Genomics 117 (2025) 111102

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Response of primary mammary epithelial cells to pathogen challenge in dairy cows with divergent genomic breeding values for udder health☆

Terhi Iso-Touru [a,*], Daniel Fischer [a], Frank Panitz [a], Suvi Taponen [b], Zexi Cai [c], Goutam Sahana [c], Ilma Tapio [a], Johanna Vilkki [a]

[a] Natural Resources Institute Finland (Luke), Jokioinen, Finland
[b] Faculty of Veterinary Medicine, University of Helsinki, Helsinki, Finland
[c] Center for Quantitative Genetics and Genomics, Aarhus University, Aarhus, Denmark

ARTICLE INFO

Keywords:
Mastitis
Dairy cattle
Udder health
Transcriptomics
Gene networks
eQTL

ABSTRACT

This study focuses on the initial cellular response to a causal pathogen in the mammary gland of dairy cows aiming to uncover genomic features associated with mastitis resistance. We analyzed transcriptomes from *Escherichia coli* -challenged primary bovine mammary epithelial cells derived from milk of cows with high or low genomic index value for udder health. The most striking difference was the markedly higher number of differentially expressed genes post-challenge in the low mastitis resistance group. Gene enrichment analysis indicates a slightly delayed adaptive immune response and early modification of primary metabolic processes in the low mastitis resistance group. Notably, there was no overlap between the DEGs or their potential cis-eQTL with the location of candidate SNPs from association studies of the same cow population. This implies that the genes and pathways involved in the early immune response in the mammary epithelium may not be central for phenotypic mastitis resistance.

1. Introduction

Mastitis in dairy cows is a complex disease impacting animal health, milk production, farm economics, and public health, making its prevention and management a critical concern in dairy farming. To meet the challenges of the global food system while simultaneously reducing the environmental impact of food production, improvements in the health and welfare of farmed animals is essential [1]. Preventive measures such as improved farm hygiene, optimized milking practices, and enhanced cow welfare serve as the foundation for reducing mastitis incidences [2]. Also, vaccinations are seen as strategy to prevent bacterial mastitis, however vaccines are mostly in their developmental stage [3,4]. Other strategies, such as stem cell therapy [5], to treat mastitis, are still in the early research stage. Thus, utilizing genomic research to deepen the understanding of mastitis and applying genomic selection to breed animals with higher mastitis resistance offers a proactive and sustainable approach to disease prevention. Prediction accuracies in genomic selection remain limited, but could be enhanced by identifying causative variants, thereby enabling a more reliable genotype to phenotype link [1,6].

Traditionally, association analyses using microsatellites, single nucleotide polymorphism (SNP) arrays or more recently, whole genome sequencing (WGS), have been used to identify genomic regions containing genetic factors that influence mastitis resistance [7]. For genome wide association studies (GWAS), phenotypes most used are breeding values of clinical mastitis (CM) and somatic cell count (SCC) from which the somatic cell score (SCS) is derived (log-transformed SCC). The AnimalQTLdb, a publicly available database collecting data for QTL (accessed 16th of December 2024 [8]), includes 571 and 1870 quantitative trait loci (QTL) for CM and SCS, respectively, but only a few QTL are consistently reported across studies. Even if the reported QTL regions overlapped, the most significant SNPs vary or are scattered in a wide region, due to long-range linkage disequilibrium in cattle. One approach to potentially increase the power and precision of the identification of genetic variants affecting mastitis resistance is to use meta-analyses that combine GWAS summary statistics from multiple studies and dairy cow breeds [9]. Using this method and various post-GWAS analyses, such as predicted regulatory elements and expression

☆ This article is part of a Special issue entitled: '10 years of FAANG' published in Genomics.

* Corresponding author.
E-mail address: terhi.iso-touru@luke.fi (T. Iso-Touru).

https://doi.org/10.1016/j.ygeno.2025.111102
Received 28 May 2025; Received in revised form 15 August 2025; Accepted 25 August 2025
Available online 26 August 2025

quantitative trait loci (eQTL) from multiple tissues, Cai et al. 2024 [9] were able to pinpoint 31 candidate genes and 14 possible causal variants that affect mastitis resistance. However, further functional validation is necessary to confirm the predicted candidates.

Hence, in line with the goals of the Functional Annotation of ANimal Genomes project (FAANG) to advance understanding of the genotype-phenotype relationship in domesticated animals (https://data.faang.org/home), our aim was to employ an *in vitro* functional approach to identify genomic features underlying mastitis resistance. To achieve this, we chose dairy cows genetically divergent for udder health and focused on the host initial cellular response differences to causal pathogens in the mammary gland. The index for udder health describes the genetic ability of the cow to resist mastitis. The breeding goal is to reduce the frequency of clinical mastitis. Records on CM and SCC from 1st to 3rd lactation and fore udder attachment (UA) and udder depth (UD) from 1st lactation are used in the genetic evaluation of udder health [10]. We analyzed transcriptomes from *Escherichia coli* -challenged primary bovine mammary epithelial cells derived from milk. These cells, as demonstrated in previous research [11], serve as a reliable tool for assessing innate immune responses to pathogens. *E. coli* is one of the major mastitis pathogens causing acute clinical mastitis, often with severe symptoms, and therefore a major concern in dairy cattle.

The transcriptomic data was integrated with WGS of the donor cows and further analyzed using eQTL mapping in genome regions enriched with differentially expressed genes after pathogen challenge or with significant association to clinical mastitis in GWAS from the Nordic Red dairy cattle (NRC) population.

2. Materials and methods

2.1. Cows

For the study, six NRC cows were selected from the experimental research herd of Natural Resources Institute Finland (Minkiö, FI). The selection criterion was their genome based estimated breeding value indices (GEBV) for udder health (UH). Predisposition to good udder health is being suggested by an index value over 115 ($UH_{>115}$) and for poor udder health below or equal to 90 ($UH_{\leq 90}$) when the population mean is 100, 10 points indicating one standard deviation. Three of the selected cows had an index value >115 and three an index value ≤90 (Table 1).

The cows of the herd have been included in various experiments and their full health records from birth onwards are available. At the time of milk collection, the cows were in their second to fifth lactation, not included in any experiment, clinically healthy, and all udder quarters were non-infected. The three cows with $UH_{>115}$ had no clinical mastitis incidence through their lifetimes (ranging from years 2014 to 2023). The three cows with $UH_{\leq 90}$ had 2 to 9 records of treatments for clinical mastitis (from 2012 to 2022, Table 1).

2.2. Pathogen challenge and RNA sequencing

Primary bovine mammary epithelial cells (pbMECs) were collected and challenged with *E. coli* using the methods described in Iso-Touru et al. 2024 [11]. Shortly, milk from the six healthy NRC cows was collected by standard machine milking at the research barn. pbMECs were collected by centrifugation, cultured until the second passage, and stored in the vapor phase of liquid nitrogen before the pathogen challenge. Cells were challenged with a Finnish strain of *Escherichia coli* FT238 isolated from bovine clinical mastitis [12] and harvested 3 h and 24 h after exposure to *E. coli*. Bacteria free control cells were harvested at the same timepoints (Fig. 1). RNA was extracted from three cell samples per time point per animal (72 samples altogether), RNA sequencing libraries were prepared, and three biological/technical replicates per timepoint and treatment per animal were sequenced with 2 × 150 bp read length on an Illumina NovaSeq 6000 platform at the Finnish Functional Genomics Centre (Turku, Finland).

2.3. Whole genome sequencing

DNA for whole genome sequencing was extracted from blood samples collected from jugular vein (ethical permission ESAVI/16348/2019) using DNeasy Blood & Tissue Kit (Qiagen, Germany) following manufacturer's protocol. The WGS of cow samples were sequenced at the Finnish Functional Genomics Centre (Turku, Finland) using TruSeq® DNA PCR-Free Library kit (Illumina, San Diego, CA, USA) and PE sequencing (2 × 150 bp) on an Illumina NovaSeq 6000 (Illumina, San Diego, CA, USA) platform.

The cattle WGS variant calling was performed following the GATK4 best practices [13] implemented as Snakemake [14] workflow called Snakebite-WGS [15]. Implemented steps contain, among others, the GATK base recalibrator as well as a model to adjust the base quality scores and a base recalibration step. Variant calling is done *via* haplotype caller. The pipeline utilizes BWA-mem to align the data but includes a refinement step using Picard before the GATK4 software suite is used for the final variant calling with applied default filters.

2.4. RNA-seq alignments and quantification

Raw sequencing reads underwent adapter removal and low-quality base trimming using fastp [16] with default parameters. Read quality, both pre- and post-trimming, was assessed using FastQC and summarized with MultiQC [17]. The trimmed reads were then mapped to the *Bos taurus* reference genome ARS-UCD1.3 (Ensembl 111) using STAR [18], incorporating the corresponding Ensembl 111 annotation for guidance. Gene-level read quantification was performed with FeatureCounts [19]. The entire data processing workflow is implemented in Snakemake [14] and is available online called Snakebite-RNAsEq. [20].

2.5. Differential expression analysis

Post-processing analyses were conducted using R Statistical Software

Table 1
Genomic breeding value indexes and individual data from cows used as cell donors. DIM = days in milk, Avg. MY = Average milk yield in kg/day/lactation period pbMECs were collected.

Genomic breeding value index*									Cow data			
Yield	Milk-kg	Fat-kg	Fat-%	Prot-kg	Prot-%	UH	SCC	Fertility	n^{th} lactation	DIM	Avg. MY	n of treatments for clinical mastitis
103	100	101	98	103	104	117	111	104	2	208	31	0
97	100	95	93	99	105	119	118	109	5	191	29	0
103	100	101	98	103	104	117	111	104	3	177	26	0
102	101	100	95	103	99	85	93	86	2	175	30	6
101	115	98	79	108	82	90	98	100	4	222	36	9
113	108	106	91	116	106	90	103	96	5	220	38	2

* March 2019.

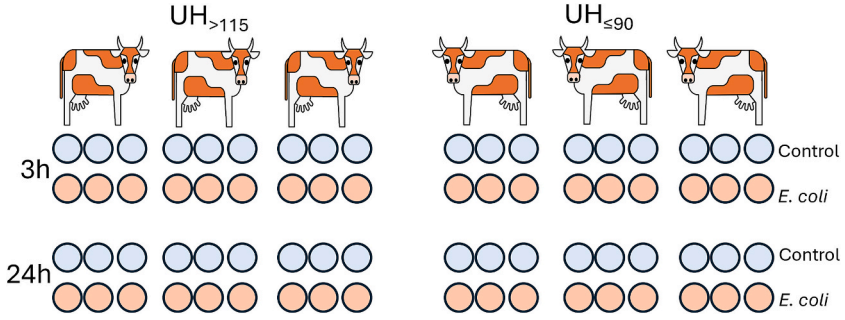

Fig. 1. The experimental setup for the transcriptomic study. Cows with predicted low ($UH_{\leq 90}$) or high ($UH_{>115}$) mastitis resistance were used as donors for milk derived pbMECs. Each sample, collected either at 3 h or 24 h post-challenge, is indicated as a sphere, blue indicating control and red indicating challenge with *E. coli*. Three biological/technical replicates per timepoint and treatment per animal were done leading to 72 separate samples for sequencing. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(version 4.3.1) and the tidyverse package [21]. Exploratory data analysis (data not shown) utilised the vsn package (version 3.70.0) [22] for variance stabilisation and pheatmap (version 1.0.12) [23] for heat maps of sample distances and clustering. Data visualisation was enhanced with ggplot2 (version 3.4.4) [24], ggrepel (version 0.9.5) [25], patchwork (version 1.2.0) [26] and RColorBrewer (version 1.1.3) [27]. The DESeq2 package (version 1.42.0) [28] was used to analyse differential gene expression. Raw read counts were pre-filtered to remove genes with little or no expression across samples. This filtering was run for each gene by counting how many samples have at least 10 reads; if fewer than 3 samples meet this threshold the gene was removed. For each time point, challenged samples were compared to controls. Count outliers were detected and handled by DESeq2 as described by [28]. Automatic independent filtering was applied, and multiple testing adjustments were performed using the Benjamini–Hochberg method with an FDR cut-off (alpha) of 0.05 to identify differentially expressed genes. Differentially expressed genes were visualized as volcano plots indicating the relationship between the magnitude of fold-change and statistical significance using EnhancedVolcano (version 1.20.0) [29]; the Ensembl gene IDs were converted to entrez gene ID symbols.

2.6. Annotation and enrichment analysis

Gene annotation was performed using biomaRt (version 2.58.2) [30] with the *btaurus_gene_ensembl* dataset (based on Ensembl release version 111) of the bovine genome version ARS-UCD1.3. Over-representation analysis of the expressed gene sets was conducted with clusterProfiler (version 4.10.0) [31]. KEGG pathway analysis [32] for differentially expressed genes (adjusted p-value <0.05) was carried out using the enrichKEGG function in clusterProfiler and visualized as bar plots. Plots displaying genomic regions with statistically enriched genes compared to the background gene density were generated using ShinyGO (version 0.81) [33], using the bovine genome version ARS-UCD1.2 in the ShinyGO platform. For each bin or region, a statistical method (hypergeometric test) is performed to test if the number of significantly DEGs in this bin is significantly higher than expected and p-values are computed and corrected for multiple testing. If enriched regions exist, more stringent FDR cutoffs will decrease the number of enriched regions while more relaxed cutoffs increase the number. The analysis was performed with a window size of 6 Mb, two steps per window, and a FDR cutoff of 0.01. Significant terms were visualized as pathway maps using the pathview package (version 1.42.0) [34]. Graphs were created based on all expressed genes, with gene symbols coloured according to fold change.

2.7. eQTL analysis

Despite the low sample size, we performed an eQTL analysis to find genomic candidate loci that regulate expression levels of genes, based on the Ensembl annotation version 111. For that, gene expression values for each variant were assigned to the corresponding genotype groups and tested for directional alternatives using a tailored non-parametric generalized Mann-Whitney test implemented in the R package gMWT [35]. The method is based on probabilistic indices, where the pairwise index $\hat{P}_{ij}$ denotes the estimated probability that an expression value from a genotype group j exceeds one from genotype group i. For three ordered groups G_1, G_2 and G_3 the triple probabilistic index is defined as

$$\hat{P}_{123} = \frac{1}{n\_1\, n\_2\, n\_3} \sum_{x \in G\_1} \sum_{y \in G\_2} \sum_{z \in G\_3} I(x < y < z),$$

where n_k is the number of samples in group G_k and $I(\cdot)$ is the indicator function. The test statistic evaluates then the orders $\hat{P}_{123}$ and $\hat{P}_{321}$ to assess the directional hypothesis $G\_1 \leq G\_2 \leq G3$ or $G\_3 \leq G\_2 \leq G\_1$ with at least one strict inequality, thus detecting a consistent ordered relationships across the expression levels of all three ordered genotypes. This rank-based approach does not assume normality and is robust to outliers. We first imported the gene expression counts, calculated the TPM expressions and imported the variants from the WGS data. We then performed for all variants within a 1 MB window around a gene (±0.5 MB up−/ and downstream) the eQTL analysis and visualized the test results using the R-package GenomicTools [36]. Statistical significance was assessed using asymptotic p-values, and a threshold of 0.05 was applied to detect a directional order of expression values grouped by genotypes. The detailed eQTL protocol is described in [37].

For the downstream eQTL analysis, we then imported the existing Ensembl *Bos taurus* ARS-UCD1.3.111 gene annotation as well as a list of genes that we identified to be differentially expressed in our four comparison groups $UH_{>115}$ and $UH_{\leq 90}$ at timepoints 3 h and 24 h, and in addition, genomic regions of particular interest. These regions were the windows obtained from the ShinyGO enrichment analysis as well as the previously reported clinical mastitis GWAS genomic regions from a single population analysis used in a mastitis meta-GWAS study [9]. The SNPs significantly associated with clinical mastitis incidences in the GWAS from the NRC population by Cai et al. [9] cluster in 11 regions ranging in length from 1 bp to 47,964 bp (BTA1:131,376,695-131,533,707; BTA3:32,985,900-33,009,094; BTA5:56,290,203-56,516,159; BTA6:56,583,649-56,624,695; BTA6:86,940,862-87,407,925; BTA9:10435,362-10,882,115; BTA16:25,116,770-25,116,771 BTA18:65,257,581-65,283,715; BTA20:22,214,845-22,247,993; BTA21:56,967,311-57,160,681; BTA23:39,541,648-39,589,612, BTA = *Bos taurus* autosome). For interpretation, we then grouped and annotated our eQTL results to see if the GWAS genomic regions from NRC overlap with differentially expressed genes (DEGs), or if eQTL (in DE gene regions) would overlap with GWAS results from the same population.

3. Results

3.1. RNA and whole genome sequencing

Altogether 72 separate RNA sequencing libraries were sequenced (Fig. 1). For downstream analyses, cows were grouped according to their UH index either to group $UH_{\leq 90}$ or to group $UH_{>115}$. Details on RNA sequencing data creation, alignment and gene assignment are shown in Supplementary File 1. For WGS, there were on average 331.2/325.8 M raw/trimmed reads for $UH_{\leq 90}$ and 322.0/317.0 M for $UH_{>115}$, respectively. The average aligned percentage was 99.8 % for both groups. The number of variants called from the WGS data varied from 17.6 M to 17.7 M per cow and the call rate was >95 % for the variants on autosomes.

3.2. Differential gene expression analyses

Among both groups ($UH_{>115}$ and $UH_{\leq 90}$), in total 1780 genes were detected to be differentially expressed ($P_{adj} < 0.05$, Supplementary File 2): 497 genes 3 h post-challenge and 1283 24 h post-challenge. Moreover, more DEGs were observed in the poor udder health index group ($UH_{\leq 90,}$ 1606 DEGs) compared to the good udder health index group ($UH_{>115}$,724 DEGs). Also, DEGs are partly overlapping between the timepoints and cow groups (Fig. 2). A majority of DEGs were upregulated; 418 genes 3 h post-challenge and 930 genes 24 h post-challenge. In the $UH_{>115}$ group, 7 % and 15 % of DEGs were downregulated at 3- and 24-h post-challenge, respectively whereas, the $UH_{\leq 90}$ group showed 5 % and 28 % downregulated DEGs at the same time points. All DEGs in $UH_{>115}$ and $UH_{\leq 90}$ groups are shown in Supplementary File 3.

Comparison of the top significantly associated genes ($-Log_{10}P > 100$) between the two udder health index groups revealed both shared and group-specific DEGs. The genes *CXCL2, RND1, NFKBIA,* and *GRO1* are differentially expressed in both groups and at both timepoints. Group-specific DEGs for timepoint 3 h are *IL1B, CD83, NFKBIZ, TNF, ZC3H12A,* and *TNFAIP3* in the $UH_{\leq 90}$ group*;* and *CXCL3* in the $UH_{>115}$ group (Supplementary File 1). After 24 h, group specific DEGs are *CCL5* and *IL6* in the $UH_{\leq 90}$ group*;* and *CXCL8, RIG1* and *NFKBIA* in the $UH_{>115}$ group.

The genes with largest fold changes (from Log_2 fold change 8 to 10), Fig. 3) at 3 h are *IL1B* and a novel gene (*ENSBTAG00000067195*, WikiGene match interferon beta-2) in both groups, and in the $UH_{\leq 90}$ group *CCL4, IFNL3* and *CXCL10*, in the $UH_{>115}$ group *SLC6A12, CCL20* and *TNF*. At 24 h the largest fold changes (from Log_2 fold change 8 to 10) in the $UH_{\leq 90}$ group are for *SAA3, M-SAA3.2, TAP,* novel gene (*ENSBTAG00000046944*, coding for a WAP domain containing protein), *CCL4, PLAC8B, IDO1, IL1B* and *CXCL10*, and in the $UH_{>115}$ group *M-SAA3.2, SAA3, IFIT3, IFIT2, CCL5* and *MX2*.

3.3. Over-representation analyses, GO and KEGG

3.3.1. Enrichment analysis on DEGs genes

When applying the cut-off $p < 0.05$ (*P*-value) and $q < 0.01$ (FDR) for enriched GO terms, 301 (3 h post-challenge) and 141 (24 h post-challenge) GO terms remained for $UH_{\leq 90}$ group. For $UH_{>115}$ group, 280 (3 h post-challenge) and 214 (24 h post-challenge) GO terms remained (Supplementary File 4).

The top 10 ranking terms were almost the same for both groups, (7/10) at 3 h and (10/10) 24 h post-challenge. Among the top 10 biological process GO-terms, group-specific GO-terms at 3 h were positive regulation of biosynthetic process and positive regulation of macromolecule biosynthesis in the $UH_{\leq 90}$ group, and cytokine-mediated signaling pathway and cellular response to cytokine stimulus in the $UH_{>115}$ group. Among the next high-ranking GO-terms (up to 25), there is more variation between the groups. At 3 h, in the $UH_{\leq 90}$ group we uniquely see regulation of signal transduction, regulation of cell differentiation, regulation of RNA biosynthetic process, positive regulation of cell differentiation, regulation of RNA metabolic process; and in the $UH_{>115}$ group we see inflammatory response, regulation of immune effector process, positive regulation of immune system process, response to lipopolysaccharide, response to molecule of bacterial origin, regulation of adaptive immune response, and innate immune response terms enriched. At 24 h, regulation of adaptive immune response terms (*e.g.* GO:0006955, GO:0006952, GO:0002376, and GO:0045087) appear enriched in the $UH_{\leq 90}$ group, thus later than in the $UH_{>115}$ group.

Most of the down-regulated genes were observed in the $UH_{\leq 90}$ group, thereby GO-terms associated with down-regulated genes were statistically significantly enriched only in that group. The enrichment analysis for down-regulated genes at 3 h post- challenge indicated three GO terms (GO:0090092, GO:0007178, GO:0090100). The GO terms

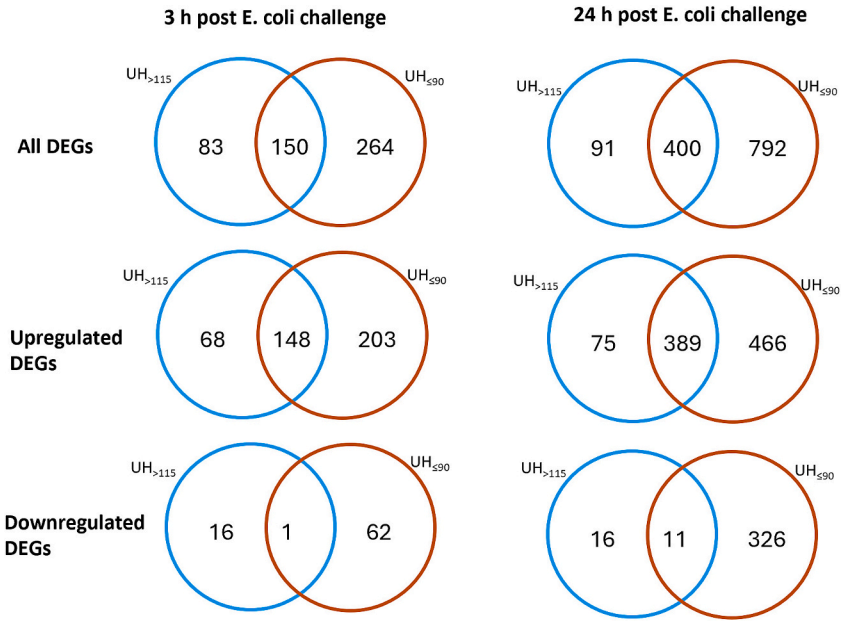

Fig. 2. Venn diagrams showing the overlap of differentially expressed genes (upregulated and downregulated DEGs) identified in two study groups ($UH_{>115}$, represented by blue circles and $UH_{\leq 90}$, represented by red circles) and post-challenge timepoints (3 h and 24 h). Overlapping regions represent shared genes between the groups and timepoints. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

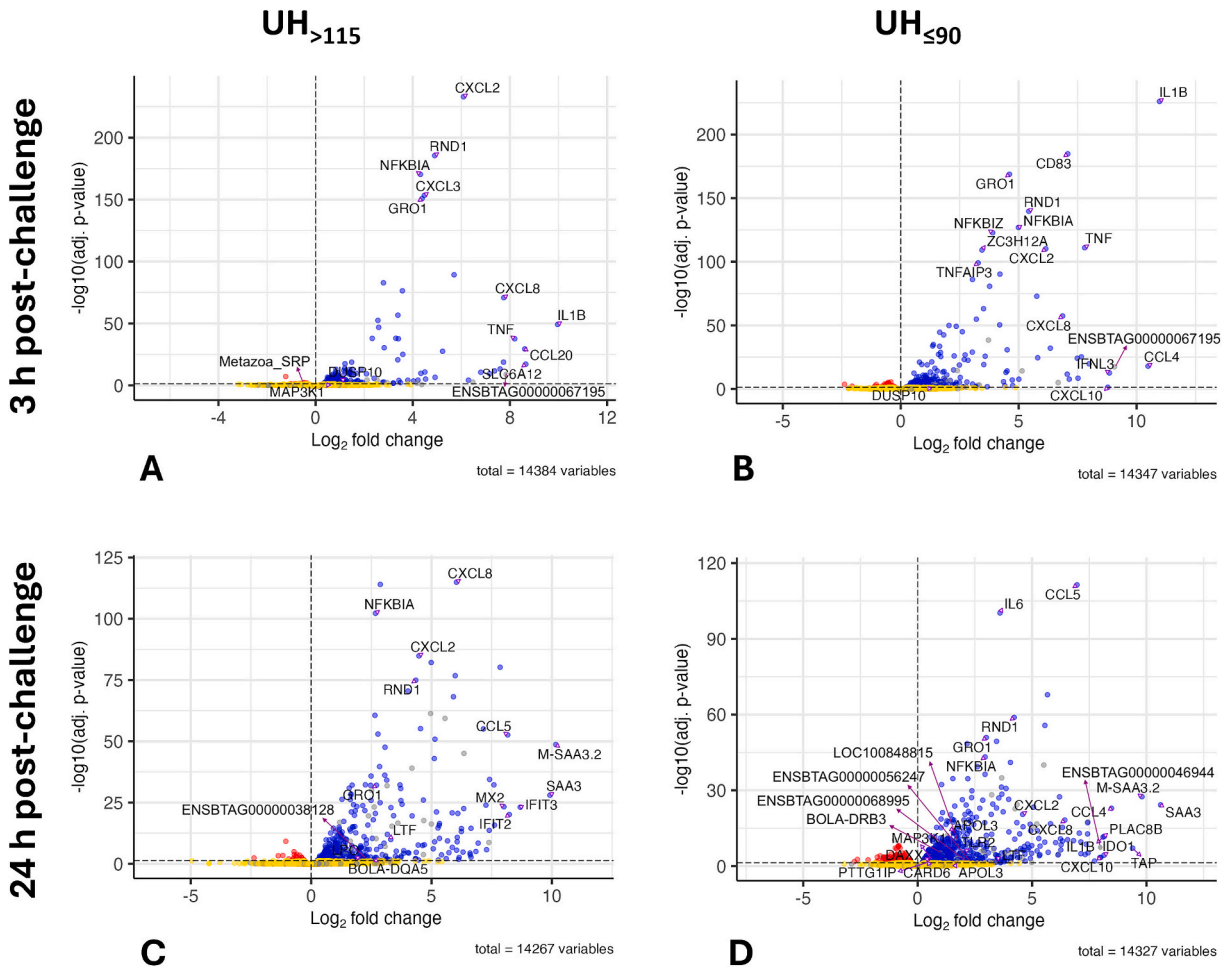

Fig. 3. Volcano plots showing the results of differential expression analyses. The DEGs referred to in this paper are indicated by their names or IDs. Fig. 3A and B 3 h post *E. coli* challenge, and 3C and 3D 24 h post *E. coli* challenge. Fig. 3A and C are for cows belonging to $UH_{>115}$ group, and Fig. 3B and D are for cows from $UH_{\leq 90}$ group. X-axis (Log_2 fold change) represents the magnitude of change in gene expression between challenged and unchallenged cells. Y-axis ($-log_{10}$ adj. *p*-value) represents the statistical significance of the change in gene expression. Blue dots: upregulated DEGs with $P_{adj} < 0.05$, red dots: downregulated DEGs with $P_{adj} < 0.05$, yellow dots: genes having $P_{adj} > 0.05$; grey dots: expressed genes with Ensembl annotation where no entrez gene symbol is available. Horizontal dashed line shows the threshold for significance ($P_{adj} < 0.05$). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

GO:0090092 and GO:0090100 are child terms of GO:0007178 (https://www.ebi.ac.uk/QuickGO/term/GO:0007178). The core pathway GO:0007178 is important, for example in developmental biology and immune regulation. At 24 h, processes related to cell cycle control, DNA replication and repair, and cellular responses to stress were enriched.

The top 15 enriched KEGG pathways based on all DE genes do not differ markedly between the $UH_{\leq 90}$ and $UH_{>115}$ groups, at either time point after challenge, with TNF signaling and cytokine-cytokine receptor interaction topping the list at 3 h post-challenge, and Influenza A and Epstein-Barr virus infection pathways getting more emphasis at 24 h (Fig. 4, Supplementary File 4). When enriched KEGG pathways are looked at by timepoints, at 3 h unique $UH_{\leq 90}$ KEGG pathways include themes like hormone signaling, drug resistance, and viral infection and unique $UH_{>115}$ KEGG pathways highlight immune specialization and cancer-related processes. Likewise, at 24 h, $UH_{>115}$ leans toward cellular processes, replication, and innate immune responses whereas $UH_{\leq 90}$ emphasizes adaptive immunity, viral mechanisms, and cancer-specific pathways.

3.4. Genomic regions with DEGs

No exact match between the significant GWAS regions for NRC [9] and location of DE genes from the two UH index groups and timepoints was observed. However, not all GWAS regions include genes (the regions on BTA1, BTA20 and BTA16), or have a limited number of genes (from 1 to 17 genes/region).

ShinyGO enrichment analysis indicated that one genomic region (BTA6: 84 Mb – 93 Mb, BTA = *Bos taurus* chromosome) at 3 h post-challenge for both the groups is statistically significantly enriched with DEGs (when using 6 Mb as window size, 2 steps in a window, and cut-off value of 0.01 for FDR, Supplementary File 5). After 24 h of challenge, no chromosomal enrichment of the genes was detected for the group $UH_{\leq 90}$ whereas seven chromosomal regions showed enrichment of the DE genes in the group $UH_{>115}$ (BTA1: 138 Mb – 147 Mb; BTA5: 73 Mb – 75 Mb; BTA6: 84 Mb – 93 Mb; BTA15: 3 Mb – 9 Mb; BTA19: 9 Mb – 18 Mb; BTA23: 6 Mb – 12 Mb; BTA23: 24 Mb – 33 Mb).

3.5. eQTL analysis

In the eQTL analysis, we calculated for each variant the potential association between the genotypes and the gene expression in a 0.5 MB window up- and downstream around the center of each protein coding gene. In total we tested 35,643 genes against 17,294,457 variants, with an average of 6285 variants (min: 1976, max: 24,497) within the cis-region of a gene. A chromosome-wise summary table can be found in Supplementary File 6. In total 9266 genes had at least one significant (*p*-value $\leq$0.05) variant in the $UH_{>115}$ group, while there were 8926 genes with at least one significant variant in the $UH_{\leq 90}$ group. On average

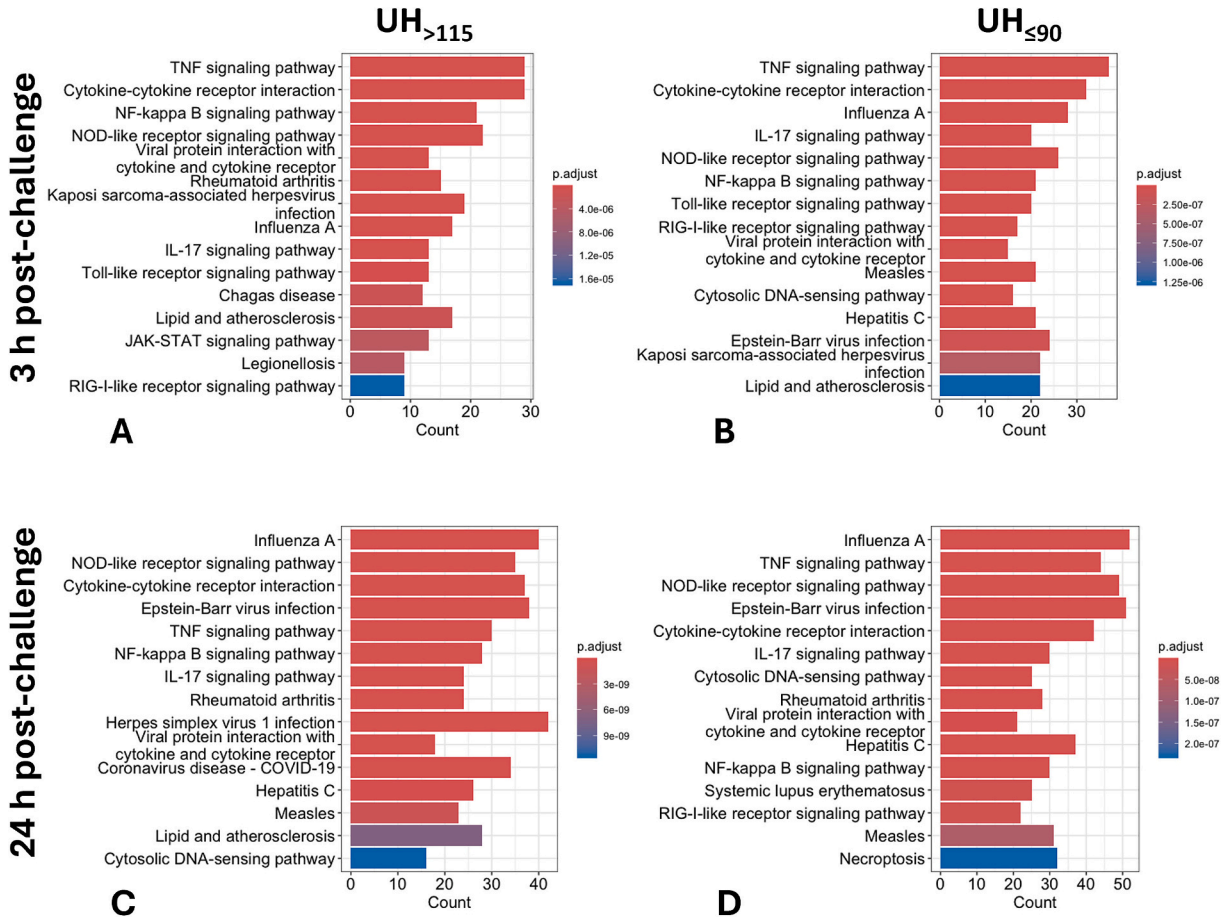

Fig. 4. Enriched KEGG pathways for 3 h and 24 h after *E. coli* challenge. Figs. A and C are from $UH_{>115}$ group, and figs. B and D are from $UH_{\leq 90}$ group.

there were 205 and 217 significant directional tests between a variant and the associated gene in the $UH_{>115}$ and $UH_{\leq 90}$ groups, respectively. The eQTL results from the DE enriched regions and GWAS regions were specifically searched for eQTL matching with time and group of the differential expression. Using this approach, we are highlighting some of the identified eQTL regions and genes below. However, the results need to be considered with caution because of the small number of samples included.

A QTL for CM and SCS at approximately 88–89 Mb on BTA6 has been reported in many studies and multiple cow breeds [8,38,39], including in Nordic Red dairy cattle breed [9,40] with the best plausible candidate gene being the *GC* gene (*ENSBTAG00000013718*, [41]) but also other genes in this region have been suggested [2]. In our study, that region was statistically significantly enriched with DEGs both 3 h and 24 h post-challenge in the $UH_{>115}$ group and enriched with DEGs at 3 h post-challenge in the $UH_{\leq 90}$ group. The gene *GC* was not significantly differentially expressed after pathogen challenge in either of the groups at either time point. Interestingly, the eQTL analysis indicates cis-eQTL in both groups overlapping the genomic region of *GC* gene (BTA6: 86,963,822-87,007,062). For the $UH_{>115}$, 24 h post-challenge, the eQTL (BTA6: 86,846,986-87,290,510) affects the gene expression of a novel gene, *ENSBTAG00000066835*, overlapping *GC* but on the opposite strand. In $UH_{\leq 90}$, 24 h post-challenge group the eQTL (BTA6: 86,464,022- 87,507,043) effect is for the *GC* gene (Fig. 5).

Besides the genomic region on BTA6, the eQTL analysis revealed other interesting genomic regions (Table 2). On BTA5, in the region (BTA5: 73 Mb – 75 Mb) showing enrichment of DE genes in the group $UH_{>115}$, two neighboring genes, *APOL3* and *ENSBTAG00000056247*, both up-regulated at 24 h post-challenge in the $UH_{\leq 90}$ group, shared the same cis-eQTL (*p*-value 0.05, BTA5: 75,030,550- 75,160,876).

The closest eQTL to GWAS regions was identified in the $UH_{\leq 90}$ group at 24 h post-challenge on BTA9 (eQTL at BTA9:10,221,349-10,435,099, GWAS at BTA9:10,435,362-10,882,115), affecting the lncRNA gene *ENSBTAG00000068995*. That gene was significantly differentially upregulated in the $UH_{\leq 90}$ group at 24 h post-challenge. Variants within the eQTL and GWAS on BTA9:10,435,362-10,882,115 were not shared but the two closest were 265 bp apart from each other (intergenic variant rs378332774 significant for eQTL and intergenic variant rs109579327 significant for GWAS).

4. Discussion

In this study, we used a non-invasive *in vitro* functional approach to identify genomic features underlying mastitis resistance. *Escherichia coli* -challenged primary bovine mammary epithelial cells (pbMECs) derived from milk served as a tool for evaluating innate immune response of dairy cows. Our study population consisted of cell donor cows that were divided in two contrasting groups (udder health index over 115 ($UH_{>115}$) and udder health index below or equal to 90 ($UH_{\leq 90}$). We hypothesized that the contrast would allow us to pinpoint differences in gene expression between cows being genetically more or less prone to mastitis. The negative genetic correlation of high milk production with susceptibility to clinical mastitis is well known [43]. The average of the index for milk yield in kg for $UH_{\leq 90}$ was 108 (SD =5.7) whereas for cows having $UH_{>115}$ it was 100 (SD = 0, Table 1). Also, cows in the two groups deviated in fertility index. For $UH_{\leq 90}$ group, the average fertility index was 94 (SD = 5.9) and for $UH_{>115}$ group that was 106 (SD = 2.4). Several studies have found that clinical mastitis is negatively correlated with fertility and longevity (reviewed by [44]).

Approximately twice as many genes were differentially expressed

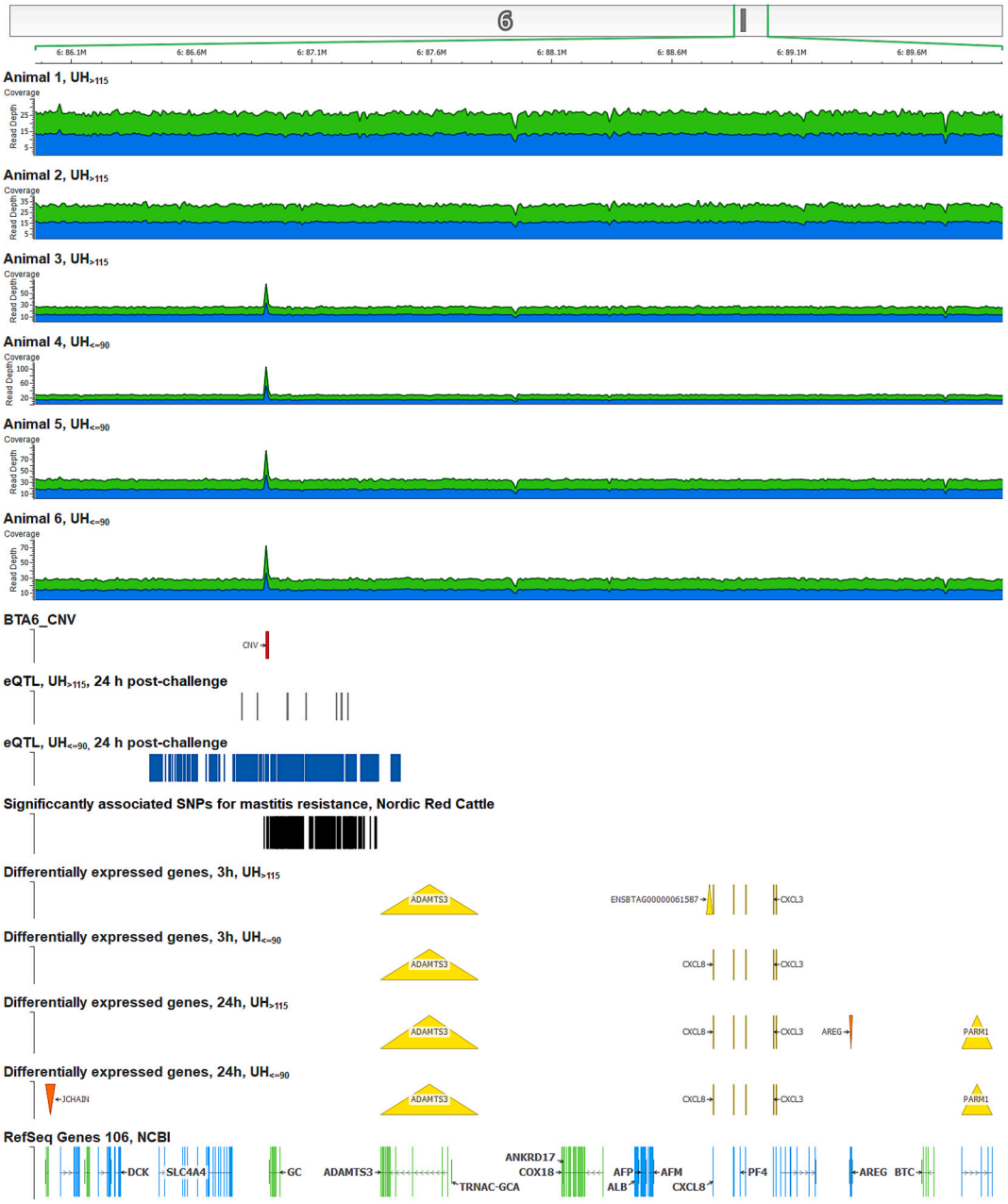

Fig. 5. Genomic region on BTA6 that harbors QTL region for clinical mastitis [9] marked in black (Significantly associated SNPs for mastitis resistance, Nordic Red Cattle), a 12 kb multi-allelic CNV associated with mastitis resistance [42] marked in red (BTA6_CNV), and two eQTL regions identified in this study marked in grey (eQTL, $UH_{>115}$, 24 h post-challenge) and blue (eQTL, $UH_{\leq 90}$, 24 h post-challenge). Yellow triangles are showing the DEGs upregulated, and orange triangles are representing the DEGs downregulated. First six tracks show the sequencing coverage for cows used as pbMEC donors. The coverage for four cows (one cow from $UH_{>115}$ group and three cows from $UH_{\leq 90}$ group) is increased at the region of CNV proving the existence of the multiple copies of that region (BTA6 region in figure: BTA6:85,988,051-90,018,028). Figure was created using Golden Helix GenomeBrowse 3.1.0. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

both at 3 h and 24 h post-challenge within the group $UH_{\leq 90}$ and almost ten times more genes were downregulated in the group $UH_{\leq 90}$ at 24 h post-challenge (Fig. 2). The fact that most DEGs were upregulated in both groups indicates a strong activation of biological pathways in response to the *E. coli* challenge. The $UH_{>115}$ group shows a modest increase in downregulated genes from 3 h (7 %) to 24 h (15 %). This may indicate a gradual suppression of certain genes, possibly as part of a regulatory or recovery process. The proportion of downregulated DEGs increased more in the $UH_{\leq 90}$ group, rising from 5 % at 3 h to 28 % at 24 h. This could indicate a stronger or more prolonged suppression of gene activity, potentially reflecting a different or delayed regulatory mechanism. In general, these differences between the groups may reflect distinct responses that might be due to different genetic backgrounds of the cows.

Enrichment analyses indicate similar GO terms and KEGG pathways to be enriched (Supplementary File 3, Fig. 2), besides the association of GO-terms with the regulation of cell cycle and primary metabolic processes at both 3 h and 24 h in the $UH_{\leq 90}$. It seems that the mammary epithelial cells from cows with lower mastitis resistance are reacting more strongly to the pathogen challenge, especially by modifying the cell metabolism. The reason and consequences of this phenomenon are not directly evident, but it may relate to lower innate immune defense

Table 2
Genomic regions with eQTL identified in this study. The table lists cis-eQTLs and the genes found in the eQTL regions being either up- or downregulated in either of our study groups, 3 h or 24 h post-challenge.

BTA	cis-eQTL, start	cis-eQTL, end	Gene	Gene ID	Gene expression	Group	h post-challenge
1	144,213,826	144,928,988	*PTTG1IP*	ENSBTAG00000021771	upregulated	UH ≤ 90	24
5	75,030,550	75,160,876	*APOL3*	ENSBTAG00000000667	upregulated	UH ≤ 90	24
5	75,030,550	75,160,876	*ENSBTAG00000056247*	ENSBTAG00000056247	upregulated	UH ≤ 90	24
6	86,963,822	87,007,062	*ENSBTAG00000066835*	ENSBTAG00000066835	not DEG	UH > 115	24
6	86,963,822	87,007,062	*GC*	ENSBTAG00000013718	not DEG	UH ≤ 90	24
9	10,221,349	10,435,099	*ENSBTAG00000068995*	ENSBTAG00000068995	upregulated	UH ≤ 90	24
19	8,703,851	9,745,960	*LPO*	ENSBTAG00000012780	upregulated	UH > 115	24
23	25,192,305	26,188,540	*LOC100848815*	ENSBTAG00000037605	upregulated	UH ≤ 90	24
23	25,223,501	26,188,470	*BOLA-DRB3*	ENSBTAG00000013919	upregulated	UH ≤ 90	24
23	30,854,073	31,855,925	*H2AC12*	ENSBTAG00000063383	upregulated	UH ≤ 90	24
23	6,999,974	7,920,165	*DAXX*	ENSBTAG00000016830	upregulated	UH ≤ 90	24
23	25,136,913	26,143,871	*ENSBTAG00000038128*	ENSBTAG00000038128	upregulated	UH > 115	24
23	29,519,501	30,517,312	*Metazoa SRP*	ENSBTAG00000047789	downregulated	UH > 115	3

making them compensate that with metabolic reprogramming or it may indicate that the cells are under higher metabolic stress.

The current results are based on the bovine genome build ARS-UCD1.3. In our previous study [11] conducted among the $UH_{>115}$ group, we used the version ARS-UCD1.2 and detected less DEGs, with a total of 590 DEGs identified — 150 at 3 h post-challenge and 440 at 24 h post-challenge. The differences in results when using different genome versions are not discussed in detail here. However, the larger number of genes identified with version ARS-UCD1.3 suggests that it is an improvement over ARS-UCD1.2. There was no exact match between the significant GWAS regions for NRC [9]and the locations of DEGs from the two UH index groups and timepoints. Some GWAS regions contain no annotated genes or only a limited number of genes and this may partly explain the lack of overlaps. Also, regulatory variants, not genes, that we were not able to pinpoint in this study can exist GWAS regions, potentially explain the existing GWAS regions.

In a systematic review of GWAS studies for genetic architecture underlying resistance to mastitis, Narayana et al. 2023 [45] pinpointed 427 significant genes by prioritization analysis. Most top-ranked genes were from the cytokine superfamily, including chemokines, interleukins, transforming growth factors, and tumor necrosis factor genes. Most prioritized genes (397) were associated with somatic cell count related traits. They further reported 54 genes associated with clinical mastitis related traits (not given in the results), and 24 genes associated with both somatic cell count and clinical mastitis related traits [45]. None of these prioritized 24 genes were indicated in our data. Among the top 8 enriched KEGG pathways in Narayana et al. 2023 [45] for the prioritized genes were three pathways that were also enriched among the DE genes in this study: cytokine-cytokine receptor interaction, Kaposi sarcoma associated herpesvirus infection and JAK-STAT signaling pathway. Brajnik and Ogorevc [46] used data integration approach to identify most promising candidate loci for mastitis resistance. In total, they included 2448 genes, of which 157 were from association studies, 2300 from expression studies, and six genes from mouse model studies. Of those, they further prioritized 22 genes that were found in multiple independent studies to represent the most promising candidate genes for mastitis resistance. Of these genes, *CXCL8 on BTA6* (*ENSBTAG00000019716*, DEG in the $UH_{\leq 90}$ at 3 h and 24 h post-challenge, DEG in the $UH_{>115}$ at 3 h and 24 h post-challenge), *LTF* on BTA22 (*ENSBTAG00000001292*, DEG in the $UH_{\leq 90}$ and $UH_{>115}$ groups at 24 h post-challenge), *TLR2* on BTA17 (*ENSBTAG00000008008*, DEG in the $UH_{\leq 90}$ at 24 h post-challenge), *BOLA-DRB3 on BTA23* (*ENSBTAG00000013919*, DEG in the $UH_{\leq 90}$ at 24 h post-challenge), and *CARD6* on BTA 20 (*ENSBTAG00000014374*, DEG in the $UH_{\leq 90}$ at 24 h post-challenge) are also among DEGs in our study. *BOLA-DRB3* gene has been associated with infectious diseases and different production traits in several studies [47]. The list of 32 candidate genes revealed by meta-analysis of six dairy cattle breeds [9] have some overlaps with genes responding to *E. coli* in our study. At 3 h post-challenge in both the $UH_{\leq 90}$ and in the $UH_{>115}$ groups, *DUSP10* (BTA16:25,186,203-25,227,307, *ENSBTAG00000001729*) was upregulated. In addition, in the $UH_{>115}$ group at 3 h post-challenge, *MAP3K1* (BTA20:22,340,163-22,417,627, *ENSBTAG00000013790*) was upregulated. After 24 h, *MAP3K1* was upregulated also in the $UH_{\leq 90}$ group but not anymore in the $UH_{>115}$ group. Furthermore, from the candidate gene list in [9] and in [46], *LTF* (BTA22:52,934,900-52,986,625, ENSBTAG00000001292) was upregulated in both groups 24 h post-challenge.

4.1. BTA6 region

For the well-known QTL for CM and SCS on BTA6, Lee et al. 2021 [42] proposed that a copy number variant (CNV) in *GC* could be the potential causal mutation overlapping an enhancer region of the gene. The multiplicated alleles of that CNV are associated with increased *GC* gene expression and low CM resistance [42]. The genome sequences of the cows used to extract pbMECs from milk showed that all three cows in the $UH_{\leq 90}$ group have the CNV as several copies, whereas 2 out of 3 in the $UH_{>115}$ group did not have CNV in that location (Fig. 4). As noted, the gene *GC* was not significantly differentially expressed after pathogen challenge in either of the groups at either time point, but we did observe cis-eQTL in both groups. In $UH_{>115}$ group it affects the expression of the novel gene, *ENSBTAG00000066835* that is overlapping *GC* but on the opposite strand, and in $UH_{\leq 90}$ group the effect is for the *GC* gene. It seems that the *GC* gene region is not involved in the first response to the pathogen in the mammary epithelium. It has been reported earlier that the gene is not expressed in the mammary gland, but predominantly in the liver [41]. In that report, no differences were detected between alternative genotypes in the expression of *GC* mRNA in adult liver in healthy cows. As in our study all cows in the $UH_{\leq 90}$ group also have the CNV, it is likely that the *GC* gene may be associated with mastitis resistance, but in a different tissue and/or at a later stage of defense against pathogens.

4.2. BTA5 region

There was a cis-eQTL at BTA5 on 75,030,550-75,160,876. Within 1 Mb of this region there are two DEGs, *APOL3* and *ENSBTAG00000056247*, belonging to the apolipoprotein L (*APOL*) gene family, which in humans is composed of six genes within 619 kb on human chromosome 22. In humans, *APOL* genes have been implicated in different diseases from arthritis to cancer, but the specific roles of the *APOL* proteins are not known. The genes are upregulated by pro-inflammatory signaling molecules which indicate a role in the immune system [44]. The two *APOL* genes in this study, sharing the same eQTL, were both upregulated at 24 h post-challenge in the $UH_{\leq 90}$ group. It could be hypothesized that the eQTL mediates the inflammatory response leading to up-regulation of the *APOL* genes. The *APOL3* gene is located within a CNVR that encompasses quantitative trait loci for

economic traits like meat quality [48], in beef breeds. This CNV was not present in our study cows. *APOL3* resides within a well-known cattle milk trait QTL region (75–76 Mbp), and it is expressed in the mammary gland, with cis-eQTL affecting the expression. It has not been prioritized as a causal gene for the QTL effect on milk traits. The top eQTL SNP, rs433710540 [49] is not among the significant eQTL variants in this study.

4.3. BTA23 region

It is noteworthy that we detected many differences in gene expression on BTA23 with eQTL in both genetic groups after pathogen challenge. In the group $UH_{\leq 90,}$ both immune (*BoLADRB3* and *LOC100848815* SLA class II histocompatibility antigen) and transcription/chromatin modification related genes (Histone genes, *DAXX*) were upregulated. In the $UH_{>115}$ group, *BOLA-DQA5* was upregulated at 24 h and *Metazoa SRP* downregulated at 3 h post-challenge. The DE genes/eQTL did not overlap the known GWAS regions on BTA23, but the enrichment of DE genes and the associated eQTL indicate a major role in the first response in the mammary epithelium.

The eQTL results are based on a small sample size and should therefore be interpreted with caution. For example, the CattleGTEx consortium has outlined that a sufficient sample size to identify eQTLs and alternative splicing (sQTLs) could be more than 40 animals [50]. However, scaling up eQTL analysis is costly due to the expense of collecting matched transcriptomic and genomic information [51] and to our knowledge, no large-scale eQTL studies have been conducted with challenged cattle mammary epithelial cells. Rather than presenting definitive conclusions, our results are better viewed as indicative or exploratory. To mitigate the limitations associated with the small sample size, we applied a non-parametric test instead of the standard linear model approach. This choice was made to avoid the dependency on distributional assumptions inherent to parametrical models, which are likely to be violated in a small-sample setting. Further, given the small sample size, we prioritized sensitivity over specificity and omitted a multiple-testing adjustment in the eQTL analysis. Consequently, these results are primarily data-driven and are presented as a pilot-scale study that paves the way for future research at population level that enable the necessary validation of our findings.

Our transcriptomic approach did not provide evidence supporting potential SNPs identified in the GWAS results. Such evidence-supported variants could be emphasized in genomic selection [52]. It is possible that the early response to the pathogen in the mammary gland epithelium is not a determining factor in phenotypic mastitis resistance. As an example, the *GC* gene is a major candidate for an effect on mastitis resistance, but it is not differentially expressed in the mammary epithelial cells. However, our results indicate eQTL in the region that could be important in another tissue or timepoint. The main difference between the groups with contrasting genetic predisposition to mastitis was the higher number of genes responding to pathogen challenge in the low mastitis resistance group. Many of the additional DE genes specific to the low mastitis resistance group are associated with biological processes relating to regulation of primary metabolic processes, but some also to the same inflammatory pathways (Supplementary File 4) as seen in the high mastitis resistance group. It may be speculated that the selection for better udder health has led to a more uniform initial inflammatory response to the pathogen in the mammary gland epithelium.

5. Conclusions

Our approach revealed a difference in the transcriptomic response to *E. coli* challenge between primary bovine mammary epithelial cells derived from two cow groups genetically contrasting for udder health index/mastitis resistance. The most striking difference is the markedly higher number of DEGs in the $UH_{\leq 90}$ group. Based on gene enrichment analyses, this difference may reflect a slightly delayed adaptive immune response and early modification of primary metabolic processes in that group. It could also indicate that the group with lower udder health index has other genetic factors causing stronger transcriptomic response to *E. coli* challenge. Another possibility is that cows with lower udder health are more susceptible to infection due to stress caused by higher milk production. Many previously identified candidate genes (such as *LTF*, *CXCL8, TLR2*, *BOLA-DRB3, CARD6, DUSP10, MAP3K1*) responded to *E. coli* challenge in both groups supporting a role in immune response but with no difference between the UH index groups in the early stage of infection. There was no overlap between the DEGs or their potential cis-eQTL with the location of candidate SNPs from GWAS studies from the same NRC population. This implies that the genes and pathways involved in the early immune response in the mammary epithelium may not be the primary ones for phenotypic mastitis resistance.

To our knowledge, our study is the first to utilize genome-wise and phenotypically deeply characterized animals for a specific trait (udder health) for assessing *in vitro* potential differences in gene expression and genotype as a base for differences in the phenotype. This is in line with the initial aims and current research priorities of the FAANG project ([1]) to improve our ability to more accurately use genotype to predict phenotype, which would directly address the challenges for sustainable and responsible food production in the future.

CRediT authorship contribution statement

Terhi Iso-Touru: Writing – review & editing, Writing – original draft, Methodology, Conceptualization. **Daniel Fischer:** Writing – review & editing, Software, Methodology. **Frank Panitz:** Writing – review & editing, Software, Methodology. **Suvi Taponen:** Writing – review & editing. **Zexi Cai:** Writing – review & editing. **Goutam Sahana:** Writing – review & editing. **Ilma Tapio:** Writing – review & editing, Funding acquisition. **Johanna Vilkki:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition, Conceptualization.

Funding

Funded by Academy of Finland, Finland grant No. 317998 and the Horizon 2020 project HoloRuminant (Grant Agreement No 101000213).

Declaration of competing interest

The authors declare that they have no competing financial interests or personal relationships that may have influenced the work reported in this study.

Acknowledgements

We thank Finnish Functional Genomics Centre, supported by University of Turku, Åbo Akademi University and Biocenter Finland, for sequencing services. The authors wish also to acknowledge CSC—IT Center for Science, Finland, for computational resources. BSc Anneli Virta and BSc Jonna Tabell are acknowledged for their excellent work in the laboratory.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ygeno.2025.111102.

Data availability

All the sequencing data of this study are publicly available and have been deposited into public repositories. All raw sequencing data and associated metadata are available in the FAANG Data portal (https://data.faang.org/home) and ENA under accession number

PRJEB88821 with their associated metadata.

References

[1] E.L. Clark, A.L. Archibald, H.D. Daetwyler, M.A.M. Groenen, P.W. Harrison, R. D. Houston, C. Kühn, S. Lien, D.J. Macqueen, J.M. Reecy, D. Robledo, M. Watson, C.K. Tuggle, E. Giuffra, From FAANG to fork: application of highly annotated genomes to improve farmed animal production, Genome Biol. 21 (2020) 285, https://doi.org/10.1186/s13059-020-02197-8.

[2] P.L. Ruegg, A 100-year review: mastitis detection, management, and prevention, J. Dairy Sci. 100 (2017) 10381–10397, https://doi.org/10.3168/jds.2017-13023.

[3] Z. Xu, Y. Zhang, Y. Wang, A. Wu, C. Meng, W. Li, J. Yi, C. Chen, Evaluation of the safety and immune protection of OMPAC, PAPF, and EBPSs recombinant subunit vaccines developed for Escherichia coli, Staphylococcus aureus, and Streptococcus agalactiae in mice, Int. Immunopharmacol. 148 (2025) 114151, https://doi.org/10.1016/j.intimp.2025.114151.

[4] H. Zengin, Z. Keskin Alkaç, Y. Vezir, S. Tanyıldızı, F.A. Korkak, B. Dağoğlu Hark, G. Dagoglu, Evaluation of the efficacy of a new inactivated vaccine against Staphylococcus aureus, Echerchia coli and *Mycoplasma bovis* mastitis in cows, Vet. Ital. 60 (2024), https://doi.org/10.12834/VetIt.3083.21855.3.

[5] J. Pokorska, S. Sawicki, J. Gabryś, D. Kułaj, E.A. Bauer, A. Lenart-Boroń, K. Bulanda, M. Kuchta-Gładysz, A. Grzesiakowska, J. Kemilew, P.M. Barton, O. Lasek, M. Bugno-Poniewierska, The use of stem cells in the treatment of mastitis in dairy cows, Sci. Rep. 14 (2024) 10349, https://doi.org/10.1038/s41598-024-61051-0.

[6] L. Andersson, A.L. Archibald, C.D. Bottema, R. Brauning, S.C. Burgess, D.W. Burt, E. Casas, H.H. Cheng, L. Clarke, C. Couldrey, B.P. Dalrymple, C.G. Elsik, S. Foissac, E. Giuffra, M.A. Groenen, B.J. Hayes, L.S. Huang, H. Khatib, J.W. Kijas, H. Kim, J. K. Lunney, F.M. McCarthy, J.C. McEwan, S. Moore, B. Nanduri, C. Notredame, Y. Palti, G.S. Plastow, J.M. Reecy, G.A. Rohrer, E. Sarropoulou, C.J. Schmidt, J. Silverstein, R.L. Tellam, M. Tixier-Boichard, G. Tosser-Klopp, C.K. Tuggle, J. Vilkki, S.N. White, S. Zhao, H. Zhou, FAANG Consortium, Coordinated international action to accelerate genome-to-phenome with FAANG, the Functional Annotation of Animal Genomes project, Genome Biol. 16 (2015) 57, https://doi.org/10.1186/s13059-015-0622-4.

[7] N. Ghavi Hossein-Zadeh, An overview of recent technological developments in bovine genomics, vet, Anim. Sci. 25 (2024) 100382, https://doi.org/10.1016/j.vas.2024.100382.

[8] Z.-L. Hu, C.A. Park, X.-L. Wu, J.M. Reecy, Animal QTLdb: an improved database tool for livestock animal QTL/association data dissemination in the post-genome era, Nucleic Acids Res. 41 (2013) D871–D879, https://doi.org/10.1093/nar/gks1150.

[9] Z. Cai, T. Iso-Touru, M.-P. Sanchez, N. Kadri, A.C. Bouwman, P.K. Chitneedi, I. M. MacLeod, C.J. Vander Jagt, A.J. Chamberlain, B. Gredler-Grandl, M. Spengeler, M.S. Lund, D. Boichard, C. Kühn, H. Pausch, J. Vilkki, G. Sahana, Meta-analysis of six dairy cattle breeds reveals biologically relevant candidate genes for mastitis resistance, Genet. Sel. Evol. 56 (2024) 54, https://doi.org/10.1186/s12711-024-00920-8.

[10] NAV, NAV Official Genetic Evaluation of Dairy Cattle – Data and Genetic Models, n.d. https://nordicebv.info/wp-content/uploads/2024/02/NAV-routine-genetic-evaluation-feb-2024.pdf, 2025.

[11] T. Iso-Touru, F. Panitz, D. Fischer, M.K. Kyläniemi, S. Taponen, J. Tabell, A. Virta, J. Vilkki, Genes and pathways revealed by whole transcriptome analysis of milk derived bovine mammary epithelial cells after Escherichia coli challenge, Vet. Res. 55 (2024) 13, https://doi.org/10.1186/s13567-024-01269-y.

[12] S. Pyörälä, L. Kaartinen, H. Käck, V. Rainio, Efficacy of two therapy regimens for treatment of experimentally induced Escherichia coli mastitis in cows, J. Dairy Sci. 77 (1994) 453–461, https://doi.org/10.3168/jds.S0022-0302(94)76973-3.

[13] R. Poplin, V. Ruano-Rubio, M.A. DePristo, T.J. Fennell, M.O. Carneiro, G.A.V. der Auwera, D.E. Kling, L.D. Gauthier, A. Levy-Moonshine, D. Roazen, K. Shakir, J. Thibault, S. Chandran, C. Whelan, M. Lek, S. Gabriel, M.J. Daly, B. Neale, D. G. MacArthur, E. Banks, Scaling Accurate Genetic Variant Discovery to Tens of Thousands of Samples, 2018, p. 201178, https://doi.org/10.1101/201178.

[14] F. Mölder, K.P. Jablonski, B. Letcher, M.B. Hall, C.H. Tomkins-Tinch, V. Sochat, J. Forster, S. Lee, S.O. Twardziok, A. Kanitz, A. Wilm, M. Holtgrewe, S. Rahmann, S. Nahnsen, J. Köster, Sustainable data analysis with snakemake, F1000Res 10 (2021) 33, https://doi.org/10.12688/f1000research.29032.2.

[15] D. Fischer, fischuu/Pipeline-WGS-VariantCalling: Stable Pre-release Version, 2023, https://doi.org/10.5281/zenodo.8401423.

[16] S. Chen, Y. Zhou, Y. Chen, J. Gu, Fastp: an ultra-fast all-in-one FASTQ preprocessor, Bioinformatics 34 (2018) i884–i890, https://doi.org/10.1093/bioinformatics/bty560.

[17] P. Ewels, M. Magnusson, S. Lundin, M. Käller, MultiQC: summarize analysis results for multiple tools and samples in a single report, Bioinformatics 32 (2016) 3047–3048, https://doi.org/10.1093/bioinformatics/btw354.

[18] A. Dobin, C.A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T.R. Gingeras, STAR: ultrafast universal RNA-seq aligner, Bioinformatics 29 (2013) 15–21, https://doi.org/10.1093/bioinformatics/bts635.

[19] Y. Liao, G.K. Smyth, W. Shi, featureCounts: an efficient general purpose program for assigning sequence reads to genomic features, Bioinformatics 30 (2014) 923–930, https://doi.org/10.1093/bioinformatics/btt656.

[20] D. Fischer, fischuu/Snakebite-RNAseq: Release Version 0.2, 2023, https://doi.org/10.5281/ZENODO.7743750.

[21] H. Wickham, M. Averick, J. Bryan, W. Chang, L.D. McGowan, R. François, G. Grolemund, A. Hayes, L. Henry, J. Hester, M. Kuhn, T.L. Pedersen, E. Miller, S. M. Bache, K. Müller, J. Ooms, D. Robinson, D.P. Seidel, V. Spinu, K. Takahashi, D. Vaughan, C. Wilke, K. Woo, H. Yutani, Welcome to the Tidyverse, J. Open Source Softw. 4 (2019) 1686, https://doi.org/10.21105/joss.01686.

[22] W. Huber, A. von Heydebreck, H. Sültmann, A. Poustka, M. Vingron, Variance stabilization applied to microarray data calibration and to the quantification of differential expression, Bioinformatics 18 (Suppl. 1) (2002) S96–104, https://doi.org/10.1093/bioinformatics/18.suppl_1.s96.

[23] R. Kolde, pheatmap: Pretty Heatmaps. https://CRAN.R-project.org/package=pheatmap, 2019. (Accessed 6 March 2023).

[24] H. Wickham, ggplot2: Elegant Graphics for Data Analysis, Springer, New York, NY, 2009, https://doi.org/10.1007/978-0-387-98141-3.

[25] K. Slowikowski, A. Schep, S. Hughes, T.K. Dang, S. Lukauskas, J.-O. Irisson, Z. N. Kamvar, T. Ryan, D. Christophe, Y. Hiroaki, P. Gramme, A.M. Abdol, M. Barrett, R. Cannoodt, M. Krassowski, M. Chirico, P. Aphalo, ggrepel: Automatically Position Non-overlapping Text Labels With "ggplot2". https://CRAN.R-project.org/package=ggrepel, 2023. (Accessed 6 March 2023).

[26] T.L. Pedersen, patchwork: The Composer of Plots. https://CRAN.R-project.org/package=patchwork, 2022. (Accessed 6 March 2023).

[27] E. Neuwirth, RColorBrewer: ColorBrewer Palettes. https://CRAN.R-project.org/package=RColorBrewer, 2022. (Accessed 6 March 2023).

[28] M.I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2, Genome Biol. 15 (2014) 550, https://doi.org/10.1186/s13059-014-0550-8.

[29] K. Blighe, EnhancedVolcano: Publication-Ready Volcano Plots with Enhanced Colouring and Labeling. https://github.com/kevinblighe/EnhancedVolcano, 2023. (Accessed 6 March 2023).

[30] S. Durinck, P.T. Spellman, E. Birney, W. Huber, Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt, Nat. Protoc. 4 (2009) 1184–1191, https://doi.org/10.1038/nprot.2009.97.

[31] T. Wu, E. Hu, S. Xu, M. Chen, P. Guo, Z. Dai, T. Feng, L. Zhou, W. Tang, L. Zhan, X. Fu, S. Liu, X. Bo, G., Yu, clusterProfiler 4.0: a universal enrichment tool for interpreting omics data, Innovation (Camb) 2 (2021) 100141, https://doi.org/10.1016/j.xinn.2021.100141.

[32] M. Kanehisa, M. Furumichi, Y. Sato, M. Kawashima, M. Ishiguro-Watanabe, KEGG for taxonomy-based analysis of pathways and genomes, Nucleic Acids Res. 51 (2023) D587–D592, https://doi.org/10.1093/nar/gkac963.

[33] S.X. Ge, D. Jung, R. Yao, ShinyGO: a graphical gene-set enrichment tool for animals and plants, Bioinformatics 36 (2020) 2628–2629, https://doi.org/10.1093/bioinformatics/btz931.

[34] W. Luo, C. Brouwer, Pathview: an R/Bioconductor package for pathway-based data integration and visualization, Bioinformatics 29 (2013) 1830–1831, https://doi.org/10.1093/bioinformatics/btt285.

[35] D. Fischer, H. Oja, J. Schleutker, P.K. Sen, T. Wahlfors, Generalized Mann–Whitney type tests for microarray experiments, Scand. J. Stat. 41 (2014) 672–692.

[36] D. Fischer, The R-package GenomicTools for multifactor dimensionality reduction and the analysis of (exploratory) Quantitative Trait Loci, Comput. Methods Prog. Biomed. 151 (2017) 171–177.

[37] D. Fischer, Performing QTL and eQTL analyses with the R-package GenomicTools, in: X.M. Shi (Ed.), eQTL Analysis: Methods and Protocols, Springer US, New York, NY, 2020, pp. 15–38, https://doi.org/10.1007/978-1-0716-0026-9_2.

[38] H. Nilsen, H.G. Olsen, B. Hayes, T. Nome, E. Sehested, M. Svendsen, T.H. E. Meuwissen, S. Lien, Characterization of a QTL region affecting clinical mastitis and protein yield on BTA6, Anim. Genet. 40 (2009) 701–712, https://doi.org/10.1111/j.1365-2052.2009.01908.x.

[39] G. Sahana, B. Guldbrandtsen, B. Thomsen, M.S. Lund, Confirmation and fine-mapping of clinical mastitis and somatic cell score QTL in Nordic Holstein cattle, Anim. Genet. 44 (2013) 620–626, https://doi.org/10.1111/age.12053.

[40] G. Sahana, B. Guldbrandtsen, B. Thomsen, L.-E. Holm, F. Panitz, R.F. Brøndum, C. Bendixen, M.S. Lund, Genome-wide association study using high-density single nucleotide polymorphism arrays and whole-genome sequences for clinical mastitis traits in dairy cattle, J. Dairy Sci. 97 (2014) 7258–7275, https://doi.org/10.3168/jds.2014-8141.

[41] H.G. Olsen, T.M. Knutsen, A.M. Lewandowska-Sabat, H. Grove, T. Nome, M. Svendsen, M. Arnyasi, M. Sodeland, K.K. Sundsaasen, S.R. Dahl, B. Heringstad, H.H. Hansen, I. Olsaker, M.P. Kent, S. Lien, Fine mapping of a QTL on bovine chromosome 6 using imputed full sequence data suggests a key role for the group-specific component (GC) gene in clinical mastitis and milk production, Genet. Select. Evol. 48 (2016) 79, https://doi.org/10.1186/s12711-016-0257-2.

[42] Y.-L. Lee, H. Takeda, G. Costa Monteiro Moreira, L. Karim, E. Mullaart, W. Coppieters, GplusE Consortium, R. Appeltant, R.F. Veerkamp, M.A.M. Groenen, M. Georges, M. Bosse, T. Druet, A.C. Bouwman, C. Charlier, A 12 kb multi-allelic copy number variation encompassing a GC gene enhancer is associated with mastitis resistance in dairy cattle, PLoS Genet. 17 (2021) e1009331, https://doi.org/10.1371/journal.pgen.1009331.

[43] J. Pösö, E.A. Mäntysaari, Relationships between clinical mastitis, somatic cell score, and production for the first three lactations of Finnish Ayrshire, J. Dairy Sci. 79 (1996) 1284–1291, https://doi.org/10.3168/jds.S0022-0302(96)76483-4.

[44] E.E. Smith, H.S. Malik, The apolipoprotein L family of programmed cell death and immunity genes rapidly evolved in primates at discrete sites of host-pathogen interactions, Genome Res. 19 (2009) 850–858, https://doi.org/10.1101/gr.085647.108.

[45] S.G. Narayana, E. de Jong, F.S. Schenkel, P.A.S. Fonseca, T.C.S. Chud, D. Powell, G. Wachoski-Dark, P.E. Ronksley, F. Miglior, K. Orsel, H.W. Barkema, Underlying genetic architecture of resistance to mastitis in dairy cattle: a systematic review and gene prioritization analysis of genome-wide association studies, J. Dairy Sci. 106 (2023) 323–351, https://doi.org/10.3168/jds.2022-21923.

[46] Z. Brajnik, J. Ogorevc, Candidate genes for mastitis resistance in dairy cattle: a data integration approach, J. Anim. Sci. Biotechnol. 14 (2023) 10, https://doi.org/10.1186/s40104-022-00821-0.
[47] T.E. Galdino Andrade, M. Scavassini Peña, J. Fiorotti, R. de Souza Bin, A. Rodrigues Caetano, T. Connelley, I.K. Ferreira de Miranda Santos, Graduate student literature review: the DRB3 gene of the bovine major histocompatibility complex-discovery, diversity, and distribution of alleles in commercial breeds of cattle and applications for development of vaccines, J. Dairy Sci. 107 (2024) 11324–11341, https://doi.org/10.3168/jds.2023-24628.
[48] S.-J. Peng, X.-K. Cao, D. Dong, M. Liu, D. Hao, X.-M. Shen, Y.-Z. Huang, C.-Z. Lei, Y. Ma, Y.-Y. Bai, L.-Y. Hu, X.-L. Qi, B. Chaogetu, H. Chen, Integrative analysis of APOL3 gene CNV for adult cattle stature, Anim. Biotechnol. 31 (2020) 440–446, https://doi.org/10.1080/10495398.2019.1615933.
[49] T.J. Lopdell, K. Tiplady, C. Couldrey, T.J.J. Johnson, M. Keehan, S.R. Davis, B. L. Harris, R.J. Spelman, R.G. Snell, M.D. Littlejohn, Multiple QTL underlie milk phenotypes at the CSF2RB locus, Genet. Sel. Evol. 51 (2019) 3, https://doi.org/10.1186/s12711-019-0446-x.
[50] The CattleGTEx atlas reveals regulatory mechanisms underlying complex traits, Nat. Genet. 54 (2022) 1273–1274, https://doi.org/10.1038/s41588-022-01155-3.
[51] M.T. Parker, S. Amar, J.A. Campoy, K. Krause, S. Tusso, M. Marek, B. Huettel, K. Schneeberger, Scalable eQTL mapping using single-nucleus RNA-sequencing of recombined gametes from a small number of individuals, PLoS Biol. 23 (2025) e3003085, https://doi.org/10.1371/journal.pbio.3003085.
[52] A. Chegini, I. Strandén, E. Karaman, T. Iso-Touru, J. Pösö, G.P. Aamand, M. H. Lidauer, Marker weighting improves single-step genomic prediction reliabilities of udder health traits in Nordic Red and Jersey dairy cattle populations, J. Dairy Sci. 108 (2025) 651–663, https://doi.org/10.3168/jds.2024-25374.