Received: 13 December 2021 | Accepted: 2 August 2022

DOI: 10.1111/jbg.12737

Journal of Animal Breeding and Genetics WILEY

ORIGINAL ARTICLE

Genome-wide scan for selective sweeps identifies novel loci associated with resistance to mastitis in German Holstein cattle

Bita Abbasi-Moshaii[1,2] | **Mohammad Hossein Moradi**[3] | **Tong Yin**[1] | **Ghodratollah Rahimi-Mianji**[2] | **Ardeshir Nejati-Javaremi**[4] | **Sven König**[1]

[1]Institute of Animal Breeding and Genetics, Justus-Liebig University Giessen, Giessen, Germany

[2]Department of Animal Science and Fisheries, Sari Agricultural Science and Natural Resources University, Sari, Iran

[3]Department of Animal Sciences, Arak University, Arak, Iran

[4]Department of Animal Science, University College of Agriculture and Natural Resources, University of Tehran, Karaj, Iran

Correspondence
Sven König, 1Institute of Animal Breeding and Genetics, Justus-Liebig University Giessen.
Email: sven.koenig@agrar.uni-giessen.de

Funding information
Förderverein Bioökonomieforschung e.v. (FBV) / German Holstein association (DHV); German Federal Ministry of Education and Research (BMBF); Federal Ministry of Education and Research

Abstract

Domestication and selection significantly changed phenotypic and behavioural traits in modern domestic animals. In this study, to identify the genomic regions associated with mastitis, genomic data of German Holstein dairy cattle were analysed. The samples were genotyped using the *Bovine 50 K SNP chip*. For each defined healthy and sick group, 133 samples from 13,276 genotyped dairy cows were selected based on mastitis random residual effects. Grouping was done to infer selection signatures based on XP-EHH statistic. The results revealed that for the top 0.01 percentile of the obtained XP-EHH values, five genomic regions on chromosomes 8, 11, 12, 14 and 26 of the control group, and four regions on chromosomes 3, 4 (two regions) and 22 of the case group, have been under selection. Also, consideration of the top 0.1 percentile of the XP-EHH values, clarified 21 and 15 selective sweeps in the control and case group, respectively. This study identified some genomic regions containing potential candidate genes associated with resistance and susceptibility to mastitis, immune system and inflammation, milk traits, udder morphology and different types of cancers. In addition, these regions overlap with some quantitative trait loci linked to clinical mastitis, immunoglobulin levels, somatic cell score, udder traits, milk fat and protein, milk yield, milking speed and veterinary treatments. It is noteworthy that we found two regions in the healthy group (on chromosomes 12 and 14) with strong signals, which were not described previously. It is likely that future research could link these identified genomic regions to mastitis. The results of the current study contribute to the identification of causal mutations, genomic regions and genes affecting mastitis incidence in dairy cows.

KEYWORDS

German Holstein cattle, mastitis, selection signatures, XP-EHH statistic

1 | INTRODUCTION

Animal selection left detectable signatures on the genome of modern cattle. Identification of these signals can help us to genetically improve species in terms of economically important traits (Qanbari et al., 2014). The genomic regions that are or have been under selection (natural or artificial) due to harbouring a beneficial mutation carry some genomic signs that are known as "selection signatures" (Sabeti et al., 2002). Selection signatures studies can provide invaluable information about the genes or genomic regions that have been under selection, contributing to a deeper understanding of genotype–phenotype relationships (Gholami, 2014; Moradi et al., 2012). "Neutral theory" in this context indicates that most of the mutations in the genome are not selectively important, without any effects on individual's genetic fitness (Kimura, 1989). Hence, a substantial increase in the allele frequency for new variants needs a long time, and recombination will break down linkage disequilibrium (LD) in the respective chromosome segments. Oppositely, if a new mutation makes a subset of the population more qualified than others, the allele frequency will increase rapidly by natural or artificial selection. In such case, recombination cannot decay LD in short time. Thus, an allele with an unusually long-range haplotype will be a signature of positive selection (Sabeti et al., 2002). Around this new allele, genetic variety decreases and LD increases. Hence, it is expected that a specific haplotype carrying some genes is unique in a population (Akey et al., 2002; Barrett & Schluter, 2008; Qanbari & Simianer, 2014).

Different statistical approaches have been designed to identify selective sweeps in the genome, based on the neutral theory predictions as null hypothesis. Departures from equilibrium-neutral expectations can indicate the presence of selection pressure on chromosomal segments under investigation or at tightly linked sites (Kreitman, 2000). The statistical power of these methods to detect selection signals depends on the allele or haplotype frequency, age and objectives of selection, SNP density and number of genotyped individuals (Sabeti et al., 2006). The cross-population extended haplotype homozygosity (XP-EHH) test detects selection footprints on the genome based on LD, i.e., considering the non-random association of alleles between two or more loci. This test searches for haplotypes with high-frequency and long-range haplotypes, whereas haplotype length is measured by the extended haplotype homozygosity (EHH). XP-EHH compares the integrated EHH between two populations at the same SNP and detects selected alleles, which have approached or achieved fixation in one population while remaining polymorphic in the populations as a whole (Cheruiyot et al., 2018; Diao et al., 2019; Sabeti, et al., 2007). Many researchers successfully applied this test and detected selection signals in different species such as pigs, sheep, horse and cattle (Chen et al., 2016; Diao et al., 2019; Manzari et al., 2019; Nolte et al., 2019).

Mastitis (OMIA 001744–9913, https://omia.org/) is one of the most important infectious diseases in lactating animals, resulting from the intrusion and proliferation of pathogenic agents. An obvious symptom of mastitis usually is an inflammation of the mammary gland. There is remarkable concern addressing farm profitability, due to involuntary culling of sick animals, quantity and quality of milk production, the costs of veterinary treatments, animal welfare, the entrance of antibiotic residues into milk and environmental aspects. The genetically antagonistic relationship between milk yield and udder health implies that selection on increasing milk yield may cause an increase in susceptibility to mastitis (Ebrahimie et al., 2018). Several factors such as invading organisms, the subsequent immune response to eliminate pathogens, inflammatory response and animal genetics influence mastitis severity and susceptibility (Izquierdo et al., 2017; Pighetti & Elliott, 2011; Tiezzi et al., 2015). Hence, it is expected that most of the identified genes of healthy animals be directly or indirectly related to immunity and resistance to mastitis, whereas some genes of the sick group may be involved in milk yield and other production traits.

Consequently, the aim of the present study was to reveal genomic regions, which are divergently selected in healthy and sick groups. Annotated potential candidate genes in the respective chromosomal segments were studied in detail, including their functions, biological pathways and networks. Both aspects in combination, i.e., identified selection signatures and associated candidate genes, will shed some light on the genetic architecture of this disease in cattle.

2 | MATERIALS AND METHODS

The data used in the present study were obtained from a project initiated by the German Holstein breeding organizations, focusing on the implementation of a cow training set for national genetic evaluations. Hence, committee approval for data recording including cow genotypes and cow traits was not necessary prior to the study.

2.1 | Genotyping data and quality control

In this project, 15,405 dairy cows from 72 large-scale herds were genotyped with the *BovineSNP50 BeadChip*. Quality control of SNP data was carried out using the software package Plink 1.07 (Purcell et al., 2007). For this purpose,

animals with more than 5% missing genotypes (animal call rate) were excluded. Afterwards, an SNP minor allele frequency (MAF) of at least 2%, an SNP call rate of at least 95% and Hardy–Weinberg disequilibrium with $p < 10^{-6}$ were considered for data mining at SNP level.

2.2 | Classification of animals to control and case groups

In the current study, the genomic information from 13,276 first-lactation dairy cows were used to investigate the selection signatures associated with mastitis. Mastitis was recorded according to the definitions of the hierarchical diagnosis key as introduced by Stock et al. (2013). All Holstein cows with at least one entry for mastitis (clinical or subclinical) (Baeza, 2016; Mbindyo et al., 2020) were allocated to the case (sick) group. On the contrary, cows without any entry for mastitis, were assigned in the control (healthy) group. The genetic background and genetic parameters for mastitis are outlined by Shabalina et al. (2021), who used the same cows from the same herds in their study. According to Bagheri et al. (2013) and Naderi et al. (2018), the most susceptible and the most resistant groups were selected based on mastitis random residual effects obtained from a single-trait animal model including herd and year-season of calving as the fixed effects through the application of the DMUv6 software package (Madsen & Jensen, 2013). Based on the extreme values for estimated residual effects, a fraction of 1% of cows representing the most negative values was defined as healthy, and a fraction of 1% of cows representing the most positive values was allocated to the sick group.

2.3 | Creation of haplotypes

The creation of haplotypes is implemented in the fastPHASE v.1.4 software package (Scheet & Stephens, 2006). This program imputes missing genotypes and defines haplotype phases based on maximum likelihood estimations. The results include a pair of created haplotypes for each chromosome per individual.

2.4 | XP-EHH calculation

The final analysis using phased haplotype data was carried out by applying the XP-EHH program at http://hgdp.uchicago.edu. Input files include map and haplotype files. For the XP-EHH calculation, the ancestral allele state of each SNP must be specified. Ancestral alleles were extracted from http://genome.jouy.inra.fr/downloads/Bovine_Ancestral_Allele/ and prepared by applying the R 3.5.1 software package.

EHH calculates haplotype diversity reduction. For a bi-allelic SNP, EHH is calculated as follows (Sabeti et al., 2002):

$$\mathrm{EHH} = \frac{\sum_{i=1}^{h_x} \binom{n_i}{2}}{\binom{n_a}{2}\binom{n_A}{2}}$$

where, n_A and n_a represents the number of haplotypes with alleles A and a, respectively; n_i is the count of the ith haplotype within a sub-population, and h_x is the number of distinct haplotypes in a genomic region up to a distance x from the core locus. XP-EHH was calculated for each SNP (SNP i) in both populations (i.e., the extreme groups) A and B. In each population, the integrated EHH for all SNPs between the core SNP and point X was assigned as I_A and I_B. Finally, unstandardized XP-EHH was calculated by $ln\frac{I_A}{I_B}$ (see Sabeti et al., 2007 for more details). These values were standard normalized using the mean and standard deviation of the obtained values. Afterwards, Manhattan plots were created using the Excel 2013 software package. The extreme positive values represent the regions that have been under selection in population A (= the healthy group), and extreme negative values indicate the selective sweeps in population B (= the sick group).

2.5 | Gene and QTL annotation, functional and network analysis

The bioinformatics investigations were carried out using the Ensembl database (Cunningham et al., 2022) for cow genes (ARS-UCD1.2), to identify potential candidate genes which already have been reported in/or surrounding genomic regions containing the peak of absolute extreme XP-EHH values. The regions corresponding to the upper and lower 0.01% of positive and negative obtained XP-EHH scores were considered as regions under selection. Gene detection was performed by introducing a window of 500 Kbp downstream and upstream of each core SNP (Moradi, 2012).

The DAVID 6.8 database (Huang et al., 2009) and QTL Animal Database (Hu et al., 2022) were utilized to search for biological processes and quantitative trait loci (QTLs), respectively.

Furthermore, gene networks for annotated genes in the top 0.1 percentile were reconstructed by applying the STRING database v11.0 (Jensen et al., 2009). The basic interaction unit in STRING is the "functional association," which is defined as a specific and meaningful interaction between two proteins that jointly contribute to the same

functional process. To reach the goal, some approaches in STRING were used including neighbourhood, gene fusion, co-occurrence, co-expression, text mining, databases and experiment analyses (Jensen et al., 2009). For accuracy improvement of defined gene relationships, high confidence (0.7) was elected (Laodim et al., 2019). Finally, engaged biological pathways in gene networks were examined ($p < 0.05$).

3 | RESULTS

Among the 13,276 genotyped and phenotyped Holstein cows in first lactation, 266 cows (133 sick animals and 133 healthy animals) were used for ongoing genomic analyses. After quality control and checking for ancestral alleles, a total of 43,678 SNP markers from the 266 cows were considered (Table 1).

3.1 | Genomic distribution of XP-EHH values in the German Holstein dataset

XP-EHH was calculated for all SNPs across the genome of sick and healthy groups to detect the genomic regions that might be targets of recent selections (Figure 1).

As shown in Figure 1, some SNPs tend to make a cluster with outlier SNPs in similar regions due to genetic hitchhiking. For the chosen threshold considering the top 0.1 percentile of negative and positive values, we identified some more candidate regions containing suggestive SNPs on chromosomes 3, 4 (2 regions), chromosomes 6, 7, 8 (2 regions), chromosomes 9, 10, 11, 12 (2 regions), chromosomes 13, 14 (2 regions), chromosome 15 (2 regions), chromosome 24 (2 regions), chromosomes 26, 28 for the healthy group, and some other regions on chromosomes 3, 4 (6 regions), chromosomes 6, 8 (2 regions), chromosome 9 (3 regions), chromosomes 14, 22 embracing selection signatures in the sick group. The corresponding XP-EHH values were larger than 5.198 for the healthy group, and smaller than −8.324 for the sick group.

The targets of recent selection also were detected as significant SNPs, considering a rigorous threshold for the top 0.01 percentile of negative or positive scores of XP-EHH (Moradi et al., 2010). These regions were located on chromosomes 8, 11, 12, 14 and 26 for the healthy, and on chromosomes 3, 4 (2 regions) and 22 for the sick group. The corresponding threshold values were + 8.475 for the healthy group and − 13.217 for the sick group. Candidate regions included one significant SNP on chromosome 8 (rs41609496), two SNPs on chromosome 11 (rs43671283, rs43672234), one SNP on chromosome 12 (rs43705112), one SNP on chromosome 14 (rs41737187) and one SNP on chromosome 26 (rs42096776). Further investigations of these regions harbouring selective sweeps revealed 21 genes, which already have been reported for these genomic positions. For the sick group, one significant SNP on chromosome 3 (rs43349513), two SNPs on chromosome 4 (rs43408665, rs43375405) and one SNP on chromosome 22 (rs41643918) harbouring 20 genes, were identified as well. The genes of the top 0.01 percentile are presented in Table S1 for healthy and sick animals.

Candidate regions were surveyed for QTL in the online animal QTL database. The results show that most regions of interest encompass QTLs which have been reported for clinical mastitis, immunoglobulin functions, somatic cell score, udder swelling score, milk fat and protein content, 305-day milk yield, milking speed and veterinary treatments (Table S1).

3.2 | Gene ontology enrichment analysis

For the gene sets found in the detected genomic regions of the healthy and sick animals according to the top 0.1 percentile, the GO terms including biological processes, cellular component and molecular functions, were surveyed using the DAVID database (Huang et al., 2009).

TABLE 1 Summary of data mining steps on genotyping data in the two populations of healthy and sick cows

	Healthy	Sick
Total number of animals	133	133
Excluding animal with call rate ≤ 95%	0	0
Remaining animals	133	133
Total number of SNPs[a]	45,613	45,613
Excluding SNPs with MAF≤2%	1180	1180
Excluding SNPs with call rate ≤ 95%	0	0
Excluding SNPs with deviation from HWE ($<10^{-6}$)	2	2
SNPs remaining	44,431	44,431
SNPs remaining with ancestral allele	43,678	43,678

[a]The number of SNPs used in the national genetic evaluations (VIT, Germany).

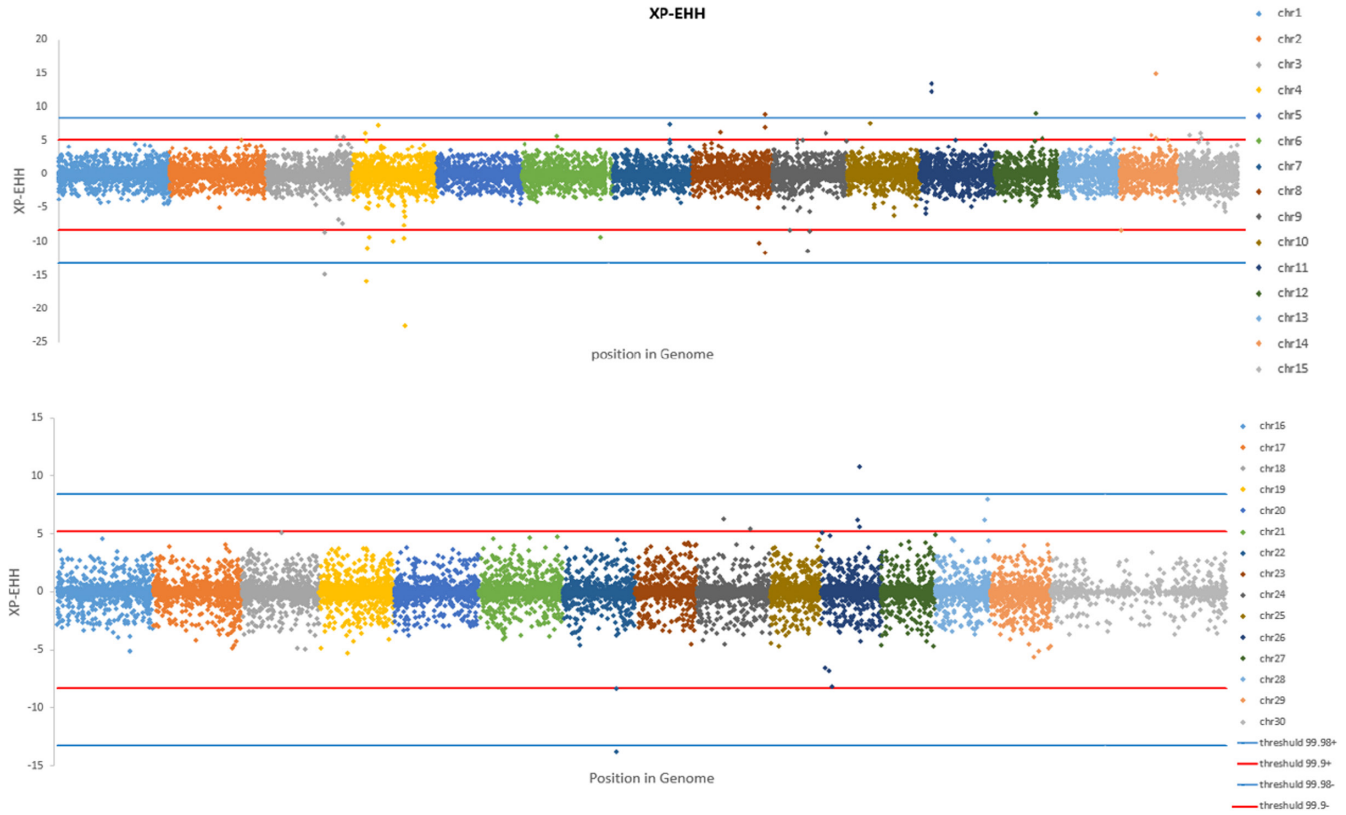

FIGURE 1 Distribution of XP-EHH values across the genome of German Holstein cows. The SNP position in the genome is shown on the x-axis, and XP-EHH values are plotted on the Y-axis. The negative values indicate selection in sick and positive values indicate selection in healthy animals. The values above the lines are in the top 0.01 (blue line) and the top 0.1 (red line) percentile. [Colour figure can be viewed at wileyonlinelibrary.com]

TABLE 2 Biological pathways in candidate regions by XP-EHH analysis in healthy animals

GO term	Term	*p*-value	Count
GO:0000303	response to superoxide	2.8E-3	4
GO:0097021	lymphocyte migration into lymphoid organs	3.0E-3	3
GO:0000305	response to oxygen radical	3.2E-3	4
GO:0010862	positive regulation of pathway-restricted SMAD protein phosphorylation	4.2E-3	5
GO:0010939	regulation of necrotic cell death	6.7E-3	4
GO:0032101	regulation of response to external stimulus	8.3E-3	18
GO:0060389	pathway-restricted SMAD protein phosphorylation	9.6E-3	5
GO:0048731	system development	1.8E-2	67
GO:0006396	RNA processing	1.8E-2	21
GO:0006979	response to oxidative stress	2.2E-2	10
GO:0050728	negative regulation of inflammatory response	2.4E-2	5
GO:0050727	regulation of inflammatory response	3.2E-2	9
GO:0019216	regulation of lipid metabolic process	4.4E-2	9
GO:0031347	regulation of defence response	4.5E-2	13
GO:0006636	unsaturated fatty acid biosynthetic process	4.9E-2	4

In the healthy animals, the main identified biological processes were related to the regulation of response to external stimulus, regulation of defence response, negative regulation of inflammatory response, regulation of pathway restricted SMAD protein, RNA processing and response to superoxide and oxygen radicals (Table 2). For sick animals, significant pathways including response to stimulus, regulation of signalling, regulation of biological processes and morphogenesis of an epithelial bud, were detected using a *p-value* <0.05 (Table 3).

The protein–protein interaction networks in sick and healthy groups are shown in Figures 2 and 3. Among all genes entered into interaction networks, some of them were connected. The biological pathways and functions of the genes involved in networks were surveyed by the use of GeneCards (Safran et al., 2022) and Gene Ontology ($p < 0.05$) (Ashburner et al., 2000), and considering results of past studies.

4 | DISCUSSION

Nowadays, mastitis is the main economically important disease in dairy cattle, possibly due to intense selection for milk production during the recent decades (Loor et al., 2011). In this study, an initial selection map was developed via whole genome scanning for selection signatures associated with mastitis in German Holstein dairy cattle. Results were verified by applying bioinformatics surveys including gene function, gene ontology and gene network analysis.

4.1 | The function of the identified genes in detected regions

Some regions embracing selection signals were detected through the top XP-EHH values. Afterwards, the genes reported in these regions were evaluated. Bioinformatics investigations discovered that functions of the identified genes in both healthy and sick groups address immune system mechanisms, diseases susceptibly for mastitis, productivity (milk yield, milk fat and protein components) as well as different types of cancers such as breast cancer and apoptosis. Results reflect divergent pressure of selection with respect to higher production and animal health. The immune system is regulated by several thousands of genes, all contributing with its effects on the genetic priority of immunity as an overall fitness trait (Mallard et al., 2015). Nowadays, a challenge is the reduction of antibiotic treatments by improving the natural resistance ability of animals against the infection. This strategy implies a better understanding of the host immune response at the early phases of infection (Brenaut et al., 2014). Healthier animals demonstrate superior immune responses, and adaptive immune response traits are heritable. Hence, identifying important chromosomal regions in this regard for selective breeding on improved immunity will improve the animal health status. In dairy cows with superior antibody-mediated and cell-mediated immune responses, incidences of many diseases including mastitis were on a quite low level (Moravčíková et al., 2018).

In the present study, the segment on chromosome 8 where the SNP rs41609496 is located, includes the genes *AKNA, MIR455, ORM1, GNG10* with functions on immunity. Previous studies indicated the key function of the *AKNA* (*AT-hook transcription factor*) gene with regard to immunity, inflammation and overall body development regulations (Moliterno & Resar, 2011; Resar, 2010). Enrichment analysis inferred the strong association between *GNG10* genotypes and *Staphylococcus aureus* infection, neutrophil activation and degranulation, neutrophil-mediated immunity and defence response (Shao et al., 2021).

MIR455 (*MicroRNA 455*), as an RNA gene, is affiliated with the miRNA class (GeneCards) and modulates epithelial immune mechanisms in the innate immune network (Zhou et al., 2011). MicroRNAs play an important role in regulating responses to infection (Eulalio et al., 2012). In the research by Li et al. (2015), the expression of *MIR455* in the infected mammary gland tissue was significantly down-regulated when compared to the control samples.

Orosomucoid 1 (*ORM1*) is one of the identified genes on chromosome 8 that is also known as alpha 1 acid

TABLE 3 Biological pathways in candidate region by XP-EHH analysis in sick animals

GO term	Term	*p*-value	Count
GO:0051716	cellular response to stimulus	2.0E-2	42
GO:0023051	regulation of signalling	2.5E-2	23
GO:0050789	regulation of biological process	3.4E-2	63
GO:0060572	morphogenesis of an epithelial bud	3.9E-02	2
GO:0050896	response to stimulus	4.6E-2	48
GO:0050794	regulation of cellular process	4.8E-2	59
GO:0060603	mammary gland duct morphogenesis	8.3E-02	2

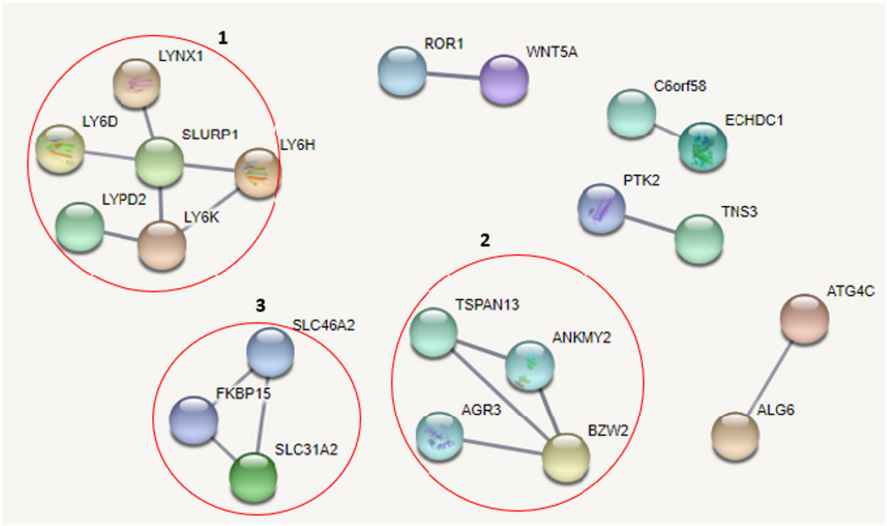

FIGURE 2 Gene network related to identified genes in candidate regions (99.9%) of sick animals. Coloured nodes and blue edges represent query proteins (genes) and protein–protein associations of genes jointly contributing to a shared function, respectively. The numbers indicate different functional clusters of a gene network. [Colour figure can be viewed at wileyonlinelibrary.com]

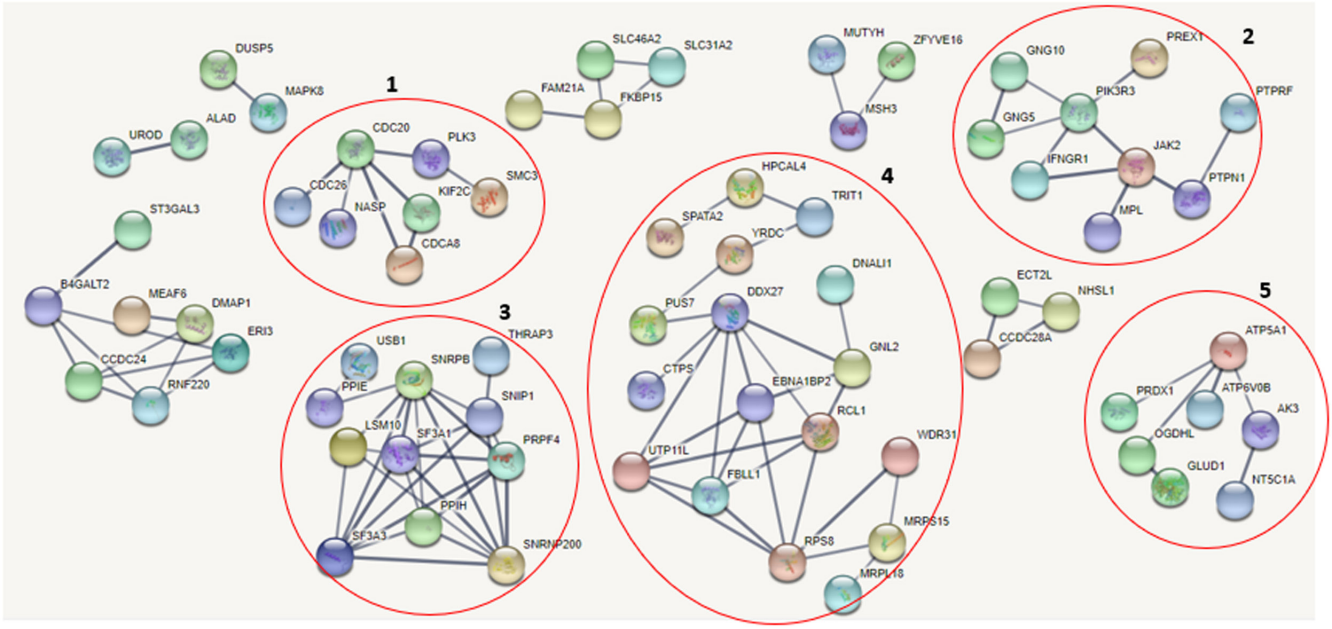

FIGURE 3 Gene network related to identified genes in candidate regions (99.9%) of healthy animals. Coloured nodes and blue edges represent query proteins (genes) and protein–protein associations of genes jointly contributing to a shared function, respectively. The numbers indicate different functional clusters of a gene network. [Colour figure can be viewed at wileyonlinelibrary.com]

glycoprotein A (*AGP-A*). This gene encodes a key acute phase protein of plasma with anti-inflammatory and immune-modulating properties. Gemelli et al. (2013) identified an association between *ORM1* gene expressions and vitamin D3 levels. Vitamin D3 concentrations in the mammary gland were associated with the response of the immune system against udder infections due to *Streptococcus uberis* (Lippolis et al., 2011).

Infected mammary glands affected milk yield and its composition, such as remarkable high concentrations of whey protein and albumin, an increase in fat content and a reduction in lactose percentage (Kobayashi et al., 2016). *LTBP1* and *GPAM* contribute to the synthesis of unsaturated fatty acids (Lung et al., 2019) and to the fatty acid metabolic process (GO: 0006631) (Yang et al., 2015), respectively. The function of *ACSL5* in fatty acids synthesis and *UGCG* in milk fat synthesis was reported by Wathes et al. (2021) and Mumtaz et al. (2022). *ACSL5* has a central function in the occurrence of ketosis and is associated with negative energy balance (NEB) (Soares et al., 2021; Wathes et al., 2021) which occurs in early lactation, especially in cows with high genetic merit for milk production. Non-esterified fatty acids are one of the metabolic indicators for cows in NEB (de Vries & Veerkamp, 2000; Kaneene et al., 1997) and are related to diminished immune response. Hence, the risk of common infectious

diseases such as mastitis will increase (Kaupe et al., 2007; Loor et al., 2011; Naderi et al., 2018).

Among the identified genes in the case group, *PGM1* and *ARL4A* are related to milk production traits such as lactose synthesis and glucose metabolism (Lemay et al., 2013), milk yield and/or composition (Raschia et al., 2018). When the mammary glands are exposed to mastitis infections, inflammatory cytokines decrease the expression of *PGM1* in mammary epithelial cells, consequently causing a down-regulation of lactose synthesis in mammary glands (Kobayashi et al., 2016). In addition, *PGM1* plays a role in inverted teat defections, triggering the risk of mastitis infections (Chomwisarutkun et al., 2012). Lashneva et al. (2021) mentioned *ROR1* on chromosome 3 as a causal gene for somatic cell count, which is close to the identified significant SNP (rs43349513) in the case group of this study.

Furthermore, some other genes such as *RASGRP3* on chromosome 11 are involved in immune system mechanisms, cytokine signalling and B-cell receptor signalling (Li et al., 2019; Parker Gaddis et al., 2018), *CFAP69, STEAP2, ITGB3BP* (chromosome 4) are related to immune system mechanisms and stimulate apoptosis in breast cancer cells (Gomes et al., 2012; van Roosmalen et al., 2015; Yang et al., 2017), *FAM98A* (chromosome 11) has a partial role in the organization of stress granules (Ozeki et al., 2019), which is a factor in tumour proliferation (Zheng et al., 2018).

Previous studies reported some genes and genomic regions as identified in the present study. Gutiérrez-Gil et al. (2015) indicated that the genes *TNS3*, *HUS1*, *GTPBP10* and *STEAP2* are located in genomic regions under selection in dairy breeds. Mokhber (2015) and Fu et al. (2016) introduced *CSMD3* as a candidate gene under selection in Iranian buffaloes and broiler chicken. Korkuć et al. (2021) stated that *PTGR1* on chromosome 8 at 101 Mb is associated with enzymes or receptors targeting milk components. Likewise, using selective sweep approaches, *TNS3, HUS1, GTPBP10, CFAP69* and *STEAP2* (74.9–75.9 Mb on chromosome 4) were reported as candidate genes by Mokhber (2015) and *ACSL5* as a candidate gene by Soares et al. (2021). The same genes and regions we identified on chromosomes 4 and 26 in this study, respectively.

The QTLs in the proximity of the detected SNPs in the control group (rs41609496, rs43671283, rs43672234, rs43705112, rs41737187, rs42096776) are relevant for udder swelling (chromosome 8), clinical mastitis, immunoglobulin G level and somatic cell count (chromosome 11), 305-day milk yield, milk fat and protein yield (chromosome 26). Previous studies have detected the same QTLs affecting clinical mastitis and milk yield on chromosomes 11 and 26 (Holmberg & Andersson-Eklund, 2004; Lund et al., 2008).

Likewise, we clarified some QTLs detected in the case group for somatic cell score and somatic cell count (chromosomes 3 and 4), milk yield and milking speed (chromosome 3), immunoglobulin G level (chromosome 4), milk protein percentage and veterinary treatments (chromosome 22) were reported in previous studies. For example, Schulman et al. (2004) and Tribout et al. (2020) identified QTLs associated with veterinary treatments and milk protein content on chromosome 22, and Cohen-Zinder et al. (2011) reported QTLs for milk yield on chromosome 3 as well.

In the current study, we addressed the *CSMD3* gene on chromosome 14 (51,566,496 bp), but we did not infer related QTLs. However, Lund et al. (2007) applied single-trait analyses, and detected QTLs for SCS on chromosome 14 at 51 cM. Also, the detected region (56,601,235 bp) on chromosome 12 (with a strong signal of selection in the healthy group) has not been reported in previous studies, and there is no information addressing the existing genes in this region. Therefore, it can be valuable to study these segments in more detail as novel candidate gene regions associated with resistance to mastitis.

4.2 | Biological pathways

RNA processing is one of the detected biological pathways linked to the genes of healthy animals. Ansel (2013) announced that RNA processing plays essential role in the development and functions of the mammalian immune system. Inflammatory response acts as a quick defence in reaction to infection or injuries due to chemical or physical agents (Bannerman, 2009). In this way, after pathogen controlling and deletion, the immune response should be controlled to avoid ongoing tissue damage and chronic inflammation (Vigano et al., 2012). Therefore, negative regulation of inflammatory response will be activated which involves any process that prevents or decreases the inflammatory response rate (Bult et al., 2019).

Response to superoxide and oxygen radical are the other annotated pathways in the healthy group. High production of reactive oxygen species (ROS) causes oxidative stress, and in further consequence, inflammation. Under some conditions, ROS is released: (i) As mentioned earlier, there is a relationship between the risk for a mastitis infection and milk yield. Particularly in high-yielding cows, excessive production of ROS due to metabolic stress was observed, which induces inflammation in alveolar cells of the mammary gland (Bae et al., 2017). (ii) In the transition period of the dry period to early lactation owing to metabolic challenge and NEB, the antioxidative status is impaired and both the oxidative stress and inflammatory response prepare predisposing conditions for higher

susceptibility to intramammary infections and mastitis. (iii) When pathogens enter into the udder, polymorphonuclear leukocytes release enzymes, inflammatory mediators and ROS to eliminate bacteria. Excessive production of ROS causes oxidative changes in the cell that induce lipid peroxidation, inhibition of protein function, mutation and cell death (Abd Ellah, 2013; Turk et al., 2017). The relation between ROS production with inflammatory response and mastitis indicates the importance of defined pathways and their function in the pathogenesis of mastitis.

Mammary gland duct morphogenesis was one of the identified pathways in the sick group. If stenosis of ductus papillaris as the first barrier against invading pathogens is damaged, pathogens can enter easily into the udder and cause infection and inflammation (Turk et al., 2017). Moreover, we can point to the importance of epithelial bud morphogenesis in mammary gland development and health (Sternlicht, 2005). Pathway analysis was performed by Ogorevc et al. (2009), considering a cattle database with 943 candidate genes and genetic markers associated with milk production and mastitis. As a result, they identified that some loci are involved in inflammatory response, development and function of connective tissue, muscle development, cell-mediated immune response and cancer.

Pathway-restricted SMAD protein is an effector protein that acts directly downstream of the transforming growth factor family receptor. Malhotra and Kang (2013) indicated the relationship between SMAD regulatory network and immune system mechanisms. In exposure to pathogens, SMAD proteins are activated to initiate a protective inflammatory immune response. Additionally, the SMAD-dependent biological pathway participates in tumour suppression by TGF-beta (Nagaraj & Datta, 2010).

4.3 | Gene networks

As depicted in Figures 2 and 3, some networks of the candidate genes under selection pressure, directly or indirectly, are related to mastitis. In sick animals, three networks were detected. Most of the genes implicates in: cluster (1) cancer and progression of inflammation, cluster (2) regulation of cell development and breast cancer and cluster (3) cytosolic immune recognition (GeneCards; Paik et al., 2017). In healthy animals, five networks were identified with the main functions as outlined in the following. Cluster (1) cell cycle progression, stress responses, tumour-suppressing activity, immunity and overall survival (Li et al., 2017; Xu et al., 2010). A significant effect on immune response was reported by Behdani and Bakhtiarizadeh (2017) for *CDC20* due to its influence on the proliferation of immune cells. *SMC3* is a gene in this network that has a role as a regulator in the differentiation and activation of B cell function (Zhao et al., 2021). The cluster (2) group involves *JAK2* which is directly related to mastitis resistance in bovine (Tiezzi et al., 2015). Other genes in this category implicate in the innate immune system, mammary development and milk secretion (Gene Ontology; Chen et al., 2018). Cluster (3) pre-mRNA processing and development, breast cancer, regulation of lipid metabolism and immune system (Cieśla et al., 2021; Godfrey et al., 2009). Meier et al. (2020) identified *PPIE* as a gene related to immunity and mastitis. They also cited its influence on the adaptive immune system and its responsibility for immune system activation in cattle. Cluster (4) RNA processing, immune system, (Gene Ontology; GeneCards). In the cluster (5) group, *PRDX1* plays an antioxidant protective role in cells and may contribute to the antiviral activity of CD8(+) T-cells (GeneCards). *ATP6V0B* implicates in innate immune response (Wee et al., 2012) and *OGDHL* suppresses tumours indirectly through apoptosis activation (Maijaroen et al., 2022).

Ibeagha-Awemu et al. (2010) focused on proteomics, genomics and pathway analyses on infected milk whey with *S. aureus* and *Escherichia coli*. They revealed some molecular pathways and networks involved in mastitis such as lipid metabolism, cancer and immunological diseases, supporting the results from the present study. Ogorevc et al. (2009) utilized a database of candidate genes and genetic markers related to milk production and mastitis. They clustered genes of the mammary gland into five functional networks linked to a variety of biological functions such as mammary gland phenotypes, milk and mastitis traits. The results of network analysis by Zhao et al. (2015) indicated the differentially expressed proteins in cows infected with *E. coli* are related to mammary epithelial cells, immune defence and structural molecule activities.

5 | CONCLUSION

Our research provides a selection map to survey the genetic architecture of mastitis in the genome of German Holstein cows. To reach this goal, the XP-EHH approach was used and most of the regions with a strong signal of selection in both groups represented a relationship with mastitis. These detected loci have likely been influenced by the different intensities of selection to achieve various commercial breeding goals. Although some inferred candidate regions overlap with reports of previous studies, most of the identified regions were detected for the first time suggesting novel loci connected to mastitis. Given that the LD-based methods depend on ancestral allele information, frequency and the distance among SNPs,

utilizing high-density marker arrays chiefly full sequencing data, will bring additional value to the overall topic. Nevertheless, the information generated from this study may shed new light on the genomic regions associated with mastitis in dairy cows.

ACKNOWLEDGEMENTS

The authors acknowledge funding from the German Federal Ministry of Education and Research (BMBF) and from the Förderverein Bioökonomieforschung e.v. (FBV)/ German Holstein association (DHV) for the collaborative project "KMU-innovativ-10: Kuh-L—cow calibration groups for the implementation of selection strategies based on high-density genotyping in dairy cattle," grant no. 031A416C.

CONFLICT OF INTEREST

There are no known conflicts of interest associated with this publication.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available in [Abbasi et al., 2021] at [https://doi.org/10.22029/jlupub-273].

ORCID

Bita Abbasi-Moshaii https://orcid.org/0000-0002-6929-8656
Sven König https://orcid.org/0000-0003-4226-3696

REFERENCES

[Dataset] Abbasi Moshaii B., Moradi M. H., Rahimi-Mianji, Gh, Nejati-Javaremi, A., & König, S. (2021) Data for "Genome-wide scan for selective sweeps identifies novel loci associated with resistance to mastitis"; JLUpub; https://doi.org/10.22029/jlupub-273.

Abd Ellah, M. R. (2013). Role of free radicals and antioxidants in mastitis. *Journal of Advanced Veterinary Research*, *3*(2013), 1–7.

Akey, J. M., Zhang, G., Zhang, K., Jin, L., & Shriver, M. D. (2002). Interrogating a high-density SNP map for signatures of natural selection. *Genome Research*, *12*, 1805–1814. https://doi.org/10.1101/gr.631202

Ansel, K. M. (2013). RNA regulation in the immune system. *Immunological Reviews*, *253*(1), 5–11. https://doi.org/10.1111/imr.12062

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., & Sherlock, G. (2000). Gene ontology: Tool for the unification of biology. *Nature Genetics*, *25*(1), 25–29. https://doi.org/10.1038/75556

Bae, H., Jeong, C. H., Cheng, W. N., Hong, K., Seo, H. G., & Han, S. G. (2017). Oxidative stress-induced inflammatory responses and effects of N-acetylcysteine in bovine mammary alveolar cells. *Journal of Dairy Research*, *84*, 418–425. https://doi.org/10.1017/S002202991700067X

Baeza, C. (2016). Acute, subclinical, and subacute mastitis: Definitions, etiology, and clinical management. *Clinical Lactation*, *7*(1), 7–10. https://doi.org/10.1891/2158-0782.7.1.7

Bagheri, M., Mirai-Ashtiani, R., Moradi-Shahrbabak, M., Nejati-Javaremi, A., Pakdel, A., Von Borstel, U. U., Pimentel, E. C. G., & König, S. (2013). Selective genotyping and logistic regression analyses to identify favorable SNP-genotypes for clinical mastitis and production traits in Holstein dairy cattle. *Livestock Science*, *151*(S2-3), 140–151. https://doi.org/10.1016/j.livsci.2012.11.018

Bannerman, D. D. (2009). Pathogen-dependent induction of cytokines and other soluble inflammatory mediators during intra mammary infection of dairy cows. *Journal of Animal Science*, *87*(13 Suppl), 10–25. https://doi.org/10.2527/jas.2008-1187

Barrett, R. D. H., & Schluter, D. (2008). Adaptation from standing genetic variation. *Trends in Ecology and Evolution*, *23*(1), 38–44. https://doi.org/10.1016/j.tree.2007.09.008

Behdani, E., & Bakhtiarizadeh, M. R. (2017). Construction of an integrated gene regulatory network link to stress-related immune system in cattle. *Genetica*, *145*, 441–454. https://doi.org/10.1007/s10709-017-9980-z

Brenaut, P., Lefèvre1, L., Rau, A., Laloë, D., Pisoni, G., Moroni, P., Bevilacqua, C., & Martin, P. (2014). Contribution of mammary epithelial cells to the immune response during early stages of a bacterial infection to Staphylococcus aureus. *Veterinary Research*, *45*, 16–35. https://doi.org/10.1186/1297-9716-45-16

Bult, C. J., Blake, J. A., Smith, C. L., Kadin, J. A., Richardson, J. E., & the Mouse Genome Database Group. (2019). Mouse genome database (MGD) 2019. *Nucleic Acids Research*, *8*(47), D801–D806. https://doi.org/10.1093/nar/gky1056

Chen, K., Hou, J., Song, Y., Zhang, X., Liu, Y., Zhang, G., Wen, K., Ma, H., Li, G., Cao, B., & An, X. (2018). Chi-miR-3031 regulates beta-casein via the PI3K/AKT-mTOR signaling pathway in goat mammary epithelial cells (GMECs). *BMC Veterinary Research*, *14*, 369–378. https://doi.org/10.1186/s12917-018-1695-6

Chen, M., Pan, D., Ren, H., Fu, J., Li, J., Su, G., Wang, A., Jiang, L., Zhang, Q., & Liu, J. F. (2016). Identification of selective sweeps reveals divergent selection between Chinese Holstein and Simmental cattle populations. *Genetics Selection Evolution*, *48*, 76–87. https://doi.org/10.1186/s12711-016-0254-5

Cheruiyot, E. K., Bett, R. C., Amimo, J. O., Zhang, Y., Mrode, R., & Mujibi, F. D. N. (2018). Signatures of selection in admixed dairy cattle in Tanzania. *Frontiers in Genetics*, *9*, 607–621. https://doi.org/10.3389/fgene.2018.00607

Chomwisarutkun, K., Murani, E., Brunner, R., Ponsuksili, S., & Wimmer, K. (2012). QTL region–specific microarrays reveal differential expression of positional candidate genes of signaling pathways associated with the liability for the inverted teat defect. *Animal Genetics*, *44*, 139–148. https://doi.org/10.1111/j.1365-2052.2012.02378.x

Cieśla, M., Thi Ngoc, P. C., Cordero, E., Martinez, A. S., Morsing, M., Muthukumar, S., Beneventi, G., Madej, M., Munita, R., Jönssen, T., Lövgren, K., Ebbesson, A., Nodin, B., Hedenfalk, I., Jirström, K., Vallon-Christersson, J., Honeth, G., Staaf, J., Incarnato, D., … Bellodi, C. (2021). Oncogenic translation directs spliceosome dynamics revealing an integral role for SF3A3 in breast cancer. *Molecular Cell*, *81*, 1453–1468. https://doi.org/10.1016/j.molcel.2021.01.034

Cohen-Zinder, M., Donthu, R., Larkin, D. M., Kumar, C. G., Rodriguez-Zas, S. L., Andropolis, K. E., Oliveira, R., & Lewin, H. A. (2011). Multisite haplotype on cattle chromosome 3 is associated with quantitative trait locus effects on lactation traits. *Physiological Genomics*, *43*, 1185–1197. https://doi.org/10.1152/physiolgenomics.00253.2010

Cunningham, F., Allen, J. E., Allen, J., Alvarez-Jarreta, J., Amode, M. R., Armean, I. M., Austine-Orimoloye, O., Azov, A. G., Barnes, I., Bennett, R., Berry, A., Bhai, J., Bignell, A., Billis, K., Boddu, S., Brooks, L., Charkhchi, M., Cummins, C., da Rin Fioretto, L., … Flicek, P. (2022). Ensembl 2022. *Nucleic Acids Research*, *50*, D988–D995. https://doi.org/10.1093/nar/gkab1049

de Vries, M. J., & Veerkamp, R. F. (2000). Energy balance of dairy cattle in relation to Milk production variables and fertility. *Journal of Dairy Science*, *83*, 62–69. https://doi.org/10.3168/jds.S0022-0302(00)74856-9

Diao, S., Huang, S., Chen, Z., Teng, J., Ma, Y., Yuan, X., Chen, Z., Zhang, H., Li, J., & Zhang, Z. (2019). Genome-wide signatures of selection detection in three South China indigenous pigs. *Genes*, *10*, 346–360. https://doi.org/10.3390/genes10050346

Ebrahimie, E., Ebrahimi, F., Ebrahimi, M., Tomlinson, S., & Petrovski, K. R. (2018). Hierarchical pattern recognition in milking parameters predicts mastitis prevalence. *Computers and Electronics in Agriculture*, *147*(C), 6–11. https://doi.org/10.1016/j.compag.2018.02.003

Eulalio, A., Schulte, L., & Vogel, J. (2012). The mammalian microRNA response to bacterial infections. *RNA Biology*, *9*, 742–750. https://doi.org/10.4161/rna.20018

Fu, W., Lee, W. R., & Abasht, B. (2016). Detection of genomic signatures of recent selection in commercial broiler chickens. *BMC Genetics*, *17*, 122–131. https://doi.org/10.1186/s12863-016-0430-1

Gemelli, C., Martello, A., Montanari, M., Marani, T. Z., Salsi, V., Zappavigna, V., Parenti, S., Vignudelli, T., Selmi, T., Ferrari, S., & Grande, A. (2013). The Orosomucoid 1 protein is involved in the vitamin D – Mediated macrophage de-activation process. *Science Direct (Experimental Cell Research)*, *319*, 3201–3213. https://doi.org/10.1016/j.yexcr.2013.08.017

Gholami, M. (2014). Selection signature detection in a diverse set of chicken breeds. Ph.D. dissertation, Faculty of Agricultural Sciences, Georg-August-University, 150 p.

Godfrey, A. C., White, A. E., Tatomer, D. C., Marzluff, W. F., & Duronio, R. J. (2009). The drosophila U7 snRNP proteins Lsm10 and Lsm11 are required for histone pre-mRNA processing and play an essential role in development. *RNA*, *15*, 1661–1672. https://doi.org/10.1261/rna.1518009

Gomes, I. M., Maia, C. J., & Santos, C. R. (2012). STEAP proteins: From structure to applications in cancer therapy. *Molecular Cancer Research*, *10*, 573–587. https://doi.org/10.1158/1541-7786.MCR-11-0281

Gutiérrez-Gil, B., Arranz, J. J., & Wiener, P. (2015). An interpretative review of selective sweep studies in *Bos taurus* cattle populations: Identification of unique and shared selection signals across breeds. *Frontiers in Genetics*, *6*, 167. https://doi.org/10.3389/fgene.2015.00167

Holmberg, M., & Andersson-Eklund, L. (2004). Quantitative trait loci affecting health traits in Swedish dairy cattle. *Journal of Dairy Science*, *87*(8), 2653–2659. https://doi.org/10.3168/jds.s0022-0302(04)73391-3

Hu, Z.-L., Park, C. A., & Reecy, J. M. (2022). Bringing the animal QTLdb and CorrDB into the future: Meeting new challenges and providing updated services. *Nucleic Acids Research*, *50*, D956–D961. https://doi.org/10.1093/nar/gkab1116

Huang, D. W., Sherman, B. T., & Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, *4*(1), 44–57. https://doi.org/10.1038/nprot.2008.211

Ibeagha-Awemu, E. M., Ibeagha, A. E., Messier, S., & Zhao, X. (2010). Proteomics, genomics, and pathway analyses of Escherichia coli and Staphylococcus aureus infected milk whey reveal molecular pathways and networks involved in mastitis. *Journal of Proteome Research*, *9*, 4604–4619. https://doi.org/10.1021/pr100336e

Izquierdo, A. C., Liera, J. E. G., Cervantes, R. E., Castro, J. F. I., Mancera, E. A. V., Crispín, R. H., Mosqueda, M. L. J., Vázquez, A. G., Pérez, J. O., Aparicio, P. S., & Denis, B. E. R. (2017). Production of milk and bovine mastitis. *Journal of Advances in dairy research*, *5*(2), 2–5. https://doi.org/10.4172/2329-888X.1000174

Jensen, L. J., Kuhn, M., Stark, M., Chaffron, S., Creevey, C., Muller, J., Doerks, T., Julien, P., Roth, A., Simonovic, M., Bork, P., & von Mering, C. (2009). STRING 8—A global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Research*, *37*, D412–D416. https://doi.org/10.1093/nar/gkn760

Kaneene, J. B., Miller, R. A., Herdt, T. H., & Gardiner, J. C. (1997). The association of serum nonesterified fatty acids and cholesterol, management and feeding practices with peripartum disease in dairy cows. *Elsevier: Preventive Veterinary Medicine*, *31*, 59–72. https://doi.org/10.1016/s0167-5877(96)01141-5

Kaupe, B., Brandt, H., Prinzenberg, E.-M., & Erhardt, G. (2007). Joint analysis of the influence of CYP11B1 and DGAT1 genetic variation on milk production, somatic cell score, conformation, reproduction, and productive lifespan in German Holstein cattle. *Journal of Animal Science*, *85*, 11–21. https://doi.org/10.2527/jas.2005-753

Kimura, M. (1989). The neutral theory of molecular evolution and the world view of neutralists. *Genome*, *31*(1), 24–31. https://doi.org/10.1139/g89-009

Kobayashi, K., Kuki, C., Shoki, O., & Kumura, H. (2016). Pro-inflammatory cytokine TNF-alpha is a key inhibitory factor for lactose synthesis pathway in lactating mammary epithelial cells. *Experimental Cell Research*, *340*, 295–304. https://doi.org/10.1016/j.yexcr.2015.10.030

Korkuć, P., Arends, D., May, K., König, S., & Brockmann, G. A. (2021). Genomic loci affecting Milk production in German black pied cattle (DSN). *Frontiers in Genetics*, *12*, 640039. https://doi.org/10.3389/fgene.2021.640039

Kreitman, M. (2000). Methods to detect selection in populations with applications to the human. *In Annual review of genomics and human genetics*, *1*, 539–559. https://doi.org/10.1146/annurev.genom.1.1.539

Laodim, T., Elzo, M. A., Koonawootrittriron, S., Suwanasopee, T., & Jattawa, D. (2019). Pathway enrichment and protein interaction network analysis for milk yield, fat yield and age at first calving in a Thai multibreed dairy population. *Asian-Australasian Journal of Animal Sciences*, *32*, 508–518. https://doi.org/10.5713/ajas.18.0382

Lashneva, I., Sermyagin, A. A., Ignatieva, L. P., Gladyr, E., Ermilov, A., & Zinovieva, N. A. (2021). PSXII-7 Milk somatic cells monitoring in Russian Holstein cattle population as a

base for determining genetic and genomic variability. *Journal of Animal Science*, *99*, 252. https://doi.org/10.1093/jas/skab235.460

Lemay, D. G., Hovey, R. C., Hartono, S. R., Hinde, K., Smilowitz, J. T., Ventimiglia, F., Schmidt, K. A., Lee, J. W. S., Islas-Trejo, A., Silva, P., Korf, I., Medrano, J. F., Barry, P. A., & German, J. (2013). Sequencing the transcriptome of milk production: Milk trumps mammary tissue. *BMC Genomics*, *14*, 872–888. https://doi.org/10.1186/1471-2164-14-872

Li, C., Luo, L., Wei, S., & Wang, X. (2017). Identification of the potential crucial genes in invasive ductal carcinoma using bioinformatics analysis. *Oncotarget*, *9*, 6800–6813. https://doi.org/10.18632/oncotarget.23239

Li, R., Zhang, C. L., Liao, X. X., Chen, D., Wang, W. Q., Zhu, Y., Geng, X. H., Ji, D. J., Mao, Y. J., Gong, Y. C., & Yang, Z. P. (2015). Transcriptome MicroRNA profiling of bovine mammary glands infected with Staphylococcus aureus. *International Journal of Molecular Sciences*, *16*, 4997–5013. https://doi.org/10.3390/ijms16034997

Li, T., Gao, J., Zhao, X., & Ma, Y. (2019). Digital gene expression analyses of mammary glands from meat ewes naturally infected with clinical mastitis. *Royal Society Open Science*, *6*, 181604–181617. https://doi.org/10.1098/rsos.181604

Lippolis, J. D., Reinhardt, T. A., Sacco, R. A., Nonnecke, B. J., & Nelson, C. D. (2011). Treatment of an intramammary bacterial infection with 25-hydroxyvitamin D3. *PLoS One*, *6*, e25479. https://doi.org/10.1371/journal.pone.0025479

Loor, J. J., Moyes, K. M., & Bionaz, M. (2011). Functional adaptations of the transcriptome to mastitis-causing pathogens: The mammary gland and beyond. *Journal of Mammary Gland Biology Neoplasia*, *16*, 305–322. https://doi.org/10.1007/s10911-011-9232-2

Lund, M. S., Guldbrandtsen, B., Buitenhuis, A. J., Thomsen, B., & Bendixen, C. (2008). Detection of quantitative trait loci in Danish Holstein cattle affecting clinical mastitis, somatic cell score, udder conformation traits, and assessment of associated effects on milk yield. *Journal of Dairy Science*, *91*(10), 4028–4036. https://doi.org/10.3168/jds.2007-0290

Lund, M. S., Sahana, G., Andersson-Eklund, L., Hastings, N., Fernandez, A., Schulman, N., Thomsen, B., Viitala, S., Williams, J. L., Sabry, A., Viinalass, H., & Vilkki, J. (2007). Joint analysis of quantitative trait loci for clinical mastitis and somatic cell score on five chromosomes in three Nordic dairy cattle breeds. *Journal of Dairy Science*, *90*, 5282–5290. https://doi.org/10.3168/jds.2007-0177

Lung, L. H. S., Petrini, J., Ramirez-Diaz, J., Salvian, M., Rovadoscki, G. A., Pilonetto, F., Dauria, B. D., Machado, P. F., Coutinho, L. L., Wiggans, G. R., & Mourao, G. B. (2019). Genome-wide association study for milk production traits in a Brazilian Holstein population. *Journal of Dairy Science*, *102*, 5305–5314. https://doi.org/10.3168/jds.2018-14811

Madsen, P., & Jensen, J. (2013). A user's guide to DMU: A package for analysing multivariate mixed models. https://www.researchgate.net/publication/291444592

Maijaroen, S., Klaynongsruang, S., Roytrakul, S., Konkchaiyaphum, M., Taemaitree, L., & Jangpromma, N. (2022). An integrated proteomics and bioinformatics analysis of the anticancer properties of RT2 antimicrobial peptide on human colon cancer (Caco-2) cells. *Molecules*, *27*(4), 1426. https://doi.org/10.3390/molecules27041426

Malhotra, N., & Kang, J. (2013). SMAD regulatory networks construct a balanced immune system. *Immunology*, *139*(1), 1–10. https://doi.org/10.1111/imm.12076

Mallard, B. A., Emam, M., Paibomesai, M., Thomson-Crispi, K., & Wagter-Lesperance, L. (2015). Genetic selection of cattle for improved immunity and health. *Japanese Journal of Veterinary Research*, *63*, S37–S44. https://doi.org/10.14943/jjvr.63.suppl.s37

Manzari, Z., Mehrabani-Yeganeh, H., Nejati-Javaremi, A., Moradi, M. H., & Gholizadeh, M. (2019). Detecting selection signatures in three Iranian sheep breeds. *Animal Genetics*, *50*, 298–302. https://doi.org/10.1111/age.12772

Mbindyo, Ch. M., Gitao, G. C., & Mulei, Ch. M. (2020). Prevalence, etiology, and risk factors of mastitis in dairy cattle in Embu and Kajiado counties, Kenya. *Veterinary Medicine International 2020*, Article ID 8831172, 1–12. https://doi.org/10.1155/2020/8831172

Meier, S., Arends, D., Korkuć, P., Neumann, G. B., & Brockmann, G. A. (2020). A genome-wide association study for clinical mastitis in the dual-purpose German black pied cattle breed. *Journal of Dairy Science*, *103*, 10289–10298. https://doi.org/10.3168/jds.2020-18209

Mokhber, M. (2015). A genome-wide scan for selective signatures in Iranian buffalo breeds. Ph.D. dissertation, Faculty of Genetic and animal breeding Tehran University, Iran 145 p.

Moliterno, A. R., & Resar, L. M. (2011). AKNA: Another AT-hook transcription factor "hooking-up" with inflammation. *Cell Research*, *21*, 1528–1530. https://doi.org/10.1038/cr.2011.96

Moradi, M. H. (2012). Genomic scan of selection signatures in some sheep breeds and identification of candidate regions associated with fat deposition. Ph.D. thesis. Faculty of Genetic and animal breeding Tehran University, Iran.

Moradi, M. H., Nejati-Javaremi, A., Moradi-Shahrbabak, M., Dodds, K. G., & McEwan, J. C. (2010). Whole-genome scan of population differentiation in Zel and Lori-Bakhtiari sheep breeds. *Agricultural Biotechnology Journal*, *2*(2), 57–70. https://doi.org/10.22103/JAB.2012.369

Moradi, M. H., Nejati-Javaremi, A., Moradi-Shahrbabak, M., Dodds, K. G., & McEwan, J. C. (2012). Genomic scan of selective sweeps in thin and fat tail sheep breeds for identifying of candidate regions associated with fat deposition. *BMC Genetics*, *13*, 10–24. https://doi.org/10.1186/1471-2156-13-10

Moravčíková, N., Simcic, M., Meszaros, G., Sölkner, J., Kukučková, V., Vlček, M., Trakovická, A., Kadlečik, O., & Kasarda, R. (2018). Genomic response to natural selection within alpine cattle breeds. *Journal of Animal Science*, *63*, 136–143. https://doi.org/10.17221/62/2017-CJAS

Mumtaz, P. T., Bhat, B., Ibeagha-Awemu, E. M., Taban, Q., Wang, M., Dar, M. A., Bhat, S. A., Shabir, N., Shah, R. A., Ganie, N. A., Velayutham, D., Haq, Z., & Ahmad, S. M. (2022). Mammary epithelial cell transcriptome reveals potential roles of lncRNAs in regulating milk synthesis pathways in Jersey and Kashmiri cattle. *BMC Genomics*, *23*, 176. https://doi.org/10.1186/s12864-022-08406-x

Naderi, S., Bohlouli, M., Yin, T., & König, S. (2018). Genomic breeding values, SNP effects and gene identification for disease traits in cow training sets. *Animal Genetics*, *49*, 178–192. https://doi.org/10.1111/age.12661

Nagaraj, N. S., & Datta, P. K. (2010). Targeting the transforming growth factor-β signaling pathway in human cancer. *Expert*

Opinion on Investigational Drugs, *19*, 77–91. https://doi.org/10.1517/13543780903382609

Nolte, W., Thaller, G., & Kuehn, C. (2019). Selection signatures in four German warmblood horse breeds: Tracing breeding history in the modern sport horse. *PLoS One*, *14*, e0215913. https://doi.org/10.1371/journal.pone.0215913

Ogorevc, J., Kunej, T., Razpet, A., & Dovc, P. (2009). Database of cattle candidate genes and genetic markers for milk production and mastitis. *Animal Genetics*, *40*, 832–851. https://doi.org/10.1111/j.1365-2052.2009.01921.x

Online Mendelian Inheritance in Animals (OMIA). Sydney School of Veterinary Science, University of Sydney. MIA Number: OMIA 001744–9913: January 11, 2022. World Wide Web URL: https://omia.org/

Ozeki, K., Sugiyama, M., Akter, K. A., Nishiwaki, K., Asano-Inami, E., & Senga, T. (2019). FAM98A is localized to stress granules and associates with multiple stress granule-localized proteins. *Molecular and Cellular Biochemistry*, *451*, 107–115. https://doi.org/10.1007/s11010-018-3397-6

Paik, D., Monahan, A., Caffrey, D. R., Elling, R., Goldman, W. E., & Silverman, N. (2017). SLC46 family transporters facilitate cytosolic innate immune recognition of monomeric peptidoglycans. *The Journal of Immunology*, *199*(1), 263–270. https://doi.org/10.4049/jimmunol.1600409

Parker Gaddis, K. L., Megonigal, J. H., Jr., Clay, J. S., & Wolfe, C. W. (2018). Genome-wide association study for ketosis in US jerseys using producer-recorded data. *Journal of Dairy Science*, *101*, 413–424. https://doi.org/10.3168/jds.2017-13383

Pighetti, G. M., & Elliott, A. A. (2011). Gene polymorphisms: The keys for marker assisted selection and unraveling core regulatory pathways for mastitis resistance. *Journal of Mammary Gland Biology and Neoplasia*, *16*, 421–432. https://doi.org/10.1007/s10911-011-9238-9

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I. W., Daly, M. J., & Sham, P. C. (2007). PLINK: A toolset for whole genome association and population-based linkage analysis. *American Journal of Human Genetics*, *81*, 575. http://pngu.mgh.harvard.edu/purcell/plink/

Qanbari, S., Pausch, H., Jansen, S., Somel, M., Strom, T. M., Fries, R., Nielsen, R., & Simianer, H. (2014). Classic selective sweeps revealed by massive sequencing in cattle. *PLoS Genetics*, *10*, e1004148. https://doi.org/10.1371/journal.pgen.1004148

Qanbari, S., & Simianer, H. (2014). Mapping signatures of positive selection in the genome of livestock. *Livestock Science*, *166*, 133–143. https://doi.org/10.1016/j.livsci.2014.05.003

Raschia, M. A., Nani, J. P., Maizon, D. O., Beribe, M. J., Amadio, A. F., & Poli, M. A. (2018). Single nucleotide polymorphisms in candidate genes associated with milk yield in Argentinean Holstein and Holstein x Jersey cows. *Journal of Animal Science and Technology*, *60*, 31–40. https://doi.org/10.1186/s40781-018-0189-1

Resar, L. M. (2010). The high mobility group A1 gene: Transforming inflammatory signals into cancer? *Cancer Research*, *70*, 436–439. https://doi.org/10.1158/0008-5472.CAN-09-1212

Sabeti, P. C., Reich, D. E., Higgins, J. M., Levine, H. Z. P., Richter, D. J., Schaffner, S. F., Gabriel, S. B., Platko, J. V., Patterson, N. J., McDonald, G. J., Ackerman, H. C., Campbell, S. J., Altshuler, D., Cooper, R., Kwiatkowski, D., Ward, R., & Lander, E. S. (2002). Detecting recent positive selection in the human genome from haplotype structure. *Nature*, *419*, 832–837. https://doi.org/10.1038/nature01140

Sabeti, P. C., Schaffner, S. F., Fry, B., Lohmueller, J., Varilly, P., Shamovsky, O., Palma, A., Mikkelsen, T. S., Altshuler, D., & Lander, E. S. (2006). Positive natural selection in the human lineage. *Science*, *3129*, 1614–1620. https://doi.org/10.1126/science.1124309

Sabeti, P. C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., Xie, X., Byrne, E. H., McCarroll, S., Gaudet, R., Schaffner, S. F., Lander, E. S., International HapMap Consortium, Frazer, K. A., Ballinger, D. G., Cox, D. R., Hinds, D. A., Stuve, L. L., Gibbs, R. A., … Stewart, J. (2007). Genome-wide detection and characterization of positive selection in human populations. *Nature*, *449*, 913–918. https://doi.org/10.1038/nature06250

Safran, M., Rosen, N., Twik, M., BarShir, R., Iny Stein, T., Dahary, D., Fishilevich, S., & Lancet, D. (2022). The GeneCards suite chapter. *Practical Guide to Life Science Databases. Springer*, 27–56. https://doi.org/10.1007/978-981-16-5812-9_2

Scheet, P., & Stephens, M. (2006). A fast and flexible statistical model for large-scale population genotype data. Applications to inferring missing genotypes and haplotypic phase. *American Journal of Human Genetics*, *78*, 629–644. https://doi.org/10.1086/502802

Schulman, N. F., Viitala, S. M., de Koning, D. J., Virta, J., Mäki-Tanila, A., & Vilkki, J. H. (2004). Quantitative trait loci for health traits in Finnish Ayrshire cattle. *Journal of Dairy Science*, *87*, 443–449. https://doi.org/10.3168/jds.S0022-0302(04)73183-5

Shabalina, T., Yin, T., May, K., & König, S. (2021). Proofs for genotype by environment interactions considering pedigree and genomic data from organic and conventional cow reference populations. *Journal of Dairy Science*, *104*, 4452–4466. https://doi.org/10.3168/jds.2020-19384

Shao, M., Wu, F., Zhang, J., Dong, J., Zhang, H., Liu, X., Liang, S., Wu, J., Zhang, L., Zhang, C., & Zhang, W. (2021). Screening of potential biomarkers for distinguishing between latent and active tuberculosis in children using bioinformatics analysis. *Medicine (Baltimore)*, *100*(5), e23207. https://doi.org/10.1097/MD.0000000000023207

Soares, R. A. N., Vargas, G., Muniz, M. M. M., Soares, M. A. M., Cánovas, A., Schenkel, F., & Squires, E. J. (2021). Differential gene expression in dairy cows under negative energy balance and ketosis: A systematic review and meta-analysis. *Journal of Dairy Science*, *104*, 602–615. https://doi.org/10.3168/jds.2020-18883

Sternlicht, M. D. (2005). Key stages in mammary gland development: The cues that regulate ductal branching morphogenesis. *Breast Cancer Research*, *8*, 201–211. https://doi.org/10.1186/bcr1368

Stock, K. F., Cole, J., Pryce, J., Gengler, N., Bradley, A., Andrews, L., Heringstad, B., & Egger-Danner, C. (2013). Standardization of health data – ICAR guidelines including health key. *ICAR Technical Series*, *17*, 75–81.

Tiezzi, F., Parker-Gaddis, K. L., Cole, J. B., Clay, J. S., & Maltecca, C. (2015). A genome-wide association study for clinical mastitis in first parity US Holstein cows using single-step approach and genomic matrix re-weighting procedure. *PLoS One*, *10*, e0114919. https://doi.org/10.1371/journal.pone.0114919

Tribout, T., Croiseau, P., Lefebvre, R., Barbat, A., Boussaha, M., Fritz, S., Boichard, D., Hoze, C., & Sanchez, M. P. (2020). Confirmed effects of candidate variants for milk production, udder health, and udder morphology in dairy cattle. *Genetic*

Selection Evolution, *52*(1), 55–80. https://doi.org/10.1186/s12711-020-00575-1

Turk, R., Koledić, M., Maćešić, N., Benić, M., Dobranić, V., Duričić, D., Cvetnić, L., & Samardžija, M. (2017). The role of oxidative stress and inflammatory response in the pathogenesis of mastitis in dairy cows. *Mljekarstvo*, *67*, 91–101. https://doi.org/10.15567/mljekarstvo.2017.0201

van Roosmalen, W., Le Dévédec, S. E., Golani, O., Smid, M., Pulyakhina, I., Timmermans, A. M., Look, M. P., Zi, D., Pont, C., Graauw, M., Naffar-Abu-Amara, S., Kirsanova, C., Rustici, G., Hoen, P. A. C., Martens, J. W. M., Foekens, J. A., Geiger, B., & Water, B. (2015). Tumor cell migration screen identifies SRPK1 as breast cancer metastasis determinant. *The Journal of Clinical Investigation*, *125*, 1648–1664. https://doi.org/10.1172/JCI74440

Vigano, S., Perreau, M., Pantaleo, G., & Harari, A. (2012). Positive and negative regulation of cellular immune responses in physiologic conditions and diseases. *Clinical and Developmental Immunology*, *2012*(485781), 1–11. https://doi.org/10.1155/2012/485781

Wathes, D. C., Cheng, Z., Salavati, M., Buggiotti, L., Takeda, H., Tang, L., Becker, F., Ingvartsen, K. L., Ferris, C., Hostens, M., & Crowe, M. A. (2021). Relationships between metabolic profiles and gene expression in liver and leukocytes of dairy cows in early lactation. *Journal of Dairy Science*, *104*, 3596–3616. https://doi.org/10.3168/jds.2020-19165

Wee, Y. S., Roundy, K. M., Weis, J. J., & Weis, J. H. (2012). Interferon-inducible transmembrane proteins of the innate immune response act as membrane organizers by influencing clathrin and v-ATPase localization and function. *Innate Immunity*, *18*(6), 834–845. https://doi.org/10.1177/1753425912443392

Xu, D., Yao, Y., Lu, L., Costa, M., & Dai, W. (2010). Plk3 functions as an essential component of the hypoxia regulatory pathway by direct phosphorylation of HIF-1. *Journal of Biological Chemistry*, *285*, 38944–38950. https://doi.org/10.1074/jbc.M110.160325

Yang, C., Zhao, X., Cui, N., & Liang, Y. (2017). Cadherins associate with distinct stem cell-related transcription factors to coordinate the maintenance of stemness in triple-negative breast cancer. *Stem Cells International*, *2017*(2), 1–13. https://doi.org/10.1155/2017/5091541

Yang, J., Jiang, J., Liu, X., Wang, H., Guo, G., Zhang, Q., & Jiang, L. (2015). Differential expression of genes in milk of dairy cattle during lactation. *Animal Genetics*, *47*, 174–180. https://doi.org/10.1111/age.12394

Zhao, H., Tang, X., Wu, M., Li, Q., Yi, X., Liu, Sh., Jiang, J., Wang, Sh., & Sun, X. (2021). Transcriptome characterization of short distance transport stress in beef cattle blood. *Frontiers in Genetics*, *12* (Article 616388), 1–15. https://doi.org/10.3389/fgene.2021.616388

Zhao, X. W., Yang, Y. X., Huang, D. W., Cheng, G. L., & Zhao, H. L. (2015). Comparative proteomic analysis of proteins expression changes in the mammary tissue of cows infected with Escherichia coli mastitis. *Journal of Veterinary Science*, *16*, 253–263. https://doi.org/10.1038/cmi.2011.19

Zheng, R., Liu, Q., Wang, T., Wang, L., & Zhang, Y. (2018). FAM98A promotes proliferation of non-small cell lung cancer cells via the P38-ATF2 signaling pathway. *Cancer Management and Research*, *10*, 2269–2278. https://doi.org/10.2147/CMAR.S163323

Zhou, R., O'Hara, S. P., & Chen, X. M. (2011). MicroRNA regulation of innate immune responses in epithelial cells. *Cellular & Molecular Immunology*, *8*, 371–379. https://doi.org/10.1038/cmi.2011.19

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Abbasi-Moshaii, B., Moradi, M. H., Yin, T., Rahimi-Mianji, G., Nejati-Javaremi, A., & König, S. (2023). Genome-wide scan for selective sweeps identifies novel loci associated with resistance to mastitis in German Holstein cattle. *Journal of Animal Breeding and Genetics*, *140*, 92–105. https://doi.org/10.1111/jbg.12737