GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

Deep sequencing-based analysis of gene expression in bovine mammary epithelial cells after *Staphylococcus aureus*, *Escherichia coli*, and *Klebsiella pneumoniae* infection

L. Xiu, Y.B. Fu, Y. Deng, X.J. Shi, Z.Y. Bian, A. Ruhan and X. Wang

College of Life Sciences, Inner Mongolia University, Hohhot, China

Corresponding author: X. Wang
E-mail: wxiao2008@gmail.com

Genet. Mol. Res. 14 (4): 16948-16965 (2015)
Received January 14, 2015
Accepted May 25, 2015
Published December 15, 2015
DOI http://dx.doi.org/10.4238/2015.December.15.1

ABSTRACT. The goal of this study was to characterize the transcriptome of primary bovine mammalian epithelial cells (pBMECs) and to identify candidate genes for response and resistance to *Staphylococcus aureus* (strain S108), *Escherichia coli* (strain E23), and *Klebsiella pneumoniae* (strain K96) infection. Using Solexa sequencing, approximately 4.9 million total sequence tags were obtained from each of the three infected libraries and the control library. Gene Ontology (GO) analysis of the S108-infected pBMECs showed differentially expressed genes (DEGs) were significantly involved in metabolic processes. In E23-infected pBMECs, DEGs were predominantly associated with cell death and programmed cell death GO terms, while in K96-infected pBMECs, DEGs were primarily involved in metabolic processes and in utero embryonic development. Analysis of the cluster of orthologous groups of proteins showed that the S108-infected, E23-infected and K96-infected pBMECs were significantly involved in "Translation, ribosomal structure and biogenesis", "General function prediction only" and "Replication, recombination and repair". The transcriptome sequences were also annotated for KEGG orthology, and it was found that DEGs in S108-infected pBMECs were significantly

involved in oxidative phosphorylation and Parkinson's disease. The clustered pathway terms of the DEGs of the E23-infected pBMECs were found to involve the NOD-like receptor signaling pathway and oxidative phosphorylation, while those of the K96-infected pBMECs were primarily involved in oxidative phosphorylation and apoptosis. Our results have identified a number of immune-related genes that showed changes in gene expression after bacterial infection, and provided insight into the interactions between pBMECs and the bacteria.

Key words: Bovine mammary epithelial cells; Deep sequencing; *Escherichia coli*; Gene expression; *Klebsiella pneumoniae*; *Staphylococcus aureus*

INTRODUCTION

Mastitis is the most economically important disease facing dairy producers worldwide costing an estimated $2 billion annually in the United States alone (Viguier et al., 2009; Jensen et al., 2013). A variety of different bacteria and fungi can cause mastitis and the etiology of the infection influences its outcome. Gram-negative coliform bacteria most frequently cause acute inflammation and, eventually, severe mastitis (Burvenich et al., 2003). Gram-positive bacteria, such as *Staphylococcus aureus* and *Streptococcus uberis* cause persistent, almost chronic, infections with the pathogens surviving inside host cells (Hensen et al., 2000); these pathogens often cause only mild symptoms.

Current mastitis control programs, which are based on milking time hygiene, antibiotic therapy, and culling of persistently infected cows, have resulted in considerable reduction of contagious mastitis pathogens such as *Streptococcus agalactiae* and *S. aureus*. However, these same procedures have had only a marginal effect on environmental pathogens such as *Escherichia coli* and *Klebsiella pneumoniae* because of their ability to survive in the environment of dairy cows. Thus, environmental mastitis pathogens have become a major problem of well-managed dairy farms following the reduction in contagious mastitis pathogens (Hogan et al., 1989). The mechanisms underlying host responses to the pathogen species *S. aureus*, *E. coli*, and *K. pneumoniae* are largely unknown. The persistent infections cause large financial losses (Swinkels et al., 2005; Huijps et al., 2008) and, hence, it would be highly desirable to understand the mechanisms that prevent an effective immune defense against pathogens causing subclinical mastitis.

The on-going developments in mammalian genome resources and in high-throughput deep-sequencing technologies have expanded the potential range of analyses of gene expression changes, such as those *in vivo* as a result of mastitis due to *S. aureus*, *E. coli*, or *K. pneumoniae* infection. Digital gene expression (DGE) tag profiling allows one to identify millions of short RNAs and differentially expressed genes (DEGs) in a sample without the need for prior annotations (Wang et al., 2009). Sequencing-based methods measure absolute gene expression and avoid many inherent limitations of earlier microarray-based assays (Wang et al., 2011).

Bovine mammary epithelial cells (BMECs) produce milk and contribute significantly to immunity in the mammary glands (Aitken et al., 2011). Cell models of host-pathogen interactions are much easier to standardize, promise better reproducibility, and eventually

allow the analytical dissection of molecular mechanisms. With regard to immune defenses against mastitis, the immune relevance and capacity of milk cells, including sentinel cells and granulocytes, have been extensively characterized (Paape et al., 2000; Dosogne et al., 2001; Rainard et al., 2000; Burvenich et al., 2003). However, the immune relevance of MECs has only recently been recognized (Lahouassa et al., 2007). The principle function of MECs is milk formation during lactation. Thus, MECs are the dominant cell type in healthy, uninfected milk parenchyma and are most likely the first cells to be confronted with a pathogen following its entry into the milk parenchyma. BMECs express many inflammatory mediators, such as cytokines and chemokines capable of mobilizing the appropriate defense strategies against invading pathogens (Alva-Murillo et al., 2013); their behavior is reminiscent of other epithelial tissues such as intestinal and respiratory epithelial tissues in which inflammatory responses have been demonstrated to mobilize neutrophils against microbial invaders (Dwinell et al., 2003; Moreilhon et al., 2004).

Analysis of primary isolates of BMECs (pBMECs) allows examination of the specific contribution of MECs to the immune capacity of the udder. The principle immune relevance of such cultures has recently been established by the demonstration of inflammation dependent enhanced expression of a variety of candidate genes for immune reactions, such as cytokines, chemokines, toll-like receptors and beta-defensins (Pareek et al., 2005; Strandberg et al., 2005; Yang et al., 2008). However, there are no reports describing the global transcriptional response of pBMECs after a pathogen challenge.

In the current study, to avoid variations in gene expression due to age, gender, and individual variability of cows (Jensen et al., 2013), we examined the responses of pBMECs after infection with strains of three bacterial species (S108, E23 or K96), using the Illumina Genome Analyzer platform. Moreover, we screened the cells at 4 h after infection to identify differentially expressed transcripts in order to evaluate the early host response to the bacteria. Our data provide new insights into the complex bovine molecular pathways elicited by different bacterial species and to the role of these pathways in establishing host immune responses to mastitis.

MATERIAL AND METHODS

Bacteria and cell strain

E. coli (*E23*), *S. aureus* (*S56*) and *K. pneumoniae* (*K96*) (Capsular serotyping K2) were isolated originally from milk of cows with acute or persistent mastitis and were kindly provided by Professor YongQing Hao of Inner Mongolia Agricultural University. The three stains were grown on brain-heart infusion agar plates at 37°C. pBMECs were prepared from the udders of healthy, uninfected lactating cows and were grown in cell growth media in T75 tissue culture flasks as described previously and incubated at 37°C under 5% CO_2 in supplemented RPMI 1640 (10% fetal bovine serum, 0.2 mM L-glutamine, 100 U penicillin/mL, and 100 gstreptomycin/mL).

Cell infection

pBMECs were plated in 75 cm^2 flasks (6 x 10^6 cells per flask) in supplemented RPMI 1640 without antibiotics 1 day prior to infection. To infect the cells, they were co-cultured with the *E. coli*, *S. aureus*, or *K. pneumoniae* strains at 37°C under 5% CO_2 for 4 h at a multiplicity of infection of 100; no bacteria were added to negative control cultures.

RNA extraction for DGE

RNA samples were collected at 4 h post infection from three independent cultures of each treatment group. The cells were collected and centrifuged at 1500 rpm for 15 min. Control cells and adherent cells were homogenized in 1 mL TRIzol reagent (Tiangen, China) and total RNA was extracted following the manufacturer instructions. The RNA samples were incubated for 20 min at 37°C with 10 U of DNaseI (Takara, Japan) to remove residual genomic DNA. RNA integrity was determined by gel electrophoresis and RNA concentrations were quantified by measuring absorbance at 260 and 280 nm.

DGE design

We added Oligo (dT) magnetic beads to the total RNA to purify mRNA, and then used the Oligo (dT) as a primer to synthesize cDNA (Figure 1). The 5' ends of tags were generated using *Nla*III, which recognizes and excises CATG sites. Illumina adaptor 1 was ligated to the sticky 5' end of the digested bead-bound cDNA fragments; the junction of Illumina adaptor 1 and the CATG site acts as the recognition site for *Mme*I, which cuts 17 bp downstream of the CATG site and produces tags with adaptor 1. After removing the 3' fragments using magnetic bead precipitation, Illumina adaptor 2 was ligated to the 3' ends of the tags to generate tags with different adaptors of each end to form the tag library. The fragments were subjected to linear PCR amplification, purified by PAGE, and used for Illumina sequencing.

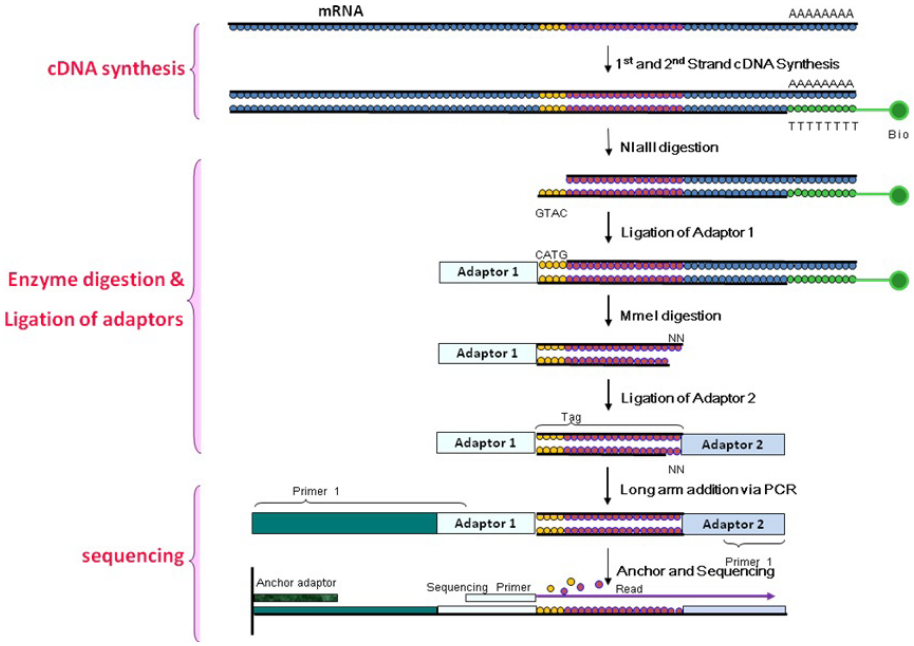

Figure 1. Principle and procedure of Tag preparation. Beads of Oligo(dT) are used to enrich mRNA from the total RNA, and then are transferred into double-stranded cDNA through reverse transcription. Four base-recognition enzyme *Nla*III is used to digest this cDNA, and Illumina adaptor 1 is ligated. *Mme*I is used to digest at 17 bp downstream of CATG site; Illumina adaptor 2 is ligated at 3' end. After linear PCR amplification, fragments are purified by PAGE.

DGE sequencing evaluation

Sequencing-received raw image data were transformed by base calling into sequence data, termed raw data or raw reads. Raw reads have 3' adaptor fragments, some low-quality sequences, and several types of impurities. Raw sequences were transformed into Clean Tags after data-processing: 3' adaptor sequence removal (since the tags are 21 nt while the sequencing reads are 49 nt, raw reads have 3' adaptor sequences); empty read removal (reads with only 3' adaptor sequences but no tags); low quality tags removal (tags with unknown sequences, "N"); removal of tags that are too long or short, leaving only tags of 21 nt; removal of tags with a copy number of 1 (probably sequencing errors). These various processing steps generated the clean tags. Following "Sequence Quality Assessment" and "Saturation Analysis of Sequencing", we could assess the sequencing data.

Screening and annotation of DEGs

Our partners provided us with virtual libraries containing all possible CATG+17 base length sequences of the reference gene sequences. All clean tags were mapped to the reference sequences and only 1-bp mismatch was allowed. Clean tags mapped to reference sequences from multiple genes were filtered. The remaining clean tags were designated unambiguous clean tags. The number of unambiguous clean tags for each gene was calculated and then normalized to the number of transcripts per million clean tags. The other four gene expression annotations of symbol, description, blast nr, and transcript ID do not appear in every result; part of these four gene expression annotations are integrated into reference gene information. Referring to "The significance of digital gene expression profiles", we developed a rigorous algorithm to identify DEGs between two samples. P value corresponds to a differential gene expression test. False discovery rate (FDR) is a method to determine the threshold of P value in multiple tests and analyses through manipulating the FDR value. Detection of DEGs across samples was performed according to the previous description (Yoav et al., 2001). Genes were deemed to show significant differential expression with an "FDR $\leq$ 0.001 and absolute value of log2 ratio $\geq$ 1" in DGE profiles. Further stringent criteria with smaller FDRs and bigger fold-change values can be used to identify DEGs. The data were statistically analyzed using the WEGO software. WEGO is a useful tool for plotting Gene Ontology (GO) annotation results. It has been widely used in many important biological research projects, such as the rice genome project and the silkworm genome project. It has become one of the daily tools for downstream gene annotation analysis, especially when performing comparative genomics tasks.

GO functional enrichment analysis of DEGs

GO is an internationally standardized system for classifying gene function; it offers a controlled vocabulary and strictly defined concepts to describe the properties of genes and their products in any organism. GO has three divisions: molecular function, cellular component, and biological process. In gene expression profiling analysis, GO enrichment analysis of functional significance applies a hypergeometric test to map all DEGs to terms in the GO database, looking for significantly enriched GO terms compared to the whole genome. The calculating formula is:

$$P = 1 - \sum_{i=0}^{m-1} \frac{\binom{m}{i}\binom{N-M}{n-i}}{\binom{N}{n}}$$

where N is the number of all genes with GO annotation; *n* is the number of DEGs in N; *M* is the number of all genes that are annotated to particular GO terms; *m* is the number of DEGs in M. In the current study, we used a corrected P value = 0.05 as the threshold to judge significantly enriched GO terms in DEGs.

Pathway enrichment analysis of DEGs

Genes generally cooperate with each other to produce biological function. Pathway-based analysis helps to provide insights into the biological functions of genes. Pathway enrichment analysis identifies significantly enriched metabolic pathways or signal transduction pathways in DEGs compared to the whole genome. The calculating formula is the same as that used in the GO analysis. Pathways with a Q value ≤ 0.05 were considered to be significantly enriched in DEGs.

Quantitative real-time PCR

cDNA from pBMECs was reverse transcribed using TRIzol reagent following the manufacturer protocol (Tiangen). Real-time RT-PCR was then performed with SYBR Premix Ex Taq (Takara) using the ABI Prism 7500 Sequence Detection System (Applied Biosystems, Foster City, USA). The total reaction volume of 20 μL included 10 μL 2X SYBR Premix Ex TaqII, 0.8 μL 10 μM sense forward primers, 0.8 μL 10 μM antisense forward primers, 0.4 μL 50X ROX Reference Dye, 2.0 μL cDNA template, and 6.0 μL dH_2O. Sense and antisense primers for the tested genes are shown in Table 1. The reaction conditions were as follows: 95°C for 30 s, followed by 40 cycles at 94°C for 5 s, 60°C for 34 s, and 95°C for 60 s. Relative expression levels were normalized against glyceraldehyde-3-phosphate dehydrogenase.

RESULTS

Analysis of DGE libraries

To identify changes in gene transcription in pBMECs after bacterial infection, we used a tag-based transcriptome sequencing method with the Solexa/Illumina DGE system. Four DGE libraries were produced: control library and one library from each of the bacterial species used for infection. The chief characteristics of these libraries are summarized in Table 2. We obtained approximately 4.9 million total sequence tags per library and approximately 0.22 million distinct tag sequences. "Tags Containing N" indicates the number of tags that contained N of the total tags and distinct tags. The number of tags that contained N was approximately 0.015 million and 0.007 million. We filtered out the 3' adaptor sequences, empty reads, low-quality tags, and tags with a copy number of 1, leaving approximately 4.8 million clean tags per library.

Heterogeneity and redundancy are significant characteristics of mRNA expression. In analyzing the depth of the distribution of the ratio of distinct tag copy numbers between two

Table 1. Sense and antisense primers for qPCR.

GenBank ID	Forward primer	Reverse primer	Product size (bp)
BC105484	ATCACTAACCAGCCAGAAATCG	TCCACGATGCCCAGGTAG	242
BC151264	AACACCAATGCCAACTTCTAC	GGTTCTGCCTGACGATGCT	259
BC105161	AGGATGCCCAGGAACTCA	ACACTCTTGGCTGCTTCTC	270
BC134510	GGAGAAGTTCGGTTTAGT	GTTTCCCGCTCAATACGA	287
BC105194	CATTGAGACCCTCGTTGC	CACCACCGTCGTCCAGAA	212
BT030701	ACCGCCATCAGCAGGAAC	TCATAGGCGTTGTCTGCTGTT	205
BC103310	GACTTCCAAGCTGGCTGTT	AGCAGACCTCGTTTCCAT	230
AB004275	CTTCTGCTCTGTGGCGTGTT	CAATATTCAGGGTAGAGCACAGTCA	222
BC123446	ACGTGGATGACAGCGAGGGC	CAAAGGAGGCGGTGGTAAA	207
NM_001045884	TGAGAACGACCGCACCAT	AGTCGGCATCGTTGGGCACAGGGA	225
NM_001080242	AGAGGAGCACGGACACCA	GCCTCATAGAAGCCATCCC	160
NM_173966	TCTCCTTCCTCCTGGTTGC	AGGGCATTGGCATACGAG	232
NM_001166608	CTGCTGACGGGCTTTACC	TCGGCATAGTCCAGGTAG	286
BC123753	CTTCTACGACGATGCCCTCA	ATGCCCGTTGCTGGACTG	238
BC120191	TCGTCGTCGTCGTCCTCGTT	CTCCGCCGGGTGTTTGCCCGGGG	183
BC133404	CTCCGCCGCTTCTTCTTCT	CTCCACTTGGCTTTGCATACTACTC	156
BC103346	AAACCTGCTACAACCCTCT	TGCTCTTCAGCCCAAA	192
XM_869007	CAACAAAATCTATGACAAAATGAAG	AGCCACTGCCGCACAACTC	273
BC153216	CGCCTGCTGCTGACGGGCACACCA	GCAACTGAGCCTCCACTTCCTTCTT	250
U25688	AAACCGCAGAGGAAGAAGA	GAGGGCGTCATAGGTGTCCTTGGT	180
BC142216	GCTGTGGCGAGTGTTAG	GATTGTGAGCGGTGTCG	237
U95811	AATGCTGCTCCTGCTCC	GTGGCTATGACTTCGGTTT	174
M37210	AGTCCCTGACCTCTTT	TGCCACCATCACCACA	217
BC123885	TCCTCGCCTGTTGATA	TTTGATGCCAGTGTTC	148
X58256	AGATGCTGAAACCCTG	GTGAAGGTCAAGGGTG	213
BC140659	TGGTGGTGGTTGGAGC	AGCCTTCGCCTGTCCTC	216
BC123466	TCCTTCCACCAATCAA	AGCCCTGTAGTTAGCC	312
BC148142	TGCGGCTTCAGCAACG	ACTTGCGGGATGGTGGG	137
BC151327	TTCAGCAGAAGGTCCGAGTG	GTCCTCCGTCTCCGACCTGGAAAAC	188
BC142430	ACCACCGCCAACTACGACCAGGA	CTAAGCCCAGCGCGTGGCCAAAC	283

Table 2. Major characteristics of DGE libraries and tag mapping to the Unigenes transcript database.

Summary	C		S108		K96		E23	
	Total	Distinct tag	Total	Distinct tag	Total	Distinct tag	Total	Distinct tag
Raw Data	4,960,570	229,566	4,846,444	238,591	4,610,296	225,435	4,789,230	230,882
Tags Containing N	14,805	7,473	14,955	7,978	14,241	7,592	13,794	7,387
Adaptors	304	255	234	208	291	252	380	332
Tag CopyNum <2	127,384	127,384	128,558	128,558	118,783	118,783	123,080	123,080
Clean Tag	4,818,077	94,454	4,702,697	101,847	4,476,981	98,808	4,651,976	100,083
Clean Tag								
CopyNum >=2	4,818,077	94,454	4,702,697	101,847	4,476,981	98,808	4,651,976	100,083
CopyNum >5	4,664,652	40,314	4,537,053	43,513	4,316,249	41,815	4,487,338	41,816
CopyNum >10	4,564,770	27,131	4,428,433	29,174	4,210,644	27,876	4,381,670	27,881
CopyNum >20	4,430,881	17,986	4,281,060	19,112	4,073,235	18,529	4,242,162	18,360
CopyNum >50	4,164,877	9,733	4,000,871	10,385	3,799,921	10,048	3,975,661	10,049
CopyNum >100	3,882,107	5,740	3,693,603	6,052	3,497,944	5,768	3,674,873	5,807
Tag Mapping	4,818,077	94,454	4,702,697	101,847	4,476,981	98,808	4,651,976	100,083
All Tag Mapping to Sense Gene	3,443,100	39,095	3,369,068	40,992	3,189,778	39,590	3,363,396	40,738
Unambiguous Tag Mapping to Sense Gene	2,567,012	34,311	2,531,486	36,129	2,384,750	34,546	2,509,775	35,912
All Tag Mapping to Anti-Sense Gene	605,435	21,212	605,051	23,180	554,422	22,294	579,375	22,888
Unambiguous Tag Mapping to Anti-Sense Gene	522,790	19,385	517,360	21,257	472,304	20,405	493,862	21,013
All Tag Mapping to Gene (Sense & Anti-Senese)	4,048,535	60,307	3,974,119	64,172	3,744,200	61,884	3,942,771	63,626
Unambiguous Tag Mapping to Gene (Sense & Anti-Sense)	3,089,802	53,696	3,048,846	57,386	2,857,054	54,951	3,003,637	56,925
Unknown Tag	206,532	8,827	163,055	9,276	204,081	10,104	154,430	9,566

All Mapping represents the number of all tags mapped to the Unigenes virtual tag database, Unambiguous Mapping represents the number of unambiguous tags mapped to the Unigenes virtual tag database, and Unambiguous Tags indicates tags matched only to 1 gene.

libraries, we determined that the number of distinct tags within 5 was approximately 99.05% of the total distinct tags (Figure 2).

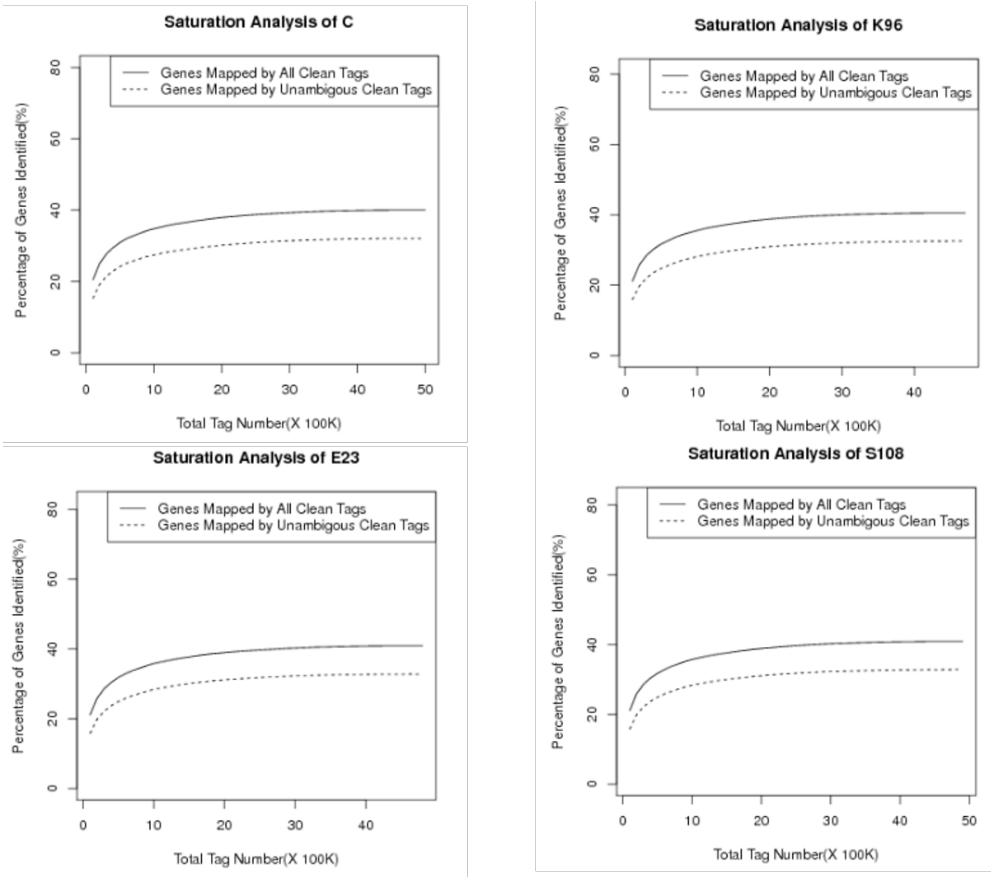

Figure 2. Saturation evaluation of different expression. K96: pPBME cells infected with *Klebsiella pneumoniae* strain; S108: pPBME cells infected with *Staphylococcus aureus* strain; E23: pBMECs infected with *Escherichia coli* strain, C: the non-infected pBMECs. Saturation analysis of capacity of libraries showed that new emerging distinct tags were gradually reduced with increasing of total sequence tags when the number of sequencing tags was big enough.

Identification of DEGs following bacterial infection

Initially, we compared gene expression profiles between uninfected controls and pBMECs infected with E23. Using the criteria of P value < 0.005, FDR ≤ 0.001, and absolute log2 ratio > 1 (Figure 3), we identified 16,019 transcripts that were differentially expressed; these transcripts represent 587 unique genes. Of these transcripts, 9283 (representing 464 unique genes) were upregulated and 6736 (representing 123 unique genes) were downregulated.

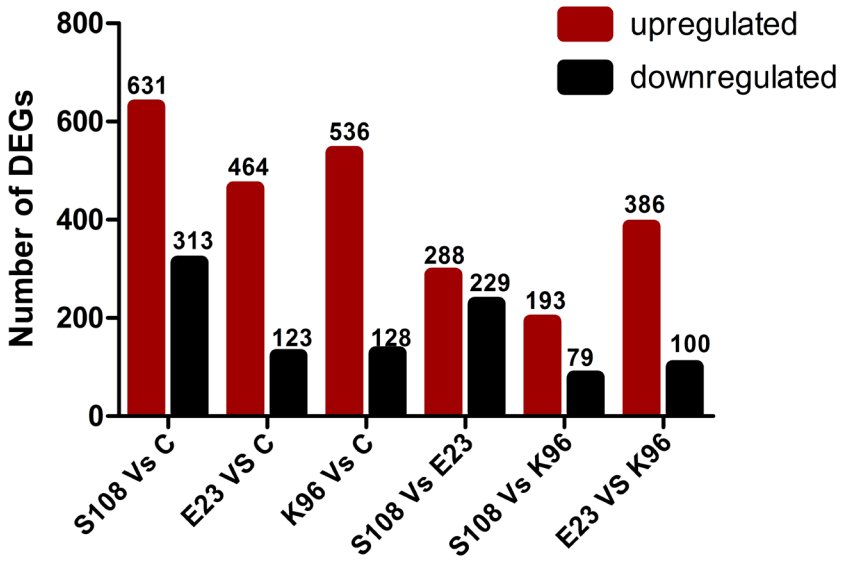

Figure 3. Differentially expressed genes between libraries. K96: pPBME cells infected with *Klebsiella pneumoniae pneumoniae* strain; S108: pPBME cells infected with *Staphylococcus aureus* strain; E23: pBMECs infected with *Escherichia coli* strain, C: the non-infected pBMECs. "FDR ≤0.001 and the absolute value of log2 ratio ≥1" was used as the threshold to judge the significance of gene expression difference.

The numbers of significant DEGs in the comparisons between controls and cells infected with either K96 or S108 are summarized in Figure 3. S108-infected cells had the largest number of DEGs, while E23-infected cells had the fewest.

Comparison of gene expression profiles in cells infected with S108 or E23 identified 16,186 differentially expressed transcripts, comprising 517 unique genes (Figure 3). Of these transcripts, 8306 (representing 288 unique genes) were upregulated and 7880 (representing 229 unique genes) were downregulated.

The comparison of S108- *vs* K96-infected cells and of E23- *vs* K96-infected cells is also shown in Figure 3. A larger number of DEGs was found in the comparison E23 *vs* K96 compared to S108 *vs* K96.

GO and cluster of orthologous groups of proteins (COG) functional enrichment analysis of DEGs

GO is an internationally standardized classification system of gene function and covers three domains: cellular component, molecular function, and biological process. We analyzed the changes in the functional domains of DEGs between all pairs of treatment groups (C *vs* E23, C *vs* K96, C *vs* S108) based on the GO database (**Table S1**). We focused on GO terms with a corrected P value ≤ 0.05 (Table 3 and Figure 4). Comparison of C *vs* K96 and C *vs* S108 identified DEGs with similar biological processes with regard to cellular metabolic process and metabolic process, and C *vs* E23 had non-common biological processes with C *vs* K96 and C *vs* S108 (Figure 5).

To more accurately annotate the gene functions, we aligned the unigenes to the COG database to identify homologous genes in C *vs* E23, C *vs* K96, and C *vs* S108. In C *vs* E23, 3525 unigenes were annotated and assigned to 23 COG classifications (Figure 6 and **Table S2**). Among the functional classes, the "Translation, ribosomal structure and biogenesis" cluster constituted the largest group (180, 20.9%) followed by "General function prediction only" (115, 13.3%) and "Replication, recombination and repair" (72, 8.4%); the two smallest groups were "RNA processing and modification" (1, 0.1%) and "Defense mechanisms" (1, 0.1%).

In C *vs* K96, 3560 unigenes were annotated and assigned to 23 COG classifications (Figure 6 and **Table S2**). Among the functional classes, the "Translation, ribosomal structure and biogenesis" cluster constituted the largest group (202, 22.7%) followed by "General function prediction only" (125, 14.0%) and "Replication, recombination and repair" (73, 8.2%); the two smallest groups were "Extracellular structures" (1, 0.1%) and "Defense mechanisms" (2, 0.2%).

In C *vs* S108, 5641 unigenes were annotated and assigned to 24 COG classifications (Figure 6 and **Table S2**). Among the functional classes, the "Translation, ribosomal structure and biogenesis" cluster constituted the largest group (263, 18.30%) followed by "General function prediction only" (211, 14.7%) and "Replication, recombination and repair" (120, 8.4%); the two smallest groups were "Extracellular structures" (2, 0.1%) and "RNA processing and modification" (3, 0.2%).

GO functional enrichment analysis of DEGs in infected cells

We used the WEGO software to perform the GO functional enrichment analyses of E23 *vs* K96, S108 *vs* E23, and S108 *vs* K96 (**Table S3**). We found that genes in S108 *vs* K96 were enriched with regard to macromolecule catabolic process, while genes from E23 *vs* K96 and S108 *vs* E23 were not enriched in any GO function.

Pathway enrichment analysis of DEGs in control and infected cells

The functional effects of the gene expression changes associated with bacterial infection were investigated by a pathway analysis of DEGs based on the KEGG database (Figure 5 and **Table S4**). We filtered the data for DEGs that were significantly enriched with a Q value ≤ 0.05. In C *vs* E23, we assigned 385 DEGs to 183 KEGG pathways (Figure 5 and **Table S4**). The enriched pathways included the nucleotide-binding oligomerization domain (NOD)-like receptor signaling pathway (12, 3.1%), which is linked to immune responses. Other enriched pathways were oxidative phosphorylation (20, 5.2%), Parkinson's disease (20, 5.2%), metabolic pathways (67, 17.4%), shigellosis (11, 2.9%), Alzheimer's disease (16, 4.2%) and hepatitis C (12, 3.12%), which are related to metabolism.

In C *vs* K96, 431 DEGs were assigned to 184 KEGG pathways, 4 of which were significantly enriched: oxidative phosphorylation (20, 4.6%), Parkinson's disease.

In C *vs* S108, 626 DEGs were assigned to 199 KEGG pathways, 2 of which were significantly enriched: oxidative phosphorylation (22, 3.5%) and Parkinson's disease.

Table 3. GO functional enrichment analysis of each *Staphylococcus aureus* strain infection.

Terms for C *vs* E23_P			Terms for C *vs* K96_P			Terms for C *vs* S108_P		
Gene Ontology term	Cluster frequency	Corrected P value	Gene Ontology term	Cluster frequency	Corrected P value	Gene Ontology term	Cluster frequency	Corrected P value
Cell death	46 of 391 genes	0.00655	Cellular metabolic process	235 of 403 genes	0.0002	Cellular metabolic process	312 of 561 genes	0.0012
Death	46 of 391 genes	0.00655	Metabolic process	270 of 403 genes	0.0005	Metabolic process	357 of 561 genes	0.01764
Programmed cell death	44 of 391 genes	0.01027	In utero embryonic development	10 of 403 genes	0.02941	Primary metabolic process	307 of 561 genes	0.04733
						Cellular respiration	14 of 561 genes	0.01098
						de novo posttranslational protein folding	7 of 561 genes	0.03658
						Nucleobase-containing compound metabolic process	152 of 561 genes	0.04129
						Nitrogen compound metabolic process	164 of 561 genes	0.04615
						Cellular nitrogen compound metabolic process	160 of 561 genes	0.04747
						Cellular macromolecule metabolic process	221 of 561 genes	0.03669

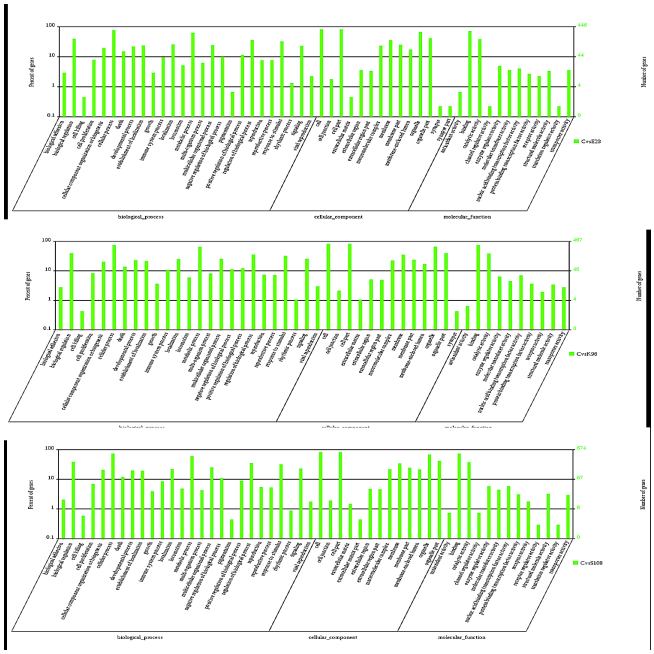

Figure 4. Gene Ontology classification of the unigenes. K96: pPBME cells infected with *Klebsiella pneumoniae* strain; S108: pPBME cells infected with *Staphylococcus aureus* strain; E23: pBMECs infected with *Escherichia coli* strain. The unigenes were annotated by three categories: biological process, cellular component, and molecular function. The left and right y-axes denote separately the percent and number of genes in the category.

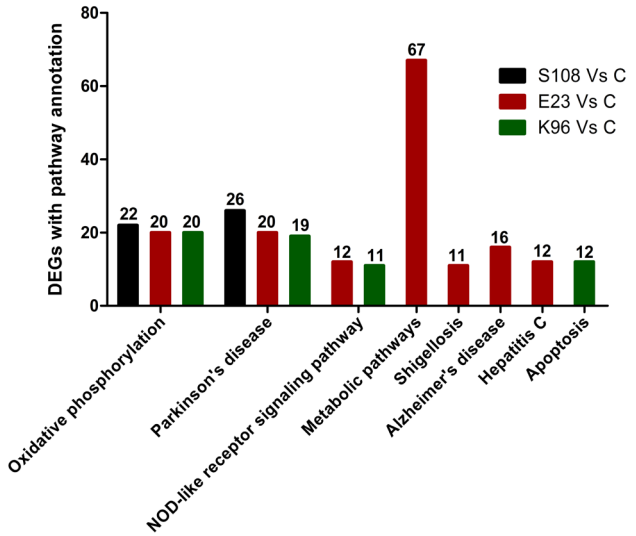

Figure 5. Pathway enrichment analysis of DEGs of each strain infection. K96: pPBME cells infected with *Klebsiella pneumoniae* strain; S108: pPBME cells infected with *Staphylococcus aureus* strain; E23: pBMECs infected with *Escherichia coli* strain, C: the non-infected pBMECs. Pathways with Q value ≤0.05 is significantly enriched in DEGs. X-axis represents signaling pathways. Y-axis indicates the gene number.

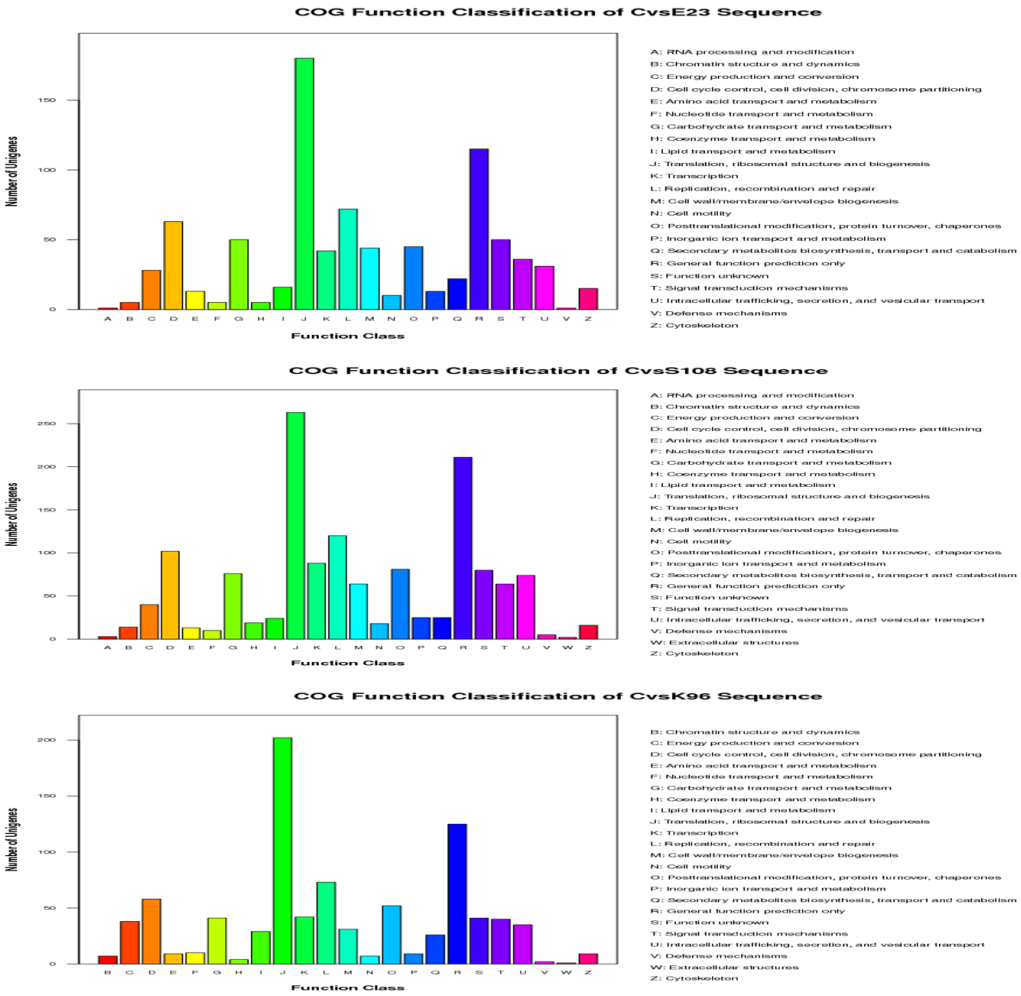

Figure 6. Clusters of orthologous group (COG) classification of the unigenes. K96: pPBME cells infected with *Klebsiella pneumoniae* strain; S108: pPBME cells infected with *Staphylococcus aureus* strain; E23: pBMECs infected with *Escherichia coli* strain.

DISCUSSION

E. coli and *K. pneumoniae* are the most important pathogenic bacteria in bovine clinical mastitis, while *S. aureus* generally induces subclinical mastitis. However, little is known about the molecular mechanisms underlying the different host response patterns caused by these bacteria. In this study, we analyzed the global gene expression changes in bovine mammary epithelial cells infected with *S. aureus*, *E. coli*, or *K. pneumoniae*. *S. aureus* can provoke clinical mastitis but more frequently causes subclinical infections that tend to become chronic and difficult to eradicate by conventional antimicrobial therapies. The frequent failure of both the innate immune system and antibiotics to prevent infection and destroy the pathogens in the

mammary glands is the reason why *S. aureus* bovine mastitis poses such a major challenge to dairy producers.

Infection with S108, K96, and E23 yielded 944, 664, and 587 DEGs from pBMECs, respectively (Figure 5). Of the 944 DEGs after S108 infection, we found upregulation of nuclear respiratory factor 1 (NRF1), *Bos taurus* interleukin 8 (IL-8), *B. taurus* solute carrier family 2, *B. taurus* chemokine (C-X-C motif) ligand 5 (CXCL5), *B. taurus* nuclear factor I/B, *B. taurus* E74-like factor 3 (ets domain transcription factor, epithelial-specific), bovine interleukin 1-alpha (IL-1-alpha), *B. taurus* chemokine (C-C motif) ligand 5, *B. taurus* chemokine (C-X-C motif) receptor 7 (CXCL7), *B. taurus* glucose transporter type I, *B. taurus* Josephin domain containing 1, *B. taurus* target of EGR1, *B. taurus* APEX nuclease (multifunctional DNA repair enzyme) 1, *B. taurus* IK cytokine, and *B. taurus* plakophilin 3; we also found downregulation of *B. taurus* transforming growth factor beta regulator 1, *B. taurus* programmed cell death 2-like (PDCD2L), *B. taurus* RAB3A, *B. taurus* RAB1B, and *B. taurus* praja 1.

NRF1 encodes a protein that can homodimerize and function as a transcription factor to activate expression of key metabolic genes involved in a wide range of processes, such as cell growth, and respiration, heme biosynthesis, mitochondrial DNA transcription and replication, and neurite outgrowth. Alternate transcriptional splice variants, which encode the same protein, have been characterized. IL-8 is a chemokine produced by macrophages and other cell types such as epithelial cells, airway smooth muscle cells (Hedges et al., 2000) and endothelial cells. Endothelial cells store IL-8 in their storage vesicles, the Weibel-Palade bodies (Utgaard et al., 1998; Wolff et al., 1998). IL-8 can be secreted by any cells with toll-like receptors that are involved in the innate immune response. Macrophages are usually the first cells to respond first to an antigen and, thus, to release IL-8 to recruit other cells. Both monomeric and homodimeric forms of IL-8 are potent inducers of the chemokine receptors CXCR1 and CXCR2. CXCL5 is a small cytokine belonging to the C-X-C chemokine family and is also known as epithelial-derived neutrophil-activating peptide 78. It is produced following stimulation of cells with the inflammatory cytokines interleukin-1 or tumor necrosis factor (TNF)-alpha (Chang et al., 1994). The expression of CXCL5 has also been observed in eosinophils, and can be inhibited with the type II interferon IFN-γ (Persson et al., 2003). This chemokine stimulates chemotaxis in neutrophils possessing angiogenic properties. These genes are related to metabolic and innate immune responses and were both found to be upregulated here. Over-expression of *B. taurus* PDCD2L suppresses AP1, CREB, NFAT, and NF-kB transcriptional activation, and delays cell cycle progression at S phase (Chen et al., 2008); *PDCD2L* was found to be downregulated here. Thus, in pBMECs infected by S108, metabolic and innate immunity response related genes showed upregulation and cell cycle related genes were downregulated.

In the comparison C *vs* S108, 58 immune system process genes showed differential expression: 40 were upregulated and 18 were downregulated (**Table S5**). The greatest upregulation was observed for *B. taurus IL-8*; *NOD1* was the most downregulated gene. NOD1 is a member of the NOD-like receptor protein family and is a close relative of NOD2. It acts as an intracellular pattern recognition receptor, and is similar in structure to the disease resistance proteins of plants. NOD1 mediates innate and acquired immunity by recognizing bacterial molecules containing a D-glutamyl-meso-diaminopimelic acid (iE-DAP) moiety. Stimulation of NOD1 by iE-DAP containing molecules results in activation of the transcription factor NF-κB.

In the comparison C *vs* K96, 664 DEGs were identified: we found upregulation of

B. taurus TNF, *B. taurus* IL-8, bovine IL-1-alpha, CXCL5, *B. taurus* dynactin 5 (p25), *B. taurus* DNA-damage-inducible transcript 4, *B. taurus* tumor protein p53 inducible protein 3, CXL7, interleukin 3-regulated, *B. taurus* transforming growth factor beta 1 (TGF-β1), *B. taurus* inhibitor of growth family. *B. taurus* early growth response 1 and *B. taurus* T-cell receptor were downregulated.

TNF is a monocyte-derived cytotoxin that has been implicated in tumor regression, septic shock, and cachexia (Kriegler et al., 1988). The protein is synthesized as a prohormone with an unusually long and atypical signal sequence, which is absent from the mature secreted cytokine. A short hydrophobic stretch of amino acids serves to anchor the prohormone in lipid bilayers. Both the mature protein and a partially processed form can be secreted after cleavage of the propeptide (Cseh and Beutler, 1989). TGF-β1 is a polypeptide member of the transforming growth factor beta superfamily of cytokines. It is a secreted protein that performs many cellular functions, including the control of cell growth, cell proliferation, cell differentiation, and apoptosis. TGF-β1 plays an important role in controlling the immune system, and shows different activities on different types of cell, or cells at different developmental stages. Most immune cells (or leukocytes) secrete TGF-β1. Some T cells (e.g., regulatory T cells) release TGF-β1 to inhibit the actions of other T cells. IL-1 and IL-2 dependent proliferation of activated T cells (Wahl et al., 1988; Tiemessen et al., 2003), and the activation of quiescent helper T cells and cytotoxic T cells, is prevented by the activity of TGF-β1. Similarly, TGF-β1 can inhibit the secretion and activity of many other cytokines including IFN-γ, TNF-alpha, and various interleukins. It can also decrease the expression levels of cytokine receptors, such as the IL-2 receptor to downregulate the activity of immune cells. However, TGF-β1 can also increase the expression of certain cytokines in T cells and promote their proliferation, particularly if the cells are immature. TGF-β1 has similar effects on B cells that also vary according to the differentiation state of the cell. It inhibits proliferation and stimulates apoptosis of B cells, and plays a role in controlling the expression of antibodies, transferrin and major histocompatibility complex (MHC) class II proteins on immature and mature B cells (Lebman and Edmiston, 1999). These genes are related to apoptosis and the immune system, and both were found to be upregulated in our study. T cell receptor (TCR) is found on the surface of T lymphocytes (or T cells) and is responsible for recognizing antigens bound to MHC molecules; TCR was found here to be downregulated. Our analyses indicate that pBMECs infected by K96 showed upregulation of genes associated with innate immunity responses, while cell cycle related genes were downregulated.

In the comparison C *vs* K96, 51 DEGs associated with immune system processes were identified: 43 were upregulated and 8 were downregulated (**Table S5**). The highest level of upregulation was observed in *B. taurus* TNF, while *B. taurus* ferritin heavy chain-like was the most severely downregulated. *B. taurus* ferritin heavy chain-like binds to the intracellular membranes of organelles and plays a role in transition metal ion binding and oxidoreductase activity.

In the comparison C *vs* E23, 587 DEGs were identified. *B. taurus* IL-8, bovine IL-1-alpha, CXCL5, CXCL7, *B. taurus* TNF-receptor I, interleukin 3 (IL-3), and *B. taurus* nuclear factor (erythroid-derived 2)-like 2 were upregulated and *B. taurus* early growth response 1 and *B. taurus* eukaryotic translation initiation factor 5 (EIF5) were downregulated.

IL-3 is an interleukin, a type of biological signal (cytokine), that can improve the natural responses to disease as part of the immune system. It acts by binding to the IL-3 receptor to

stimulate the differentiation of multi-potent hematopoietic stem cells into myeloid progenitor cells or, with the addition of IL-7, into lymphoid progenitor cells. In addition, IL-3 stimulates proliferation of all cells in the myeloid lineage (granulocytes, monocytes, and dendritic cells), in conjunction with other cytokines, e.g., erythropoietin, granulocyte macrophage colony-stimulating factor, and IL-6. It is secreted by basophils and activated T cells to support growth and differentiation of T cells from the bone marrow in an immune response. EIF are proteins involved in the initiation phase of eukaryotic translation. They function by forming a complex with the 40S ribosomal subunit and Met-tRNA$_i$, the 43S preinitiation complex (PIC), that recognizes the 5' cap structure of mRNA and recruits the 43S PIC to mRNA. Recruitment promotes ribosomal scanning of mRNA and regulates the recognition of the AUG initiation codon and the joining of the 60S ribosomal subunit to create the 80S ribosome. There are more EIFs than prokaryotic initiation factors due to greater biological complexity of eukaryotic cells. pBMECs infected by E23 showed upregulation of genes related to innate immunity responses and downregulation of cell cycle related genes.

In the comparison C *vs* E23, 43 DEGs were identified: 34 showed upregulation and 9 showed downregulation (**Table S5**). The highest level of upregulation was observed for IL-8, while ferritin heavy chain-like was the most downregulated.

The DEGs induced in pBMECs by infection with *E. coli*, *K. pneumoniae*, or *S. aureus* strains were found to be associated with a number of biochemical pathways implicated in various aspects of metabolism, signal transduction, and pathophysiology. The comparison C *vs* S108 showed that genes associated with Parkinson's disease and oxidative phosphorylation were enriched. Parkinson's disease results from greatly reduced activity of dopamine-secreting cells due to cell death (Obeso et al., 2008). Oxidative phosphorylation is the metabolic pathway through which mitochondria generate ATP using membranous structures, enzymes, and energy released by the oxidation of nutrients. Our results indicate that pBMECs infected by *S. aureus* had significant changes to their metabolic pathways.

In the comparison C *vs* E23, genes for oxidative phosphorylation, Parkinson's disease, NOD-like receptor signaling pathway, metabolic pathways, shigellosis, Alzheimer's disease and hepatitis C were enriched. NOD-like receptors (NLRs) are intracellular sensors of pathogen-associated molecular patterns that enter the cell via phagocytosis or through pores and of damage-associated molecular patterns associated with cell stress. They are part of the pattern recognition receptor system and play key roles in regulation of innate immune responses. NLRs can cooperate with toll-like receptors to regulate inflammatory and apoptotic responses. They are found in lymphocytes, macrophages, dendritic cells, and also in nonimmune cells, e.g., the epithelium (Franchi et al., 2009). Our results showed that pBMECs infected by *E. coli* were not only changed in their metabolic pathways, but also in their innate immune response pathways.

In the comparison C *vs* K96, genes for oxidative phosphorylation, Parkinson's disease, apoptosis and the NLR signaling pathway were enriched. Apoptosis (programmed cell death) is an essential part of normal development in multicellular organisms but is also induced by disease (Green, 2011). Biochemical events lead to characteristic changes in cell morphology and to their eventual death. These changes include blebbing, cell shrinkage, nuclear fragmentation, chromatin condensation, and chromosomal DNA fragmentation. In contrast to necrosis, a form of traumatic cell death that results from acute cellular injury, apoptosis confers advantages during an organism's life cycle. TNF is a cytokine mainly produced by activated macrophages and is the major extrinsic mediator of apoptosis; in addition, caspases play a central role in

the transduction of death receptor apoptotic signals. Caspases, cysteine-dependent aspartate-specific proteases, are highly conserved. There are two types of caspase: initiator caspases (caspases 2, 8, 9, and 10) and effector caspases (caspases 3, 6, and 7). Our results showed that pBMECs infected with *K. pneumoniae* not only showed metabolic pathway and innate immune response pathway changes but also changes to the apoptosis pathway.

In summary, we performed a comprehensive gene expression analysis in pBMECs. Although the treatment of mastitis caused by *E. coli*, *K. pneumoniae* or *S. aureus* strains remains unclear, the transcriptome and DGE data produced here offer valuable insights regarding the immune response, apoptosis and metabolic pathway changes that are induced by agents causing mastitis and will assist in elucidation of the underlying molecular mechanisms of this livestock disease.

Conflicts of interest

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

We thank the Beijing Genomics Institute at Shenzhen for Illumina sequencing and raw data analysis. Research supported by the National Natural Science Foundation of China (#30901066) and the Natural Science Foundation of Inner Mongolia Autonomous Region (#2010MS0514).

Supplementary material

REFERENCES

Aitken SL, Corl CM and Sordillo LM (2011). Immunopathology of mastitis: insights into disease recognition and resolution. *J. Mammary Gland Biol. Neoplasia* 16: 291-304.

Alva-Murillo N, Ochoa-Zarzosa A and López-Meza JE (2013). Effects of sodium octanoate on innate immune response of mammary epithelial cells during *Staphylococcus aureus* internalization. *BioMed Res. Int*. Vol. 2013 Article ID 927643.

Burvenich C, Van Merris V, Mehrzad J, Diez-Fraile A, et al. (2003). Severity of *E. coli* mastitis is mainly determined by cow factors. *Vet. Res*. 34: 521-564.

Chang MS, McNinch J, Basu R and Simonet S (1994). Cloning and characterization of the human neutrophil-activating peptide (ENA-78) gene. *J. Biol. Chem*. 269: 25277-25282.

Chen Q, Yan C, Yan Q, Feng L, et al. (2008). The novel MGC13096 protein is correlated with proliferation. *Cell Biochem. Funct*. 26: 141-145.

Cseh K and Beutler B (1989). Alternative cleavage of the cachectin/tumor necrosis factor propeptide results in a larger, inactive form of secreted protein. *J. Biol. Chem*. 264: 16256-16260.

Dosogne H, Vangroenweghe F, Barrio B, Rainard P, et al. (2001). Decreased number and bactericidal activity against *Staphylococcus aureus* of the resident cells in milk of dairy cows during early lactation. *J. Dairy Res*. 68: 539-549.

Dwinell MB, Johanesen PA and Smith JM (2003). Immunobiology of epithelial chemokines in the intestinal mucosa. *Surgery* 133: 601-607.

Franchi L, Warner N, Viani K and Nuñez G (2009). Function of Nod-like receptors in microbial recognition and host defense. *Immunol. Rev*. 227: 106-128.

Green DR (2011). Means to an end: Apoptosis and other cell death mechanisms. Cold Spring Harbor Laboratory Press, New York.

Hedges JC, Singer CA and Gerthoffer WT (2000). Mitogen-activated protein kinases regulate cytokine gene expression in human airway myocytes. *Am. J. Respir. Cell Mol. Biol.* 23: 86-94.

Hensen SM, Pavicic MJ, Lohuis JA, de Hoog JA, et al. (2000). Location of *Staphylococcus aureus* within the experimentally infected bovine udder and the expression of capsular polysaccharide type 5 *in situ*. *J. Dairy Sci.* 83: 1966-1975.

Hogan JS, Smith KL, Hoblet KH, Schoenberger PS, et al. (1989). Field survey of clinical mastitis in low somatic cell count herds. *J. Dairy Sci.* 72: 1547-1556.

Huijps K, Lam TJ and Hogeveen H (2008). Costs of mastitis: facts and perception. *J. Dairy Res.*75: 113-120.

Jensen K, Günther J, Talbot R., Petzl W, et al. (2013). *Escherichia coli*- and *Staphylococcus aureus*-induced mastitis differentially modulate transcriptional responses in neighbouring uninfected bovine mammary gland quarters. *BMC Genomics* 14: 36.

Kriegler M, Perez C, DeFay K, Albert I, et al. (1988). A novel form of TNF/cachectin is a cell surface cytotoxic transmembrane protein: ramifications for the complex physiology of TNF. *Cell* 53: 45-53.

Lahouassa H, Moussay E, Rainard P and Riollet C (2007). Differential cytokine and chemokine responses of bovine mammary epithelial cells to *Staphylococcus aureus* and *Escherichia coli*. *Cytokine* 38: 12-21.

Lebman DA and Edmiston JS (1999). The role of TGF-beta in growth, differentiation, and maturation of B lymphocytes. *Microbes Infect.* 1: 1297-1304.

Moreilhon C, Gras D, Hologne C, Bajolet O, et al. (2004). Live *Staphylococcus aureus* and bacterial soluble factors induce different transcriptional responses in human airway cells. *Physiol. Genomics* 20: 244-255.

Obeso JA, Rodríguez-Oroz MC, Benitez-Temino B, Blesa FJ, et al. (2008). Functional organization of the basal ganglia: therapeutic implications for Parkinson's disease. *Mov. Disord.* 23 (Suppl 3): S548-S559.

Paape MJ, Shafer-Weaver K, Capuco AV, Van Oostveldt K, et al. (2000). Immune surveillance of mammary tissue by phagocytic cells. *Adv. Exp. Med. Biol*. 480: 259-277.

Pareek R, Wellnitz O, Dorp R, Burton J, et al. (2005). Immunorelevant gene expression in LPS-challenged bovine mammary epithelial cells. *J. Appl. Genet.* 46: 171-177.

Persson T, Monsef N, Andersson P, Bjartell A, et al. (2003). Expression of the neutrophil-activating CXC chemokine ENA-78/CXCL5 by human eosinophils. *Clin. Exp. Allergy* 33: 531-537.

Rainard P, Riollet C, Poutrel B and Paape MJ (2000). Phagocytosis and killing of *Staphylococcus aureus* by bovine neutrophils after priming by tumor necrosis factor-alpha and the des-arginine derivative of C5a. *Am. J. Vet. Res*. 61: 951-959.

Strandberg Y, Gray C, Vuocolo T, Donaldson L, et al. (2005). Lipopolysaccharide and lipoteichoic acid induce different innate immune responses in bovine mammary epithelial cells. *Cytokine* 31: 72-86.

Swinkels JM, Hogeveen H and Zadoks RN (2005). A partial budget model to estimate economic benefits of lactational treatment of subclinical *Staphylococcus aureus mastitis*. *J. Dairy Sci.* 88: 4273-4287.

Tiemessen MM, Kunzmann S, Schmidt-Weber CB, Garssen J, et al. (2003). Transforming growth factor-beta inhibits human antigen-specific CD4+ T cell proliferation without modulating the cytokine response. *Int. Immunol.* 15: 1495-1504.

Utgaard JO, Jahnsen FL, Bakka A, Brandtzaeg P, et al. (1998). Rapid secretion of prestored interleukin 8 from Weibel-Palade bodies of microvascular endothelial cells. *J. Exp. Med*. 188: 1751-1756.

Viguier C, Arora S, Gilmartin N, Welbeck K, et al. (2009). Mastitis detection: current trends and future perspectives. *Trends Biotechnol.* 27: 486-493.

Wahl SM, Hunt DA, Wong HL, Dougherty S, et al. (1988). Transforming growth factor-beta is a potent immunosuppressive agent that inhibits IL-1-dependent lymphocyte proliferation. *J. Immunol.* 140: 3026-3032.

Wang F, Hu S, Liu W, Qiao Z, et al. (2011). Deep-sequencing analysis of the mouse transcriptome response to infection with *Brucella melitensis* strains of differing virulence. *PLoS One* 6: e28485.

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

Wolff B, Burns AR, Middleton J and Rot A (1998). Endothelial cell "memory" of inflammatory stimulation: human venular endothelial cells store interleukin 8 in Weibel-Palade bodies. *J. Exp. Med*. 188: 1757-1762.

Yang W, Zerbe H, Petzl W, Brunner RM, et al. (2008). Bovine TLR2 and TLR4 properly transduce signals from *Staphylococcus aureus* and *E. coli*, but *S. aureus* fails to both activate NF-kappaB in mammary epithelial cells and to quickly induce TNFalpha and interleukin-8 (CXCL8) expression in the udder. *Mol. Immunol*. 45: 1385-1397.

Yoav B and Daniel Y (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Stat.* 29: 1165-1188.