	Biological regulation	BP	8.53E-10
GO:0050789	regulation of BP	BP	3.66E-09
GO:0050794	Regulation of cellular process	BP	3.66E-09
GO:0051716	Cellular response to stimulus	BP	6.09E-08
GO:0019222	Regulation of metabolic process	BP	2.26E-07
GO:0051171	Regulation of nitrogen compound metabolic process	BP	2.26E-07
GO:0080090	Regulation of primary metabolic process	BP	2.26E-07
GO:0031323	Regulation of cellular metabolic process	BP	2.63E-07
GO:0005488	Binding	MF	3.90E-10
GO:0043167	Ion binding	MF	2.55E-06
GO:0005515	Protein binding	MF	1.30E-05
GO:0003779	Actin binding	MF	0.00011
GO:0043169	Cation binding	MF	0.00024
GO:0046872	Metal ion binding	MF	0.00029
GO:0008092	Cytoskeletal protein binding	MF	0.00034
GO:1901363	Heterocyclic compound binding	MF	0.0014
GO:0097159	Organic cyclic compound binding	MF	0.0019
GO:0036094	Small molecule binding	MF	0.002
GO:0005622	Intracellular	CC	7.27E-13
GO:0005623	Cell	CC	8.74E-13
GO:0005737	Cytoplasm	CC	8.74E-13
GO:0043229	Intracellular organelle	CC	1.39E-09
GO:0043226	Organelle	CC	1.74E-09
GO:0043227	Membrane-bound organelle	CC	5.37E-09
GO:0043231	Intracellular membrane-bound organelle	CC	7.48E-09
GO:0005829	Cytosol	CC	2.03E-07
GO:0070013	Intracellular organelle lumen	CC	1.28E-05
GO:0005634	Nucleus	CC	3.30E-05

Only the significantly enriched ($p < 0.05$) GO terms are presented.

TABLE 4 | The significant KEGG metabolic pathways associated with the common genes.

Pathway name	p value	Total genes in pathway	Strength
Peroxisome	0.0012	10	0.84
NOD-like receptor signaling pathway	0.0058	12	0.61
IL-17 signaling pathway	0.0058	9	0.76
TNF signaling pathway	0.008	9	0.68
Salmonella infection	0.008	8	0.76
Viral carcinogenesis	0.008	13	0.55
Human papillomavirus infection	0.0089	16	0.46
Necroptosis	0.0194	10	0.56
Autophagy—animal	0.0216	9	0.59

TNF, tumor necrosis factor.

visualization of common meta-genes. The top genes were STAT1, RTPRC, SOD2, and VCP (Supplementary Material 5).

Weighted Gene Coexpression Network Construction

A WGCNA was performed to identify genes with a high correlation and classified the common genes into four modules. The turquoise module ($n = 214$ genes) and the brown module ($n = 57$ genes) were identified as critical functional modules associated with mastitis through WGCNA analysis (Figure 6A). The remaining modules, such as the blue module ($n = 84$ genes) plus the gray module ($n = 5$ genes), were not notable. Figure 6B illustrates the hierarchical clustering of common genes.

The correlation coefficient and p value for the significant modules in the mastitis and healthy groups were $r = 0.28$, $p = 0.002$, and $r = 0.36$, $p = 4e - 05$, respectively, for the turquoise and brown modules (Figure 7).

Turquoise had a negative correlation with mastitis disease, whereas brown had a positive correlation. Table 5 lists the top five hub genes in brown and turquoise modules. Supplementary Material 6 contains a list of the more significant modules identified.

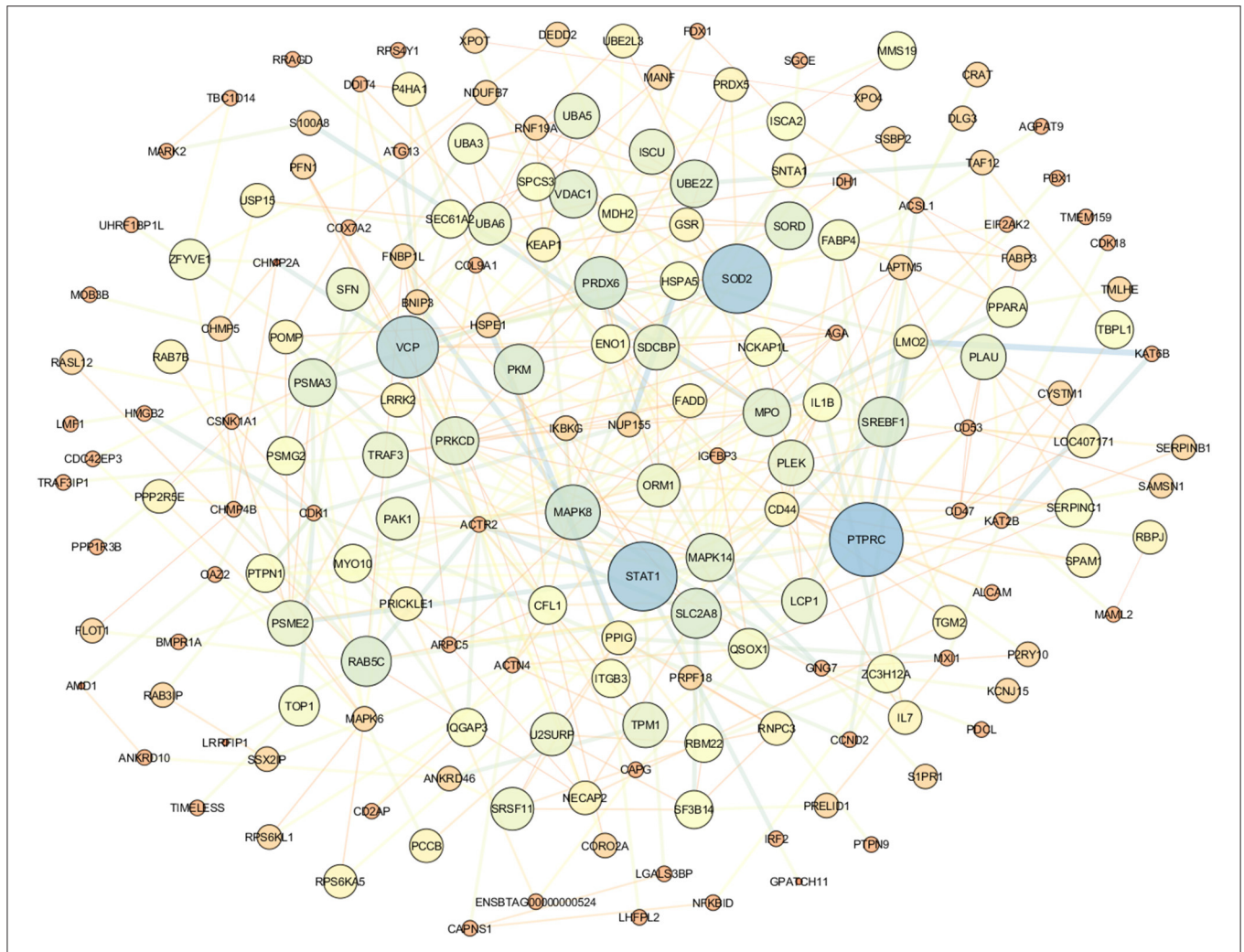


FIGURE 5 | Protein-protein interaction network for the common genes using Cytoscape.

“Peroxisome,” “viral carcinogenesis,” and “arginine and proline metabolism” were determined as the most significantly enriched pathways based on the enriched functional analysis in these modules that were potentially associated with mastitis development. These modules enriched for genes involved in “negative regulation of peptidyl-serine phosphorylation,” “response to stimulus,” “cell process regulation,” “protein hydroxylation,” “actin filament-based process,” and “cellular process.” These modules carried out critical MFs such as “actin binding,” “binding,” “transcription factor binding,” and “peroxidase activity.” “Intracellular,” “organelle,” “cytoplasm,” “cell,” “cortical actin cytoskeleton,” and “microvillus” were identified as CCs.

Attribute Weighting

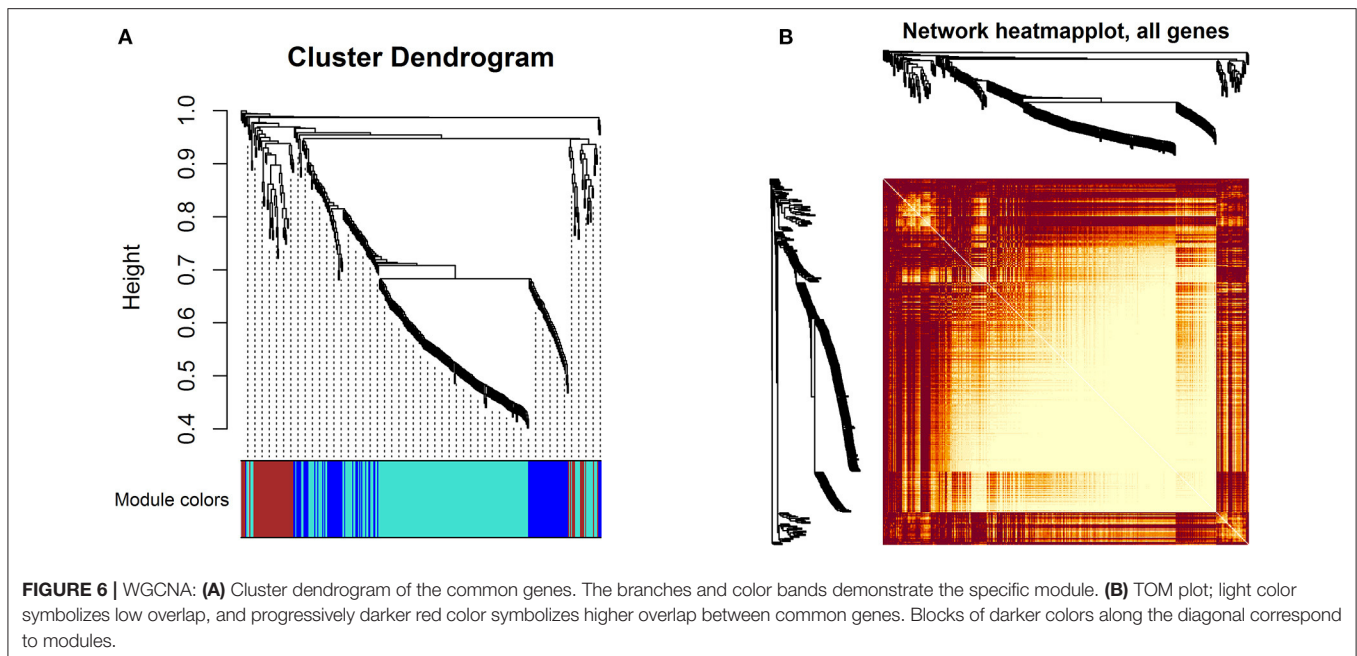
The data-cleaning process was used to eliminate redundant and highly correlated (>95%) attributes. Finally, modeling was performed on the 360 genes. If an attribute was assigned a weight >0.5 by a specific attribute weighting algorithm, it

was considered essential. **Supplementary Material 7** contains the results of 10 different attribute weighting algorithm applications. **Table 6** summarizes the number of attribute weighting algorithms that supported the selected DEGs.

Validation Hub Genes in Coexpressed Modules

The DT technique was used to validate the identified hub genes. Thus, the accuracy of various models was calculated and presented in **Supplementary Material 8** using four different criteria, namely, information gain ratio, information gain, Gini index, and accuracy. According to the results, the DT with the gain ratio criterion achieved the highest accuracy (75%) (**Table 7**). The DT validated the role of the top-ranked genes in mastitis classification using the expression values of common meta-genes.

As illustrated in **Figure 8**, because the *LOC407171* gene is located at the root of the constructed tree, it can be considered a biomarker for mastitis. When the *LOC407171* gene value



exceeded 8.119, and the *SFN* gene value exceeded 5.291, the samples were classified as having mastitis. When the *LOC407171* gene value is equal to or <8.119, the sample is considered healthy. When *LOC407171* exceeded 8.119, *SFN* was equal to or <5.291, and *PTPRC* was equal to or <14.390, the sample was classified as healthy. In addition, if the last feature exceeded 14.390 and the expression of *IDH1* was present, *PTPRC* would be classified as having mastitis.

The significance of the *LOC407171*, *PTPRC*, *ABCG2*, and *IDH1* genes in the turquoise module was confirmed using DT models and attribute weighting, highlighting their critical roles in mastitis.

DISCUSSION

Mastitis is a significant disease involving multiple genes that may interact to enrich specific signaling pathways. We performed a meta-analysis of RNA-seq and microarray transcriptome data to gain a comprehensive understanding of the master/key genes during mastitis disease that may play a significant role in response to *E. coli* mastitis. As individual studies have limitations in statistical power and reproducibility, several small impact genes remain unknown. Meta-analysis has been suggested as a practical approach for resolving this issue (Farhadian et al., 2018b; Sharifi et al., 2018). The BP, biological regulation, and reaction to a stimulus, the study's primary enriched GO terms, have been described as BPs in mastitis disease (Asselstine et al., 2019). These words include various activities, including cell proliferation, cell growth, biochemical processes, and signaling pathways (Long et al., 2001; Arnellos, 2018). The terminology used to describe MF in this research, such as binding, ion binding, protein binding, actin binding, cation binding, and catalytic activity, has been

previously described in immune response and protein transport (Swanson et al., 2009; Asselstine et al., 2019).

Enrichment analysis of metabolic KEGG pathways was used to identify metabolic pathways that were significantly overrepresented among 360 common genes. Several significant pathways were enriched, including peroxisomes and three subcategories of signaling pathways [interleukin 17 (IL-17) signaling pathway, nucleotide-binding and oligomerization domain (NOD)-like receptor signaling pathway, TNF signaling pathway]. Peroxisomes are required to oxidize specific biomolecules and the inflammatory response to environmental stress (Trindade Da Rosa, 2016; Su et al., 2019). Mammary epithelial cells have been shown to have immune activity, activating signaling pathways during mastitis (Song et al., 2014). TNF plays a role in various pathological processes, including immune cell regulation and immune response modulation (Shah et al., 2012; Gao et al., 2015). The NOD-like receptor regulates the immune and inflammatory responses in mammals' innate immune systems (Saxena and Yeretssian, 2014). IL-17 expression in milk peaked 24 to 48 h after pathogen challenge. These findings indicated that IL-17 was a significant cytokine in the development of dairy goat mastitis and played a critical role in mastitis development (Jing et al., 2012). Previously published research indicated that mastitis involves the NOD-like receptor, IL-17, and TNF signaling pathways (Asselstine et al., 2019). As a result of their function, we can deduce that these pathways are involved in immune system responses to mastitis disease.

The PPI networks constructed using Cytoscape revealed that the hub genes are *PTPRC*, *SOD2*, and *STAT1*. As a result, these hub genes may affect mastitis and thus warrant further validation. The *PTPRC* gene is required to signal T- and B-cell antigen receptors (Mitterski et al., 2002; Porcu et al., 2012). *PTPRC* is a highly connected gene in PPI networks and is

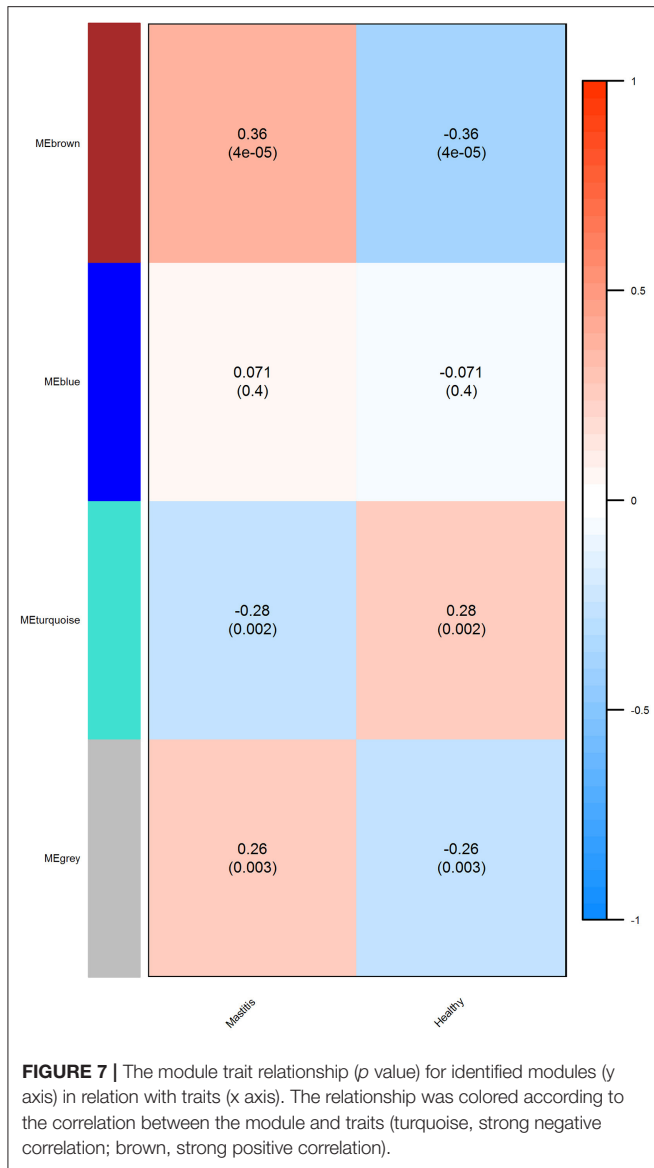


TABLE 5 | Top hub genes in significant modules in mastitis disease.

Mastitis disease	
Brown	Turquoise
<i>PRDX5</i>	<i>CD53</i>
<i>RAB5C</i>	<i>NCKAP1L</i>
<i>ACTN4</i>	<i>ARHGEF2</i>
<i>SLC25A16</i>	<i>COL9A1</i>
<i>MAPK6</i>	<i>PTPRC</i>

involved in the development of mastitis (Bakhtiarzadeh et al., 2020). *SOD2* and *IDH1* genes have been up-regulated in ewes' mammary glands using functional enrichment analysis (Gao et al., 2018, 2019). *SOD2* gene expression increased in mammary tissue of cows and ewes with mastitis caused by *S. aureus*

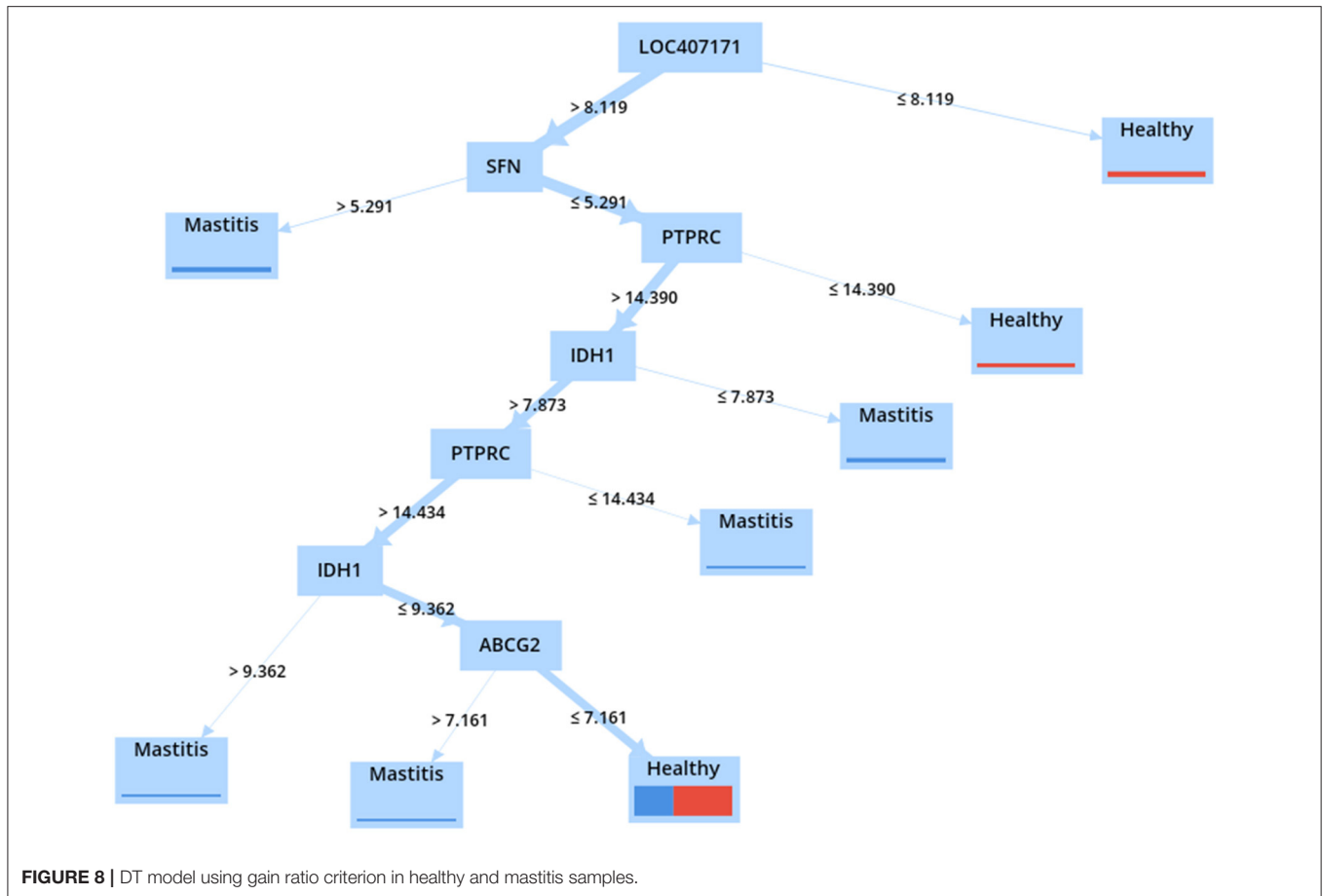
TABLE 6 | Results of different attribute weighting algorithms confirmed the most important genes.

Attribute	No. of weighting models
<i>LOC407171</i>	5
<i>MT2A</i>	4
<i>PTPRC</i>	4
<i>LPCAT2</i>	4
<i>SAMSN1</i>	4
<i>IL1B</i>	4
<i>SELPLG</i>	4
<i>CD53</i>	4
<i>PLEK</i>	4
<i>SFN</i>	3
<i>KCNJ15</i>	3
<i>SPCS3</i>	3
<i>SOD2</i>	3
<i>IDH1</i>	3
<i>SYNGR1</i>	3
<i>TANC2</i>	3
<i>CXCL3</i>	3

TABLE 7 | Accuracy comparison of constructed DT models by different criteria.

Criteria	Accuracy (%)
Gain ratio	75
Information gain	63.89
Gini index	58.33
Accuracy	63.89

and *E. coli* (Mitterhuemer et al., 2010; Jensen et al., 2013). Also, in addition, *STAT1* regulates genes involved in milk protein synthesis, fat metabolism, and immune cell activation (Cobanoglu et al., 2006). The analysis of common meta-gene coexpression networks identified four modules, two of which were significant. These modules were the most significant in the current study based on the enriched functional terms related to mastitis development. The brown module's most essential genes included *PRDX5*, *RAB5C*, *ACTN4*, and *MAPK6*. The *PRDX5* gene is expressed ubiquitously in tissues and protects cells from oxidative stress by detoxifying peroxides (Knoops et al., 2011). *PRDX5* has been shown to play a critical role in inflammation in mice by protecting cells from oxidative stress (Argyropoulou et al., 2016). In addition, the *PRDX5* gene expression is increased in mastitis sheep milk (Pisanu et al., 2015). *RAB5C* and *MAPK6* genes were identified as candidate genes for mastitis in dairy cattle following intramammary infection with *E. coli* or *S. uberis* using a combination of GWAS and DEG data analyses (Chen et al., 2015). The *ACTN4* gene was identified as the DEG in mastitis vs. healthy samples of sheep milk by transcriptomic analysis (Bonfont et al., 2011). Furthermore, *ACTN4* was identified as a hub gene in mastitis-related modules (Bakhtiarzadeh et al., 2020). On the other hand, the turquoise module's master genes were the *CD53*,



ARHGEF2, and *COL9A1* genes. *CD53* regulates cell development, and its function has been implicated in mastitis disease (Rinaldi et al., 2010). The results of a high-throughput analysis on infected bovine mammary glands with *E. coli* indicated the *ARHGEF2* gene's importance (Bagnicka et al., 2021). The *COL9A1* gene has been implicated in research involving identifying genomic regions and expression analysis of mastitis (Lu et al., 2020).

Several genes, including *LOC407171*, *MT2A*, *LPCAT2*, *CXCL3*, *SFN*, *IDH1*, and *ABCG2*, were confirmed as essential genes based on the outcome of the attribute weighting algorithm. The *LOC407171* gene is associated with the innate immune response in beef cattle and has been identified as an up-regulated gene in a dairy cow with *E. coli* mastitis (Li et al., 2019). *MT2A* plays a role in stimulus response in the pathogenesis of bovine *E. coli* in early lactation cows (Cheng et al., 2021). *LPCAT2* regulates the glycerophospholipid metabolism in periparturient dairy cattle (Bakhtiarizadeh et al., 2020). *CXCL3* is recognized as a proinflammatory cytokine in dairy cows with experimentally induced *S. aureus* clinical mastitis (Peralta et al., 2020). *SFN* was reported to regulate cell cycle progression in bovine mastitis via genome-wide association (Miles et al., 2021). *IDH1* was identified as a candidate gene in the milk transcriptome of dairy cattle implicated in innate immunity by pathway and network analysis (Banos et al., 2017). *ABCG2* gene, which is regulated by the mammary gland, responsible for the

active secretion of several compounds into milk (Otero et al., 2015).

The DT model identified the *LOC407171* gene as a critical player in mastitis disease in this study. *LOC407171* has been validated using an attribute weighting algorithm and a machine-learning algorithm. In addition, *SFN* and *IDH1* were identified using attribute weighting and machine-learning techniques, with *IDH1*, validated using WGCNA. Furthermore, *ABCG2* is recognized using weighted attributes, machine learning, and WGCNA. In addition, machine learning, attribute weighting, PPI network, and WGCNA were used to confirm *PTPRC*.

We examined possible changes in gene expression and connectivity during mastitis, and it was concluded that genes involved in the development, proliferation, and differentiation of cells in the mammary gland, as well as genes involved in immune system improvement, were primarily altered in their expression.

CONCLUSION

Because of the complexity of mastitis disease in dairy animals, far more relevant research is required to identify biomarkers associated with mastitis. The current study's findings from meta-analysis, WGCNA, and machine-learning approach allow us to represent the primary contribution to our understanding

of the most valuable genes for *E. coli* mastitis, which may provide a more robust biosignature and thus serve as reliable biomarker candidates in future studies. Our study suggests that all identified genes affect mastitis disease *via* their immune system-related functions.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

NG: research concept and design, data analysis and interpretation, wrote the article, and final

REFERENCES

- Anders, S., McCarthy, D. J., Chen, Y., Okoniewski, M., Smyth, G. K., Huber, W., et al. (2013). Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. *Nat. Protoc.* 8:1765. doi: 10.1038/nprot.2013.099
- Argyropoulou, V., Goemaere, J., Clippe, A., Lefort, C., Tissir, F., Schakman, O., et al. (2016). Peroxiredoxin-5 as a novel actor in inflammation and tumor suppression. *Free Radic. Biol. Med.* 100:S92. doi: 10.1016/j.freeradbiomed.2016.10.229
- Arnellos, A. (2018). From organizations of processes to organisms and other biological individuals. *Everything Flows: Toward. Process. Philo. Biol.* 3, 199–221. doi: 10.1093/oso/9780198779636.003.0010
- Asselstine, V., Miglior, F., Suárez-Vega, A., Fonseca, P., Mallard, B., Karrow, N., et al. (2019). Genetic mechanisms regulating the host response during mastitis. *J. Dairy Sci.* 102, 9043–9059. doi: 10.3168/jds.2019-16504
- Bagnicka, E., Kawecka-Grochocka, E., Pawlina-Tyszko, K., Zalewska, M., Kapusta, A., Kościuczuk, E., et al. (2021). MicroRNA expression profile in bovine mammary gland parenchyma infected by coagulase-positive or coagulase-negative staphylococci. *Vet. Res.* 52, 1–20. doi: 10.1186/s13567-021-00912-2
- Bakhtiarzadeh, M. R., Mirzaei, S., Norouzi, M., Sheybani, N., and Vafaei Sadi, M. S. (2020). Identification of gene modules and hub genes involved in mastitis development using a systems biology approach. *Front. Genet.* 11:722. doi: 10.3389/fgene.2020.00722
- Banos, G., Bramis, G., Bush, S., Clark, E., Mcculloch, M., Smith, J., et al. (2017). The genomic architecture of mastitis resistance in dairy sheep. *BMC Genom.* 18, 1–18. doi: 10.1186/s12864-017-3982-1
- Bansal, M., and Di Bernardo, D. (2007). Inference of gene networks from temporal gene expression profiles. *IET Syst. Biol.* 1, 306–312. doi: 10.1049/iet-syb:20060079
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Statistic. Soc. Series B* 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x
- Blum, S. E., Goldstone, R. J., Connolly, J. P., Réperant-Ferter, M., Germon, P., Inglis, N. F., et al. (2018). Postgenomics characterization of an essential genetic determinant of mammary pathogenic *Escherichia coli*. *MBio* 9:18. doi: 10.1128/mBio.00423-18
- Bolstad, B. M., Irizarry, R. A., Åstrand, M., and Speed, T.P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 19, 185–193. doi: 10.1093/bioinformatics/19.2.185
- Bonnefont, C. M., Toufeer, M., Caubet, C., Foulon, E., Tasca, C., Aurel, M.-R., et al. (2011). Transcriptomic analysis of milk somatic cells in mastitis resistant and susceptible sheep upon challenge with *Staphylococcus epidermidis* and *Staphylococcus aureus*. *BMC Genom.* 12, 1–16. doi: 10.1186/1471-2164-12-208

approval of the article. JS, SR, and KH: wrote the article. BP: data analysis, interpretation, wrote the article, and final approval of the article. All authors contributed to the article and approved the submitted version.

ACKNOWLEDGMENTS

We are grateful to Dr. Mohammad Farhadian and Dr. Sana Farhadi for their kindly help.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2021.712306/full#supplementary-material>

- Botía, J. A., Vandrovцова, J., Forabosco, P., Guelfi, S., D'sa, K., Hardy, J., et al. (2017). An additional k-means clustering step improves the biological features of WGCNA gene co-expression networks. *BMC Syst. Biol.* 11, 1–16. doi: 10.1186/s12918-017-0420-6
- Brand, B., Hartmann, A., Reipsilber, D., Griesbeck-Zilch, B., Wellnitz, O., Kühn, C., et al. (2011). Comparative expression profiling of *E. coli* and *S. aureus* inoculated primary mammary gland cells sampled from cows with different genetic predispositions for somatic cell score. *Gene. Select. Evol.* 43, 1–17. doi: 10.1186/1297-9686-43-24
- Buitenhuis, B., Røntved, C. M., Edwards, S. M., Ingvarstsen, K. L., and 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.* 12, 1–10. doi: 10.1186/1471-2164-12-130
- Burvenich, C., Van Merris, V., Mehrzad, J., Diez-Fraile, A., and Duchateau, L. (2003). Severity of *E. coli* mastitis is mainly determined by cow factors. *Veterinary Res.* 34, 521–564. doi: 10.1051/vetres:20030323
- Cebon, N., Maman, S., Walachowski, S., Gausserès, B., Cunha, P., Rainard, P., et al. (2020). Th17-related mammary immunity, but not a high systemic Th1 immune response is associated with protection against *E. coli* mastitis. *NPJ vaccines* 5, 1–13. doi: 10.1038/s41541-020-00258-4
- Chen, X., Cheng, Z., Zhang, S., Werling, D., and Wathes, D. C. (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. doi: 10.4236/ojas.2015.54040
- Cheng, Z., Buggiotti, L., Salavati, M., Marchitelli, C., Palma-Vera, S., Wylie, A., et al. (2021). Global transcriptomic profiles of circulating leucocytes in early lactation cows with clinical or subclinical mastitis. doi: 10.21203/rs.3.rs-204708/v1
- Cobanoglu, O., Zaitoun, I., Chang, Y., Shook, G., and Khatib, H. (2006). Effects of the signal transducer and activator of transcription 1 (STAT1) gene on milk production traits in Holstein dairy cattle. *J. Dairy Sci.* 89, 4433–4437. doi: 10.3168/jds.S0022-0302(06)72491-2
- Compton, C. W., Cursons, R. T., Barnett, C. M., and Mcdougall, S. (2009). Expression of innate resistance factors in mammary secretion from periparturient dairy heifers and their association with subsequent infection status. *Vet. Immunol. Immunopathol.* 127, 357–364. doi: 10.1016/j.vetimm.2008.10.331
- Dennis, G., Sherman, B. T., Hosack, D. A., Yang, J., Gao, W., Lane, H. C., et al. (2003). DAVID: database for annotation, visualization, and integrated discovery. *Genome Biol.* 4, 1–11. doi: 10.1186/gb-2003-4-9-r60
- Duarte, C. M., Freitas, P. P., and Bexiga, R. (2015). Technological advances in bovine mastitis diagnosis: an overview. *J. Vet. Diagn. Investig.* 27, 665–672. doi: 10.1177/1040638715603087

- Ebrahimi, M., Lakizadeh, A., Agha-Golzadeh, P., Ebrahimie, E., and Ebrahimi, M. (2011). Prediction of thermostability from amino acid attributes by combination of clustering with attribute weighting: a new vista in engineering enzymes. *PLoS ONE* 6:e23146. doi: 10.1371/journal.pone.0023146
- Ebrahimi, M., Esmaili, F., Cheraghi, S., Houshmand, F., Shabani, L., and Ebrahimi, 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. doi: 10.1371/journal.pone.0090885
- Evangelou, E., and Ioannidis, J. P. (2013). Meta-analysis methods for genome-wide association studies and beyond. *Nat. Rev. Genet.* 14, 379–389. doi: 10.1038/nrg3472
- Farhadian, M., Rafat, S. A., Hasanpur, K., Ebrahimi, M., and Ebrahimi, E. (2018a). Cross-species meta-analysis of transcriptomic data in combination with supervised machine learning models identifies the common gene signature of lactation process. *Front. Genet.* 9:235. doi: 10.3389/fgene.2018.00235
- Farhadian, M., Rafat, S. A., Hasanpur, K., and Ebrahimi, E. (2018b). Transcriptome signature of the lactation process, identified by meta-analysis of microarray and RNA-Seq data. *BioTechnologia. J. Biotechnol. Comput. Biol. Bionanotechnol.* 99:75659. doi: 10.5114/bta.2018.75659
- Farhadian, M., Rafat, S. A., Panahi, B., and Ebrahimi, E. (2020). Transcriptome signature of two lactation stages in Ghezel sheep identifies using RNA-Sequencing. *Anim. Biotechnol.* 20, 1–11. doi: 10.1080/10495398.2020.1784185
- Farhadian, M., Rafat, S. A., Panahi, B., and Mayack, C. (2021). Weighted gene co-expression network analysis identifies modules and functionally enriched pathways in the lactation process. *Sci. Rep.* 11, 1–15. doi: 10.1038/s41598-021-97893-1
- Gao, J., Li, T., Lu, Z., Wang, X., Zhao, X., and Ma, Y. (2019). Proteomic analyses of mammary glands provide insight into the immunity and metabolism pathways associated with clinical mastitis in meat sheep. *Animals* 9:309. doi: 10.3390/ani9060309
- Gao, Y., Theng, S. S., Mah, W.-C., and Lee, C. G. (2015). Silibinin down-regulates FAT10 and modulate TNF- α /IFN- γ -induced chromosomal instability and apoptosis sensitivity. *Biol. Open* 4, 961–969. doi: 10.1242/bio.011189
- Gao, J., Ma, Y., Li, T., Lu, Z., Chen, C., and Zhao, X. (2018). Expression and localization of SOD2 gene in breast tissue of clinical mastitis sheep (*Ovis aries*). *J. Agric. Biotechnol.* 26, 246–252. <https://pubmed.ncbi.nlm.nih.gov/12154050/>
- Gelasakis, A., Mavrogiani, V., Petridis, I., Vasileiou, N., and Fthenakis, G. (2015). Mastitis in sheep—The last 10 years and the future of research. *Vet. Microbiol.* 181, 136–146. doi: 10.1016/j.vetmic.2015.07.009
- Goldman, B., Nierman, W., Kaiser, D., Slater, S., Durkin, A. S., Eisen, J. A., et al. (2006). Evolution of sensory complexity recorded in a myxobacterial genome. *Proc. Nat. Acad. Sci.* 103, 15200–15205. doi: 10.1073/pnas.0607331103
- Gorji, A. E., Roudbari, Z., Sadeghi, B., Javadmanesh, A., and Sadkowski, T. (2019). Transcriptomic analysis on the promoter regions discover gene networks involving mastitis in cattle. *Microb. Pathog.* 137:103801. doi: 10.1016/j.micpath.2019.103801
- Günther, J., Esch, K., Poschadel, N., Petzl, W., Zerbe, H., Mitterhuemer, S., et al. (2011). Comparative kinetics of *Escherichia coli*-and *Staphylococcus aureus*-specific activation of key immune pathways in mammary epithelial cells demonstrates that *S. aureus* elicits a delayed response dominated by interleukin-6 (IL-6) but not by IL-1A or tumor necrosis factor alpha. *Infect. Immun.* 79, 695–707. doi: 10.1128/IAI.01071-10
- Günther, J., Petzl, W., Zerbe, H., Schuberth, H.-J., Koczan, D., Goetze, L., et al. (2012). Lipopolysaccharide priming enhances expression of effectors of immune defence while decreasing expression of pro-inflammatory cytokines in mammary epithelia cells from cows. *BMC Genom.* 13, 1–13. doi: 10.1186/1471-2164-13-17
- Haidich, A.-B. (2010). Meta-analysis in medical research. *Hippokratia* 14, 29.
- Jensen, K., Günther, J., Talbot, R., Petzl, W., Zerbe, H., Schuberth, H.-J., et al. (2013). *Escherichia coli*-and *Staphylococcus aureus*-induced mastitis differentially modulate transcriptional responses in neighbouring uninfected bovine mammary gland quarters. *BMC Genom.* 14, 1–19. doi: 10.1186/1471-2164-14-36
- Jing, X., Zhao, Y., Shang, C., Yao, Y., Tian, T., Li, J., et al. (2012). Dynamics of cytokines associated with IL-17 producing cells in serum and milk in mastitis of experimental challenging with *Staphylococcus aureus* and *Escherichia coli* in dairy goats. *J. Anim. Veterin. Adv.* 11, 475–479. doi: 10.3923/javaa.2012.475.479
- Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S. L. (2013). TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 14, 1–13. doi: 10.1186/gb-2013-14-4-r36
- Knoops, B., Goemaere, J., Van Der Eecken, V., and Declercq, J.-P. (2011). Peroxiredoxin 5: structure, mechanism, and function of the mammalian atypical 2-Cys peroxiredoxin. *Antioxid. Redox Signal.* 15, 817–829. doi: 10.1089/ars.2010.3584
- Kulkarni, A. G., and Kaliwal, B. (2013). Bovine mastitis: a review. *Int. J. Recent Sci. Res.* 4, 543–548. <http://recentscientific.com/bovine-mastitis-review>
- Lai, Y.-C., Fujikawa, T., Maemura, T., Ando, T., Kitahara, G., Endo, Y., et al. (2017). Inflammation-related microRNA expression level in the bovine milk is affected by mastitis. *PLoS ONE* 12:e0177182. doi: 10.1371/journal.pone.0177182
- Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinform.* 9, 1–13. doi: 10.1186/1471-2105-9-559
- Li, L., Chen, X., and Chen, Z. (2019). Identification of key candidate genes in dairy cow in response to *Escherichia coli* mastitis by bioinformatical analysis. *Front. Genet.* 10:1251. doi: 10.3389/fgene.2019.01251
- Long, E., Capuco, A., Wood, D., Sonstegard, T., Tomita, G., Paape, M., et al. (2001). *Escherichia coli* induces apoptosis and proliferation of mammary cells. *Cell Death Differ.* 8, 808–816. doi: 10.1038/sj.cdd.4400878
- Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 1–21. doi: 10.1186/s13059-014-0550-8
- Lu, X., Duan, A., Liang, S., Ma, X., and Deng, T. (2020). Genomic identification, evolution, and expression analysis of collagen genes family in water buffalo during lactation. *Genes* 11:515. doi: 10.3390/genes11050515
- Marot, G., Jaffrézic, F., and Rau, A. (2020). metaRNASeq: Differential meta-analysis of RNA-seq data. *dim (param)* 1:3. <https://cran.r-project.org/web/packages/metaRNASeq/vignettes/metaRNASeq.pdf>
- Miles, A. M., Posbergh, C. J., and Huson, H. J. (2021). Direct phenotyping and principal component analysis of type traits implicate novel QTL in bovine mastitis through genome-wide association. *Animals* 11:1147. doi: 10.3390/ani11041147
- Miterski, B., Sindern, E., Haupts, M., Schimrigk, S., and Epplen, J. T. (2002). PTPRC (CD45) is not associated with multiple sclerosis in a large cohort of German patients. *BMC Med. Genet.* 3, 1–3. doi: 10.1186/1471-2350-3-3
- 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 Genom.* 11, 1–16. doi: 10.1186/1471-2164-11-138
- Moyes, K., Sorensen, P., and Bionaz, M. (2016). The impact of intramammary *Escherichia coli* challenge on liver and mammary transcriptome and cross-talk in dairy cows during early lactation using RNAseq. *PLoS ONE* 11:e0157480. doi: 10.1371/journal.pone.0157480
- Nami, Y., Imeni, N., and Panahi, B. (2021). Application of machine learning in bacteriophage research. *BMC Microbiol.* 21, 1–8. doi: 10.1186/s12866-021-02256-5
- Otero, J., Barrera, B., De La Fuente, A., Prieto, J., Marqués, M., Álvarez, A., et al. (2015). The gain-of-function Y581S polymorphism of the ABCG2 transporter increases secretion into milk of danofloxacin at the therapeutic dose for mastitis treatment. *J. Dairy Sci.* 98, 312–317. doi: 10.3168/jds.2014-8288
- Panahi, B., Farhadian, M., and Hejazi, M. A. (2020). Systems biology approach identifies functional modules and regulatory hubs related to secondary metabolites accumulation after transition from autotrophic to heterotrophic growth condition in microalgae. *PLoS ONE* 15:e0225677. doi: 10.1371/journal.pone.0225677
- Panahi, B., Farhadian, M., Dums, J. T., and Hejazi, M. A. (2019a). Integration of cross species RNA-Seq meta-analysis and machine-learning models identifies the most important salt stress-responsive pathways in microalga *Dunaliella*. *Front. Genet.* 10:752. doi: 10.3389/fgene.2019.00752
- Panahi, B., and Hejazi, M. A. (2021). Weighted gene co-expression network analysis of the salt-responsive transcriptomes reveals novel hub genes in green halophytic microalgae *Dunaliella salina*. *Sci. Rep.* 11, 1–11. doi: 10.1038/s41598-020-80945-3

- Panahi, B., Mohammadi, S. A., and Doulati-Baneh, H. (2019b). Characterization of Iranian grapevine cultivars using machine learning models. *Proceed. Nat. Acad. Sci. India Sect. B: Biol. Sci.* 19, 1–7. doi: 10.1007/s40011-019-01131-8
- Peralta, O., Carrasco, C., Vieytes, C., Tamayo, M., Muñoz, I., Sepulveda, S., et al. (2020). Safety and efficacy of a mesenchymal stem cell intramammary therapy in dairy cows with experimentally induced *Staphylococcus aureus* clinical mastitis. *Sci. Rep.* 10, 1–12. doi: 10.1038/s41598-020-59724-7
- Pisanu, S., Cubeddu, T., Pagnozzi, D., Rocca, S., Cacciotto, C., Alberti, A., et al. (2015). Neutrophil extracellular traps in sheep mastitis. *Vet. Res.* 46, 1–14. doi: 10.1186/s13567-015-0196-x
- Porcu, M., Kleppe, M., Gianfelici, V., Geerdens, E., De Keersmaecker, K., Tartaglia, M., et al. (2012). Mutation of the receptor tyrosine phosphatase PTPRC (CD45) in T-cell acute lymphoblastic leukemia. *Blood* 119, 4476–4479. doi: 10.1182/blood-2011-09-379958
- Rhodes, D. R., Barrette, T. R., Rubin, M. A., Ghosh, D., and Chinnaiyan, A. M. (2002). Meta-analysis of microarrays: interstudy validation of gene expression profiles reveals pathway dysregulation in prostate cancer. *Cancer Res.* 62, 4427–4433. <https://pubmed.ncbi.nlm.nih.gov/12154050/>
- Rinaldi, M., Li, R. W., Bannerman, D. D., Daniels, K. M., Evock-Clover, C., Silva, M. V., et al. (2010). A sentinel function for teat tissues in dairy cows: dominant innate immune response elements define early response to *E. coli* mastitis. *Funct. Integrat. Genom.* 10, 21–38. doi: 10.1007/s10142-009-0133-z
- Robinson, M. D., and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* 11, 1–9. doi: 10.1186/gb-2010-11-3-r25
- Saxena, M., and Yeretsian, G. (2014). NOD-like receptors: master regulators of inflammation and cancer. *Front. Immunol.* 5:327. doi: 10.3389/fimmu.2014.00327
- Shah, N., Dhar, D., Mohammed, F. E. Z., Habtesion, A., Davies, N. A., Jover-Cobos, M., et al. (2012). Prevention of acute kidney injury in a rodent model of cirrhosis following selective gut decontamination is associated with reduced renal TLR4 expression. *J. Hepatol.* 56, 1047–1053. doi: 10.1016/j.jhep.2011.11.024
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504. doi: 10.1101/gr.1239303
- Sharifi, S., Pakdel, A., Ebrahimi, M., Reecy, J. M., Fazeli Farsani, S., and Ebrahimi, E. (2018). Integration of machine learning and meta-analysis identifies the transcriptomic bio-signature of mastitis disease in cattle. *PLoS ONE* 13:e0191227. doi: 10.1371/journal.pone.0191227
- Sipka, A., Klaessig, S., Duhamel, G. E., Swinkels, J., Rainard, P., and Schukken, Y. (2014). Impact of intramammary treatment on gene expression profiles in bovine *Escherichia coli* mastitis. *PLoS ONE* 9:e85579. doi: 10.1371/journal.pone.0085579
- Smyth, G. K., Ritchie, M., Thorne, N., and Wettenhall, J. (2005). LIMMA: linear models for microarray data. in *Bioinformatics and computational biology solutions using R and bioconductor*. *Stat. Biol. Health* 6:23. doi: 10.1007/0-387-29362-0_23
- Song, X., Guo, M., Wang, T., Wang, W., Cao, Y., and Zhang, N. (2014). Geniposide inhibited lipopolysaccharide-induced apoptosis by modulating TLR4 and apoptosis-related factors in mouse mammary glands. *Life Sci.* 119, 9–17. doi: 10.1016/j.lfs.2014.10.006
- Su, T., Li, W., Wang, P., and Ma, C. (2019). Dynamics of peroxisome homeostasis and its role in stress response and signaling in plants. *Front. Plant Sci.* 10:705. doi: 10.3389/fpls.2019.00705
- Swanson, K., Stelwagen, K., Dobson, J., Henderson, H., Davis, S., Farr, V., et al. (2009). Transcriptome profiling of *Streptococcus uberis*-induced mastitis reveals fundamental differences between immune gene expression in the mammary gland and in a primary cell culture model. *J. Dairy Sci.* 92, 117–129. doi: 10.3168/jds.2008-1382
- Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, J., et al. (2015). STRING v10: protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* 43, D447–D452. doi: 10.1093/nar/gku1003
- Takeshima, S., Matsumoto, Y., Chen, J., Yoshida, T., Mukoyama, H., and Aida, Y. (2008). Evidence for cattle major histocompatibility complex (BoLA) class II DQA1 gene heterozygote advantage against clinical mastitis caused by *Streptococci* and *Escherichia* species. *Tissue Antigens* 72, 525–531. doi: 10.1111/j.1399-0039.2008.01140.x
- Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., et al. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* 7, 562. doi: 10.1038/nprot.2012.016
- Trindade Da Rosa, F. (2016). Role of peroxisome proliferator-activated receptor gamma on prevention/cure of mastitis. *Review* 54(8):2460–2470. doi: 10.2337/diabetes.54.8.2460
- Tseng, G. C., Ghosh, D., and Feingold, E. (2012). Comprehensive literature review and statistical considerations for microarray meta-analysis. *Nucleic Acids Res.* 40, 3785–3799. doi: 10.1093/nar/gkr1265
- Vangroenweghe, F., Duchateau, L., Boutet, P., Lekeux, P., Rainard, P., Paape, M., et al. (2005). Effect of carprofen treatment following experimentally induced *Escherichia coli* mastitis in primiparous cows. *J. Dairy Sci.* 88, 2361–2376. doi: 10.3168/jds.S0022-0302(05)72914-3
- Vasudevan, P., Nair, M. K. M., Annamalai, T., and Venkitanarayanan, K. S. (2003). Phenotypic and genotypic characterization of bovine mastitis isolates of *Staphylococcus aureus* for biofilm formation. *Vet. Microbiol.* 92, 179–185. doi: 10.1016/S0378-1135(02)00360-7
- Wang, G., Muschelli, J., and Lindquist, M. A. (2021). Moderated t-tests for group-level fMRI analysis. *Neuroimage* 237:118141. doi: 10.1016/j.neuroimage.2021.118141
- Wang, X., Kang, D. D., Shen, K., Song, C., Lu, S., Chang, L.-C., et al. (2012). An R package suite for microarray meta-analysis in quality control, differentially expressed gene analysis and pathway enrichment detection. *Bioinformatics* 28, 2534–2536. doi: 10.1093/bioinformatics/bts485

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Ghahramani, Shodja, Rafat, Panahi and Hasanpur. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.