

RESEARCH ARTICLE

Integration of machine learning and meta-analysis identifies the transcriptomic bio-signature of mastitis disease in cattle

Somayeh Sharifi^{1,2}, Abbas Pakdel^{1*}, Mansour Ebrahimi³, James M. Reecy², Samaneh Fazeli Farsani⁴, Esmaeil Ebrahimie^{5,6,7,8}

1 Department of Animal Science, College of Agriculture, Isfahan University of Technology, Isfahan, Iran, **2** Department of Animal Science, Iowa State University, Ames, Iowa, United States of America, **3** School of Basic Sciences, University of Qom, Qom, Iran, **4** Department of Chemical Engineering, Ferdowsi University of Mashhad, Mashhad, Iran, **5** School of Medicine, The University of Adelaide, Adelaide, Australia, **6** Institute of Biotechnology, Shiraz University, Shiraz, Iran, **7** Division of Information Technology, Engineering and the Environment, School of Information Technology and Mathematical Sciences, University of South Australia, Adelaide, South Australia, Australia, **8** School of Biological Sciences, Faculty of Science and Engineering, Flinders University, Adelaide, South Australia, Australia

* pakdel@cc.iut.ac.ir



OPEN ACCESS

Citation: Sharifi S, Pakdel A, Ebrahimi M, Reecy JM, Fazeli Farsani S, Ebrahimie E (2018) Integration of machine learning and meta-analysis identifies the transcriptomic bio-signature of mastitis disease in cattle. PLoS ONE 13(2): e0191227. <https://doi.org/10.1371/journal.pone.0191227>

Editor: Juan J Loor, University of Illinois, UNITED STATES

Received: July 25, 2017

Accepted: December 29, 2017

Published: February 22, 2018

Copyright: This is an open access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the [Creative Commons CC0](https://creativecommons.org/licenses/by/4.0/) public domain dedication.

Data Availability Statement: All transcripts datasets are available from the GEO database with following accession numbers: GSE15025 GSE24217 GSE24560. GSE25413 GSE32186 GSE50685

Funding: The authors received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

Abstract

Gram-negative bacteria such as *Escherichia coli* (*E. coli*) are assumed to be among the main agents that cause severe mastitis disease with clinical signs in dairy cattle. Rapid detection of this disease is so important in order to prevent transmission to other cows and helps to reduce inappropriate use of antibiotics. With the rapid progress in high-throughput technologies, and accumulation of various kinds of ‘-omics’ data in public repositories, there is an opportunity to retrieve, integrate, and reanalyze these resources to improve the diagnosis and treatment of different diseases and to provide mechanistic insights into host resistance in an efficient way. Meta-analysis is a relatively inexpensive option with good potential to increase the statistical power and generalizability of single-study analysis. In the current meta-analysis research, six microarray-based studies that investigate the transcriptome profile of mammary gland tissue after induced mastitis by *E. coli* infection were used. This meta-analysis not only reinforced the findings in individual studies, but also several novel terms including responses to hypoxia, response to drug, anti-apoptosis and positive regulation of transcription from RNA polymerase II promoter enriched by up-regulated genes. Finally, in order to identify the small sets of genes that are sufficiently informative in *E. coli* mastitis, the differentially expressed gene introduced by meta-analysis were prioritized by using ten different attribute weighting algorithms. Twelve meta-genes were detected by the majority of attribute weighting algorithms (with weight above 0.7) as most informative genes including *CXCL8* (*IL8*), *NFKBIZ*, *HP*, *ZC3H12A*, *PDE4B*, *CASP4*, *CXCL2*, *CCL20*, *GRO1* (*CXCL1*), *CFB*, *S100A9*, and *S100A8*. Interestingly, the results have been demonstrated that all of these genes are the key genes in the immune response, inflammation or mastitis. The Decision tree models efficiently discovered the best combination of the meta-genes as bio-signature and confirmed that some of the top-ranked genes -*ZC3H12A*, *CXCL2*, *GRO*, *CFB*- as biomarkers for *E. coli* mastitis (with the accuracy 83% in average). This research

properly indicated that by combination of two novel data mining tools, meta-analysis and machine learning, increased power to detect most informative genes that can help to improve the diagnosis and treatment strategies for *E. coli* associated with mastitis in cattle.

Introduction

Bovine mastitis is an inflammatory disease with clinical and subclinical forms which result in significant economic losses due to negative impacts on animal welfare [1–3], productive [4–6] and reproductive performances [7, 8], poor milk quality [9], increased workload [10], early culling [1, 11], and high treatment costs [12]. Clinical mastitis was detected in almost 25% of the 9.3 million dairy cows present in the USA every year; a quarter of them were removed/sold from the herd, and approximately less than 5% of all cows died as a result of mastitis [13]. Environmental pathogens including coliforms are the major contributors to clinical mastitis causing acute inflammation with clinical signs in dairy cows, which however may be self-healing by eventually eradicating the invader [14], are occasionally fatal [15]. Nevertheless, self-care is often associated with a longer duration of infection, lower milk yield, and the potential for pathological changes in the mammary gland [16].

There is evidence that mastitis-causing pathogens use various mechanisms to induce cell pathways. Hence, the identification of pathogens is of major importance in order to correct actions, prevent transmission to other cows, reduce the risk of appearance of chronic infections, and helps to reduce inappropriate use of antibiotics, antimicrobial resistance and cost of treatment [17–19]. Disease-causing genes [20] and biomarkers help to improve diagnosis, prognosis, and monitoring of responses to therapy [21]. Genes coding for proteins such as Haptoglobin (HP), Serum Amyloid A (SAA) [22], Cathelicidin antimicrobial peptide (CAMP) [23], and Lingual antimicrobial peptide (LAP) [24] have been identified as potential biomarkers for mastitis detection. The performance of the most mastitis detection systems do not satisfy the high accuracy required for practical clinical mastitis detection systems [25, 26]. Potential to include several biomarkers on one test strip to enhance the diagnostic efficiency is an aim of developmental research. Antibiotic therapy should be chosen based on mastitis pathogen and the type of mastitis [27, 28]; therefore, biomarker discovery with the focus on specific pathogens will be useful. The efficacy of antibiotic and/or anti-inflammatory treatment in mastitis is still a topic of scientific debate, and studies on treatment value in clinical cases show conflicting results [29, 30]. Moreover, efforts to find other therapy methods such as homeopathic treatment had no success in this disease [31]. Identification of disease-causing genes that underlie complex traits such as susceptibility to mastitis is the goal of many genetic and biomedical studies, which provides mechanistic insights into host resistance in addition to improving the diagnosis and treatment of the disease. The amplitude of the inflammatory response is mainly dependent on individual cow factors, and different animals will respond inconsistently to *Escherichia coli* (*E. coli*) infection [32, 33]. Combining the results of independent studies with a related hypothesis using meta-analysis, as a relatively inexpensive option with good potential to increase the statistical power and the generalizability of single-study analysis, can bypass the challenges associated with individual variations, and strengthen the mildest data perturbations [34, 35]. In the previous meta-analysis studies, different aspects of mastitis disease have been investigated. Genini *et al.* (2011) identified a common transcriptional response to different pathogens in the mammary glands of several species [36]. Younis *et al.* (2016) investigated differences in transcriptional response between *E. coli* and

Staphylococcus aureus strains infections and also between lipopolysaccharide (LPS), and *E. coli*-induced mastitis [37].

In the current study, for the first time, two novel data mining tools, meta-analysis and machine learning, were integrated to detect differentially expressed gene (DE)s and prioritize them to identify the most informative genes in response to *E. coli* mastitis. Attribute weighting algorithm (AW)s and Decision tree model (DT)s are the most widely used approaches in machine learning. Various algorithms of AW or feature selection give weight to features and allow the variable set to be reduced in size, thereby creating a more manageable set of attributes for modeling and attribute ranking [38, 39]. Decision tree models predict the value of a discrete dependent variable within a finite set of independent variables [40]. We used various DTs to classify samples in datasets for confirmation of AWs. The high efficiency and applicability of several well-known AWs and DTs have been demonstrated previously [41–44].

Material and methods

The following steps were performed in this article: 1. Identifying the suitable microarray studies of bovine mammary gland infected with *E. coli*, extracting the data from studies, preparing, normalizing, and annotating the individual studies; 2. Analyzing individual studies and then combining the studies-specific *p-values* with rOP meta-analysis method; 3. Fulfilling the functional enrichment analysis on the DEs introduced by meta-analysis; 4. Applying 10 different AWs on standardized expression values of meta-genes in all samples to rank and select the most important genes and making 10 new datasets based on the selection of attributes; and 5. Utilizing various DTs to classify samples in datasets for confirmation of AWs

Microarray datasets

PubMed central ("<https://www.ncbi.nlm.nih.gov/pubmed/>" Accessed January 2016) and Google Scholar ("<https://scholar.google.com/>" Accessed January 2016) were searched by using “*Bos Taurus* [organism]”, “Mastitis” and “*Escherichia coli*” keywords. Microarray gene expression data were retrieved from either, GEO of NCBI ("<https://www.ncbi.nlm.nih.gov/gds/>" Accessed January 2016) or ArrayExpress of EMBL_EBI ("<https://www.ebi.ac.uk/arrayexpress/>" Accessed January 2016). Twelve studies matched these search criteria. Upon additional review, only six studies were selected for further analysis as they all used the Affymetrix bovine GeneChip™ ("<http://www.affymetrix.com/index.affx>" Accessed February 2016). Information of these studies are shown in Table 1. Studies were excluded from the meta-analysis for the following reasons: had non-commercial platforms, which incompletely overlap the Affymetrix arrays, therefore would significantly reduce the number of genes after matching and/or they had incomplete annotation or no valid citation. Affymetrix Bovine Genome Array platform contains 24,128 probe sets to measure global transcript abundance (Bovine.na.36, March 2016). From these probe sets, 19,192 ones, which had an associated gene symbol, were used in the analysis reported here. The Bovine Genome Array annotation is available from NetAffx Analysis Centre ("<http://www.affymetrix.com/support/technical/annotationfilesmain.affx>" Accessed December 2016"). Only samples infected by *E. coli* without any treatment and appropriate controls were used in this analysis. The study by Brand *et al.* which was mentioned in Table 1, had samples from animals with either high or low susceptibility to mastitis [45]. Only data from the highly susceptible animals were used in this analysis. As sampling times after infection differed among experiments, each sampling time was considered as a separate study. A total of 130 mammary gland samples (57 healthy and 73 infected) of 15 retrieved datasets from 6 studies were included in the differential expression analysis (Table 1).

Table 1. Summary of the microarray datasets employed in meta-analysis in this study.

Accession number	Citation	Treatment time ^a (h)	Pathogen	Challenge/ Inoculum dose	Kind of experiment	Preparation of bacteria	Samples (ctr:tr) ^b
GSE15025	[46]	6	<i>E. coli</i> 1303	500 CFU ^c	in vivo	Live	5:5
		24					5:5
GSE24217	[47]	24	<i>E. coli</i> K2BH2	20–40 CFU	in vivo	Live	9:12
		192					14:14
GSE24560	[45]	1	<i>E. coli</i> 1303	100 µL solution	in vitro	heat-inactivated	3:5
		6					3:5
		24					4:4
GSE25413	[48]	1	<i>E. coli</i> 1303	10 ⁷ particles/ml	in vitro	heat-inactivated	3:3
		3					3:3
		6					3:3
		24					3:3
GSE32186	[49]	6	<i>E. coli</i> 1303	10 ⁷ particle/ml	in vitro	heat-inactivated	3:3
		6 ^d					3:3
GSE50685	[29]	24	<i>E. coli</i> ECC-Z	100 CFU	in vivo	Live	2:2
		48					3:3

^a Time of sampling after infection

^b Number of healthy samples: number of treatment samples

^c Colony Forming Unit

^d Cells were harvested either 30h (short waiting experiment) or 60 h (long waiting experiment) after the start of the trial.

<https://doi.org/10.1371/journal.pone.0191227.t001>

Pre-processing of microarray datasets

The quality of each dataset was explored by PCA analysis and box plots before and after normalization, as previously described [50–52]. Quartile normalization and summarization were performed on individual datasets by log scale Robust Multi-array Average (RMA) [53] as implemented in R Affy package [54]. The Affymetrix Bovine GeneChip™ has multiple probes (or probe sets) that represent the same genes. Therefore, gene matching was necessary for these probe sets/genes. Among all possible probe IDs for a given gene, the probe ID with the largest Inter-Quartile Range (IQR) of expression value was selected to represent that gene. In order to reduce the false discovery rate of microarray data analysis, we removed approximately 10% of the non-expressed genes based on the small average expression values across the majority of studies, and approximately 10% of the non-informative genes that had minimal amounts of variation. Final dataset (Fd) was used for the next meta-analysis process. The MetaDE package in R (version 1.0.5) was used for matching and filtering procedures [55].

Meta-analysis

Here, we utilized transcriptome data from 6 independent studies that were different in employed techniques (in vivo versus in vitro), methods of bacterial preparation (live *E. coli* versus heat-inactivated *E. coli*), strains of *E. coli* (1303, K2BH2 and ECC-Z) and also different doses of Challenge (see Table 1). Differences in the response to bacterial challenge of the mammary epithelial cells in vivo and in vitro have been characterized previously [56, 57]. It has been shown that virulence factors of heat-inactivated pathogens are different from those of active pathogens [45]. It has been illustrated that phenotypic properties of strains from different phylogroups are likely to be different [58]. For this reasons, we used meta-analysis based on *p-values* because this method permits us to join related studies with heterogeneous data [59]. For each meta-analysis, it is possible to apply different purposes with different approaches. In the current study, we

considered investigating genes, which commonly up/down expressed in all studies related to *E. coli* mastitis.

Here, at first, expression levels of mastitis and healthy samples for each gene were compared by using a moderated Student's t-test implemented to run on Fd by MetaDe package [55]. We used a one-tailed *p-value* analysis in each study to specify the direction of the alternative hypothesis to identify up- and down-regulated genes after meta-analysis. The *p-values* of each dataset were used in the r^{th} ordered *p-value* (rOP) meta-analysis method. We used $r^{\text{th}} = 5$ to combine *p-values* in order to detect DEs in 5 smallest *p-values* among all datasets (out of 15 datasets) [60]. A separated meta-analysis performed on right-sided *p-values*, and left-sided *p-values* offer up- and down-regulated genes, respectively. A false discovery rate adjustment for multiple testing with cut off value of 0.005 (one tailed) was performed as described by Benjamini and Hochberg [61]. All individual data analyses and meta-analyses were performed in R program (version 3.3.1) using the MetaDE package (version 1.0.5). Differentially expressed gene(s) identified by meta-analysis (meta-gene(s)) were used for machine-learning process. A flow diagram has been prepared to better understanding of all processes in an attempt to achieve meta-genes (see Fig 1).

Functional enrichment analysis

The meta-genes were submitted to functional annotation tool of Dataset for Annotation, Visualization and Integrated Discovery program, version 6.8, (DAVID, <http://david.abcc.ncifcrf.gov/home.jsp>) in order to identify the biological processes, cellular components and molecular functions [62, 63].

We analyzed the gene ontologies for up- and down-regulated meta-genes separately. The gene ontology (GO) terms generated by modified Fisher Exact test and terms with *p-values* better than 0.05 were selected.

Attribute weighting algorithms

After meta-analysis, 885 genes showed DEs between healthy and induced mastitis samples, based on Benjamini & Hochberg adjustment *p-value* correction ($q < 0.01$). To improve the ability to detect the most informative genes, we used a two-step standardization procedure proposed by Yoon *et al.* (2006) on meta-genes including within-array standardization (array-specific Z-score calculation) followed by the gene-specific multi-array standardization (gene-specific Z-score calculation) [64]. Subject feature (categorized as healthy and mastitis) was set as the target or label variable and standard expression value of meta-genes was set as feature or attribute, which were classified as continuous data. This new dataset (Metad), was used to import into RapidMiner Studio software (RapidMiner 7.0.001 GmbH). A supplemental spreadsheet file shows this dataset (see S1 Table).

Ten different AWs consisting of PCA, Uncertainty, Relief, Chi Squared, Gini Index, Deviation, Rule, Gain Ratio, Information Gain, and SVM [65] were applied on the list of meta-genes. We ranked meta-genes based on the number of AT algorithms which indicate that gene (attribute) is important (weight above 0.7) with respect to the subject (mastitis and healthy). Ten new datasets produced by trimming the Metad based on a weight above 0.7 given by each AW (Attribute Selection), as well as the Metad (11 datasets in total), used as input for DT models.

Decision tree models

Sixteen Tree Induction models including: Decision Tree, Random Tree, Tree Stump, Tree and Random Forest models, each model with 4 different criteria Accuracy, Gain Ratio, Gini Index

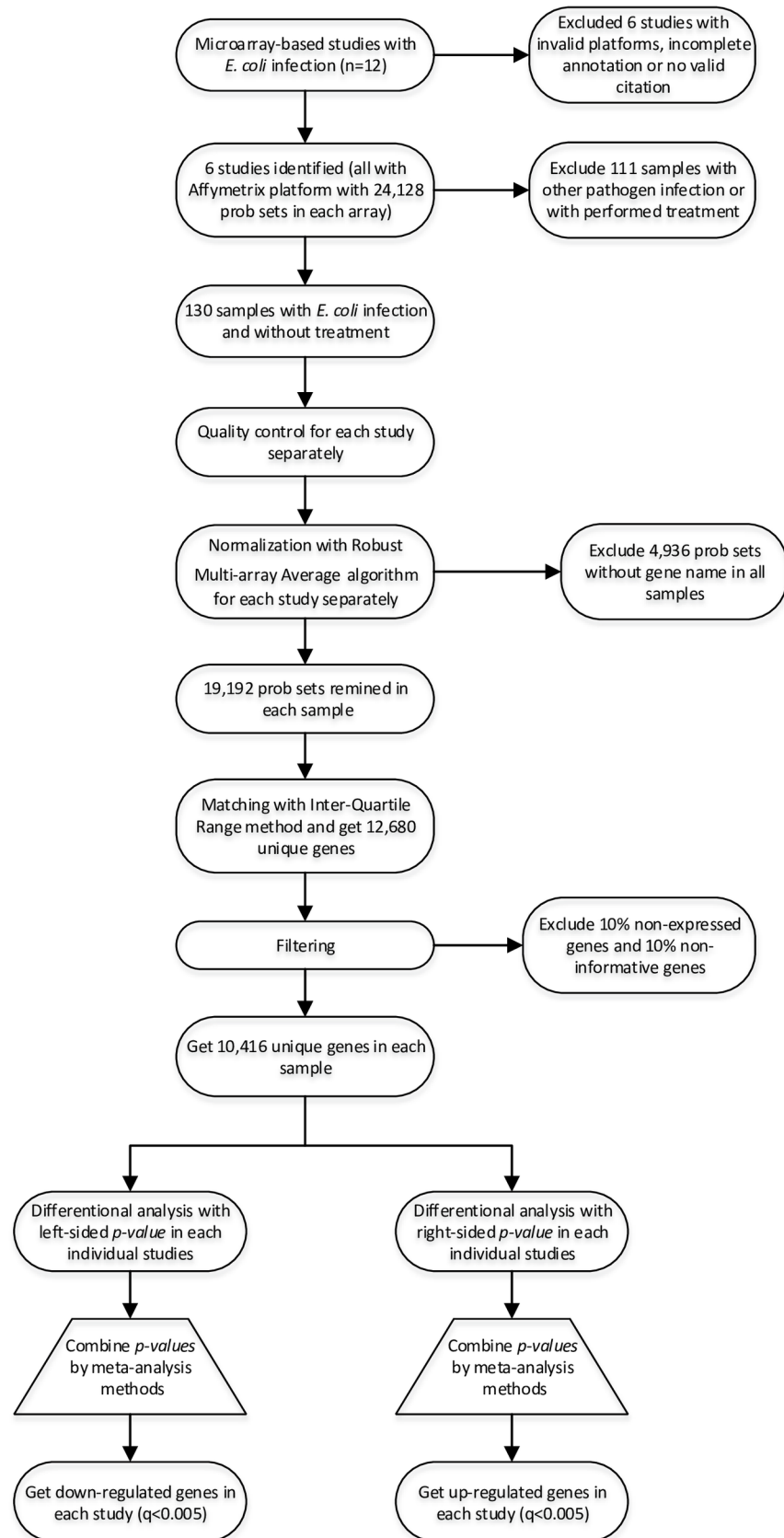


Fig 1. All processes including selection of studies, pre-processing of datasets (quality control, normalization, summarization and preparation of datasets), individual differential analysis and performing a meta-analysis to achieve differentially expressed genes (meta-genes).

<https://doi.org/10.1371/journal.pone.0191227.g001>

and Information Gain were applied on eleven datasets including the original Metad and 10 datasets generated by the 10 AWs as described previously [65]. The Decision tree model was applied to find patterns between important genes. The models were run with a minimal size of two for all leaves, a minimal gain of 0.1 to produce a split, and a maximal tree depth of 20. A confidence level of 0.25 was selected for the pessimistic error calculation for pruning [42]. The performance of different models in prediction of the target variable (healthy and mastitis) based on attribute variables (standardized expression of meta-genes) was used to calculate model efficiency. Accuracy was calculated by taking the percentage of correct predictions over the total number of samples (130 samples). A ten-fold cross-validation algorithm with stratified sampling was used to build the trees. Furthermore, an average of ten runs were used to calculate the performance percentage [66].

The PRISMA checklist is included as [S2 Table](#).

Results

Meta-analysis increase power to detect DEs

From the 19,192 probe sets on the Affymetrix Bovine Genome Array, which contained annotation, 12,860 unique genes were identified after matching. Following the filtering step, the meta-analysis was applied on 10,416 probes. 885 meta-genes were differentially expressed, of which 143 genes were down-regulated and 742 genes were up-regulated (one-tailed, $q < 0.005$). We prepared a supplemental spreadsheet file that contains more additional information (see [S3 Table](#)). In these meta-genes, 291 genes never showed a significant q -value in any of the individual studies, most likely due to the relatively small sample sizes of those individual studies (see [S4 Table](#)). The results provide a strong evidence that meta-analysis has improved the ability of DEs detection.

Functional annotation clustering analysis of meta-genes revealed several novel themes

In order to understand the functional significance of the identified meta-genes, GO enrichment analysis was performed using the DAVID program. We had more focus on biological process pathways. The p -values < 0.05 were used to determine statistically significant categories. Up-regulated genes mainly enriched the biological processes terms were associated with the immune response, defense responses, inflammation, chemotaxis, acute phase protein, protein degradation and proteolysis, growth and death of cell, response to wounding and cell signaling pathways. Product of up-regulated genes was mostly localized in plasma membrane and extracellular region based on cellular component analysis.

Down-regulated genes mainly enriched terms related to fatty acid metabolism and lipid biosynthesis including cholesterol, sterol, terpenoid biosynthesis and metabolic process. All components of GO terms related to up- and down-regulated genes were shown in supplemental spreadsheet files (see [S5](#) and [S6](#) Tables respectively).

Attribute weighting algorithms were used to rank meta-genes

Various AWs were employed to identify the important genes. In the AWs, normalized data were used to run the models. It was expected that all weights would be between 0 and 1.0 value, closer to 1 is an indication that a given gene is an important attribute. *CXCL2* gene

(Chemokine (C-X-C motif) ligand 2) was the most important gene pointed out by 70% of the AWs (7 from 10 AWs); followed by *CXCL8*, *CFB*, *ZC3H12A*, *CCL20*, *NFKBIZ*, *S100A9*, *S100A8*, *PDE4B*, *CASP4* and *HP*. A table containing the meta-genes with all weights given by 10 AWs was shown in a supplemental spreadsheet file (see [S7 Table](#)). A complete list of high relevant genes that were confirmed by the majority of AWs (with a weight above 0.7) is presented in [Table 2](#). In order to run DTs, 10 new datasets based on attribute selection with weights above 0.7 in each AWs were also generated.

Decision tree models identified gene bio-signatures that can discriminate mastitis from healthy samples. Sixteen different DTs were applied to eleven datasets. The minimum and maximum performances were 53.08% and 86.5%, respectively ([Table 3](#)).

The architecture of selected threes generated by DTs has shown in [Fig 2](#). This selection was based on the size of tree, display the role of top-ranked genes in the classification of samples and performance percentages of trees in prediction of label of samples as healthy or mastitis based on standard expression value of meta-genes. We generated these trees by performing of Random Forest models with Gini Index, Accuracy, Information Gain and Gain Ratio criterion run on SVM (A), Gini Index(B) Relief (C) and SVM (D) datasets respectively. As shown in [Fig 2](#) (A), *ZC3H12A* gene has potential biomarker performance. When the value of *ZC3H12A* gene was greater than -0.100, the cases fell into the mastitis class. Moreover, when the value was equal to or lower than -0.100, and the value of *NFKBIZ* gene was lower than -1.204, a sample fell into the healthy class. In contrast, when the value of last feature was equal or higher than -1.138, the sample fell into the healthy class. Otherwise, a sample fell into the mastitis class with an accuracy of 83.85%, indicating that from the 130 samples, 110.5 were correctly categorized between mastitis and healthy class. In [Fig 2](#), in the same way, *CXCL2* in B part, *CFB* in C part and *GRO1* in D part were at the peak of trees and have potential biomarker performance with 83.85%, 82.31%, and 83.85% accuracy respectively.

Discussion

With the rapid progress in high-throughput technologies and accumulation of various kinds of ‘-omics’ data in public repositories, there is an opportunity to retrieve, integrate, and re-

Table 2. The most important attributes (differentially expressed genes introduced by meta-analysis) ranked based on 10 attribute weighting algorithms (AWs), including PCA, Uncertainty, Relief, Chi-Squared, Gini Index, Deviation, Rule, Gain Ratio, Information Gain, and SVM.

Attribute (Gene symbol)	Gene name (alias)	The number of AWs that indicate the attribute is important (weight above 0.7)
<i>CXCL2</i>	chemokine (C-X-C motif) ligand 2 (GRO3)	7
<i>CXCL8</i>	C-X-C motif chemokine ligand 8 (IL-8, IL8)	6
<i>GRO1</i>	chemokine (C-X-C motif) ligand 1 (CXCL1, MGSA)	6
<i>CFB</i>	complement factor B (BF)	6
<i>ZC3H12A</i>	zinc finger CCCH-type containing 12A	6
<i>CCL20</i>	C-C motif chemokine ligand 20	5
<i>NFKBIZ</i>	NFKB inhibitor zeta (MAIL)	5
<i>S100A9</i>	S100 calcium binding protein A9	5
<i>S100A8</i>	S100 calcium binding protein A8	5
<i>PDE4B</i>	phosphodiesterase 4B	5
<i>CASP4</i>	caspase 4, apoptosis-related cysteine peptidase (CASP13)	5
<i>HP</i>	haptoglobin	5

<https://doi.org/10.1371/journal.pone.0191227.t002>

Table 3. Comparison of performance percentage of 16 Decision tree induction models run on 11 datasets (10 datasets generated by trimming the Metad based on a weight above 0.7 given by each AW's plus the Metad) of differentially expressed genes introduced by meta-analysis response to mastitis disease.

Models Datasets	Random Forest Accuracy (%)	Random Forest Gini index (%)	Random Forest Information Gain (%)	Random Forest Gain Ratio (%)	Random Tree Accuracy (%)	Random Tree Gini index (%)	Random Tree Information Gain (%)	Random Tree Gain Ratio (%)	Tree Stump Accuracy (%)	Tree Stump Gini index (%)	Tree Stump Information Gain (%)	Tree Stump Gain Ratio (%)	Decision Tree Accuracy (%)	Decision Tree Gini index (%)	Decision Tree Information Gain (%)	Decision Tree Gain Ratio (%)
Chi Squared	78.46	83.08	86.15	84.62	70.00	70.77	76.92	74.62	73.85	83.08	79.23	79.23	83.08	83.85	81.54	83.08
Deviation	66.92	63.08	58.46	58.46	60.00	64.62	62.31	53.08	60.77	62.31	56.92	56.92	64.62	62.31	58.46	55.38
Gini Index	83.85	82.31	78.46	82.31	70.77	76.15	80.00	73.85	73.85	83.08	79.23	79.23	86.92	84.62	83.08	82.31
Information Gain	82.31	84.62	82.31	83.08	72.31	76.15	75.38	74.62	73.85	83.08	79.23	79.23	86.92	86.15	83.08	81.54
Gain Ratio	83.85	84.62	84.62	82.31	73.85	70.77	75.38	70.00	79.23	83.08	79.23	79.23	84.62	79.23	78.46	80.77
PCA	77.69	78.46	78.46	73.85	58.46	66.15	63.08	66.92	70.00	73.08	72.31	72.31	71.54	80.00	83.08	79.23
Relief	80.77	80.77	82.31	81.54	70.77	71.54	83.85	75.38	82.31	83.08	82.31	82.31	83.08	80.00	75.38	81.54
Rule	75.38	79.23	75.38	80.00	63.85	64.62	66.92	66.92	68.46	80.00	64.62	64.62	76.92	76.92	73.85	81.54
SVM	83.08	83.85	78.46	83.85	74.62	72.31	73.85	66.92	82.31	83.08	82.31	82.31	79.23	76.92	80.00	80.00
Uncertainty	83.85	85.38	84.62	80.77	72.31	76.15	77.69	73.08	73.85	83.08	79.23	79.23	83.08	78.46	80.77	79.23
Metad	78.46	79.23	78.46	75.38	76.92	67.69	67.69	67.69	73.85	83.08	79.23	79.23	79.23	80.00	76.92	76.92

<https://doi.org/10.1371/journal.pone.0191227.t003>

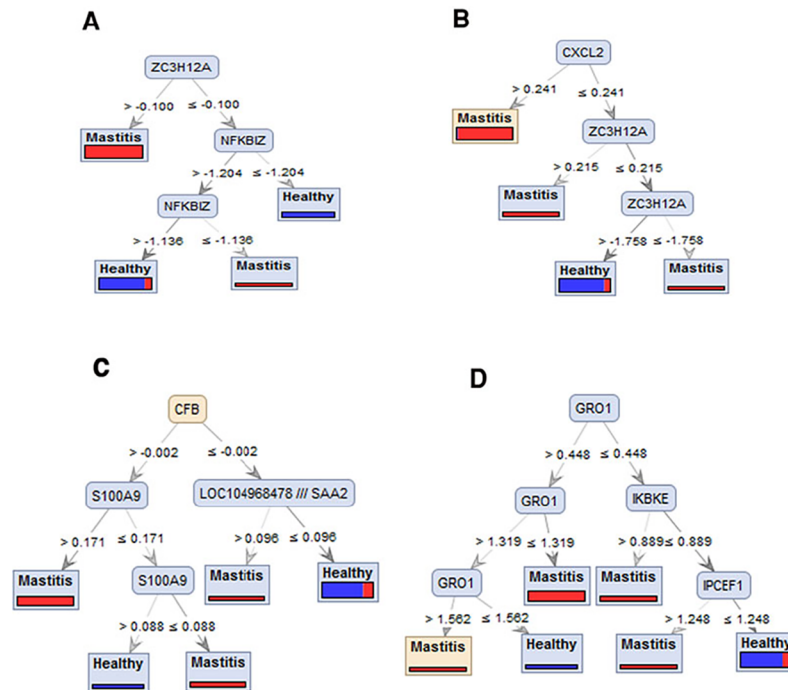


Fig 2. The architecture of different Decision tree models in predicting mastitis and healthy samples, based on the differentially expressed genes introduced by meta-analysis (A) Random Forest model with Gini Index criterion run on SVM dataset. (B) Random Forest model with accuracy criterion run on Gini Index dataset (C) Random Forest with Information Gain criterion run on Relief dataset and (D) Random Forest model with Gain Ratio criterion run on SVM dataset.

<https://doi.org/10.1371/journal.pone.0191227.g002>

analyze them to identify the most important genes and biomarker candidates in an efficient way [67–70]. Based on definition of biomarker, a “good” biomarker as an indicator must be specific for a disease and should remain unchanged by unrelated disorders. Moreover, reliable and reproducible biomarker quantifications must be demonstrated [17].

Here, we performed a meta-analysis on series of microarray gene expression datasets in order to enhance the power of analysis to identify genes that may be significantly involved in response to *E. coli* mastitis in dairy cows. Meta-analysis confirmed the most important findings in individual studies such as induction of the pathways related to immune response, inflammation, cytokines and chemokines signaling, acute phase proteins, proteolysis, response to wounding, apoptosis and cell signaling. It also suppressed several aspects of basic epithelial biology including extracellular matrix biosynthesis, mammary gland development markers and epidermis morphogenesis such as cholesterol, sterol and terpenoid biosynthesis [29, 45–49]. Importantly, based on our results, *E. coli* infection causes down-regulation of genes encoding lipid biosynthesis enzymes including *ALOX15*, *FASN*, *GPAM*, *TM7SF2* that are involved in milk production [37]. Generally, in infection, host metabolism is suppressed because the tissue has to divert energy to fight infection. Moreover, up-regulated meta-genes enriched novel biological pathways including responses to hypoxia, positive regulation of transcription from RNA polymerase II promoter and anti-apoptosis agents.

Low oxygen (O₂) environments are created by pathophysiological conditions including sites of infection and inflammation. In addition, pyruvate accumulation caused by inhibition of lipid metabolism has indeed been shown to stimulate hypoxia signaling in mastitis disease in dairy cattle [71]. In the previous studies, the results have demonstrated that stress-response

genes such as those responsible to immune-response pathways were enriched in paused RNA polymerase II [72]. For this reason, and due to this point that RNA polymerase II is essential for the transcription of many genes which up-regulated genes during *E. coli* infection, induced expression of genes related to positive regulation of transcription from RNA polymerase II promoter is necessary. Macrophages are the key players in innate immunity, and because of their crucial role in immunity, regulation of monocyte/macrophage lifespan is important in both physiological and pathological processes. Anti-apoptotic genes such as *Bcl2* family has been shown to be involved in the survival of monocytes/macrophages through enhancing the resistance of macrophages against various apoptotic stimuli [73].

In the current research, for the first time, the machine-learning approach were used to prioritize meta-genes to find the most important genes in response to *E. coli*-induced mastitis. The top-ranked meta-genes- *CXCL8 (IL8)*, *NFKBIZ*, *HP*, *CXCL2*, *CCL20*, *GRO1*, *ZC3H12A*, *PDE4B*, *CASP4*, *CFB*, *SA00A9*, *SA00A8*- that were listed in Table 2 play an important role in the immune defense, inflammation, and/or chemotaxis. Inflammatory chemokine interleukin-8 (IL-8), one of the most widely studied chemokines, is a critical inflammatory mediator and plays an important role in neutrophil migration into bovine mammary glands during mastitis [74, 75]. Furthermore, previous studies demonstrated IL-8 as an antibody therapeutic target in inflammatory diseases in human [76] and bovine mastitis [74].

I κ B ζ (also known as Molecule possessing ankyrin-repeats induced by lipopolysaccharide (MAIL) and INAP), encoded by the *NFKBIZ* gene, is a member of the nuclear I κ B family of proteins that act as transcriptional regulators via association with nuclear factor kappa B (NF- κ B)) [77]. The critical role of I κ B ζ signaling in the regulation of immune responses has been revealed previously [78, 79]. Like other I κ B proteins, I κ B ζ has inhibitory effects on the transcription of inflammatory genes regulated by NF- κ B such as tumor necrosis factor (TNF)- α , interleukin-1 (IL-1) [77, 80] and IL-17A production from CD4+ T cells [81]. Furthermore, it has been demonstrated that I κ B ζ is indispensable for the expression of a subset of genes activated in TLR/IL-1R signaling pathways [77]. Toll-like receptors (TLRs) recognize various bacterial cell wall components such as LPS, peptidoglycan (PGN) and lipopeptides, and trigger the inflammatory and immune responses against pathogens [82]. Investigations have revealed that function and gene polymorphisms of *NFKBIZ* can be introduced as potential markers of mastitis resistance in dairy heifers [83].

Already abbreviated Haptoglobin (HP), an acute phase protein mostly secreted by the liver, is synthesized within the mammary gland through stimulation by pro-inflammatory stimuli as it is in the liver [84]. HP has been introduced as a sensitive inflammatory marker for acute mastitis by numerous studies [84–86].

Pro-inflammatory cytokines, chemokines such as *CXCL2*, *CCL20*, and *GRO1*(*CXCL1*) have important roles in immune responses due to modulation of leukocyte infiltration (neutrophils and monocytes). *CXCL2* has been determined as a biomarker of the inflammatory reaction previously [87]. It has been suggested that *CXCL1* can be used as therapeutic targets, therapeutics, or biomarkers in mastitis [88]. According to our result and validation with DTs, as shown in Fig 2, *CXCL2* and *GRO1* have good abilities to separate mastitis and healthy samples with 83.85% accuracy; and they are good candidates to distinguish *E. coli* mastitis as a biomarker.

Zinc finger protein, *ZC3H12A*, has been shown as TLR-inducible gene to modulate LPS-induced inflammatory response [89]. It is also an RNase essential for the control of immune responses by regulating mRNA decay [90]. As shown in Fig 2 and based on DTs, *ZC3H12A* also has been identified as a potential biomarker for *E. coli* mastitis with 83.85% accuracy. However it needs more investigations at the protein level to be considered as a biomarker.

The *PDE4B2* is the short isoform of *PDE4* isoenzyme family. *PDE4* is cAMP-specific and the dominant *PDE* in inflammatory cells. Inhibition of *PDE4* elevates intracellular cAMP

levels, which inhibit the activity of promoters such as NF- κ B and down-regulation of the inflammatory responses by reducing the expression of TNF- α and other pro-inflammatory cytokines, while increasing anti-inflammatory cytokines such as IL-10 [91]. Interestingly, PDE4 inhibition is used as therapeutics for the treatment of inflammatory diseases in numerous studies [92, 93]

Caspases are a family of cysteine proteases that are highly conserved in multicellular organisms, functioning as central regulators of apoptosis [73]. Caspase-4 is classified as inflammatory caspases [94]. CASP4 has been shown to bind with LPS with high specificity and affinity directly and it is an innate immune receptor for intracellular LPS [94, 95]. It has been demonstrated that caspase-4 plays an important role in the classical LPS induced TLR4-signaling pathway, leading to NF- κ B dependent transcriptional up-regulation and secretion of important cytokines and chemokines in innate immune signaling in human monocytic cell [94]. Remarkably, CASP4 represents a new candidate for pattern recognition in immunity [95].

Complement factor B (CFB) an acute phase plasma protein is central to the action of the innate immune system in response to inflammation and infection and plays a role in B-cell activation and the cytotoxic reaction [86, 96, 97]. Research in bovine has demonstrated that the complete complement system can be found in colostrum, and components of the system are also present in the milk [97]. At present, attention is being focused on using acute phase proteins such as haptoglobin, serum amyloid A. [85, 98, 99], as biomarkers for the diagnosis of mastitis. However they are non-specific markers of the inflammatory process. CFB has been confirmed by DTs with 82.31% accuracy (Fig 2) and it may be a good candidate for the diagnosis of *E. coli* mastitis.

The role of last two top-ranked genes, SA00A9 and SA00A8, are inducing chemotaxis and adhesion of neutrophils [100] and play an important role in the innate immunity and tissue repair [101]. Moreover, these genes were identified as biomarkers for acute inflammation in infused and autoimmune disease [102, 103].

Due to the fact that the performances of the most mastitis detection systems do not satisfy the high accuracy required for practical clinical mastitis detection [25, 26], potential to include several biomarkers on one test strip or commercial kit might enhance the diagnostic efficiency of mastitis. Therefore, antibiotic therapy can, therefore, be chosen based on the mastitis pathogen and the type of mastitis. These results are valuable bioinformatics findings that need more laboratory based-studies to confirm.

Conclusions

This finding showed that the meta-analysis based on a large amount of original data represents an important contribution to our understanding of most informative genes for *E. coli* mastitis in cattle. Furthermore, this research properly indicated that the combination of machine learning with meta-analysis provides an opportunity to obtain a better resolution of the most important genes that might provide a more robust bio-signature and thereby may be good bio-marker candidates. Our results provide the basis for strategies to improve the diagnosis and treatment of the *E. coli* mastitis in the dairy cow.

Supporting information

S1 Table. Standard expression value of differentially expressed genes achieved by meta-analysis (Meta-genes) were set as features or attributes, this new dataset (Metad) was used to import into RapidMiner Studio software.

(XLSX)

S2 Table. Table is the PRISMA Checklist.

(DOC)

S3 Table. Differentially expressed genes identified after meta-analysis (one-tailed $q < 0.005$).

(XLSX)

S4 Table. Investigation of count of individual studies which show significant q -value (Benjamini & Hochberg corrected p -value, one tailed $q < 0.005$) for differentially expressed genes achieved after meta-analysis (meta-genes).

(XLSX)

S5 Table. The gene ontology (GO) information based on modified Fisher Exact test analysis (p -value < 0.05), revealed on up-regulated genes achieved by meta-analysis (meta-genes).

(XLSX)

S6 Table. The gene ontology (GO) information based on modified Fisher Exact test analysis (p -value < 0.05), revealed on down-regulated genes achieved by meta-analysis (meta-genes).

(XLSX)

S7 Table. All weights give after applying 10 attribute weighting algorithms (AW)s on differentially expressed genes achieved by meta-analysis (meta-genes) and count of algorithms with given weight above 0.7.

(XLSX)

Author Contributions

Conceptualization: Samaneh Fazeli Farsani.

Project administration: Somayeh Sharifi.

Supervision: Abbas Pakdel, Mansour Ebrahimi, James M. Reecy, Esmaeil Ebrahimie.

References

1. Heikkila AM, Nousiainen JI, Pyorala S. Costs of clinical mastitis with special reference to premature culling. *J Dairy Sci.* 2012; 95(1):139–50. <https://doi.org/10.3168/jds.2011-4321> PMID: 22192193
2. Fogsgaard KK, Bennedsgaard TW, Herskin MS. Behavioral changes in freestall-housed dairy cows with naturally occurring clinical mastitis. *J Dairy Sci.* 2015; 98(3):1730–8. <https://doi.org/10.3168/jds.2014-8347> PMID: 25547306
3. Peters MD, Silveira ID, Fischer V. Impact of subclinical and clinical mastitis on sensitivity to pain of dairy cows. *Animal.* 2015; 9(12):2024–8. <https://doi.org/10.1017/S1751731115001391> PMID: 26220469
4. Grohn YT, Wilson DJ, Gonzalez RN, Hertl JA, Schulte H, Bennett G, et al. Effect of pathogen-specific clinical mastitis on milk yield in dairy cows. *J Dairy Sci.* 2004; 87(10):3358–74. [https://doi.org/10.3168/jds.S0022-0302\(04\)73472-4](https://doi.org/10.3168/jds.S0022-0302(04)73472-4) PMID: 15377615
5. Bar D, Grohn YT, Bennett G, Gonzalez RN, Hertl JA, Schulte HF, et al. Effects of repeated episodes of generic clinical mastitis on mortality and culling in dairy cows. *J Dairy Sci.* 2008; 91(6):2196–204. <https://doi.org/10.3168/jds.2007-0460> PMID: 18487642
6. Schukken YH, Hertl J, Bar D, Bennett GJ, Gonzalez RN, Rauch BJ, et al. Effects of repeated gram-positive and gram-negative clinical mastitis episodes on milk yield loss in Holstein dairy cows. *J Dairy Sci.* 2009; 92(7):3091–105. <https://doi.org/10.3168/jds.2008-1557> PMID: 19528587
7. Kumar N, Manimaran A, Kumaresan A, Jeyakumar S, Sreela L, Mooventhan P, et al. Mastitis effects on reproductive performance in dairy cattle: a review. *Trop Anim Health Prod.* 2017; 49(4):663–73. <https://doi.org/10.1007/s11250-017-1253-4> PMID: 28283873

8. Ahmadzadeh A, Frago F, Shafii B, Dalton JC, Price WJ, McGuire MA. Effect of clinical mastitis and other diseases on reproductive performance of Holstein cows. *Anim Reprod Sci.* 2009; 112(3–4):273–82. <https://doi.org/10.1016/j.anireprosci.2008.04.024> PMID: 18554826
9. Hortet P, Seegers H. Loss in milk yield and related composition changes resulting from clinical mastitis in dairy cows. *Prev Vet Med.* 1998; 37(1–4):1–20. PMID: 9879576
10. Kossaibati MA, Esslemont RJ. The costs of production diseases in dairy herds in England. *Vet J.* 1997; 154(1):41–51. PMID: 9265852
11. Rajala-Schultz PJ, Grohn YT. Culling of dairy cows. Part I. Effects of diseases on culling in Finnish Ayrshire cows. *Prev Vet Med.* 1999; 41(2–3):195–208. PMID: 10448946
12. Pyorala S. Treatment of mastitis during lactation. *Ir Vet J.* 2009; 62 Suppl 4:S40–4.
13. USDA.Dairy. Milk Quality, Milking Procedures, and Mastitis in the United States, Fort Collins. USDA–APHIS–VS–CEAH–NAHMS; 2016.
14. Bannerman DD, J PM, Jai-Wei L, Xin Z, Hope JC, P R. *Escherichia coli* and *Staphylococcus aureus* Elicit Differential Innate Immune Responses following Intramammary Infection. *Clin Diagn Lab Immunol.* 2004; 11(3):463–72. <https://doi.org/10.1128/CDLI.11.3.463-472.2004> PMID: 15138171
15. Hagiwara S, Mori K, Nagahata H. Predictors of fatal outcomes resulting from acute *Escherichia coli* mastitis in dairy cows. *J Vet Med Sci.* 2016; 78(5):905–8. <https://doi.org/10.1292/jvms.15-0400> PMID: 26875836
16. Bramley AJ, Dodd FH. Reviews of the progress of dairy science: mastitis control—progress and prospects. *J Dairy Res.* 1984; 51(3):481–512. PMID: 6381562
17. Duarte CM, Freitas PP, Bexiga R. Technological advances in bovine mastitis diagnosis: an overview. *J Vet Diagn Invest.* 2015; 27(6): 665–72. <https://doi.org/10.1177/1040638715603087> PMID: 26450837
18. Lai YC, Fujikawa T, Maemura T, Ando T, Kitahara G, Endo Y, et al. Inflammation-related microRNA expression level in the bovine milk is affected by mastitis. *PLoS One.* 2017; 12(5)(5):e0177182. <https://doi.org/10.1371/journal.pone.0177182> PMID: 28520748
19. Oliver SP, Murinda SE. Antimicrobial resistance of mastitis pathogens. *Vet Clin North Am Food Anim Pract.* 2012; 28(2):165–85. <https://doi.org/10.1016/j.cvfa.2012.03.005> PMID: 22664201
20. Mohammadi A, Saraee MH, Salehi M. Identification of disease-causing genes using microarray data mining and Gene Ontology. *BMC Med Genomics.* 2011; 4:12. <https://doi.org/10.1186/1755-8794-4-12> PMID: 21269461
21. Lewis JD. The utility of biomarkers in the diagnosis and therapy of inflammatory bowel disease. *Gastroenterology.* 2011; 140(6):1817–26 e2. <https://doi.org/10.1053/j.gastro.2010.11.058> PMID: 21530748
22. Ceciliani F, Ceron JJ, Eckersall PD, Sauerwein H. Acute phase proteins in ruminants. *J Proteomics.* 2012; 75(14):4207–31. <https://doi.org/10.1016/j.jprot.2012.04.004> PMID: 22521269
23. Smolenski GA, Wieliczko RJ, Pryor SM, Broadhurst MK, Wheeler TT, Haigh BJ. The abundance of milk cathelicidin proteins during bovine mastitis. *Vet Immunol Immunopathol.* 2011; 143(1–2):125–30. <https://doi.org/10.1016/j.vetimm.2011.06.034> PMID: 21774993
24. Kawai K, Akamatsu H, Obayashi T, Nagahata H, Higuchi H, Iwano H, et al. Relationship between concentration of lingual antimicrobial peptide and somatic cell count in milk of dairy cows. *Vet Immunol Immunopathol.* 2013; 153(3–4):298–301. <https://doi.org/10.1016/j.vetimm.2013.03.002> PMID: 23528609
25. Hogeveen H, Kamphuis C, Steeneveld W, Mollenhorst H. Sensors and clinical mastitis—the quest for the perfect alert. *Sensors (Basel).* 2010; 10(9):7991–8009.
26. Jensen DB, Hogeveen H, De Vries A. Bayesian integration of sensor information and a multivariate dynamic linear model for prediction of dairy cow mastitis. *J Dairy Sci.* 2016; 99(9):7344–61. <https://doi.org/10.3168/jds.2015-10060> PMID: 27320667
27. Ehinger AM, Kietzmann M. [Pharmacokinetic aspects of mastitis therapy]. *Berl Munch Tierarztl Wochenschr.* 1998; 111(9):337–43. PMID: 9789371
28. Pieterse R, Todorov SD. Bacteriocins—exploring alternatives to antibiotics in mastitis treatment. *Braz J Microbiol.* 2010; 41(3):542–62. <https://doi.org/10.1590/S1517-83822010000300003> PMID: 24031528
29. Sipka A, Klaessig S, Duhamel GE, Swinkels J, Rainard P, Schukken Y. Impact of intramammary treatment on gene expression profiles in bovine *Escherichia coli* mastitis. *PLoS One.* 2014; 9(1):e85579. <https://doi.org/10.1371/journal.pone.0085579> PMID: 24454893
30. Suojala L, Simojoki H, Mustonen K, Kaartinen L, Pyorala S. Efficacy of enrofloxacin in the treatment of naturally occurring acute clinical *Escherichia coli* mastitis. *J Dairy Sci.* 2010; 93(5):1960–9. <https://doi.org/10.3168/jds.2009-2462> PMID: 20412909

31. Ebert F, Staufenbiel R, Simons J, Pieper L. Randomized, blinded, controlled clinical trial shows no benefit of homeopathic mastitis treatment in dairy cows. *J Dairy Sci.* 2017; 100(6):4857–67. <https://doi.org/10.3168/jds.2016-11805> PMID: 28342609
32. Burvenich C, Van Merris V, Mehrzad J, Diez-Fraile A, Duchateau L. Severity of *E. coli* mastitis is mainly determined by cow factors. *Vet Res.* 2003; 34(5):521–64. <https://doi.org/10.1051/vetres:2003023> PMID: 14556694
33. Wenz JR, Barrington GM, Garry FB, Ellis RP, Magnuson RJ. *Escherichia coli* isolates' serotypes, genotypes, and virulence genes and clinical coliform mastitis severity. *J Dairy Sci.* 2006; 89(9):3408–12. [https://doi.org/10.3168/jds.S0022-0302\(06\)72377-3](https://doi.org/10.3168/jds.S0022-0302(06)72377-3) PMID: 16899673
34. Ramasamy A, Mondry A, Holmes CC, Altman DG. Key issues in conducting a meta-analysis of gene expression microarray datasets. *PLoS Med.* 2008; 5(9):e184. <https://doi.org/10.1371/journal.pmed.0050184> PMID: 18767902
35. Mimoso C, Lee DD, Zavadil J, Tomic-Canic M, Blumenberg M. Analysis and meta-analysis of transcriptional profiling in human epidermis. *Methods Mol Biol.* 2014; 1195:61–97. https://doi.org/10.1007/7651_2013_60 PMID: 24297317
36. Genini S, Badaoui B, Sclep G, Bishop SC, Waddington D, Pinard van der Laan MH, et al. Strengthening insights into host responses to mastitis infection in ruminants by combining heterogeneous microarray data sources. *BMC Genom.* 2011; 12(1):225.
37. Younis S, Javed Q, Blumenberg M. Meta-Analysis of Transcriptional Responses to Mastitis-Causing *Escherichia coli*. *PLoS One.* 2016; 11(3):e0148562. <https://doi.org/10.1371/journal.pone.0148562> PMID: 26933871
38. Hall MA, Holmes G. Benchmarking Attribute Selection Techniques for Discrete Class Data Mining. *IEEE Trans Knowl Data Eng (TKDE)*,. 2003; 15(3):1041–57.
39. Thai KM, Ecker GF. Similarity-based SIBAR descriptors for classification of chemically diverse hERG blockers. *Mol Divers.* 2009; 13(3):321–36. <https://doi.org/10.1007/s11030-009-9117-0> PMID: 19219559
40. Dancey D, Bandar ZA, McLean D. Logistic model tree extraction from artificial neural networks. *IEEE Trans Syst Man Cybern B Cybern.* 2007; 37(4):794–802. PMID: 17702280
41. Ebrahimi M, Ebrahimie E, Ebrahimi M. Searching for patterns of thermostability in proteins and defining the main features contributing to enzyme thermostability through screening, clustering, and decision tree algorithms. *EXCLI J.* 2009; 8:218–33.
42. Ashrafi E, Alemzadeh A, Ebrahimi M, Ebrahimie E, Dadkhodaei N, Ebrahimi M. Amino Acid Features of P1B-ATPase Heavy Metal Transporters Enabling Small Numbers of Organisms to Cope with Heavy Metal Pollution. *Bioinform Biol Insights.* 2011; 5:59–82. <https://doi.org/10.4137/BBI.S6206> PMID: 21573033
43. Bakhtiarzadeh MR, Moradi-Shahrbabak M, Ebrahimi M, Ebrahimie E. Neural network and SVM classifiers accurately predict lipid binding proteins, irrespective of sequence homology. *J Theor Biol.* 2014; 356:213–22. <https://doi.org/10.1016/j.jtbi.2014.04.040> PMID: 24819464
44. Zinati Z, Zamansani F, Hossein KayvanJoo A, Ebrahimi M, Ebrahimi M, Ebrahimie E, et al. New layers in understanding and predicting alpha-linolenic acid content in plants using amino acid characteristics of omega-3 fatty acid desaturase. *Comput Biol Med.* 2014; 54:14–23. <https://doi.org/10.1016/j.combiomed.2014.08.019> PMID: 25199845
45. Brand B, Hartmann A, Reipsilber D, Griesbeck-Zilch B, Wellnitz O, Kuhn C, et al. 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. *Genet Sel Evol.* 2011; 43:24. <https://doi.org/10.1186/1297-9686-43-24> PMID: 21702919
46. Mitterhuemer S, Petzl W, Krebs S, Mehne D, Klanner A, Wolf E, et al. *Escherichia coli* infection induces distinct local and systemic transcriptome responses in the mammary gland. *BMC Genom.* 2010; 11:138.
47. Buitenhuis B, Rontved CM, Edwards SM, Ingvarsen KL, Sorensen P. In depth analysis of genes and pathways of the mammary gland involved in the pathogenesis of bovine *Escherichia coli*-mastitis. *BMC Genom.* 2011; 12:130.
48. Gunther J, Esch K, Poschadel N, Petzl W, Zerbe H, Mitterhuemer S, et al. 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.* 2011; 79(2):695–707. <https://doi.org/10.1128/IAI.01071-10> PMID: 21115717
49. Gunther J, Petzl W, Zerbe H, Schuberth HJ, Koczan D, Goetze L, et al. Lipopolysaccharide priming enhances expression of effectors of immune defence while decreasing expression of pro-inflammatory cytokines in mammary epithelia cells from cows. *BMC Genom.* 2012; 13:17.

50. Alisoltani A, Fallahi H, Ebrahimi M, Ebrahimi M, Ebrahimie E. Prediction of potential cancer-risk regions based on transcriptome data: towards a comprehensive view. *PLoS One*. 2014; 9(5):e96320. <https://doi.org/10.1371/journal.pone.0096320> PMID: 24796549
51. Bakhtiarzadeh MR, Moradi-Shahrbabak M, Ebrahimie E. Underlying functional genomics of fat deposition in adipose tissue. *Gene*. 2013; 521(1):122–8. <https://doi.org/10.1016/j.gene.2013.03.045> PMID: 23523858
52. Hosseinpour B, Bakhtiarzadeh MR, Khosravi P, Ebrahimie E. Predicting distinct organization of transcription factor binding sites on the promoter regions: a new genome-based approach to expand human embryonic stem cell regulatory network. *Gene*. 2013; 531(2):212–9. <https://doi.org/10.1016/j.gene.2013.09.011> PMID: 24042128
53. Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP. Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res*. 2003; 31(4):e15. PMID: 12582260
54. Gautier L, Cope L, Bolstad BM, Irizarry RA. Affy- analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics*. 2004; 20(3):307–15. <https://doi.org/10.1093/bioinformatics/btg405> PMID: 14960456
55. Wang X, Kang DD, Shen K, Song C, Lu S, Chang LC, et al. An R package suite for microarray meta-analysis in quality control, differentially expressed gene analysis and pathway enrichment detection. *Bioinformatics*. 2012; 28(19):2534–6. <https://doi.org/10.1093/bioinformatics/bts485> PMID: 22863766
56. Gunther J, Koczan D, Yang W, Nurnberg G, Reipsilber D, Schuberth HJ, et al. Assessment of the immune capacity of mammary epithelial cells: comparison with mammary tissue after challenge with *Escherichia coli*. *Vet Res*. 2009; 40(4):31. <https://doi.org/10.1051/vetres/2009014> PMID: 19321125
57. Swanson KM, Stelwagen K, Dobson J, Henderson HV, Davis SR, Farr VC, et al. 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*. 2009; 92(1):117–29. <https://doi.org/10.3168/jds.2008-1382> PMID: 19109270
58. Kempf F, Slugocki C, Blum SE, Leitner G, Germon P. Genomic Comparative Study of Bovine Mastitis *Escherichia coli*. *PLoS One*. 2016; 11(1):e0147954. <https://doi.org/10.1371/journal.pone.0147954> PMID: 26809117
59. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. *Introduction to Meta-Analysis*. The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom: John Wiley & Sons, Ltd; 2009.
60. Song C, Tseng GC. Hypothesis Setting and Order Statistic for Robust Genomic Meta-Analysis. *Ann Appl Stat*. 2014; 8(2):777–800. PMID: 25383132
61. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate—a Practical and Powerful Approach to Multiple Testing. *J R Stat Soc Series B Stat Methodol*. 1995; 57(1):289–300.
62. Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res*. 2009; 37(1):1–13. <https://doi.org/10.1093/nar/gkn923> PMID: 19033363
63. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009; 4:44–57. <https://doi.org/10.1038/nprot.2008.211> PMID: 19131956
64. Yoon S, Yang Y, Choi J, Seong J. Large scale data mining approach for gene-specific standardization of microarray gene expression data. *Bioinformatics*. 2006; 22(23):2898–904. <https://doi.org/10.1093/bioinformatics/btl500> PMID: 17032674
65. Ebrahimi E, Ebrahimi M, Sarvestani NR, Ebrahimi M. Protein attributes contribute to halo-stability, bioinformatics approach. *Saline Systems*. 2011; 7:1. <https://doi.org/10.1186/1746-1448-7-1> PMID: 21592393
66. Ebrahimi M, Lakizadeh A, Agha-Golzadeh P, Ebrahimie E, Ebrahimi M. Prediction of thermostability from amino acid attributes by combination of clustering with attribute weighting: a new vista in engineering enzymes. *PLoS One*. 2011; 6(8):e23146. <https://doi.org/10.1371/journal.pone.0023146> PMID: 21853079
67. Moreau Y, Aerts S, De Moor B, De Strooper B, Dabrowski M. Comparison and meta-analysis of microarray data: from the bench to the computer desk. *Trends Genet*. 2003; 19(10):570–7. PMID: 14550631
68. Joyce AR, Palsson BO. The model organism as a system: integrating ‘omics’ data sets. *Nat Rev Mol Cell Biol*. 2006; 7:198–210. <https://doi.org/10.1038/nrm1857> PMID: 16496022
69. Ebrahimie E, Nurollah Z, Ebrahimi M, Hemmatzadeh F, Ignjatovic J. Unique ability of pandemic influenza to downregulate the genes involved in neuronal disorders. *Mol Biol Rep*. 2015; 42(9):1377–90. <https://doi.org/10.1007/s11033-015-3916-4> PMID: 26246405

70. Pashaiasl M, Ebrahimi M, Ebrahimie E. Identification of the key regulating genes of diminished ovarian reserve (DOR) by network and gene ontology analysis. *Mol Biol Rep.* 2016; 43(9):923–37. <https://doi.org/10.1007/s11033-016-4025-8> PMID: 27324248
71. Intiyaz HZ, Simon MC. Hypoxia-inducible factors as essential regulators of inflammation. *Curr Top Microbiol Immunol.* 2010; 345:105–20. https://doi.org/10.1007/82_2010_74 PMID: 20517715
72. Guenther MG, Levine SS, Boyer LA, Jaenisch R, Young RA. A chromatin landmark and transcription initiation at most promoters in human cells. *Cell.* 2007; 130(1):77–88. <https://doi.org/10.1016/j.cell.2007.05.042> PMID: 17632057
73. Busca A, Saxena M, Kryworuchko M, Kumar A. Anti-Apoptotic Genes in the Survival of Monocytic Cells During Infection. *Curr Genomics.* 2009; 10:306–17. <https://doi.org/10.2174/138920209788920967> PMID: 20119528
74. Barber MR, Yang TJ. Chemotactic Activities in Nonmastitic and Mastitic Mammary Secretions: Presence of Interleukin-8 in Mastitic but Not Nonmastitic Secretions. *Clon Diagn lab Immunol.* 1998;Vol. 5 (1):82–6. PMID: 9455886
75. Lee J, Zhao X. Recombinant human interleukin-8, But not human interleukin-1, induces bovine neutrophil migration in an invitro co-culture system. *Cell Biol Int.* 2000;Vol. 24 (12) 889–95. <https://doi.org/10.1006/cbir.2000.0562> PMID: 11114238
76. Skov L, Beurskens FJ, Zachariae COC, Reitamo S, Teeling J, Satiijn D, et al. IL-8 as Antibody Therapeutic Target in Inflammatory Diseases: Reduction of Clinical Activity in Palmoplantar Pustulosis. *J Immunol.* 2008; 181:669–79. PMID: 18566434
77. Yamamoto M, Yamazaki S, Uematsu S, Sato S, Hemmi H, Hoshino K, et al. Regulation of Toll/IL-1-receptormediated gene expression by the inducible nuclear protein IκBζ. *Nature.* 8 July 2004; 430:218–22. <https://doi.org/10.1038/nature02738> PMID: 15241416
78. Okamoto K, Iwai Y, Oh-hora M, Yamamoto M, Morio T, Aoki K, et al. IκBf regulates TH17 development by cooperating with ROR nuclear receptors. *Nature.* 2010; 464:1381–5. <https://doi.org/10.1038/nature08922> PMID: 20383124
79. Yamazaki S, Matsuo S, Muta T, Yamamoto M, Akira S, Takeshige K. Gene-Specific Requirement of a Nuclear Protein, IκBζ, for Promoter Association of Inflammatory Transcription regulator. *J Biol chem.* 2008; 283:32404–11. <https://doi.org/10.1074/jbc.M802148200> PMID: 18824552
80. Motoyama M, Yamazaki S, Eto-Kimura A, Takeshige K, Muta T. Positive and Negative Regulation of Nuclear Factor-κB-mediated Transcription by IκBζ, an Inducible Nuclear Protein. *J Biol Chem.* 4 March 2005 280(9):7444–51.
81. Kim Y, Lee Y, Yang J, Lee S, Park Y, Kweon M. The resident pathobiont *Staphylococcus xylosum* in Nfbiz-deficient skin accelerates spontaneous skin infammation. *Nature.* 24 July 2017;Scientific Reports 7:Article number 6348.
82. Takeda K, Kaisho T, Akira S. Toll-like Receptors. *Annu Rev Immunol.* 2003; 21:335–76. <https://doi.org/10.1146/annurev.immunol.21.120601.141126> PMID: 12524386
83. Compton CW, Cursons RT, Barnett CM, McDougall S. Expression of innate resistance factors in mammary secretion from periparturient dairy heifers and their association with subsequent infection status. *Vet Immunol Immunopathol.* 2009;Vol. 127(3–4):357–64. <https://doi.org/10.1016/j.vetimm.2008.10.331> PMID: 19070369
84. Hiss S, Mielenz M, Bruckmaier RM, Sauerwein H. Haptoglobin Concentrations in Blood and Milk After Endotoxin Challenge and Quantification of Mammary Hp mRNA Expression. *J Dairy Sci.* 2004; 87:3778–84. [https://doi.org/10.3168/jds.S0022-0302\(04\)73516-X](https://doi.org/10.3168/jds.S0022-0302(04)73516-X) PMID: 15483161
85. Lauzon K, Zhao X, Lacasse P. Deferoxamine Reduces Tissue Damage During Endotoxin-Induced Mastitis in Dairy Cows. *J Dairy Sci.* 2006; 89(10):3846–57. [https://doi.org/10.3168/jds.S0022-0302\(06\)72427-4](https://doi.org/10.3168/jds.S0022-0302(06)72427-4) PMID: 16960060
86. Eckersall PD, Young FJ, Nolan AM, Knight CH, McComb C, Waterston MM, et al. Acute Phase Proteins in Bovine Milk in an Experimental Model of *Staphylococcus aureus* Subclinical Mastitis. *J Dairy Sci.* 2006; 89:1488–501. [https://doi.org/10.3168/jds.S0022-0302\(06\)72216-0](https://doi.org/10.3168/jds.S0022-0302(06)72216-0) PMID: 16606719
87. Yagdiran Y, Tallkvist J, Artursson K, Oskarsson A. *Staphylococcus aureus* and Lipopolysaccharide Modulate Gene Expressions of Drug Transporters in Mouse Mammary Epithelial Cells Correlation to Inflammatory Biomarkers. *PLoS One.* 2016; 11(9):e0161346. <https://doi.org/10.1371/journal.pone.0161346> PMID: 27584666
88. Johnzon C, Artursson K, Söderlund R, Guss B, Rönnerberg E, Gunnar Pejler G. Mastitis Pathogens with high Virulence in a Mouse Model Produce a Distinct cytokine Profile In Vivo. *Front Microbiol.* 2016; 7: e00368.
89. Noman AS, Koide N, Iftakhar-E-Khuda I, Dagvadorj J, Tumurkhuu G, Naiki Y, et al. Retinoblastoma protein-interacting zinc finger 1, a tumor suppressor, augments lipopolysaccharide-induced

- proinflammatory cytokine production via enhancing nuclear factor-kappaB activation. *Cell Immunol.* 2010; 264(2):114–8. <https://doi.org/10.1016/j.cellimm.2010.05.007> PMID: 20557878
90. Matsushita K, Takeuchi O, Standley DM, Kumagai Y, Kawagoe T, Miyake T, et al. Zc3h12a is an RNase essential for controlling immune responses by regulating mRNA decay. *Nature.* 2009; 458(7242):1185–90. <https://doi.org/10.1038/nature07924> PMID: 19322177
 91. Schafer P. Apremilast mechanism of action and application to psoriasis and psoriatic arthritis. *Biochem Pharmacol.* 2012; 83:1583–90. <https://doi.org/10.1016/j.bcp.2012.01.001> PMID: 22257911
 92. Houslay MD, Schafer P, Zhang KYJ. Phosphodiesterase-4 as a therapeutic target. *DDT.* 2005; 10(22):1503–19. [https://doi.org/10.1016/S1359-6446\(05\)03622-6](https://doi.org/10.1016/S1359-6446(05)03622-6) PMID: 16257373
 93. Schafer PH, Parton A, Capone L, Cedzik D, Brady H, Evans JF, et al. Apremilast is a selective PDE4 inhibitor with regulatory effects on innate immunity. *cell signal.* 2014;26.
 94. Lakshmanan U, Porter AG. Caspase-4 Interacts with TNF Receptor-Associated Factor 6 and Mediates Lipopolysaccharide-Induced NF- κ B-Dependent Production of IL-8 and CC Chemokine Ligand 4 (Macrophage-Inflammatory Protein-1). *J Immunol.* 2007; 179:8480–90. PMID: 18056395
 95. Shi J, Zhao Y, Wang Y, Gao W, Ding J, Li P, et al. Inflammatory caspases are innate immune receptors for intracellular LPS. *Nature.* 2014; 514(7521):187–92. <https://doi.org/10.1038/nature13683> PMID: 25119034
 96. Kolb WP, Morrow PR, Tamerius JD. Ba and Bb fragments of factor B activation: fragment production, biological activities, neopeptide expression and quantitation in clonical samples. *Complement Inflamm.* 1989; 6(3):175–204. PMID: 2472921
 97. Jacob T, Subramani G, Sivaprakasam P, Xavier AP, Mukhopadhyay HK. Immuno-Detection of C3a, a C3 Complement Activated Product in Mastitis Milk, a Potential Diagnostic Marker. *Vet Sci.* 2017; 4(1):2017.
 98. Eckersall PD, Young FJ, Nolan AM, Knight CH, McComb C, Waterston MM, et al. Acute Phase Proteins in Bovine Milk in an Experimental Model of *Staphylococcus aureus* Subclinical Mastitis. *J Dairy Sci.* 2006 89(5):1488–501. [https://doi.org/10.3168/jds.S0022-0302\(06\)72216-0](https://doi.org/10.3168/jds.S0022-0302(06)72216-0) PMID: 16606719
 99. Hiss S, Mielenz M, Bruckmaier RM, Sauerwein H. Haptoglobin concentrations in blood and milk after endotoxin challenge and quantification of mammary Hp mRNA expression. *J Dairy Sci* 2004 87(11):3778–84. [https://doi.org/10.3168/jds.S0022-0302\(04\)73516-X](https://doi.org/10.3168/jds.S0022-0302(04)73516-X) PMID: 15483161
 100. Boyd JH, Kan B, Roberts H, Wang Y, Walley KR. S100A8 and S100A9 mediate endotoxin-induced cardiomyocyte dysfunction via the receptor for advanced glycation end products. *Circ Res.* 2008; 102(10):1239–46. <https://doi.org/10.1161/CIRCRESAHA.107.167544> PMID: 18403730
 101. Kim AJ, Ro H, Kim H, Chang JH, Lee HH, Chung W, et al. Klotho and S100A8/A9 as Discriminative Markers between Pre-Renal and Intrinsic Acute Kidney Injury. *PLoS One.* 2016; 11(1):e0147255. <https://doi.org/10.1371/journal.pone.0147255> PMID: 26799323
 102. Drynda S, Ringel B, Kekow M, Kühne C, Drynda A, Glocker MO, et al. Proteome analysis reveals disease-associated marker proteins to differentiate RA patients from other inflammatory joint diseases with the potential to monitor anti-TNFalpha therapy. *Pathol Res Pract.* 2004; 200(2):165–71. PMID: 15237925
 103. Gebhardt C, Németh J, Angel P, Hess J. S100A8 and S100A9 in inflammation and cancer. *Biochem Pharmacol.* 2006; 72(11):1622–31. <https://doi.org/10.1016/j.bcp.2006.05.017> PMID: 16846592