Gene co-expression in response to *Staphylococcus aureus* infection reveals networks of genes with specific functions during bovine subclinical mastitis

Mengqi Wang,1,2 Nathalie Bissonnette,1 Mario Laterrière,3 Pier-Luc Dudemaine,1 David Gagné,3 Jean-Philippe Roy,4 Marc-André Sirard,2 and Eveline M. Ibeagha-Awemu1,*

1Sherbrooke Research and Development Centre, Agriculture and Agri-Food Canada, Sherbrooke, Quebec, J1M0C8, Canada
2Department of Animal Science, Laval University, Quebec City, Quebec, G1V 0A6, Canada
3Quebec Research and Development Centre, Agriculture and Agri-Food Canada, Quebec City, Quebec, G1V 2J3, Canada
4Department of Clinical Sciences, Faculty of Veterinary Medicine, Université de Montréal, St-Hyacinthe, Quebec, H3T 1J4, Canada

**ABSTRACT**

*Staphylococcus aureus* is one of the most prevalent contagious bacterial pathogen of bovine mastitis. The subclinical mastitis it causes has long-term economic implications and it is difficult to control. To further understanding of the genetic basis of mammary gland defense against *S. aureus* infection, the transcriptomes of milk somatic cells from 15 cows with persistent natural *S. aureus* infection (*S. aureus*-positive, SAP) and 10 healthy control cows (HC) were studied by deep RNA-sequencing technology. Comparing the transcriptomes of SAP to HC group revealed 4,077 differentially expressed genes (DEG; 1,616 up- and 2,461 downregulated). Functional annotation indicated enrichment of DEG in 94 Gene Ontology (GO) and 47 Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. Terms related to the immune response and disease processes were mostly enriched for by upregulated DEG, whereas biological process terms related to cell adhesion, cell movement and localization, and tissue development were mostly enriched for by downregulated DEG. Weighted gene co-expression network analysis grouped DEG into 7 modules, the most important module (colored turquoise by software and here referred to as Turquoise module) was positively significantly correlated with *S. aureus* subclinical mastitis, and were enriched in functional annotations involved in the regulation of cell migration, cell communication, metabolic process, and blood circulatory system development, respectively. Application of sparse partial least squares discriminant analysis to genes of the Turquoise module identified 5 genes (*NR2F6*, *PDLIM5*, *RAB11FIP5*, *ACOT4*, and *TMEM53*) capable of explaining the majority of the differences in the expression patterns between SAP and HC cows. In conclusion, this study has furthered understanding of the genetic changes in the mammary gland and the molecular mechanisms underlying *S. aureus* mastitis, as well as revealed a list of candidate discriminant genes with potential regulatory roles in response to *S. aureus* infection.

**Key words:** bovine *Staphylococcus aureus* subclinical mastitis, RNA sequencing of milk somatic cells, gene networks, immune response

**INTRODUCTION**

Mastitis, defined as an inflammation of the mammary gland, is the most widespread and costly disease of dairy livestock all over the world (Barkema et al., 2006). The pathogenesis of mastitis is influenced by multiple factors, including cow (host), environment and type of pathogen among others (Dego, 2020). A plethora of pathogens are implicated in the development of mastitis, and *Staphylococcus aureus* is one of the most important causal pathogens of subclinical mastitis. Mastitis caused by *S. aureus* could develop into a clinical or subclinical (mild chronic or persistent) infection (Niedziela et al., 2020). It is contagious and easily transmitted during the milking process, sometimes leading to outbreaks of mastitis cases (Gussmann et al., 2019). *Staphylococcus aureus* persistence is fa-
ciliated by its ability to invade diverse cell types in the mammary gland, to form biofilms and encapsulated lesions, and to acquire antibiotic resistance genes by horizontal transfer of mobile genetic elements (Abdi et al., 2018; Rainard et al., 2018a). These characteristics of *S. aureus* persistence prolong the treatment process resulting in the extended use of antimicrobials, thereby promoting the resistance of microbes to antibiotic treatment (Keefe, 2012; Abdi et al., 2018).

Over the past 2 decades, many strategies have been put in place to monitor, prevent and control mastitis (Ruegg, 2017). Milk SCC is generally regarded as an indicator of mammary gland health status. This phenotype is included in dairy cow breeding programs to improve mastitis resistance worldwide (Alhussien and Dang, 2018; Rainard et al., 2018a). Somatic cells are the second line of mammary gland defense after the physical barrier of the teat canal (Sordillo, 2018). During IMI, somatic cells in milk are present as a part of the innate immune system of the mammary gland with variable cell composition as influenced by infection status (Benić et al., 2012; Malik et al., 2018). The milk SCC is relatively constant and stable during the whole lactation period except the first weeks postpartum and during infection, which is usually high (Miller et al., 2004). When the presence of a pathogen is detected, there is systemic immune reaction whereby resident somatic cells react by sending signals to white blood cells in the blood resulting to their influx and increase in the number of milk somatic cells, with the consequence of increasing the proportions of leucocytes (Malik et al., 2018). Therefore, the increase in SCC during infection reflects an increase in the number of immune-related cells in milk, which work to contain mammary gland inflammation (Halasa and Kirkeyb, 2020). Genetic variation has been found to contribute to differences in SCC between individual cows and mastitis resistance (Haile-Mariam et al., 2001; Rupp and Boichard, 2003). Additionally, thousands of QTL have been identified for mammary gland health, including mastitis (Ruegg, 2017; Hu et al., 2022). Recognizing the threat posed by high prevalence of subclinical mastitis and the effect of the emergence of antibiotic resistance on animal, human and environmental health, it is important, considering the One Health concept as well as other global regulatory efforts, to further understanding of the genetic basis of mastitis, which is necessary for the development of efficient strategies such as genomic selection for udder traits and mastitis resistance (Destoumieux-Garzón et al., 2018).

A better understanding of the genetics underlying complex diseases has been found to contribute to the development of genomic prediction and effective control strategies (Edwards et al., 2016; Raymond et al., 2018). Previous studies have used RNA sequencing to explore the gene expression profiles of mammary gland tissue, milk cells, and peripheral blood in response to *S. aureus* mastitis in dairy cow, goat, and sheep (Pisoni et al., 2010; Bonnefont et al., 2011; Jensen et al., 2013). Several differentially expressed genes (DEG) were identified for mastitis caused by *S. aureus* in these studies, including *CXCL14, SLC4A11, KIT, BLA-DQB, C3AR1, FCAR, KRT10, C1R*, and *CFI*, among others (Fang et al., 2016; Wang et al., 2020). However, the reported transcriptomic responses were explored in milk somatic cells following experimental challenge of sheep, goat, and cow with various *S. aureus* strains (Pisoni et al., 2010; Bonnefont et al., 2011; Niedziela et al., 2021), or in vitro infection of cell lines, including bovine mammary gland cells (Jensen et al., 2013; Fang et al., 2016) and bovine mammary epithelial cell line (Wang et al., 2013; Han, 2019; Chen et al., 2021). Although experimental bacterial infection is a good strategy to study the direct cellular or tissue responses to *S. aureus* infection within a short period, it may not be capable of capturing the long-term effects and related genetic regulation underlying naturally occurring subclinical mastitis. The transcriptional changes during subclinical mastitis naturally caused by field *S. aureus* infection have also been studied, but mainly using specific cell types, such as peripheral blood lymphocytes (Song et al., 2016) and mononuclear cells (Tao and Mallard, 2007). Tao and Mallard (2007) identified some DEG in milk somatic cells in response to persistent natural infection by *S. aureus*, but using a reduced sequencing technology (hybridization and gene array), which restricted the scope to 167 immune and endocrine genes.

The genetic basis of the mammary gland defense system, especially the transcriptomic response of milk somatic cells, and mastitis resistance is not fully understood. Because the milk somatic cells represent a collection of the molecules responding to the presence of a pathogen, the objectives of this study therefore were to perform transcriptomic mRNA sequencing of milk somatic cells from cows with naturally occurring *S. aureus* subclinical mastitis compared with healthy cows, identify hub genes related to *S. aureus* subclinical mastitis, and build a gene co-expression network to provide additional information for understanding the genetic background of mastitis resistance.

**MATERIALS AND METHODS**

*Ethics Approval and Consent to Participate*

The animal use procedures in this study were in accordance with the guidelines of the Canadian Council on Animal Care and ethical approval to conduct the...
study was provided by the Animal Care and Ethics Committee of Agriculture and Agri-Food Canada (approval #570).

**Animals and Sample Collection**

Five commercial dairy farms (Quebec, Canada) with a history of IMI were recruited for this study. The milk SCC of the DHI records of the herds was monitored for all 386 lactating Canadian Holstein cows for a period of 6 mo. The DHI data (SCC, protein %, fat %, lactose, and other milk components) was obtained by analyzing monthly milk samples (~30 mL/cow) from each cow in the herds by Lactanet (https://lactanet.ca/). A total of 81 cows with SCC greater than 350,000 cells/mL and with or without a record of previous identification of having IMI including infection by *S. aureus* for 3 consecutive months (HSCC group), and 63 cows with less than 100,000 cells/mL and no previous history of IMI (LSCC group) for 3 consecutive months, were selected for pathogen detection. All the animals selected were not showing clinical signs of infection during the period of sampling. For cows in the HSCC group, about 5 mL of milk was aseptically collected per quarter separately, meanwhile a composite milk sample of all 4 quarters (3 mL/quarter) was collected from each cow of the LSCC group. The milk samples were transported on ice for bacteriological examination immediately after collection (Biovet Laboratories, Saint-Hyacinthe, Quebec, Canada). A total of 18 cows of the HSCC group were found positive to *S. aureus* in one or more quarters and retained to constitute the *S. aureus*-positive (SAP) group. Meanwhile, cows with one or more quarters positive for other mastitis pathogens in the HSCC group were removed from the study. A total of 15 cows in the LSCC group were negative for all mastitis pathogens and were enrolled in the healthy control (HC) group. Following receipt of the bacteriological results and cow selection, a second visit was made to collect a higher volume of milk samples from cows in the SAP and HC groups. Approximately 200 mL of milk was sampled from one quarter of each cow in the SAP group (only one quarter was sampled even if more than one quarter was positive to *S. aureus*). Meanwhile, a 200-mL composite milk sample (50 mL/quarter) was collected from each cow of the control group. Because about 3 to 5 d elapsed between bacteriological examination of mastitis pathogens and the second milk sampling, another bacteriological test was performed on the day of the second sampling to validate the first bacteriological results. Only samples with consistent results were kept for this study, including 15 *S. aureus*-positive cows as final SAP group and 10 healthy cows as HC group.

Further details on the selected cows (farm, age, parity, stage of lactation, milk components, and SCC) are shown in Supplemental Table S1 (https://doi.org/10.6084/m9.figshare.20732380.v1; Ibeagha-Awemu et al., 2022b).

Milk samples were placed on ice and immediately transferred to the Animal Genomics and Epigenomics Laboratory of the Sherbrooke Research and Development Centre, Agriculture and Agri-Food Canada. Milk somatic cells were immediately isolated from freshly collected milk (200 mL/cow) by low speed centrifugation (1,500 × g, 15 min, 4°C) followed by 2 times washing by addition of 40 mL 1× PBS and centrifugation (1,500 × g, 15 min, 4°C). The fat and whey layers were removed and the milk somatic cells were placed in Trizol reagent and stored at −80°C until RNA isolation, which was performed within one month after sample collection.

**RNA Isolation, Library Preparation, and Sequencing**

The RNeasy Mini Kit (Qiagen Inc.) was used to isolate total RNA from milk somatic cells according to manufacturer’s protocol. The quantity and integrity of total RNA were checked with Agilent Bioanalyzer 2100 (Agilent Technologies) and LabChip GXII (PerkinElmer Inc.) instruments, respectively. The RNA samples that passed quality control (>7 RNA integrity number) were further processed. First, rRNA was depleted from 125 ng of total RNA using QIAseq FastSelect rRNA HMR Kit (QIAGEN Inc.). Then, the NEBNext RNA First Strand Synthesis and NEBNext Ultra Directional RNA Second Strand Synthesis Modules from NEBNext Ultra RNA Library Prep Kit (New England BioLabs Ltd.) were used for cDNA synthesis. The library preparation was then performed by using the NEBNext Ultra II DNA Library Prep Kit for Illumina (New England BioLabs). Adapters and PCR primers were purchased from New England BioLabs. Libraries were quantified using Kapa Illumina GA with Revised Primers-SYBR Fast Universal kit (Kapa Biosystems Inc.). Average size fragment was determined using a LabChip GXII (PerkinElmer Inc.) instrument. The libraries were normalized in equimolar concentrations, pooled and loaded at 200 pM on an Illumina NovaSeq S4 lane using Xp protocol according to the manufacturer’s recommendations. The library preparation and RNA sequencing were done by Centre d’Expertise et de Services Génome Québec (https://www.genomequebec.com/).

**Identification of DEG**

The preprocessing of raw RNA sequencing reads and bioinformatics analyses were performed with nf-core
RNaseq bioinformatics version 3.3 pipeline (https://nf-co.re/rnaseq; Ewels et al., 2020). Briefly, the barcodes were removed with UMI-tools (https://github.com/CGATOxford/UMI-tools) and adapters and low-quality reads (reads having more than 15% bases with Phred quality score less than 30 or reads having more than 8% unknown bases) were removed with Trim Galore! (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). The alignment of clean reads to the bovine reference genome (ARS-UCD1.2/bosTau9) was done with STAR (https://github.com/alexdobin/STAR) and Salmon (Patro et al., 2017) software, followed by deduplication with UMI-tools (Smith et al., 2017). Gene expression differences between SAP and HC groups was analyzed with DESeq2 (version 1.34.0; Love et al., 2014). Parity, stage of lactation, and farm were included as batch factors during analysis to understand their possible influence on gene expression differences. Genes having false discovery rate (FDR) < 0.05 (Benjamini and Hochberg, 1995) and |log2 fold change| (|log2FC|) > 1 were defined as significantly DEG.

**Weighted Gene Co-expression Network Construction and Module Detection**

The normalized expression value of DEG selected by DESeq2 was used as the input for weighted gene co-expression network analysis (WGCNA) to construct the weighted gene co-expression network (Pei et al., 2017). First of all, the input data were checked to ensure that samples or genes containing too many missing values were not included by setting parameter “verbose” of function “goodSamplesGenes” as 3. All samples included in the next-step analysis were based on the sample cluster analysis. The appropriate soft thresholding power β was selected by using the pickSoftThreshold function and based on a set of powers (1–20), and “β = 14” was used (Zhang and Horvath, 2005). Then, the one-step network construction was used to cluster splitting and generate modules (networkType = “unsigned,” TOMType = “unsigned,” minModuleSize = 30, reassignThreshold = 0, mergeCutHeight = 0.25, and other default parameters). The correlation between modules, and health status (SAP or HC) and production traits (fat %, protein %, and milk yield) were evaluated by correlating the summary profile of each module (module eigengene) to each parameter (Pearson correlation coefficient). Significant correlation was defined as having a P-value <0.05 and |r| > 0.5. In addition, the correlation between individual genes and parameter of interest (health condition) was qualified by gene significance (GS), that is the Pearson correlation (|r|) between the gene and the trait (health condition). Meanwhile, the expression profile of individual genes was also correlated with the module eigengene of each module by using Pearson correlation coefficient, defined as module membership (MM), to quantify the similarity of genes to each module.

**Functional Enrichment and Visualization of Module Genes of Interest**

To explore the possible biological influences of modules that were significantly related to S. aureus infection, module genes were submitted to Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis using the R package, ClusterProfiler (Version 4.2.2; Wu et al., 2021). In the parent-child structure of GO terms, the parent term could overlap with a large percentage of child terms, resulting in redundancy of enrichment results. Therefore, a simplified strategy was used to reduce redundancy of enriched GO terms by keeping one representative term of all semantically similar ones. In addition, the pairwise similarities of enriched GO terms were calculated using “pairwise_termsim” function, and hierarchical clustering of enriched GO terms was performed with “treeplot” function using default agglomeration method (ward.D). Adjusted P-value (FDR) < 0.05 was used as cutoff to define significantly enriched GO terms or KEGG pathways. Cytoscape (Version 3.9.1) was used for network visualization of selected discriminant genes.

**Important Genes Driving the Discrimination Between SAP and Control Groups**

All genes from the module significantly associated with the SAP group (Turquoise) were subjected to sparse partial least squares discriminant analysis (sPLS-DA) from the R package MixOmics (Version 6.18.1) to identify key variables (genes) that drive the discrimination of gene expression between SAP and HC groups (Rohart et al., 2017). The key parameters (number of components and number of variables to choose) were chosen according to the authors’ recommendations (Rohart et al., 2017). The function “perf” was used to evaluate the performance of sPLS-DA for a large number of components (1–10) using 5-fold cross-validation repeated 50 times to get the optimal number of components. And then, the function ‘tune.splsd’ was used to assess the number of variables to choose by running 5-fold cross-validation, repeated 50 times. Finally, one component and 5 variables showing the best performances were chosen. The correlation between the expression level of selected genes was calculated by Pearson correlation coefficient.
Verification of RNA Sequencing Results by Real-Time Quantitative PCR

To validate the RNA sequencing results, 7 genes (3 upregulated, 3 downregulated, and 1 not differentially expressed) were selected to quantify their expression levels by real-time quantitative (q)PCR (RT-qPCR). Primer-BLAST (https://www.ncbi.nlm.nih.gov/tools/primer-blast/) was used to design primers for selected genes (Supplemental Table S2; https://doi.org/10.6084/m9.figshare.20732380.v1; Ibeagha-Awemu et al., 2022b). Reverse transcription of total RNA (1 µg per sample) from the same samples used for RNA sequencing was done by using SuperScript IV VILO Master Mix (Invitrogen). The diluted cDNA (1:15) was then used for gene specific qPCR amplification. The 10-µL qPCR reaction mix contained 1 µL (80 ng) of cDNA, 5 µL of PowerTrack SYBP Green Master Mix (Applied Biosystems), 0.25 µL of Yellow sample buffer, 0.5 µL each of forward and reverse primers (500 nM each), and 2.75 µL of nuclease-free water. The real-time amplification through a fast cycling mode of a StepOnePlus instrument (Applied Biosystems) included 2 min of enzyme activation at 95°C, 40 cycles of denaturation (95°C for 5 s) and extension (60°C for 30 s). The 2−ΔΔCt method was used to calculate the relative expression value of genes, using ACTB (Modesto et al., 2013), and RPS9 as reference genes. These reference genes were stably expressed in this study. The Student’s t-test (2 tailed) was used to compare the relative expression levels of tested genes between SAP and HC groups.

RESULTS

Sequencing Data Analysis

The sequencing of 25 libraries by high throughput RNA sequencing generated 70.56 million raw reads on average per sample (Supplemental Table S3A; https://doi.org/10.6084/m9.figshare.20732380.v1; Ibeagha-Awemu et al., 2022b). After quality control (adaptor and low-quality sequences removed), 60.04 million (85%) clean reads on average per sample were mapped to the bovine reference genome (ARS-UCD1.2). The average GC content per sample was 47.08%. The uniquely mapped reads accounted for 92.44% of the total clean reads (Supplemental Table S3A) and 47.44, 33.76, 10.72, and 1.59% of clean reads were mapped to exons, introns, 3′UTR, and 5′UTR, respectively (Supplemental Table S3B; Supplemental Figure S1, https://doi.org/10.6084/m9.figshare.20732542.v1). Similarly, 2.92% of clean reads were in the 10 kb region downstream of the transcriptional termination site, which is more than 2-folds of the reads in the 10 kb region upstream of the transcription start sites (about 1.32%). In addition, principal component analysis (PCA) was performed to ascertain sample reliability and the global relationship between samples. The PCA plot showed that the majority of the samples clustered into 2 main groups: the SAP and the HC groups (Supplemental Figure S2; https://doi.org/10.6084/m9.figshare.20732542.v1; Ibeagha-Awemu et al., 2022a). The intergroup difference showed 61% variance, whereas intragroup difference showed 15% variance, indicating that the sample phenotypes are reliable and suitable for next-step analyses.

Differentially Expressed Genes in Response to S. aureus Infection

A total of 4,077 genes were identified to have significant differential expression levels between SAP and HC groups (FDR <0.05 and |log₂FC| > 1; Supplemental Table S4, https://doi.org/10.6084/m9.figshare.20732380.v1, Ibeagha-Awemu et al., 2022b), including 1,616 upregulated and 2,461 downregulated DEG (Figure 1A). Functional enrichment of the 4,077 DEG revealed 94 significantly enriched GO terms, including 81 biological process (BP), 4 cellular component (CC), and 9 molecular function (MF) GO terms, and 47 KEGG pathways (FDR <0.05; Figure 1B and 1C; Supplemental Table S5, https://doi.org/10.6084/m9.figshare.20732380.v1, Ibeagha-Awemu et al., 2022b). The significantly enriched GO terms revealed the possible involvement of DEG in cellular activities (adhesion, movement, and localization), tissue development, metabolic processes and immune responses (Supplemental Table S5). Twenty (25%) BP-GO terms were closely related to cell adhesion, movement and communication, represented by the most significantly enriched BP-GO terms (Figure 1C), including biological adhesion (GO:0022610, FDR = 1.10 × 10−5), cell migration (GO:0016477, FDR = 2.14 × 10−5) localization (GO:0040011, FDR = 7.11 × 10−6), and the related regulations that support the biological functions of the mammary gland (such as regulation of cell adhesion, GO:0030155, FDR = 7.41 × 10−4; regulation of cell migration, GO:0030334, FDR = 1.25 × 10−3; regulation of locomotion, GO:0040012, FDR = 1.70 × 10−3). Twenty percent of significantly enriched BP-GO terms (n = 16) were related to tissue development and relevant processes, such as regulation of multicellular organismal process (GO:0051239, FDR = 4.44 × 10−7), tissue development (GO:0009888, FDR = 1.43 × 10−6) and epithelial cell differentiation (GO:0030855, FDR = 2.14 × 10−5). Moreover, 9 BP-GO terms are relevant to...
biosynthesis and metabolic processes, especially fat and protein secretion (Supplemental Table S5). In addition, 24 (30%) BP-GO terms were related to immune defense processes, for instance, response to external stimulus (GO:0009605, FDR = 1.79 × 10⁻⁸), defense response (GO:0006952, FDR = 2.63 × 10⁻⁵), and immune system process (GO:0002376, FDR = 3.38 × 10⁻⁵).

In addition, 3 quarters of the 47 KEGG pathways were directly related to disease and immune functions (Figure 1B), such as the 5 most significantly enriched pathways (cytokine-cytokine receptor interaction, FDR = 6.04 × 10⁻⁸; cell adhesion molecules, FDR = 8.19 × 10⁻⁷; toxoplasmosis, FDR = 1.46 × 10⁻⁵; intestinal immune network for IgA production, FDR = 1.72 × 10⁻⁵ and autoimmune thyroid disease, FDR = 2.21 × 10⁻⁵). In addition, S. aureus infection (bta05150) pathway was also significantly enriched (FDR = 0.01) and 83% of the DEG (24 out of the 29 DEG) were upregulated. Interestingly, the majority of GO terms (n = 76, 81%) and KEGG pathways (n = 35, 74%) were enriched for by mainly down- and upregulated DEG, respectively (Figure 1D). The majority of genes (≥50%) in ≥90% of BP-GO terms related to cell migration, metabolic processes, and tissue development were downregulated (Supplemental Table S5). Meanwhile 58% immune-related BP-GO terms and 84% immune- or disease-related KEGG pathways were mainly enriched for by upregulated DEG (increased expression in SAP group). Furthermore, 70% of genes found in locomotion (GO:0040011, FDR = 5.25 × 10⁻⁷) as well as 80% of genes found in tissue development (GO:00088488), lipid biosynthetic process (GO:0006610, FDR = 0.01), and regulation of peptide secretion (GO:0002791, FDR = 0.04) GO terms were downregulated in SAP group. This suggests that the genetic regulation of immune-related functions may be enhanced, but the
regulation of mammary gland processes/functions and milk production possibly is impaired during *S. aureus* subclinical mastitis.

**Gene Co-expression Network and Gene Modules Associated with *S. aureus* Infection**

The gene co-expression network of the 4077 DEG based on WGCNA resulted in 7 modules colored black, turquoise, red, green, blue, brown, and yellow, and here referred to as Black, Turquoise, Red, Green, Blue, Brown, and Yellow modules, respectively (Figure 2A; Supplemental Table S6, https://doi.org/10.6084/m9.figshare.20732380.v1; Ibeagha-Awemu et al., 2022b). The Turquoise module was the biggest with 1,546 DEG, including 1,480 up- and 66 downregulated DEG, followed by Blue module (1,219 up- and 60 downregulated DEG; Supplemental Table S6). As shown in Figure 2B, all modules, except the Green module, were significantly correlated with *S. aureus* subclinical mastitis (*P* < 0.05 and |r| > 0.5). However, milk production performance (milk yield, milk fat, and protein content) showed no significant correlation with any of the 7 modules. Turquoise and Black modules were positively correlated with *S. aureus* subclinical mastitis, whereas Yellow, Blue, Brown, and Red modules showed negative correlations (Figure 2B). The module similarity analysis revealed that the Turquoise module was most closely clustered with *S. aureus* infection condition (Figure 2C). In addition, there was a highly significant correlation between the GS and MM in the Turquoise module, and the vast majority of genes in this module are strongly correlated with *S. aureus* subclinical mastitis (Figure 2D). Significant positive correlations between GS and MM were also identified for Black, Red, Yellow, and Brown modules, but not as strong as with the Turquoise module (Supplemental Figure S3; https://doi.org/10.6084/m9.figshare.20732542.v1; Ibeagha-Awemu et al., 2022a). Meanwhile, the GS of genes in Blue module was negatively correlated with MM (Supplemental Figure S3D). Therefore, the Turquoise module was considered as the most important and relevant module to *S. aureus* subclinical mastitis.

**Functional Enrichment of Modules Associated with *S. aureus* Infection**

Genes of all 6 modules significantly correlated with *S. aureus* subclinical mastitis were further submitted for functional enrichment analysis separately to explore their possible biological effects. The modules positively correlated with *S. aureus* subclinical mastitis (Turquoise and Black) were enriched in immune- and disease-related functional annotations (Supplemental Table S7A and B; https://doi.org/10.6084/m9.figshare.20732380.v1; Ibeagha-Awemu et al., 2022b). The Turquoise module with 1,546 genes was significantly enriched in 48 GO terms (42 BP-, 4 MF-, and 2 CC-GO terms), with over 90% of GO terms being immune-related processes (Supplemental Table S7A). As shown in Figure 3A, 42 BP-GO terms were hierarchically clustered into 5 clusters according to their pairwise similarities. The network depicting the linkages of genes and BP-GO terms of each cluster suggest their involvement in different immune functions. The top 8 most significantly enriched BP-GO terms grouped in the first cluster and are related to defense and immune response to stimuli processes (Figure 3A). Most of the DEG found in BP-GO terms of cluster one were upregulated, suggesting their possible genetic effects on enhanced immune responses (Figure 3B). Cluster 4 was the biggest cluster and included 15 BP-GO terms related to key cellular processes required for the immune response, especially leukocytes represented by T cell and cytokine, and the relevant regulation of these immune-related cells (Figure 3A). Furthermore, 6 BP-GO terms of cluster 5 are closely related to T cell mediated immunity and cell killing. Meanwhile clusters 2 and 3 are related to response to cytokine and movement of immune-related cells, respectively. Similar with cluster one, the genes in other clusters were also mostly upregulated (Supplemental Figure S4; https://doi.org/10.6084/m9.figshare.20732542.v1; Ibeagha-Awemu et al., 2022a), further revealing the potential influence of upregulated expression of genes in the Turquoise module in enhancing mammary gland immune responses to *S. aureus* subclinical mastitis.

In addition, the Turquoise module was significantly enriched in 72 KEGG pathways with immune- and disease-related functions (Figure 3C, Supplemental Table S7A), further revealing the regulatory roles of *S. aureus* module genes in the immune defense against *S. aureus* invasion. For instance, cytokine-cytokine receptor interaction (bta04060) was the most significantly enriched immune-related pathway (FDR = 4.36 × 10^{-14}). Additionally, the *S. aureus* invasion pathway (bta05150, FDR = 1.38 × 10^{-7}) was significantly enriched by the Turquoise module. As showed in Figure 4, Turquoise module genes found in *S. aureus* infection pathway were upregulated in SAP group and may affect the key immune processes, such as inhibition of chemotaxis and phagocyte activation, inhibition of neutrophil transmigration, B cell proliferation, apoptosis, and inflammation. In addition, KRT10 and FGG from the Blue module are implicated in the colonization process by *S. aureus*, whereas MASP1 and CIQTC are involved in complement activation (Figure 4). A total of 406 genes in the Turquoise module were found
Figure 2. The weighted gene co-expression network of differentially expressed genes (DEG). (A) Heatmap of gene networks. Branches in the hierarchical clustering dendrogram correspond to modules and are marked with different colors, notably turquoise, red, blue, brown, yellow, black, and green, as shown in the left or top edge bars. Modules are consistent with the blocks in the heatmap of highly interconnected genes as revealed by gradually saturated yellow and red colors. (B) The correlation between modules and cows’ production and health traits. In each cell, the value at the top represents the correlation coefficient and the value at the bottom indicates level of significance. ME = module (MEblack = black module and so on); TD = test day; milk24kg = 24 hours milk yield (kg) of the test day; milk305kg = 305-day milk yield (kg); Fat305kg = 305-day milk fat yield (kg); Prot305kg = 305-day milk protein yield (kg). (C) Correlation between module eigengenes and trait (*Staphylococcus aureus* infection). The hierarchical clustering dendrogram is at the top and the heatmap is at the bottom. The blue color in the heatmap represents low adjacency (negative correlation), and red color represents high adjacency (positive correlation). SAP = *S. aureus* positive. (D) The scatterplot of gene significance and module membership in the Turquoise module. Cor = Pearson correlation coefficient.
in 48 GO terms and 72 KEGG pathways, including 106 genes found in ≥10 immune-related annotations (GO terms and KEGG pathways) (Supplemental Table S8; https://doi.org/10.6084/m9.figshare.20732380.v1; Ibeagha-Awemu et al., 2022b). For example, the top 10 genes with immune-related functions and found in most annotations are BOLA, IFNG, IL18, NFKB1, IL1B, CXCL8, IKBKG, IKBKB, TLR4 and NLRP3. The Black module was a small module with only 40 upregulated DEG, and significantly enriched in 1 BP-GO term and 6 KEGG pathways related to antigen processing and presentation (GO:0019882, FDR = 4.79 × 10⁻³, bta04612, FDR = 0.01) and diseases (Supplemental Table S7B).

The 4 modules (Yellow, Brown, Red, and Blue) significantly negatively correlated with S. aureus subclinical mastitis were enriched in less GO terms and KEGG pathways compared with the Turquoise module, and mostly have mammary gland development related functions (Supplemental Table S7C–F). The Yellow module (303 genes) was significantly enriched in 14 BP-, 8 CC-GO terms, and 11 KEGG pathways, mostly related to cellular activities and movement, such as cell adhesion, migration, and localization, especially of epithelial...

Figure 3. Gene Ontology (GO) terms (biological process) and pathways significantly enriched by genes in the Turquoise module. (A) Hierarchical clustering of enriched terms. (B) Networks showing the linkages between genes and GO terms of cluster one (lime color) identified in (A). (C) Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways significantly enriched by genes in the Turquoise module. The top 40 KEGG pathways ranked by significance (false discovery rate, FDR) are visualized in the plot. The detail information about all enriched GO terms and KEGG pathways for Turquoise module can be found in Supplemental Table S8 (https://doi.org/10.6084/m9.figshare.20732380.v1; Ibeagha-Awemu et al., 2022b).
Figure 4. The Kyoto Encyclopedia of Genes and Genomes (KEGG) *Staphylococcus aureus* infection pathway (bta05150). The fill color of each gene reports significant differentially expressed genes (log_{fold change}, log_{FC}): the green and red colors represent down- and upregulated expressed genes, respectively. The deeper the color, the greater the |log_{FC}| is. The outer border color represents the module a gene belongs to. For example, the LOC524810 and FGG genes with turquoise and blue borders belong to the Turquoise and Blue modules, respectively.

cells (Figure 5A; Supplemental Figure S5A, https://doi.org/10.6084/m9.figshare.20732542.v1, Ibeagha-Awemu et al., 2022a). For instance, the top 3 most significantly enriched GO terms in the Yellow module were cell adhesion (GO:0007155, FDR = 4.18 × 10^{-4}), biological adhesion (GO:0022610, FDR = 4.18 × 10^{-4}),
and epithelial cell migration (GO:0010631, FDR = 3.85 × 10^{-3}). In addition, Brown module (708 genes) was enriched in 16 BP-GO terms and 9 KEGG pathways mostly related to cell communication and signal transduction (Figure 5B, Supplemental Figure S5B), such as regulation of cell communication (GO:0010646, FDR = 0.03), and regulation of signaling (GO:0023051, FDR = 0.03). In addition, enriched BP-GO terms of the Brown module also revealed involvement in macrophage regulation, such as macrophage chemotaxis (GO:0048246, FDR = 0.03) and macrophage migration (GO:1905517, FDR = 0.04). The Blue module contained more genes (n = 1,279) and significantly enriched in 8 BP-, 1 MF-, 1 CC-, and 5 KEGG pathways related to metabolic and biosynthetic processes, suggesting its possible involvement in milk production processes (Figure 5C, Supplemental Figure S5C). The Red module (63 downregulated DEG) was significantly enriched in 9 BP-GO terms and 5 KEGG pathways related to blood circulation system including circulatory system development (GO:0072359, FDR = 3.06 × 10^{-3}), negative regulation of angiogenesis (GO:0016525, FDR = 1.11 × 10^{-3}), and negative regulation of blood vessel morphogenesis (GO:2000181, FDR = 1.11 × 10^{-3}; Figure 5D, Supplemental Figure S5D). It is worth noting that genes found in GO terms and KEGG pathways of the Yellow, Brown, Red, and Blue modules were mainly downregulated (Figure 5, Supplemental Figure S5), suggesting that these 4 modules may play roles in decreased mammary gland processes and production performance during S. aureus subclinical mastitis. In addition, compared with the Turquoise module, the Black, Yellow, Brown, Blue, and Red modules were enriched in much less GO terms and KEGG pathways with generally less significance (higher FDR values), indicating their minimal involvement in the host immune response to S. aureus subclinical mastitis, which is consistent with the weaker and less significant correlation between them and S. aureus subclinical mastitis (Figure 2B).

**Validation for RNA Sequencing by RT-qPCR**

Three upregulated DEG (TGM3, G0S2, and STEAP4), 3 downregulated DEG (PDLIM5, GLYCAM1, and CSN1S2), as well as one non-DEG (RPS15), were selected to check their expression level by RT-qPCR for validating results generated by RNA sequencing data. As showed in Figure 7, the expression levels of the non-DEG RPS15 detected by RT-qPCR also showed no significant differences between SAP and HC groups. The expression change of PDLIM5, which is one of the 5 candidate discriminant genes, was similar between RT-qPCR and RNA sequencing results. Although the log2FC of the other 5 DEG detected by RT-qPCR and RNA sequencing were slightly different, the changes were generally similar which confirms the reliability and accuracy of the RNA sequencing data.

**DISCUSSION**

Because the milk somatic cell content is an indicator of udder health, this study investigated the transcriptional profiles of milk somatic cells to S. aureus pathogen by next-generation sequencing technology. In addition to identifying genes with altered expression, the gene co-expression network technique was used to further explore the regulatory roles of DEG during S. aureus subclinical mastitis. Six modules significantly correlated with S. aureus subclinical mastitis were revealed by the gene co-expression network analysis, and their genes were found in different functional terms pointing to varied roles by the gene networks during S.
S. aureus subclinical mastitis. In particular, the Turquoise module was most significantly positively correlated with S. aureus subclinical mastitis and its genes were found in immune-related functions, among which 5 were considered important for S. aureus subclinical mastitis due to their ability to distinguish between SAP and HC cows in this study.

Comparing the transcriptome of milk somatic cells between SAP and HC groups revealed 4,077 DEG. About 40% (1,627) of the DEG identified in this study were also detected as DEG in milk somatic cells after 24 to 168 h challenge of Holstein-Friesian cows with 2 strains of S. aureus (strains MOK124 and MOK 023; Niedziela et al., 2021). Moreover, investigation of the

Figure 5. Biological process Gene Ontology (GO) terms significantly enriched by genes in 4 modules negatively correlated with Staphylococcus aureus subclinical mastitis. Panels A through D represent Yellow, Brown, Blue, and Red modules, respectively. In each figure, the straw-colored bigger dots in relatively central positions represent the significantly enriched GO terms. The small dots around the larger centrally located dots represent associated genes, whose color intensities represent the gene expression changes (log2 fold change, log2FC). The line between dots represents the interaction between genes and enriched GO terms. Detail information on all the enriched GO terms for each module can be found in Supplemental Table S7 (https://doi.org/10.6084/m9.figshare.20732380.v1; Ibeagha-Awemu et al., 2022b).
expression changes of 167 immune-related genes in milk somatic cells from Holstein cows with naturally occurring 
*S. aureus* mastitis found 16 DEG (Tao and Mallard, 2007). Six of the DEG, including 5 upregulated (TAP2, ICAM3, CTLA4, TCIRG1, and JSP.1) and one downregulated (TLR4; Tao and Mallard, 2007), were also detected, and at similar levels of expression changes in this study. These reports further strengthened the observed associations of these DEG with *S. aureus* mastitis. The functional enrichment revealed that more than half of the DEG involved in the regulation of BP related to mammary gland immune defense and immune- or disease-related KEGG pathways were upregulated in SAP group (Figure 1D). Although more than 90% of GO terms related to cell migration, signal transduction, and cellular activities, the activities of epithelial cells required for mammary gland homeostatic response to injuries induced by the continuous infiltration of immune cells, were mainly enriched for by downregulated DEG (Figure 1D, Supplemental Table S5). This suggests that the gene expression changes detected in the naturally infected samples could reflect the genetic regulation and management of the immune defense against *S. aureus* infection, and also of the possible negative effects on the normal functioning of the mammary gland during subclinical mastitis. The difference in the functional enrichment of up- and downregulated DEG also reflected that genes may have a clear division of labor to regulate the host response to subclinical mastitis and other mammary gland functions. This underlines the need to build a network of DEG to further explore their regulatory mechanisms during subclinical mastitis. Therefore, the WGCNA approach was used to build the weighted gene co-expression network of DEG responding to *S. aureus* in the mammary gland, which resulted in 6 gene modules with different possible influences.

The Turquoise module emerged as the most important module with significant and strongest positive correlation with *S. aureus* subclinical mastitis. Functional enrichment revealed notable regulatory roles of the Turquoise module genes in the mammary gland defense against *S. aureus* infection and immune-related functions (Figures 3). As showed in Figure 4, the *S. aureus* infection pathway (bta05150) was enriched for by upregulated genes in the Turquoise module, suggesting the involvement of these genes in various immune processes during *S. aureus* subclinical mastitis. For instance, half of the key genes related to inhibition of chemotaxis and phagocyte activation were upregulated, including FCAR, FCGR3A, C5AR1, and C5AR2 (Figure 4). The upregulated expression of *DEFB* and *IL37* might have affected the antimicrobial activity of neutrophils in this study, especially *IL37* (a cytokine of the *IL*-*1* family) previously found to be involved in the regulation of anti-inflammatory processes (Boraschi et al., 2011; Nold-Petry et al., 2015). In addition, toll-like receptors (TLR), which are among well-studied pattern recognition receptors in cattle, play important roles in pathogen recognition and initiation of the immune responses after the translocation of nuclear factor (NF)-κB (Brightbill and Modlin, 2000; Wellnitz et al.,

### Table 1. Five candidate discriminant genes that demonstrated the most expression variation and distinguished *Staphylococcus aureus*-positive cows from healthy control cows

<table>
<thead>
<tr>
<th>Gene</th>
<th>log2FC1</th>
<th>FDR1</th>
<th>Loading weight2</th>
<th>GS3</th>
<th>p.GS3</th>
<th>MM4</th>
<th>p.MM4</th>
<th>Function</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>NR2F6</em></td>
<td>−1.54</td>
<td>2.81 × 10⁻⁶</td>
<td>0.62</td>
<td>−0.86</td>
<td>2.88 × 10⁻⁷</td>
<td>−0.83</td>
<td>2.69 × 10⁻⁷</td>
<td>DNA-binding transcription factor activity and involved in negative regulation of transcription by RNA polymerase II.</td>
</tr>
<tr>
<td><em>PDLIM5</em></td>
<td>−1.56</td>
<td>7.24 × 10⁻⁵</td>
<td>0.52</td>
<td>−0.86</td>
<td>4.54 × 10⁻⁸</td>
<td>−0.75</td>
<td>1.45 × 10⁻⁵</td>
<td>Involved in binding activities, cell adhesion, actin cytoskeleton organization, and cell growth.</td>
</tr>
<tr>
<td><em>RAB11FIP5</em></td>
<td>−1.40</td>
<td>9.80 × 10⁻⁷</td>
<td>0.47</td>
<td>−0.85</td>
<td>5.45 × 10⁻⁸</td>
<td>−0.80</td>
<td>1.61 × 10⁻⁶</td>
<td>Participates in gamma-tubulin binding, regulation of exocytosis, cellular response to acidic pH, regulation of protein localization to cell surface.</td>
</tr>
<tr>
<td><em>ACOT4</em></td>
<td>−1.73</td>
<td>5.59 × 10⁻⁵</td>
<td>0.35</td>
<td>−0.85</td>
<td>9.17 × 10⁻⁸</td>
<td>−0.87</td>
<td>1.35 × 10⁻⁸</td>
<td>Enables acyl-CoA hydrolase and succinyl-CoA hydrolase activities. Involved in metabolic processes of carboxylic acid, saturated monocarboxylic acid, and succinyl-CoA.</td>
</tr>
<tr>
<td><em>TMEM53</em></td>
<td>−1.33</td>
<td>1.23 × 10⁻⁵</td>
<td>0.09</td>
<td>−0.83</td>
<td>2.53 × 10⁻⁷</td>
<td>−0.75</td>
<td>1.93 × 10⁻⁵</td>
<td>Conserved in different species and located in the nucleus.</td>
</tr>
</tbody>
</table>

1Gene expression changes (log₂ fold change, log₂FC) and adjusted P-values (false discovery rate, FDR) of corresponding genes in the *S. aureus*-positive (SAP) group compared with healthy control (HC) group. Negative values indicate downregulation of the gene expression (i.e., expression level in SAP was downregulated compared with HC).
2The loading weights calculated by sparse partial least squares discriminant analysis (sPLS-DA). Most important genes are ordered from top to bottom according to the absolute value of their coefficients.
3Gene significance (GS) and P-values (p.GS) indicate the correlation between each gene and the phenotype (*S. aureus* positive). The higher the absolute value of GS, the more biologically significant the gene is.
4Module membership (MM) and P-values (p.MM) revealed the correlation between each gene and the module eigengene of the Turquoise module.
Toll-like receptor signaling pathway (bta04620) and NF-κB signaling pathway (bta04064) were enriched for by 20 and 33 upregulated DEG, including the key genes TLR4, TLR9, NFκB1, and NFκB2, respectively, supporting the involvement of these pathways in mammary gland immune recognition of S. aureus and response to the pathogen. It should be noted that TLR2 was also significantly upregulated in this study (log₂FC = 0.9768, FDR = 0.039), but did not pass the filter of |log₂FC| > 1 and was not included in the WGCNA analysis. Among these key upregulated genes, TLR4 was the most highly expressed gene in the S. aureus group (log₂FC = 1.53, when compared with control group), which is consistent with the important role of TLR4 in recognizing different types of ligands, mainly LPS from gram-negative bacteria such as Escherichia coli, as well as viral proteins, polysaccharide, some endogenous proteins, and peptidoglycan, from gram-positive or gram-negative pathogens including S. aureus, Streptococcus uberis, and E. coli (Goldammer et al., 2004; Hadley et al., 2005; Ibeagha-Awemu et al., 2008; Porcherie et al., 2012). Furthermore, the activation of TLR4 was found to associate with the expression of pro-inflammatory cytokines and activa-
tion of NF-κB (Akira et al., 2001), whose upregulation was also detected in this study. The enhanced activity of NF-κB was reported in milk and mammary gland tissues of dairy cows with mastitis (Wu et al., 2015), which is consistent with the upregulation of NF-κB genes and the significant enrichment of NF-κB signaling pathway in this study.

Additionally, the upregulation of pro-inflammatory chemokines and their receptors in SAP group [such as C-X-C motif chemokine receptor family (CXCR1, CXCR2, CXCR4, CXCR6), C-C motif chemokine receptor family (CCR1, CCR2, CCR6, CCR3, CCR5, CCR7, CCR12), interferon family (IFNG, IFNAR1, IFNAR2, IFNGR1, and IFNGR2) and IL family (such as IL18, IL1B, IL12B, IL15, IL17A, IL2RG, IL2RA, and IL6R)] in this study probably contributed to increase the bacteriostatic and bactericidal activities required to kill S. aureus (Sordillo, 2018; Zhelavskyi, 2021). This is supported by the significant enrichment of related GO terms and pathways, including cell killing (GO:0001906), leukocyte mediated cytotoxicity (GO:0001909), regulation of cell killing (GO:0031341) and natural killer cell mediated cytotoxicity (bt04650), and other terms listed in Figure 3A. After sensing S. aureus presence, leukocytes are recruited to the mammary gland to contain the infection (Sordillo, 2018). As revealed by the significantly enriched related BP GO and pathways [such as leukocyte activation (GO:0045321), positive regulation of lymphocyte mediated immunity (GO:0002708), cytokine-cytokine receptor interaction (bt04060), TNF signaling pathway (bt04668) and IL-17 signaling pathway (bt04657)], the Turquoise module genes may also play roles in regulating the activities of immune-related cells (Figure 3). Cytokine-cytokine receptor interaction was the most significantly enriched KEGG pathway with immune functions and with 60 genes involved. Cytokines are essential to drive acute-phase responses at almost every key stage that promote inflammation (Mohamed Ibrahim et al., 2016). The differential expression of diverse cytokines, such as TNFα,
were enriched in functional annotations represented by the Turquoise module. The Yellow and Brown modules was in contrast to genes with upregulated expression in 4 modules were downregulated in SAP group, which mastitis. The vast majority of genes (>95%) in these found to correlate negatively with S. aureus subclinical.

The potential in the genetic regulation of the immune response with et al., 2018; Mondino et al., 2020). Therefore, the Turquoise module in the genetic regulation of potential immunity of the mammary gland. In addition, the specific immunity of the mammary gland (Bröker et al., 2016). Moreover, the upregulation of important genes such as IL18, IL12B, IFNG, and NLRP3, and enriched functional annotations related to the regulation of T cell activities, suggest that T cell activities may be enhanced to orchestrate the immune response to S. aureus infection and balance the protection and pathology during subclinical mastitis. In addition to the functional annotations related to T cell regulation, some other important pathways with key roles in specific immune processes were also enriched in the Turquoise module, such as B cell receptor signaling pathway (bta04662), antigen processing and presentation (bta04612), Th1 and Th2 cell differentiation (bta04658), further indicating the crucial involvement of Turquoise module in the genetic regulation of specific immunity of the mammary gland. In addition, a bunch of enriched disease KEGG pathways of Turquoise module, such as Epstein-Barr virus infection (bta05169), toxoplasmosis (bta05145), leishmaniasis (bta05140), Chagas disease (bta05142), and legionellosis (bta05134), are activated by pathogens known to persist for long-term in host by subverting the host immune system (Cruz-Pacheco et al., 2012; Dumuire et al., 2018; Mondino et al., 2020). Therefore, the Turquoise module being the most significantly correlated with S. aureus subclinical mastitis, has considerable potential in the genetic regulation of the immune response of the mammary gland to infection by S. aureus.

Four modules (Yellow, Brown, Blue, and Red) were found to correlate negatively with S. aureus subclinical mastitis. The vast majority of genes (>95%) in these 4 modules were downregulated in SAP group, which was in contrast to genes with upregulated expression in the Turquoise module. The Yellow and Brown modules were enriched in functional annotations represented by cell movement activities, such as cell adhesion and migration and cell communication, and signal transduction (Figure 5A and B, Supplemental Table S7C–D), respectively. In addition, the Red module was enriched in BP-GO terms closely related to blood circulation system process (Figure 5C), which was an important guarantee for maintaining the health and normal functioning of the mammary gland (Ryman et al., 2015). The downregulated expression of genes in these modules reflects the possible regulatory mechanisms underlying impaired cellular activities and decreased productivit of the mammary gland during subclinical mastitis. During long-term subclinical mastitis, mammary gland cells are easily invaded causing damage to the epithelial cell lining of the mammary gland, impairs its recovery ability and affects milk production. Functional annotations enriched by the Blue module were more related to metabolic processes, especially the metabolic and biosynthetic process of small molecule, such as lipid and steroid (Figure 5D, Supplemental Table S7E). This suggests the possible involvement of these modules in the activities leading to decreased milk production and lower milk fat and protein yields during subclinical mastitis (Asfaw and Negash, 2017). Although no significant correlations were detected between milk production traits and any module, the majority of them were negative (Figure 2B). The enrichment of metabolic pathways in Blue module indicates the possibility that the metabolic activities in the mammary glands of cows of the SAP group were affected, which might support a metabolic reprogramming during S. aureus subclinical mastitis. Therefore, the strategy of construction of gene co-expression networks revealed deeper understanding of the specific roles of highly interconnected gene clusters during S. aureus mastitis than strategies that interrogate single list of DEG. The DEG clusters indicated that during S. aureus subclinical mastitis, the Turquoise module genes are involved in immune responses, the Yellow module genes in cell movement and localization, the Brown module genes in signal transduction, the Blue module genes in metabolic processes, and the Red module genes in blood circulation system processes, and therefore provide specific reference targets for next-step functional exploration of processes of interests.

In line with respecting antimicrobial stewardship principles, animal welfare and the proactive thinking of “prevention is better than cure,” genetic breeding is one of the strategies used to improve the resistance to mastitis in dairy cows (Mallard et al., 2015; Ruegg, 2017; Cunha, 2018; Martin et al., 2018). Identifying important genes related to S. aureus infection will support the development of new control methods for subclinical mastitis and the genetic breeding strategy for resistance.
to mastitis infection. To identify such genes, the sPLS-DA method was used to identify 5 genes showing the most expression variation and also being able to clearly distinguish between the SAP and HC groups, from the Turquoise module. The 5 candidate discriminant genes selected by sPLS-DA are also the top 5 most significant genes identified by WGCNA as correlated with S. aureus infection, further supporting their association with S. aureus subclinical mastitis. All 5 candidate discriminant genes (NR2F6, PDLIM5, RAB11FIP5, ACOT4, and TMEM53) were downregulated (Figure 6 and Table 1), and their expression levels were strongly correlated with each other (Figure 6D and E). NR2F6, the most important candidate discriminant gene with highest loading weigh and most significantly correlated with S. aureus subclinical mastitis was found as a key regulator of hepatic inflammatory response in beef cattle (Alexandre et al., 2019). The genes RAB11FIP5 and TMEM53 were found to play roles during bacterial or viral infection in ruminants. RAB11FIP5 was found to be targeted by miRNAs to regulate the endocytosis pathway during the inflammatory response of alveolar macrophage to Mycobacterium bovis infection (Vegh et al., 2015). Upregulated expression of TMEM53 was reported to possibly affect small ruminant lentiviruses infection in goat (Olech et al., 2021). Reported roles for some of the candidate discriminant genes are related to livestock productivity. For example, NR2F6 and ACOT4 were identified as associated with feed efficiency in cattle or pig (Ramayo-Caldas et al., 2018; Alexandre et al., 2020; Brunes et al., 2021). The second most important candidate discriminant gene, PDLIM5, is a cytoskeleton-related protein involved in cytokines functions and cellular mechanotransduction (Elbediwy et al., 2018; Huang et al., 2020), and one QTL for milk yield has been identified in PDLIM5 in Chinese Holstein population (Liu et al., 2013). As aforementioned, many of the candidate discriminant genes have not been studied in relation to bovine mastitis and other bovine diseases. Therefore, further functional studies and validation in larger cow populations are necessary.

However, it is worth mentioning that this study still has some limitations. First, the milk somatic cells consist of multiple cell types, each of which may respond differently to the presence of S. aureus and which could not be detected by the RNA sequencing technology used. The milk somatic cell composition was changed in SAP cows compared with the HC cows in this study and because the RNA-sequencing technology analyzed the same concentration of starting RNA for all the cows, the gene expression differences were probably due to a higher proportion of immune cell populations in the SAP samples. For example, the expression level of the pan-leukocyte marker CD35 was significantly (FDR = 0.0017) increased in SAP cows (however, its log2FC = 0.85 was below the threshold of 1 and was not included in functional analyses). However, the current data did not distinguish the gene expression changes coming from the different cell populations. Therefore, further investigations with robust technologies such as single-cell sequencing (milk somatic cells) or spatial profiling of mammary gland tissues will better explore the spatiotemporal specific genetic regulation of the different cell types in response to subclinical mastitis. The cows enrolled in this study were collected from the same region (Quebec province of Canada), which may lead to possible geographic specificity of the transcription profiles constructed in this study. Also, 85% of SAP cows were of known IMI status, including being positive for S. aureus, but the exact duration of infections were not known, which may influence the gene expression profiles. Therefore, more samples with detailed health status collected from different regions of Canada and other countries could be helpful to improve the identification of the transcriptional characteristics of the mammary gland in relation to S. aureus mastitis.

In addition, the identified candidate discriminant genes require further validation of their ability to distinguish SAP cows from healthy cows, using a higher number of samples and different populations.

**CONCLUSIONS**

A total of 4,077 DEG were identified in the transcriptomes of milk somatic cells from S. aureus-positive cows compared with healthy cows. Gene co-regulation network of the DEG revealed 6 modules correlated with S. aureus subclinical mastitis. The Yellow, Brown, Blue, and Red modules correlated negatively with S. aureus infection, and their enriched processes were related to the regulation of cell movement and adhesion, signal interaction, metabolic processes and blood circulation system processes, respectively. The Turquoise module with 1,546 DEG was the most significant and positively correlated with S. aureus subclinical mastitis, and its enriched terms have immune- and disease-related functions, including a subgroup of immune-related genes (e.g., IFNG, IL18, IL1B, NFKB1, and CXCL8) suggesting important regulatory roles in mammary gland immune defense in response to S. aureus subclinical infection. In addition, 5 genes (NR2F6, PDLIM5, RAB11FIP5, ACOT4, and TMEM53) driving the most difference in gene expressions and able to distinguish between SAP and HC groups were identified. Overall, the results furthered understanding of the genetic changes in transcriptional profiles of milk somatic cells (mammary gland) during S. aureus mastitis. The described candidate discriminant genes could constitute...
ACKNOWLEDGMENTS

Funding for this study was provided by Agriculture and Agri-Food Canada (AAFC no. J-002223). The financial support from the program of China Scholarship Council during the PhD study of Mengqi Wang in Canada is acknowledged (No.202008880009). Participating producers are acknowledged for allowing access to their cows for animal sampling. EL-A conceptualized the research project with inputs from NB and J-PR. P-LD, MW, JPR, and EL-A performed field sampling. P-LD and MW prepared RNA samples for RNA sequencing. Data processing and bioinformatics analyses were performed by ML, DG, and MW. EL-A, and M-AS were responsible for MW supervision. EL-A obtained funding for the project. MW drafted the manuscript with formal interpretation of the transcriptomic data by EL-A, NB, and J-PR. All authors (MW, ML, P-LD, J-PR, NB, DG, M-AS, and EI-A) revised and accepted the final version. The raw RNA-sequencing data reported in this study have been deposited in the NCBI Sequence Read Archive (SRA) under the BioProject PRJNA878880. The authors declare that they have no conflict of interest.

REFERENCES


Porcherie, A., P. Cunha, A. Trotreau, P. Roussel, F. B. Gilbert, P. Rainard, and P. Germon. 2012. Repertoire of *Escherichia coli* ago...


ORCIDS

Mengqi Wang ♦ https://orcid.org/0000-0002-6067-6067
Nathalie Bissonnette ♦ https://orcid.org/0000-0003-3281-209X
Mario Laterrière ♦ https://orcid.org/0000-0001-5659-1951
David Gagné ♦ https://orcid.org/0000-0003-1331-8774
Jean-Philippe Roy ♦ https://orcid.org/0000-0002-0444-2303
Marc-André Sirard ♦ https://orcid.org/0000-0001-8667-6682
Eveline M. Ibeagha-Awemu ♦ https://orcid.org/0000-0001-8810-9294

Journal of Dairy Science Vol. 106 No. 8, 2023