Methylome-wide analysis of milk somatic cells upon subclinical mastitis in dairy cattle

D. Giannuzzi, E. Capra, V. Bisutti, A. Vanzin, P. Ajmone Marsan, A. Cecchinato, and S. Pegolo

Department of Agronomy, Food, Natural Resources, Animals and Environment (DAFNAE), University of Padova, viale dell’Università 16, Legnaro (PD), Italy

Institute of Agricultural Biology and Biotechnology, National Research Council (IBBA CNR), via Einstein, Lodi, Italy

Department of Animal Science, Food and Nutrition (DIANA), Università Cattolica del Sacro Cuore, via E. Parmense 84, Piacenza, Italy

ABSTRACT

Better understanding of the molecular mechanisms behind bovine mastitis is fundamental for improving the management of this disease, which continues to be of major concern for the dairy industry, especially in its subclinical form. Disease severity and progression depend on numerous aspects, such as livestock genetics, and the interaction between the causative agent, the host, and the environment. In this context, epigenetic mechanisms have proven to have a role in controlling the response of the animal to inflammation. Therefore, in this study we aimed to explore genome wide DNA methylation of milk somatic cells (SC) in healthy cows (n = 15) and cows affected by naturally occurring subclinical mastitis by Streptococcus (Strep.) agalactiae (n = 12) and Prototheca spp. (n = 11), to better understand the role of somatic cells methylome in the host response to disease. Differentially methylated regions (DMR) were evaluated comparing: i) Strep. agalactiae-infected vs. healthy; ii) Prototheca-infected vs. healthy, and iii) mastitis vs. healthy and iv) Strep. agalactiae-infected vs. Prototheca-infected. The functional analysis was performed at 2 levels. To begin with, we extracted differentially methylated genes (DMG) from promoter - DMR, which were analyzed using the Cytoscape ClueGO plug-in. Coupled with this DMG-driven approach, all the genes associated with promoter - methylated regions were fed to the Pathifier algorithm. From the DMR analysis, we identified 1,081 hypermethylated and 361 hypomethylated promoter regions in Strep. agalactiae infected animals, while 1,514 hypermethylated and 358 hypomethylated promoter regions were identified in Prototheca infected animals, when compared with the healthy controls. When considering infected animals as a whole group (regardless of the pathogen), we found 1,576 hypermethylated and 460 hypomethylated promoter regions. Both pathogens were associated with methylation differences in genes involved in pathways related to meiosis, reproduction and tissue remodeling. Exploring the whole methylome, in subclinically infected cows we observed a strong de-regulation of immune related pathways, such as nuclear factor kB and toll-like receptors signaling pathways, and of energy-related pathways such as the tricarboxylic acid cycle and unsaturated fatty acid biosynthesis. In conclusion, no evident pathogen-specific SC methylome signature was detected in the present study. Overall, we observed a clear regulation of host immune response driven by DNA methylation upon subclinical mastitis. Further studies on a larger cohort of animals are needed to validate our results and to possibly identify a unique SC methylome that signifies pathogen-specific alterations.

Keywords: MBD-seq, milk somatic cells, subclinical mastitis, Prototheca

INTRODUCTION

Mastitis is a major concern for the dairy industry with consequent animal welfare and economic losses due to the adverse effects on milk production, quality and composition (Seegers et al., 2003). Bovine mastitis is mainly caused by infection with a wide range of microorganisms, among which the most common are Escherichia coli (E. coli), Staphylococcus aureus (S. aureus), Streptococcus agalactiae (Strep. agalactiae) and Streptococcus uberis (Strep. uberis) (Ranjan et al., 2006). Strep. agalactiae is a Gram-positive bacterium mostly responsible for inducing subclinical and/or chronic forms of the disease (Lewandowska-Sabat et al., 2019), which has recently reemerged as a significant causative agent of mastitis, despite the implementation of eradication programs (Barsi et al., 2022).

On the other hand, algae of the genus Prototheca, which a few years ago were labeled as uncommon mas-
tis agents, have rapidly emerged as non-negligible pathogens (Shave et al., 2021). Due to their capacity to infect and survive within macrophages and invade mammary tissue, they cause persistent infection with intermittent shedding (Roesler and Hensel, 2003). In addition, as *Prototheca* spp. does not respond to antibiotic mastitis therapy, the only effective control method to date has been the elimination of the infected animals (Libisch et al., 2022), which represents a serious challenge for the dairy sector.

The mastitis-causing pathogens induce an inflammatory response in the mammary gland. In response to the inflammation the host secretes cytokines causing changes in the regulation of gene expression (Oviedo-Boyso et al., 2007). There is growing evidence on the epigenetic regulation of innate immunity genes expression (Chen et al., 2014). Epigenetic mechanisms seem to play a role in the trained immunity, as they establish a transcriptional profile that modifies the signaling and metabolism of innate immune cells, which onsets a long-lasting adaptation (Netea et al., 2011). DNA methylation, the most studied form of epigenetic modifications, is not only associated with gene repression, but also with gene activation (Bahar Halpern et al., 2014), splicing regulation (Shukla et al., 2011), nucleosomes positioning (Chodavarapu et al., 2010), and the recruitment of transcription factors (Fujimoto et al., 2005).

DNA methylation is involved in the response to pathogen challenge and ultimately in the occurrence and development of bovine mastitis (Vanselow et al., 2006; Chang et al., 2015; Zhang et al., 2018). Differences in the DNA methylation of peripheral blood lymphocytes have been observed between mastitis affected and healthy cows (Song et al., 2016; Ju et al., 2020).

Given the fact that epimutations can be transmitted to future generations, there is interest in studying the associations between the epigenome and the development of desired phenotypes (Ibeagha-Awemu and Zhao, 2015) as they could help understanding the outcome of complex dynamics between genotype and phenotype. Therefore, identifying epigenetic signatures associated with mastitis resistance can be useful in breeding programs aimed at improving animal welfare in dairy cattle.

To our knowledge, changes in genome-wide DNA methylation in bovine subclinical mastitis naturally induced by *Strep. agalactiae* and *Prototheca* have not yet been investigated. Hence, the present study aimed to provide the landscape of DNA methylome in bovine milk somatic cells of Holstein cows with natural subclinical mastitis induced by *Strep. agalactiae* and *Prototheca* spp. Understanding the possible regulatory roles of DNA methylation upon mastitis can lay the groundwork for mechanistic studies on susceptibility to mastitis in dairy cattle.

**METHODS**

**Animals’ Data**

This research was approved by the Ethical Animal Care and Use Committee (OPBA - Organismo Preposto al Benessere degli Animali) of the Università Cattolica del Sacro Cuore and by the Italian Ministry of Health (protocol number 510/2019-PR of 19/07/2019).

Thirty-eight Holstein cows ranging from 2 to 5 parity and between 92 and 448 DIM were selected from one commercial farm of 450 lactating cows (Veneto region, Italy) regularly monitored by the Istituto Zooprofilattico delle Venezie (IZSVe) for the presence of *Strep. agalactiae* and *Prototheca* spp. between January 2020 and February 2021. The farm was selected because of previous collaborations with the dairy farm owners and their associated veterinary practices. To reduce to a minimum the sources of variation which could affect the methylome analysis, animals were selected based on the following criteria: (i) absence of any clinical sign of disease for at least one year; ii) no medical treatment for at least 6 mo before enrollment; iii) being multiparous (parity ≥2) and in mid-late lactation (≥92 DIM), to exclude cows with negative energy balance condition, which could affect the proper activation of the immune cells metabolism in response to pathogens’ invasion (Wathes et al., 2009; Ingvartsen and Moyes, 2013). Moreover, we required that animals assigned to the healthy group had no previous history of clinical mastitis. Cows with clinical signs of mastitis or other diseases (e.g., metritis, hepatic lipidosis, ketosis, abscesses, laminitis) as well as animals under medical treatment were excluded from the trial. Animals’ data were collected from the herd management software (Dairy Comp Sata, Alta Italia Srl, Milan, Italy). Considering these criteria, the identification of healthy individuals and cows with subclinical mastitis from either *Strep. agalactiae* or *Prototheca* spp. was based on an initial bacteriological screening (time 0, T0), which was conducted on 188 lactating cows. Then, a second bacteriological examination was conducted 2 weeks after T0 (T1) to confirm the animals’ infection status. Between the T0 and T1 animals underwent daily monitoring by both the farmers and the local veterinarian to ensure they remained free from any visible signs of mastitis, thereby maintaining their subclinical condition. Following the results of the bacteriological test at T0 and T1, 3 experimental groups were defined: (i) healthy individuals (n = 15) with a negative bacteriological examination in all glands at T0 and T1; (ii) animals...
naturally infected by *Strep. agalactiae* (*n* = 12) and (iii) animals naturally infected by *Prototheca* spp. (*n* = 11), having a positive bacteriological result for either *Strep. agalactiae* or *Prototheca* spp., respectively. Animals with co-infections with either environmental or other microorganisms responsible of mastitis were excluded from the experiment. Cows were fed total mixed ration (TMR) formulated to meet or exceed the requirements of mid-lactation dairy cattle, mainly based on corn silage, sorghum silage, and concentrate (Pegolo et al., 2023). Drinking water was available in automatic water bowls, and cows were milked twice a day, from 2 a.m. to 6 a.m. and from 2 p.m. to 6 p.m. Animal welfare was managed by the farmers and local veterinarians, who intervened when needed.

**Milk Sampling**

Before morning milking, ~200 mL of milk from all quarters (pool sample) was aseptically collected from each animal according to the National Mastitis Guidelines (NMC, 2017). Briefly, the teat was disinfected pre-milking and, after discarding the first stream of foremilk from each quarter, composite milk from the 4 glands was collected. For each milk sample, 3 aliquots (~50 mL) were collected and gently mixed into sterile tubes for the following analyses: (i) bacteriological test; (ii) milk composition, somatic cell count, and differential somatic cell count (DSCC); and (iii) DNA extraction and DNA methylation analysis. All the samples were immediately refrigerated at 4°C and transferred to the different laboratories.

**Microbiological Analysis**

Microbiological examination of milk samples was conducted at the IZSVe laboratories (Legnaro, PD, Italy). Samples were frozen and analyzed within 3 d. Details of microbiological analysis are reported in Pegolo et al. (Pegolo et al., 2022). Briefly, 10 μL of every composite sample were inoculated in the following selective media: i) tallium kristalviolette tossin agar (TKT; IZSVe internal production), and ii) *Prothoteca* isolation medium (PIM; IZSVe internal production). After 24h of incubation, suspected colonies of *Strep. agalactiae* were confirmed using the Christie–Atkins–Munch-Peterson test (NMC, 2017). After 24, 48, and 72 h, PIM plates were examined for the growth of *Prothoteca*. Suspected colonies were confirmed by mass-spectrometry analysis (NMC, 2017). Furthermore, a screening of the most common microorganisms responsible for mastitis, such as *S. aureus* and *Strep. uberis*, as well as some environmental ones, such as *Streptococcus* spp., *Staphylococcus* spp., *Klebsiella* spp. and *Enterococcus* spp. was conducted to avoid possible bias in the trial.

**Milk Composition and Quality Traits**

Milk composition (protein, casein, lactose, fat, and urea content) and udder health traits (lactose, milk conductivity and pH) were assessed on fresh samples using an FT6000 Milkoscan infrared analyzer (Foss A/S, Hillerød, Denmark). Somatic cell count and DSCC were measured through the Fossomatic 7 DC analyzer (Foss A/S).

**DNA Extraction from Milk Somatic Cells**

A 50 mL aliquot from each sample was centrifuged at 2,000 rpm for 10 min at 4°C. The fat layer and the supernatant were discarded, and the cell pellet was washed with 50 mL PBS containing EDTA (EDTA) at 0.05 mM, pH 7.2. Samples were then centrifuged again at 1,500 g for 10 min at 4°C, the supernatant discarded, and the pellet stored at −80°C until the DNA extraction. Genomic DNA was isolated from somatic cells pellet using the commercial kit NucleoSpin® Tissue (Macherey-Nagel, Düren, Germany), following manufacturer instruction and an overnight proteinase K digestion. DNA quality was evaluated by agarose gel electrophoresis and DNA concentration was estimated by PicoGreen® (Thermo Fisher, Waltham, MA USA).

**Library Preparation and Sequencing**

About 200 ng of genomic DNA from each sample were sonicated to produce DNA fragments between 300 and 350 bp using Covaris® S220 instrument (Covaris, Woburn, MA). DNA was enriched for Methyl-Binding Domain (MBD) using the MethylMiner Methylated DNA Enrichment Kit (Invitrogen, Carlsbad, CA, USA), following manufacturer instructions. Library construction was performed using the TruSeq® Nano Library Preparation Kit (Illumina, San Diego, CA, USA). Libraries were quantified and quality checked with 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) using the High Sensitivity DNA Kit (Agilent Technologies). Libraries were sequenced on an Illumina HiSeq X (Macrogen, Seoul, Republic of Korea) to generate 150-base paired-end (PE) reads for high confidence mapping of captured fragments. MDB-Seq data are available at the Sequence Reads Archive (SRA), BioProject accession number, PRJNA976818.
Bioinformatic Analysis

MBD-Seq reads were analyzed as described in detail by Decock et al. (2016). Briefly, raw reads were quality checked using FASTQQC tool v.0.11.9 bowtie (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and were aligned to the Bos taurus genome ARS-UCD1.2 (Ensembl release 108) using Bowtie 2 v.2.2.9 (Langmead and Salzberg, 2012). PCR duplicates were marked with Picard v. 2.6.0 (http://broadinstitute.github.io/picard/) and the BAM files were sorted and indexed using SAMtools v.1.9 sort and index commands. The detection of enriched regions (peak calling) was performed using MACS2 v.2.1.1 (Zhang et al., 2008) using the callpeak function. BED files with location and score (linked to the P-value) of the identified peaks were generated with bedtools v.2.17.0 (https://bedtools.readthedocs.io/en/latest/). Also, WIG files for visualization of the mapping results were obtained. Due to the inconsistencies of the number of peaks called (low average number of peaks compared with other samples), 2 samples from the group of Strep. agalactiae-infected were discarded from subsequent analysis. The DiffBind package (Ross-Innes et al., 2012) in the R environment (https://www.R-project.org/) was used to create the consensus peak set matrix, which includes merged peaks that overlap in at least 2 samples and using the option summits = 300, which results in 601 bp peaks (the summit point plus 300 bp in each direction) representing the consensus sites. Then, DeSeq2 package (Love et al., 2014) was called via DiffBind package to perform the differential methylation analysis. Reads were normalized using a library size normalization based on Reads in Peaks. Four comparisons were performed: i) Strep. agalactiae-infected vs. healthy samples; ii) Proteotheca-infected vs. healthy samples, iii) mastitis vs. healthy samples and finally iv) Strep. agalactiae-infected vs. Proteotheca-infected samples. Genomic regions that were significantly differentially methylated (P < 0.05; FDR corrected) were considered as differentially methylated regions (DMR). To quantify the changes between 2 conditions log2 fold change (FC) was used. Differentially methylated regions were annotated using HOMER (http://homer.ucsd.edu/homer/) annotatePeaks function to extract differentially methylated genes (DMG). The promoter- transcription start site (TSS) was defined when the distance from the gene is comprised from −1kb to +100bp and the transcription termination site (TTS) from −100 bp to +1kb. If a region mapped to more than one gene, a priority was assigned based on the shorter distance from the promoter region. All downstream analyses did not take into account the sexual and mitochondrial chromosomes.

Exploratory and Functional Analyses

Exploratory analysis of the matrix built on the 24,916 methylated regions was conducted through a principal component analysis (PCA) after the regularized-logarithm transformation of the counts data using the DESeq2 package. Euclidean distances between samples were calculated and visualized in a heatmap. Moreover, for each experimental comparison, the R package GenomicFeatures (v.1.46.4) was applied to display the chromosomal locations of the DMR. The exonsBy() function was used to extract the coordinates of the associated genes, which were then plotted using the plotGrandlinear() function in the ggbiom package (v.1.42.0).

For functional analyses, the DMG extracted from DMR at promoters were fed to the Cytoscape (v.3.9.1, http://cytoscape.org) ClueGO plug-in (v.2.5.9; Bindea et al., 2009). Default parameters for the identification of relevant (P < 0.05, Benjamini-Hochberg corrected) biological processes (BP), molecular function (MF), cellular component (CC) and immune systems (IS) related pathways were used. The “GO term fusion” option was applied for reducing pathway redundancy and preserving the more representative terms. A second complementary approach was adopted using the whole set of genes associated with promoter - methylated regions of healthy animals and infected animals using the Pathifier algorithm from the pathifier R package (v.1.32.0, Drier et al., 2013). With this algorithm we transformed the methylated gene-level information into pathway-level information, inferring how much a specific pathway deviated from the control samples, by computation of the pathway deregulation scores (PDS). The PDS were calculated by constructing a n-dimensional space (n = number of genes in the path), where a main curve that captures the variation of a cloud of points is calculated by nonlinear regression. Each point represents each sample and its values of expression of the n genes of the pathway, and expressed as the distance of the projection to the main curve of each sample with respect to the projection of normal samples. Kyoto Encyclopedia of Genes and Genomes (KEGG) annotation was used. A heatmap was then built using the samples Euclidian distances and the Ward.D2 clustering method.

RESULTS

Milk Composition and Quality Traits

The average milk yield was 27.13 (±8.70) kg/day. Milk had 2.12 ± 0.74% of fat, 3.50 ± 0.27% of protein, and 2.74 ± 0.24% of casein, 22.03 ± 4.92 mg/100 mL
of urea, and 4.54 ± 0.42% of lactose. Milk pH and conductivity were 6.46 ± 0.08 and 9.84 ± 1.22 mS/cm, respectively (Table 1). *Streptococcus agalactiae* and *Prototheca* spp. infections decreased milk production (*P* < 0.05, Figure 1A). In addition, milk samples from *Strep. agalactiae* infected animals had a lower lactose percentage with respect to the healthy animals (*P* < 0.05, Figure 1B). Both pathogens increased SCS (*P* < 0.001, Figure 1C) and *Prototheca* also DSCC (*P* < 0.05, Figure 1D).

**Milk Somatic Cells DNA Methylation Landscape**

The MBD-Seq experiment on milk SC produced an average of over 46 million (M) reads per sample (range 30M-68M). Quality control and trimming procedures retained the majority of reads obtained (97%), with a mean percentage of uniquely mapped reads to the *Bos taurus* genome of 70% (range 53 - 90%) (Supplementary Table S1).

**DNA Methylated Regions Distribution.** After filtering steps and assigning genome annotation with HOMER, the consensus peak set gave a total of 221,910 methylated regions of which 123,835 (55.8%) intergenic, 82,969 (37.4%) intronic, 8,827 (4%) exonic, 3,413 (1.5%) in the promoter-TSS, and finally 2,866 (1.3%) in the TTS (Figure 2). In detail, the methylation autosomal maps show the genome-wide DNA methylation distribution in the studied samples (Supplementary Figure S1 and Supplementary Table S2). The methylation sites were unevenly distributed across bovine autosomes (BTA); BTA with the highest level of methylation sites were BTA3 (5.7%), BTA5 (5.7%), and BTA7 (5.6%). Additionally, methylation sites were generally more concentrated at the telomeres.

**Exploratory Analyses of DNA Whole Methyloyme Data.** The generalized PCA (Figure 3A) built with the 24,916 genes associated with methylated regions shows that the degree of methylation varied according to the status of infection and allowed a good separation between healthy and infected samples (except for 2 healthy animals that clustered with the infected ones; PC1: 60%) while no clear separation was observed between the *Strep. agalactiae* and *Prototheca* spp. infected samples. A similar finding was evidenced using a supervised clustering using the 500 highly variable methylated regions (Figure 3B), where all samples clustered perfectly between healthy and mastitic animals but failed to differentiate the type of pathogen.

**DNA Methylation Changes upon Intramammary Infection**

**Differentially Methylated Regions.** The DMR analysis focused on the 3,413 promoter-TSS regions identified in the consensus peak set. In the comparison *Strep. agalactiae*-infected vs. healthy cows, we found 1,442 significant DMR of which 1,081 were hypermethylated and 361 were hypomethylated in *Strep. agalactiae*-infected cows (Supplementary FigureS2). When annotated to genes, the most hypermethylated DMR were co-located with *ARL6* (log2FC = 1.7), *IL17REL* (log2FC = 1.3), and *AP1G1* (log2FC = 1.2), whereas the most hypomethylated DMR were co-located with *GLT8D2* (log2FC = −5.1), *ANGPTL4* (log2FC = −5.1), and *IL10RB* (log2FC = −3.2) (Supplementary Table S2).

Concerning the *Prototheca*-infected vs. healthy cows comparison, samples showed a higher number of DMR compared with *Strep. agalactiae*-infected samples, with 1,872 significant DMR of which 1,514 were hypermethylated and 358 were hypomethylated (Supplementary FigureS2). The most hypermethylated DMR were co-located with *ARL6* (log2FC = 1.8) and *HACD2* (log2FC

---

**Table 1.** Milk yield and composition of Holstein cows (n = 38)

<table>
<thead>
<tr>
<th>Metric</th>
<th>Mean</th>
<th>SD</th>
<th>P1</th>
<th>P99</th>
</tr>
</thead>
<tbody>
<tr>
<td>Milk yield, kg/d</td>
<td>27.13</td>
<td>8.70</td>
<td>10.44</td>
<td>43.90</td>
</tr>
<tr>
<td>Milk composition</td>
<td>2.12</td>
<td>0.74</td>
<td>1.02</td>
<td>3.42</td>
</tr>
<tr>
<td>Fat, %</td>
<td>3.50</td>
<td>0.27</td>
<td>2.99</td>
<td>4.16</td>
</tr>
<tr>
<td>Protein, %</td>
<td>2.74</td>
<td>0.24</td>
<td>2.31</td>
<td>3.32</td>
</tr>
<tr>
<td>Casein, %</td>
<td>22.03</td>
<td>4.92</td>
<td>12.64</td>
<td>35.39</td>
</tr>
<tr>
<td>Urea</td>
<td>4.54</td>
<td>0.42</td>
<td>3.51</td>
<td>5.20</td>
</tr>
<tr>
<td>Udder health traits</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SCC, 10^4 cells/mL</td>
<td>1,514</td>
<td>2,297</td>
<td>19.88</td>
<td>9,280</td>
</tr>
<tr>
<td>SCS</td>
<td>5.06</td>
<td>2.61</td>
<td>0.43</td>
<td>9.52</td>
</tr>
<tr>
<td>DSCC, %</td>
<td>63.59</td>
<td>13.27</td>
<td>33.73</td>
<td>83.36</td>
</tr>
<tr>
<td>pH</td>
<td>6.46</td>
<td>0.08</td>
<td>6.29</td>
<td>6.62</td>
</tr>
<tr>
<td>Milk conductivity, mS/cm</td>
<td>9.84</td>
<td>1.22</td>
<td>8.09</td>
<td>12.62</td>
</tr>
</tbody>
</table>

SCC: somatic cell count; SCS: somatic cell score calculated as log2 (SCC/100,000) +3 (Ali et al., 1980); DSCC: differential cell count.
Giannuzzi et al.: METHYLOME OF SOMATIC CELLS UPON MASTITIS

= 1.5), and PITHD1 (log2FC = 1.4), whereas the most hypomethylated DMR were co-located with ADAM11 (log2FC = −5.6), APAF1 (log2FC = −5.1), KLRG2

Figure 1. Results of Kruskal-Wallis test for milk yield and composition of Holstein cows (n = 38). Only significant traits are displayed. Healthy: clinically healthy cows, negative at the bacteriological examination; Sa+: animals infected by Strep. agalactiae; P+: cows animals infected by Prototheca spp. SCS: somatic cell score; DSCC: differential somatic cell count.
(log2FC = −5.1), and MAPKAPK2 (log2FC = −4.9) (Supplementary Table S2).

When comparing subclinically infected animals, regardless of the specific pathogen, with healthy cows, we identified 2,036 significant DMR, of which 1,576 were hypermethylated and 460 were hypomethylated (Supplementary Figure S2). Once annotated to genes, the most hypermethylated promoter regions were co-located with ARL6 (log2FC = 1.8), PLAU (log2FC = 1.7), COX5B (log2FC = 1.5), and IL17REL (log2FC = 1.3). Conversely, the most hypomethylated promoter regions were co-located with TEK (log2FC = −5.1), LEO1 (log2FC = −5.0), RALB (log2FC = −4.8) and ICAM4 (log2FC = −4.8) (Supplementary Table S2).

The last comparison for evidencing DMRs that might distinguish the 2 types of infection (i.e., Strep. agalactiae vs. Prototheca) resulted in only 4 significant DMR (hypomethylated in Strep. agalactiae-infected cows). Accordingly, we found 683 hypermethylated and 250 hypomethylated associated genes in common with the comparisons of Prototheca vs healthy, Strep. agalactiae vs healthy and mastitis vs healthy, which did not suggest a highly divergent cow response to the 2 pathogens in the investigated population (Supplementary Figures S3-S4). Pathogen-specific hypermethylated genes were 2 (Strep. agalactiae) and 34 (Prototheca) and hypomethylated genes were 10 (Strep. agalactiae) and 4 (Prototheca).

**Functional Analyses of Differentially Methylated Genes.** Using ClueGO, we found that DMG associated with promoter - DMR acted coordinately within specific biological functions (Figure 4). Strep. agalactiae infection was associated with the enrichment of several pathways related to meiosis and reproduction (e.g., meiotic cell cycle, FDR = 0.035; development-
Functional analysis of the whole methylome. Using Pathifier on the set of genes associated with the whole promoter methylome (n = 18,514), we identified 206 KEGG pathways with PDS associated with the 2 types of infection compared with healthy individuals (Supplementary Table S3, Figure 5). The most deregulated pathway upon Strep. agalactiae infection was one carbon pool by folate (2.5-fold), which was one of the top deregulated pathways triggered by Prototheca infection (2.9-fold) as well. Overall, several pathways related to energy metabolism (e.g., tricarboxylic acid cycle, glycolysis/gluconeogenesis, biosynthesis of unsaturated fatty acids) were associated with both pathogens. Concerning the immune response, in both cases we observed strong deregulation in several pathways related to cytokine production and specifically interleukin 1 (IL-1) (FDR = 0.035) (Supplementary Table S4) (Figure 4A). Finally, we observed the enrichment of non-specific terms related to cytokine secretion (Calcium ion transport, FDR = 0.038; ribonucleotide binding, FDR = 0.035) (Supplementary Table S4) (Figure 4B). When considering subclinical mastitis, regardless of the specific pathogen investigated, growth hormone (GH) secretion (FDR = 0.049), and positive regulation of interleukin 6 (IL6) production (FDR = 0.033) were enriched (Supplementary Table S4) (Figure 4C).

DNA Methylation Data

We obtained a high depth (46M reads per sample) and high-quality sequencing output with satisfying average alignment rate to the Bos taurus genome, which provided good data for subsequent analysis (Lau et al., 2011; Kim et al., 2018). Consistent with previous studies, methylated regions were mostly concentrated in intergenic regions, followed by intragenic regions (Kim et al., 2018; de Souza et al., 2022). We focused our downstream analysis on the DNA methylation in promoter-TSS regions as, to date, they have been the most closely studied due to their association with transcriptional alterations of the involved genes, which include mostly transcriptional silencing, but also gene regulation, splicing activation, and transcription factors recruitment (Tirado-Magallanes et al., 2017).

DISCUSSION

Milk SC have been widely adopted as useful indicator for mastitis screening as, being the first line of defense against mammary infections, they are suitable for capturing the host immune response and possible pathogen-specific immune-molecular signatures. Furthermore, compared with mammary gland tissue or blood, they are easier to collect and more considerate of animal welfare (Wang et al., 2022). Herein, we explored the DNA methylation landscape of milk SC during naturally occurring subclinical mastitis from Strep. agalactiae, a Gram-positive bacterium, and Prototheca spp, a unicellular alga, to identify pathogen-generic and possibly pathogen-specific signatures associated with the 2 different microorganisms, which are phylogenetically very distant from each other. It is worth noting that milk SC are composed of different types of immune cells, whose relative proportion can change according not only to the infection status but also to the type of pathogen. Indeed, flow cytometry analyses conducted on the same cohort of animals enrolled in this study evidenced changes in the distribution of the milk immune cell population in response to Strep. agalactiae and Prototheca spp. infection (Pegolo et al. 2022). In particular it was shown that while Strep. agalactiae seemed to trigger a more innate immune response by the recruitment of macrophages and PMN, Prototheca seemed to drive a more adaptive response mediated by the activation of T-cells. As an outcome, the DNA methylation profile analyzed in this study is expected to originate from the immune cell population itself and, therefore, reflect their varying proportions.
Figure 4. Results of Cluego functional analyses for A) *Strep. agalactiae* vs healthy; B) *Prototheca* spp. vs healthy; C) mastitis vs healthy. The number at the end of the bar corresponds to the number of differentially methylated genes for each of the enriched biological functions.
known about the functional consequence and regulatory implications of DNA methylation in intergenic and intronic regions, although an indirect role in regulating gene expression is hypothesized to affect alternative splicing or noncoding RNA (Deaton and Bird, 2011; Anastasiadi et al., 2018). Moreover, a study conducted by Brenet et al. (2011) highlighted that methylation of the first exon might be critical for transcriptional silencing, therefore suggesting that we should approach the DNA methylation mechanism taking into greater consideration the different regions of the gene cassette.

**Milk Somatic Cells Methylome Signatures upon Subclinical Mastitis**

*Differentially methylated genes and enrichment analyses.* When it comes to the DMR analysis, considering all the different comparisons we observed that promoter aberrant hypermethylation is more prominent than hypomethylation ($P < 0.001$) even if overall, hypomethylated DMR had higher FC. When considering immune system impairment, deregulated changes in gene expression are more attributable to aberrant hypomethylation than hypermethylation (Kushwaha et al., 2016, Upchurch et al., 2016). One of the most hypomethylated DMR found in this study in the *Strep. agalactiae*-infected group was co-located with *IL10RB* gene locus, a cytokine receptor gene which is involved in the IL-10 signal transduction (Yogev et al., 2022). The IL-10 is a key cytokine that is considered one of the master regulators of immune response, having prevalently anti-inflammatory activity and being secreted by monocytes, T-cells and macrophages (Carey et al., 2012). Notably, IL-10 signaling dysregulation has already been associated with subclinical and clinical bacterial mastitis in dairy cows (Wenz et al., 2010, Faaz and Abdullah, 2022). Another interesting finding is the hypomethylated DMR associated with

**Figure 5.** Heatmap built using the pathway deregulation scores (PDS) of the associated genes of the entire methylome at the promoter level of healthy animals and animals affected by subclinical mastitis, regardless the pathogen. Each row corresponds to a pathway, and each column to a sample. Blue corresponds to “no deregulation,” and yellow to high deregulation.
This gene encodes for a member of the glycosyltransferase family and is related to energy metabolism. The dysregulation of GLT8D2 was already reported in clinical mastitis of cows during early lactation (Cheng et al., 2021), highlighting the potential relation between mastitis and metabolic deficit, which may affect the immune response. Notably, we observed a strong hypomethylation in the first intron of CIITA, which is not only one of the master regulators in the expression of MHC complex but also involved in the transcription of several immune-related genes like IL-4 and IL-10 (Devaiah and Singer, 2013). In Prototheca-infected cows with respect to healthy animals, one of the most hypomethylated DMR was co-located with the ADAM11 gene locus, encoding a member of the disintegrin and metalloprotease (ADAM) protein family, which are transmembrane and secreted proteins involved in cell-cell and cell-matrix interactions. Other members of the ADAM family have been found to interact with microorganisms (e.g., Staphylococcus aureus) during the pathogenesis of mastitis in dairy cows, promoting necrosis of the mammary gland tissue and cell lysis (Campos et al., 2022). Further studies are needed to explore the pathogenetic mechanisms of Prototheca spp. and its potential crosstalk with transmembrane receptors of the host’s mammary gland epithelium. Hypomethylation was also reported in the first exon of IFNAG. Type I IFN not only mediate antiviral activities but also have an important influence on the adaptive immune response through the promotion of both cellular immunity and antibody responses (Bauermann et al., 2017). One of the most hypermethylated DMR in the mastitis vs healthy group was co-located with IL17REL, which is a member of the IL-17 receptor family. The IL-17 family members are produced by Th17 cells to protect the host against bacterial and fungal infections (Kolls and Lindén, 2004), and this was also observed in dairy cows with bacterial-induced mastitis (Cebron et al., 2020). Interestingly, in humans, IL17REL has been shown to often be suppressed in inflammatory conditions, leading to a preponderance of pro-inflammatory factors (Franke et al., 2010).

Overall, enrichment analyses on the DMG evidenced that the 2 pathogens were associated with the regulation of the same molecular pathways. In both Strep. agalactiae and Prototheca infection functional analyses revealed that hypermethylation was associated with the upregulation of pathways related to reproduction, which might be attributable to the fact that mastitis can affect the reproduction of cows by destroying follicles, affecting oocyte growth or function and reducing ovulation ability (Boujenane et al., 2015, Eckel and Ametaj, 2020). Inflammatory cytokines and bacterial endotoxins induced by mastitis can also lead to delayed estrus, hormonal imbalances and other related problems, resulting in a decrease in fertility (Wang et al., 2021). Several genes were associated with these functions, among which FSHR, CYP27B1 and IGF1. Follicle stimulating hormone receptor is connected with ovarian responsiveness to FSH in ovulation induction in humans (Perez Mayorga et al., 2000). The expression of FSHR has an important role in the cumulus cells’ expansion and the final maturation of cumulus-oocyte complexes (Kafi et al., 2021). Pathogen-associated molecular patterns activate toll-like receptors that stimulate immune cells such as monocytes to induce expression of CYP27B1 to convert 25-hydroxyvitamin D₃ into the more active form of vitamin D, 1,25-dihydroxyvitamin D₃ (Nelson et al., 2018) which is a direct regulator of antimicrobial innate immune responses (Liu et al., 2009). The liver modulates the reproductive function in females through secretion of IGF1. Ovarian follicles need blood derived IGFI to complete growth and maturation before ovulation (Ginther et al., 2002). For instance, elevated blood IGFI in Brahman heifers (Samadi et al., 2013) and postpartum Droughtmaster cows (Tena-Sempere, 2006) was associated with earlier puberty and shorter postpartum anestrus, respectively. Hypermethylation was also associated with the upregulation of pathways related to actin-based cytoskeleton and tissue remodeling, which were also in common between Strep. agalactiae and Prototheca spp. Pathogens develop different strategies to facilitate invasion of the host cell through manipulating the activity of Rho-GTPases (Popoff, 2014). For instance, it has been shown that primary bovine mammary epithelial cells cultured with Staphylococcus aureus have an altered, more filamentous, actin cytoskeleton, which may facilitate bacterial invasion (Günther et al., 2017). In particular, ARHGAP26 was recently identified as a candidate gene for clinical mastitis in dairy cattle (Kour et al., 2021) (Cheng et al., 2022). This is the first piece of evidence for a modulation of these pathways associated with IAI induced by an alga, i.e., Prototheca spp. Epidermal growth factor has been reported to have an important role in the normal development and homeostasis of the mammary gland (Xian, 2007). Among the hypermethylated genes, we found TGFBI, which functions as a potent proliferation inhibitor and apoptosis inducer (Ramesh et al., 2009). It is also a potent chemotactic factor, recruiting monocytes, neutrophils and lymphocytes (Berie and Moses, 2010). Recent results showed that EGF and TGFBI are key regulatory elements in resistance to mastitis in dairy cattle (Sharifi et al., 2020). One other study reported the overexpression of the polymeric immunoglobulin receptor (PIGR), which was also present among hypermethylated genes found in the current study, in a mammary transgenic model.
that may find application in mastitis resistance (De Groot et al., 2000).

Further to the above mentioned pathways, in the comparison between animals affected by subclinical mastitis (regardless of the pathogen) and healthy animals we found that hypermethylation was associated with upregulation of GH secretion which is explained by its ability to modulate neutrophil and lymphocyte function (Cirillo et al., 2017). Hypermethylation was also associated with the positive regulation of IL-6 production. Interleukin-6 is a multifunctional cytokine produced and secreted by activated macrophages, lymphocytes, and epithelial cells. In case of intramammary infection, inflammatory and immune cells can express the IL-6 gene and produce IL-6 (Shuster et al., 1993). Among the associated DMG, we found several IL-17 family cytokines such as IL-17A, IL-17D, IL-17RC. Interleukin-17A mediates the crosstalk between the immune system and different epithelial tissues, which is involved in neutrophils infiltration, activating various defense mechanisms against bacterial and fungal infections (Das and Khader, 2017). IL-17R is a heteromeric receptor comprising of IL-17RA and IL-17RC and mediates signaling of IL-17A and IL-17F, although other members of the IL-17 family such as IL-17B, IL-17C, and IL-17D can also induce the production of pro-inflammatory cytokines and chemokines (Jin and Dong, 2019). Methylation of IL-17A, as well as other interleukin genes was also observed by Wang et al., 2022, who evaluated the changes in the DNA methylome in response to Streptococcus uberis subclinical mastitis.

Functional analyses on the whole somatic cells methylome. Having analyzed the whole methylome of promoter regions, we found a clear separation between healthy and mastitic cows, highlighting that infection and the inflammatory-derived status were shown to have profound effects on the DNA methylation signature (Bayarsaihan, 2011). On the other hand, we did not observe a clear separation between the 2 groups of infected animals, consistently with the results obtained from the DMR analysis. DNA methylation in a locus can work as an inhibitor or activator for gene expression, acting on the regulatory stamping of the genome (Dhar et al., 2021). In this context, DNA methylation changes induced by environmental stimuli, including microbial infections, are targeted at the host’s innate immune response rather than being pathogen-specific, which are commonly led on a transcriptional or post-transcriptional level. However, it is important to emphasize that we found a stronger pathways deregulation in the Prototheca infected animals. This could be related to potential differences in the stage of infection, which we were not able to control as we dealt with naturally occurring subclinical mastitis (and that could also explain the differences in PDS within samples belonging to the same group) and/or possible Prototheca-specific DNA methylation marks.

In contrast to the DMR enrichment analyses, we found a large number of pathways associated with the immune response, which were deregulated in mastitic animals. Overall, our results evidenced that there is a blurred line between the DNA methylation regulation of innate and adaptive host response. Deregulated pathways included genes acting in the NF-kB signaling pathway, RIG-I-like signaling pathway, TLR signaling pathway, and T cell receptor signaling pathway, with a specific involvement of Th17 and IL-17. Activation of Th17 differentiation and antigen processing and presentation pathways was also observed by Nayan et al. (2022), who evaluated the whole genome DNA methylation profile from peripheral blood lymphocytes in water buffaloes with subclinical mastitis. The inflammatory response requires the development of a complex and dynamic regulatory network that involves specific signaling pathways for antimicrobial defense, immune response, and tissue repair and remodeling (Medzhitov, 2008). Among those pathways, NF-kB and IRF families are the major players acting as the most relevant effectors of innate immunity (Iwaszsko and Kimmel, 2015). Innate immunity is the host’s first line of defense against invading pathogens. When microbial infection occurs, pattern recognition receptors, including TLR and RIG-I-like receptors, detect pathogen-related molecular patterns and induce a variety of downstream cascades leading to the production of pro-inflammatory cytokines and chemokines by various immune cells, such as macrophages (Akira et al., 2006, Jin et al., 2014). Moreover, TLR signaling operates an indirect regulation of T-cell differentiation and proliferation by activating innate immune cells (Iwasaki and Medzhitov, 2004), but also a direct stimulation of Th17 with the production of IL-17 which act as chemoattractant on monocytes and PMN (Bird, 2010). Consistent with our findings, the altered promoter methylation of genes pertaining to TLR, RIG-I-like and NF-kB signaling pathways have been shown to be associated with the crosstalk between innate and adaptive host response to microbial agents (Bayarsaihan, 2011), underlining the inextricable connection between methylation and inflammation response to pathogens.

CONCLUSIONS

Our study has contributed to bridge the gap between genotype and phenotype by providing a view of the milk SC methylome landscape involved in the occurrence and development of subclinical mastitis in dairy cattle. Although no evident pathogen-specific signature was
detected, we evidenced that subclinical mastitis was associated with the methylation profile of genes involved in the regulation of meiosis, reproduction, and tissue remodeling. Furthermore, DNA methylation changes encompassed genes implicated in several immune-related pathways involving the interface between innate and adaptive host response. DNA methylation information could complement genomic information and provide a better understanding of the factors that shape livestock phenotypes and have a directional application in breed improvement and management practices. Further studies on a bigger sample size are needed to validate our results and to possibly identify a unique somatic cells methylene that identifies pathogenesis-specific alterations.

ACKNOWLEDGMENTS

This research was part of the LATSAN project, funded by the Ministry of Agriculture, Food Sovereignty and Forests (MASAF; Rome, Italy). Moreover, the study was carried out within the Agritech National Research Center and received funding from the European Union Next-Generation EU (PIANO NAZIONALE DI RIPRESA E RESILIENZA (PNRR) – MISSIONE 4 COMPONENTE 2, INVESTIMENTO 1.4 – D.D. 1032 17/06/2022, CN00000022). This manuscript reflects only the authors’ views and opinions, neither the European Union nor the European Commission can be considered responsible for them.

REFERENCES


Bahar Halpern, K., T. Vana, and M. D. Walker. 2014. Paradoxical role on a bigger sample size are needed to validate our phenotypes and have a directional application in breed improvement and management practices. Further studies on a bigger sample size are needed to validate our results and to possibly identify a unique somatic cells methylene that identifies pathogenesis-specific alterations.


