Animal Biotechnology

ISSN: 1049-5398 (Print) 1532-2378 (Online) Journal homepage: www.tandfonline.com/journals/labt20

# Candidate gene search for milk production and composition, milk coagulation properties, and milk protein profile in dairy sheep from a New Zealand flock

A. C. Marshall, N. Lopez-Villalobos, V. Vigolo, M. De Marchi, S. M. Loveday, M. Weeks & W. McNabb

**To cite this article:** A. C. Marshall, N. Lopez-Villalobos, V. Vigolo, M. De Marchi, S. M. Loveday, M. Weeks & W. McNabb (2025) Candidate gene search for milk production and composition, milk coagulation properties, and milk protein profile in dairy sheep from a New Zealand flock, Animal Biotechnology, 36:1, 2516688, DOI: 10.1080/10495398.2025.2516688

**To link to this article:** https://doi.org/10.1080/10495398.2025.2516688

ANIMAL BIOTECHNOLOGY
2025, VOL. 36, NO. 1, 2516688
https://doi.org/10.1080/10495398.2025.2516688



# Candidate gene search for milk production and composition, milk coagulation properties, and milk protein profile in dairy sheep from a New Zealand flock

A. C. Marshall[a,b], N. Lopez-Villalobos[a], V. Vigolo[c], M. De Marchi[c], S. M. Loveday[b,e], M. Weeks[d] and W. McNabb[b]

[a]School of Agriculture and Environment, Massey University, Palmerston North, New Zealand; [b]The Riddet Institute, Massey University, Palmerston North, New Zealand; [c]Department of Agronomy, Food, Natural resources, Animals and Environment, University of Padova, Legnaro, Italy; [d]AgResearch Ltd, Smart Foods & Bioproducts Group, Te Ohu Rangahau Kai, Massey University, Palmerston North, New Zealand; [e]Agriculture and Food, Commonwealth Scientific and Industrial Research Organisation (CSIRO), Werribee, Australia

**ABSTRACT**

The objective of this study was to perform genome-wide association analysis and thus search for candidate genes for milk production and composition, milk coagulation properties (MCP), and milk protein profile in dairy sheep from a New Zealand flock. After quality control, 45,801 single nucleotide polymorphisms (SNPs) were included in the analysis, 147 ewes, and 470 individual records. A total of 87 SNPs and 55 candidate genes were found across *Ovis aries* autosomes (OAR) 2, 3, 6, 16, 18, 20, 25, and 26. Of particular importance, were the candidate genes PDZRN4 for milk yield, and BMP2K for contents of $\alpha_{s1}$- and $\alpha_{s2}$-caseins. No SNPs were found in the casein genes, LALBA or PAEP. Only one SNP was significant for MCP, and overall, the genetic architecture of MCP was similar to that of ratio of casein to calcium, pH, lactose, and the ratio of casein to protein. Further studies with larger flocks and with genomic imputation are required to validate the findings of this study before incorporating markers or genes into breeding programmes.

**ARTICLE HISTORY**
Received 4 August 2024
Accepted 2 June 2025

**KEYWORDS**
Sheep milk; dairy sheep; milk production; milk composition; protein profile; genome-wide association

## Introduction

The New Zealand dairy sheep industry is still in its early stages and requires improved animal genetics for dairy traits and animals that are fit for the New Zealand climate and pasture-based farming systems. Genomic information could play a crucial role in accelerating genetic gains and improving the accuracy of breeding value estimations. Furthermore, identifying specific genetic markers that significantly impact dairy traits would provide an additional tool for selection.

Genome-wide association studies (GWAS) provide insights into the genetic architecture of important traits for production, sustainability, or animal resilience. In animal breeding, GWAS help identify, across the genome, markers such as single nucleotide polymorphisms (SNPs) that are associated with quantitative traits. These traits are usually polygenic, meaning they are controlled by multiple genes located in quantitative trait loci (QTLs) across the genome. Some GWAS have been performed for dairy sheep across the globe on milk production and composition traits,[1–5] and on somatic cell score.[6]

The findings of GWAS vary across studies in dairy sheep, indicating that knowledge in this field is still building. This variation is influenced by the different genetic architectures of the various breeds and populations studied. However, different GWAS findings within and across breeds are also reported. For protein content in the milk of Churra sheep, a region on *Ovis aries* autosome (OAR) 6 near the casein cluster was suggested to influence protein percentage,[7] while others have pointed to OAR3 as being more significant.[1,8] In a non-selected dairy sheep breed (Altamurana), OAR3 was also the most relevant for protein

**CONTACT** A. C. Marshall a.c.marshall@massey.ac.nz School of Agriculture and Environment, Massey University, Private Bag 11222, 4442 Palmerston North, New Zealand.

Supplemental data for this article can be accessed online at https://doi.org/10.1080/10495398.2025.2516688.

percentage.[4] In contrast, OAR1 and OAR7 were important for protein percentage in Valle del Belice sheep.[5]

In New Zealand, a recent study found significant SNP associations with milk yield in dairy sheep.[9] However, GWAS on sheep milk processability traits have been performed mainly in Mediterranean countries.[10,11] Other genomic studies have targeted the analysis of the casein genes[12] or single genes such as PRLR, GHR, and GHRHR[13] with the use of PCR. No GWAS have yet been performed on New Zealand dairy sheep for milk coagulation properties (MCP), which provide information on the quality of milk for processing into cheese.

While GWAS on protein composition have been performed on dairy cows,[14] studies are lacking in dairy sheep. It is well known that the genes CSN1S1, CSN2, CSN1S2, and CSN3 encode for the four casein fractions and contribute to their high protein polymorphism. The corresponding genes for whey proteins are LALBA and PAEP, which encode for α-lactalbumin and β-lactoglobulin, respectively. Genetic variation in these genes leads to different levels of protein expression, and consequently can lead to different physico-chemical properties of milk.[15] However, genetic variations in the milk protein profile not only are a result of these protein-encoding genes, but are also influenced by the broader polygenic background of the animal.[16] For example, the molecular mechanism and factors controlling casein phosphorylation are still largely unknown.[17]

Considering these, this study aimed to perform genome-wide analysis to identify associated SNPs and candidate genes potentially affecting milk yield and composition, milk coagulation properties, and the content of protein fractions in dairy sheep milk, thus providing a better understanding of the genetic architecture controlling these traits.

## Materials and methods

### Animals and milk samples

This study was conducted at a commercial sheep dairy farm in Masterton, Wairarapa, New Zealand. Ethics approval was obtained for this study (Massey University Animal Ethics Committee-Protocol 21/45). The breed's development started in 1996, initially using Coopworth and Border Leicester dams from New Zealand and semen from European East Friesian sires. The resulting progenies were backcrossed with East Friesian sires. The farm has been using the self-replacement of ewes and rams over the past 12 years. In this dairy sheep breed, two color variants are visually distinguished, white or black sheep. The small flock of 169 milking ewes is traditionally managed on a pasture-based system with limited supplementation, with selection based mainly on temperament, dairy conformation, and prolificacy. The age structure of the flock was 15% first-, 28% second-, 23% third- and 34% ≥ fourth-parity ewes in the studied season. Black ewes constituted 33% of the flock, while the remaining 67% were white. The median lambing date of the flock was the 20$^{th}$ of August 2021. Machine milking for cheese production started after the exclusive lamb suckling period, which lasted 57 days on average. A total of 470 test-day records from once-a-day milking were gathered from 1$^{st}$ November 2021 to 31$^{st}$ January 2022, from 147 ewes (50–182 days in milk), and this dataset was used for GWAS. A minimum of 2 test-day records were obtained from each ewe. Information on pedigree, lambing date, litter size, and age of animals were supplied to the study. The average inbreeding coefficient of the whole flock was 4%. The average inbreeding coefficient of the group of animals with some level of inbreeding was 9%.

### Milk yield and composition

Milk yield from individual ewes on a test-day was recorded as the total volume taken from the individual test buckets, and the individual milk samples were then taken for compositional and processability analyses. The milk samples were immediately refrigerated after collection and sodium azide (to final concentration of 0.025% w/w) was added for better preservation. All subsequent analyses were performed within three days following the sample collection.

Milk samples were analyzed by Milk Test NZ Ltd. (Hamilton, NZ) using a Combi FOSS instrument (Foss Analytics, Hilleroed, Denmark). The composition analysis included percentages of fat, protein, casein, lactose, and SCC (cells/mL) converted into somatic cell scores (SCS) using a $Log_2$ transformation. The analyses for the percentage of casein and content of urea (mg/100 mL) were performed using a Fourier-transform Infrared (FTIR) milk-analyzer MilkoScan FT6000 (Foss Analytics) calibrated for sheep milk samples. The ratio of casein to protein (CSN:PP) is the casein content divided by the protein content. Total calcium content (mg/100 mL) was analysed by a contract laboratory (Massey University Nutrition Lab) using the Arsenazo III method (Randox reagent kit Ca8309)

and the RX Daytona Plus clinical analyzer. The ratio of casein to calcium (CSN:Ca) was calculated as casein percentage divided by calcium content (mg/100 mL), multiplied by 100.

### Milk coagulation properties

The traditional milk coagulation properties of individual fresh milk samples were measured using a Formagraph instrument (Foss Analytics). The traits included rennet coagulation time (RCT, minutes), time to reach curd firmness of 20 mm after start of coagulation (K20, minutes), and curd firmness at 30 min after rennet addition (A30, millimetres).[18] The preparation of samples was described in Marshall et al.[19]

### Protein profile by RP-HPLC

Protein profile analyses for contents of κ-casein (κ-CN), $\alpha_{s1}$-casein ($\alpha_{s1}$-CN), $\alpha_{s2}$-casein ($\alpha_{s2}$-CN), β-casein (β-CN), α-lactalbumin (α-LA), and β-lactoglobulin (β-LG), in total protein, were obtained by reverse-phase high performance liquid chromatography (RP-HPLC). The sample preparation method and instrument setup specifications were performed as per method of Bobe et al.[20] and Bonfatti et al.,[21] respectively. The HPLC setup comprised an Agilent 1260 Infinity II LC system (Agilent Technologies) equipped with a quaternary pump (Agilent 1260 Infinity II, G7111B) and a diode array detector (Agilent 1260 Infinity II, G7115A).[22] The results were expressed as area (mAU) underneath the peaks of κ-CN, $\alpha_{s1}$-CN, $\alpha_{s2}$-CN, β-CN, α-LA, and β-LG, which represents the quantity of each protein fraction. A strong linear relationship ($r^2 > 0.90$) was observed between protein concentration in milk (g/L) and combined peak areas of protein fractions (total area). The percentage of each protein in total protein was calculated as the area of the respective fraction divided by the total area.

### Genotyping and quality control

A total of 323 ewes, which included 169 ewes and 154 ewe lambs, and 6 rams had ear tissue samples collected for DNA extraction and scanning using an iScan® at the Equine Parentage Testing Lab (Massey University, Palmerston North). Genotype information was obtained using the OvineSNP50 Beadchip array (Illumina, San Diego, CA) with a medium-density SNP panel (50k SNPs). A total of 64,734 SNPs was obtained for quality control using the SNP & Variation Suite (SVS 8.8) software. In the filtering process, genomic records were removed for 23 animals with a call rate <95% across all the SNPs, from which 4 animals were ewes that had phenotype records. Also, SNPs with >5% missing genotypes across all individuals (call rate < 95%), and that had a significant deviation from the Hardy-Weinberg equilibrium threshold of $p < 10^{-6}$ or that had minor allele frequency < 1% were also removed. After these quality control edits, a total of 306 animals with SNPs (including 147 ewes with phenotypes), and 45,801 SNPs remained for association analyses.

Coat color was identified as the first principal component (PC1) that explained >50% variation in the analysis and included as a fixed effect in the GWAS model described further ahead, the scatter plot is provided in Supplementary Figure S6.

### Statistical analyses

Descriptive statistics for milk production and composition, milk coagulation properties, and contents of κ-CN, $\alpha_{s1}$-CN, $\alpha_{s2}$-CN, β-CN, α-LA, and β-LG (in total protein) were obtained using the MEANS procedure of SAS 9.4 software (SAS Institute Inc., Cary, NC).

### Genome-wide association analyses

The phenotypic traits were pre-corrected for fixed effects using the single-trait animal repeatability model in the ASReml 4.1 software package.[23] Detailed number of animals in the pedigree file for ASReml is included in Supplementary File 3, Table S8. The model included the fixed effects of ewe coat color (black or white), litter size (1 or ≥2 lambs), parity number (1st, 2nd, 3rd or ≥4th parity). Covariates included in the model were the effect of time (days in milk), the quadratic effect of time, and deviation from the median lambing date of the flock. The random effect of ewe permanent environment effect, and the random residual were included.

Principal component analyses for population stratification correction were performed in SVS software (SVS 8.8). PC1 was highly correlated ($r = 0.95$) with the coat-color effect and therefore PCs were not included, instead, the genomic relationship matrix was used for adjustment of population structure.

The genome-wide association study (GWAS) was performed using a mixed linear model in the GCTA

software package.[24] The following model was fitted for each trait:

$$y = \mu + X\beta + g + e$$

Where $y$ was the vector of pre-corrected phenotypes for each ewe, $\mu$ was the vector of overall mean, $\beta$ was the vector of each SNP fixed effect (additive SNP effect), and $X$ was the incidence matrix of † to $y$ with the SNPs genotypes of BB, AB, or AA, respectively, $g$ was the random additive polygenic effect (the accumulated effect of all SNPs) and $e$ was the random residual error. The assumptions for the model were: $g \sim N\left(0, G\sigma_g^2\right)$, where $G$ was the genomic relationship matrix between the ewe and $\sigma_g^2$ was the additive genetic variance explained by SNPs, and $e \sim N\left(0, I\sigma_e^2\right)$, where $I$ was the identity matrix of order n = 147 and $\sigma_e^2$ was the residual variance.

Bonferroni multiple-test was used to avoid a type I error.[25] Bonferroni multiple-test correction was estimated at 5% genome-wide significance as 0.05/m ($0.05/45{,}801 = 1.09 \times 10^{-6}$), which corresponds to 5.96 on a $-\log_{10}$ ($p$-value) scale. In addition, the less conservative suggestive significance threshold[26] was used, which was 1/number of SNPs, which in this case is ($1/45{,}801 = 2.18 \times 10^{-5}$), which was 4.66 on a $-\log_{10}$ (p-value) scale.

Manhattan plots in which $-\log_{10}$ ($p$-values) were plotted against their genomic locations of the markers for each trait using the qqman package in R software 4.2.1.[27]

### Candidate genes and functional analysis

The candidate genes were searched using Ensembl Release 112,[28] based on the *Ovis aries* (sheep) reference genome assembly ARS-UI_Ramb_v2.0 (GCA_016772045.1).[29] Gene annotation boundaries were set at 100 Kb upstream or downstream from the position of each significantly associated intergenic SNP. The biological functions of the associated candidate genes were reviewed using the Gene Ontology (GO) tool in Ensembl.

## Results

### Descriptive statistics

Average milk production and composition, milk protein profile, milk coagulation properties, and milk pH for this flock of dairy sheep are presented in Table 1.

**Table 1.** Means, standard deviations (*SD*), minimum and maximum values for milk production, composition, protein profile, and milk coagulation properties (MCP) of dairy sheep from a New Zealand flock.

| Trait[a] | N | Mean | SD | Minimum | Maximum | CV (%) |
|---|---|---|---|---|---|---|
| Milk yield (L/day) | 470 | 0.57 | 0.25 | 0.10 | 1.80 | 44 |
| Fat (%) | 470 | 6.32 | 1.26 | 1.87 | 11.24 | 20 |
| Protein (%) | 470 | 5.52 | 0.68 | 4.09 | 8.80 | 12 |
| Lactose (%) | 470 | 4.72 | 0.27 | 3.66 | 5.34 | 6 |
| Casein (%) | 470 | 4.21 | 0.50 | 3.08 | 6.28 | 12 |
| Casein:Protein | 470 | 0.77 | 0.02 | 0.54 | 0.82 | 3 |
| Casein:Calcium | 470 | 2.28 | 0.29 | 1.58 | 4.02 | 13 |
| Calcium (mg/100 mL) | 470 | 185.5 | 25.3 | 104.6 | 269.4 | 14 |
| Protein profile[b] (% of total protein) | | | | | | |
| κ-casein | 470 | 7.28 | 1.30 | 2.60 | 13.80 | 18 |
| $\alpha_{s1}$-casein | 470 | 27.58 | 1.78 | 20.60 | 32.80 | 6 |
| $\alpha_{s2}$-casein | 470 | 12.78 | 1.58 | 5.90 | 17.80 | 12 |
| β-casein | 470 | 40.06 | 1.96 | 32.50 | 45.50 | 5 |
| α-lactalbumin | 470 | 1.94 | 0.61 | 0.09 | 4.60 | 31 |
| β-lactoglobulin | 470 | 10.36 | 1.25 | 6.20 | 14.30 | 12 |
| MCP | | | | | | |
| RCT (min) | 451 | 13.3 | 3.2 | 6.86 | 26.0 | 24 |
| K20 (min) | 447 | 2.7 | 1.1 | 1.30 | 10.2 | 41 |
| A30 (mm) | 451 | 53.3 | 9.5 | 11.60 | 75.4 | 18 |
| pH | 470 | 6.59 | 0.09 | 6.34 | 6.90 | 1 |

[a]Casein:Protein = ratio of casein to protein; Casein:Calcium = ratio of casein to calcium.
[b]Protein profile: % of each protein fraction in total protein. RCT: rennet coagulation time; $K_{20}$= : time to reach curd firmness of 20 mm; $A_{30}$=: = curd firmness at 30 min post rennet addition.

### Manhattan plots

Figures 1(A–C) are the Manhattan plots for contents of $\alpha_{s1}$-CN, $\alpha_{s2}$-CN, and α-LA, respectively. The contents of $\alpha_{s1}$-CN, $\alpha_{s2}$-CN, and α-LA had significant associations with SNPs at both genome-wide Bonferroni and suggestive thresholds in OAR6 ($\alpha_{s1}$-CN and $\alpha_{s2}$-CN), OAR20 ($\alpha_{s2}$-CN), OAR25 and OAR26 ($\alpha_{s1}$-CN), and OAR3 (α-LA). For $\alpha_{s2}$-CN, OAR3 also had clustering but with no significant SNP associations.

Manhattan plots for the other investigated traits that were only significantly associated at the suggestive threshold are provided in Supplementary File 1, Figures S1–S3. It was possible to observe notable clustering of SNPs for most studied traits. Despite there being a clear peak in OAR25 for κ-CN, there were no significant SNP associations for κ-CN. For β-lactoglobulin, OAR3 had the most notable clustering, but also with no significant SNP associations. Milk yield, casein content, ratio of casein to protein, calcium content, ratio of casein to calcium, pH, and rennet coagulation time (RCT) were only significantly associated with SNPs at the suggestive threshold.

The QQ-plots for all traits investigated are presented in the Supplementary File 2, Figures S4–S6, which showed proper correction for data stratification.

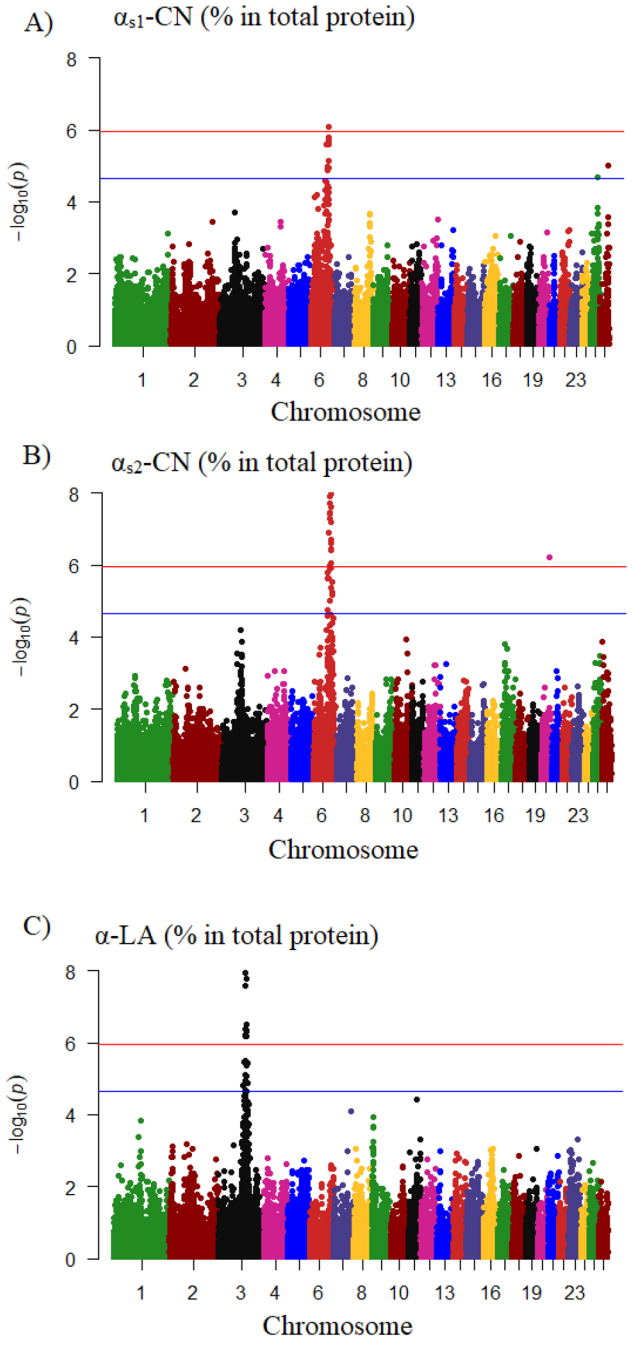


**Figure 1.** Manhattan plots displaying the $-\log_{10}$ (*p*-values) of genetic markers associated with the contents of $\alpha_{s1}$-casein (a), $\alpha_{s2}$-casein (B), and α-lactalbumin (C), plotted against their respective genomic positions. The red line indicates the Bonferroni-corrected significance threshold of 5.96 on the $-\log_{10}$ (*p*-value) scale, while the blue line represents the suggestive significance threshold of 4.66 on the $-\log_{10}$ (*p*-value) scale.

In the QQ-plots of contents of $\alpha_{s1}$-CN, $\alpha_{s2}$-CN, and α-LA, *p*-values were more significant than expected under the null hypothesis, with points moving toward the *y*-axis.

Detailed investigation into linkage disequilibrium in the breed was beyond the scope of the present study. However, it was possible to observe linkage disequilibrium in the region of the casein genes and in the region of LALBA with the use of the linkage disequilibrium map function of the SVS software (Supplementary File 2, Figures S7 and S8).

## *Candidate genes and function analyses*

The SNPs that were significant at the Bonferroni level and the respective candidate genes for contents of $\alpha_{s1}$-CN and $\alpha_{s2}$-CN are presented in Table 2, and those for α-LA content are presented in Table 3. The SNPs that were only significant at the suggestive level and the respective candidate genes for milk yield, casein percentage, ratio of casein to protein, calcium content, ratio of casein to calcium, milk pH, and rennet coagulation time are presented in Supplementary File 1, Table S1. The SNPs that were significant at the suggestive level for contents of $\alpha_{s1}$-CN, $\alpha_{s2}$-CN, and α-LA are also included in Supplementary File 1, Tables S2–S4. The full list of candidate genes, with respective annotations and Gene Ontology terms are in Supplementary File 2, Table S5.

A total of 61 candidate genes, including novel genes, were associated with the significant SNPs, at either the Bonferroni or suggestive significance thresholds. Of these, 27 candidate genes had intragenic SNPs associated. Several common genes were identified across the studied traits. Notably, the genes BMP2K, CCDC158, CFAP299, GK2, NAA11, SCL4A4, and USO1, in OAR6, were associated with both $\alpha_{s1}$-CN and $\alpha_{s2}$- CN contents. The gene TBK1, in OAR3, was significantly associated with α-LA content and casein percentage at the suggestive significance level. Additionally, CNTN1 and PDZRN4, both located in OAR3, were associated with milk yield and with α-LA content at Bonferroni and suggestive significance levels. Lastly, the gene DCK was associated with both milk yield and $\alpha_{s2}$-CN content.

For $\alpha_{s1}$-CN, a total of 12 SNPs were associated, with 10 located in OAR6 (87.1–96.4 Mb). The region between 94.2 and 95.0 Mb contained the highest concentration of associated SNPs. In contrast, the 96.1–96.4 Mb region included three SNPs, all of which were intragenic to CFAP299 gene.

For $\alpha_{s2}$-CN, 41 associated SNPs were found in OAR6 (80.9–101.7 Mb). The casein genes were located within this region (86.2–86.4 Mb), but none were within 500 Kb of any significantly associated

**Table 2.** The single nucleotide polymorphisms (SNPs) identified as significant at the Bonferroni level ($-\log_{10}$ (*p-value*) ≥ 5.96) for $\alpha_{s1}$*-casein* (% in total protein) and for $\alpha_{s2}$*-casein* (% in total protein) in dairy sheep milk.

| $\alpha_{s1}$-casein | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| SNP | Chr | Position (bp) | Ref./MA | Ref. Freq. | Effect | SE | -$\log_{10}$(*p*) | Gene | SNP Annotation | Genes within 100 Kb |
| s43490.1 | 6 | 94276957 | G/A | 0.83 | −1.07 | 0.21 | 6.09 | BMP2K | Intron | |
| $\alpha_{s2}$-casein | | | | | | | | | | |
| Lead SNP | Chr | Position (bp) | Ref./MA | Ref. Freq. | Effect | SE | -$\log_{10}$(*p*) | Gene | SNP Annotation | Genes within 100 Kb |
| ilmnseq_rs425296828 | 6 | 84420871 | A/G | 0.8 | 1.00 | 0.19 | 6.89 | TMPRSS11D | Intron | |
| ilmnseq_rs413585365 | 6 | 87126636 | A/G | 0.83 | 1.19 | 0.2 | 8.22 | – | Intergenic | SLC4A4, DCK |
| OAR6_87592155.1 | 6 | 87592155 | G/A | 0.81 | 1.14 | 0.2 | 8.25 | – | Intergenic | SLC4A4, GC |
| OAR6_88110298.1 | 6 | 88110298 | C/A | 0.8 | 0.92 | 0.19 | 5.98 | – | Intergenic | NPFFR2, ADAMTS3 |
| OAR6_88303825.1 | 6 | 88303825 | A/G | 0.8 | 0.92 | 0.19 | 5.98 | ADAMTS3 | Intron | |
| OAR6_88678679.1 | 6 | 88678679 | A/G | 0.8 | 0.92 | 0.19 | 5.98 | – | Intergenic | ENSOARG0002003468 |
| OAR6_91381996.1 | 6 | 91381996 | A/G | 0.84 | 1.23 | 0.21 | 8.47 | USO1 | Intron | |
| OAR6_91640306.1 | 6 | 91640306 | A/G | 0.85 | 1.24 | 0.21 | 8.31 | CXCL10 | Intron | |
| OAR6_91995865.1 | 6 | 91995865 | A/G | 0.81 | 1.22 | 0.19 | 9.41 | CCDC158 | Intron | |
| OAR6_92241864.1 | 6 | 92241864-92321965 | A/G | 0.79 | 1.06 | 0.19 | 7.73 | SHROOM3 | Intron | |
| OAR6_92730863.1 | 6 | 92730863 | G/A | 0.83 | 1.08 | 0.2 | 7.44 | – | Intergenic | CCNI, CCNG2 |
| ilmnseq_rs405099401 | 6 | 92837990 | G/A | 0.81 | 1.03 | 0.19 | 7.28 | – | Intergenic | CCNG2 |
| OAR6_93031365.1 | 6 | 93031365 | G/A | 0.79 | 1.08 | 0.19 | 8.19 | – | Intergenic | CXCL13 |
| OAR6_93397312.1 | 6 | 93397312 | G/A | 0.83 | 1.11 | 0.19 | 7.91 | – | Intergenic | MRPL1, FRAS1 |
| s43490.1 | 6 | 94276957 | G/A | 0.83 | 1.33 | 0.2 | 10.37 | BMP2K | Intron | |
| OAR6_94793686.1 | 6 | 94793686 | G/A | 0.82 | 1.11 | 0.2 | 7.61 | – | Intergenic | NAA11, GK2 |
| s59224.1 | 6 | 94902165 | A/G | 0.82 | 1.09 | 0.2 | 7.20 | – | Intergenic | GK2 |
| OAR6_94954855.1 | 6 | 94954855 | G/A | 0.83 | 1.16 | 0.2 | 7.98 | – | Intergenic | |
| OAR6_96130389.1 | 6 | 96130389-96431721 | A/G | 0.83 | 1.19 | 0.2 | 8.22 | CFAP299 | Intron | |
| OAR6_97085567.1 | 6 | 97085567-97185219 | A/G | 0.79 | 0.97 | 0.2 | 6.05 | RASGEF1B | Intron | |
| s52112.1 | 6 | 97682782 | G/A | 0.82 | 1.01 | 0.2 | 6.39 | – | Intergenic | RASGEF1B |
| OAR6_97927928.1 | 6 | 97927928 | A/G | 0.82 | 1.02 | 0.2 | 6.43 | HNRNPDL | Exon 1 | |
| s71640.1 | 6 | 98157876 | A/G | 0.81 | 1.02 | 0.2 | 6.60 | – | Intergenic | TMEM150C, SCD5, ENSOARG00020035089 |
| OAR6_98214895.1 | 6 | 98214895 | A/G | 0.82 | 1.04 | 0.2 | 6.71 | SCD5 | Intron | |
| OAR6_99947976.1 | 6 | 99947976 | A/G | 0.81 | 0.99 | 0.19 | 6.70 | – | Intergenic | CDS1, NKX6-1 |
| ilmnseq_rs421607124 | 20 | 45934168 | G/C | 0.84 | 0.98 | 0.19 | 6.20 | LOC105605956 | Intron | – |

**Table 3.** The single nucleotide polymorphisms (SNPs) identified as significant at the Bonferroni level ($-\log_{10}$ (*p-value*) ≥ 5.96) for *α-lactalbumin* (% in total protein) in dairy sheep milk.

| Lead SNP | Chr | Position (bp) | No. SNPs | Ref./MA | Ref. Freq. | Effect | SE | -$\log_{10}$(*p*) | Gene | SNP Annotation | Genes within 100 Kb |
|---|---|---|---|---|---|---|---|---|---|---|---|
| s31828.1 | 3 | 143359947 | 1 | G/A | 0.38 | −0.2 | 0.04 | 6.37 | ADAMTS20 | Intron | – |
| OAR3_145276010.1 | 3 | 145276010 | 1 | C/A | 0.79 | −0.28 | 0.05 | 7.57 | – | Intergenic | – |
| OAR3_145295705.1 | 3 | 145295705 | 1 | G/A | 0.66 | −0.24 | 0.04 | 7.93 | – | Intergenic | – |
| s34624.1 | 3 | 145835332-145931284 | 2 | A/G | 0.34 | −0.19 | 0.04 | 6.22 | PDZRN4 | Intron | – |
| OAR3_146042298.1 | 3 | 146042298 | 1 | A/G | 0.85 | −0.35 | 0.06 | 8.71 | – | – | CNTN1, PDZRN4 |
| s35014.1 | 3 | 146168927-146281949 | 4 | C/A | 0.71 | −0.26 | 0.05 | 7.77 | CNTN1 | Intron | – |
| OAR3_146751517.1 | 3 | 146751517 | 1 | G/A | 0.76 | −0.24 | 0.05 | 6.31 | MUC19 | Intron | |
| OAR3_147028849.1 | 3 | 147028849 | 1 | A/C | 0.82 | −0.34 | 0.05 | 9.5 | – | Intergenic | LRRK2, SLC2A13 |
| OAR3_148569270.1 | 3 | 148569270 | 1 | C/A | 0.77 | −0.24 | 0.05 | 6.33 | CPNE8 | Intron | – |

SNPs. Additionally, two SNPs were intragenic to MAPK10.

For α-lactalbumin content, 26 SNPs in OAR3 (128.5–156.6 Mb) were significantly associated. The LALBA (137.5 Mb) gene was located within this region, though the closest SNP to LALBA was 188.5 Kb away. Two significant intragenic SNPs were found in ADAMTS20 and in PDZRN4, with the latter also being associated with milk yield in the present study.

## Discussion

The average production level of this flock during the studied season has been previously discussed,[30] where the relatively low milk production was primarily attributed to limited pasture availability, also, peak production in early lactation was missed due to the lamb suckling period. The milk coagulation properties during the once-a-day milking period have also been discussed[19] and were found to be within the range previously

reported for dairy sheep. The effects of protein polymorphisms on milk composition, coagulation properties, and milk protein profile in this flock of dairy sheep were also investigated in our previous study.[31]

### Manhattan plots for yield traits and milk composition

Consistent with previous studies,[1,4,32] OAR3 emerged as a key genomic region for milk production and milk composition traits in dairy sheep. In particular, the 145–146 Mb region was important for milk yield, while the 151–156 Mb was important for fat and protein percentages.

These findings align with Moioli et al.[4] who identified significant associations between protein percentage at the 145 Mb region of OAR3. Similarly, García-Gámez et al.[1] found associations at 137.3 Mb for both fat and protein percentages, and Carta et al.[32] reported associations at 131.31 Mb for fat and protein yields and at 141.53 Mb for protein content.

Interestingly, Gutiérrez-Gil et al.[8] reported a QTL on OAR3 (131.74–134.46 Mb) for protein content, overlapping with the region associated with α-LA content in the present study (128.5–156.6 Mb). Furthermore, García-Gámez et al.[1] reported that a marker located in the third intron of LALBA (OAR3) had a strong significant association with protein percentage ($p$-value$=3.78\times10^{-26}$).

For lactose percentage, the 122–126 Mb region in OAR2 was important in the present study. This is in partial agreement with García-Gámez et al.[1] who identified associations with milk, fat, and lactose yields in broader regions of OAR2 (42.7–63.2 Mb and 53.0–58.0 Mb).

### Candidate genes for milk yield and calcium content

This section will focus on candidate genes with the strongest SNP signals, meeting either Bonferroni or suggestive significance thresholds. The candidate gene for milk yield, PDZRN4, has also been recently associated with milk yield in New Zealand dairy sheep by another research group[9] using three large flocks, evidencing that this gene's function should be further investigated in New Zealand dairy sheep. PDZRN4 is involved in protein degradation pathways. In other livestock species, this gene has been associated with heat stress of dairy cows, which can cause a decrease in the expression of genes involved in milk production.[33]

Both candidate genes associated with milk yield (CNTN1 and PDZRN4) in dairy sheep in the present study have also been previously correlated with milk production in Holstein dairy cattle.[34] Additionally, CNTN1 has also been associated with udder and teat scores in Angus cattle.[35] CNTN1 gene codes for Contactin 1, a protein involved in cell adhesion. This gene has been associated with myopathy and muscular weakness.[36] This gene is likely associated with milk production in dairy animals due to the dependence of udder structure on muscular vigor.

The candidate gene for calcium content, SNTG2, has been associated with osteoporosis in women.[37] Variations in SNTG2 might influence calcium metabolism in the organism. However, in the literature, other genes have been associated with calcium content in bovine milk, such as SLC37A1, in *Bos taurus* autosome BTA 1, and ANKH, in BTA 16.[38]

### Candidate genes for contents of $\alpha_{s1}$- and $\alpha_{s2}$-casein

No significant SNP associations in the casein genes were found, which was likely due to the limited bead chip coverage, and further investigation on the variation of casein genes with genomic imputation could not be done for the present study. However, a linkage was found between DCK, SLC4A4, and GC, involving the casein cluster. Overall, most genes associated here with protein profile have also been reported by Marina et al.[39] in the enrichment and association analysis of genes associated with milk and cheese-making traits.

The most significant SNP for $\alpha_{s1}$-CN ($-\log_{10} p=6.09$) and $\alpha_{s2}$-CN ($-\log_{10} p=10.37$) was in gene BMP2K which is involved in protein phosphorylation. This suggests a potential relationship between casein phosphorylation levels and the concentrations of $\alpha_{s1}$-CN and $\alpha_{s2}$-CN. This finding could be of great importance for the technological and functional properties of sheep milk.[40] This gene's indirect role in the phosphorylation of milk caseins should be further investigated.

Other genes identified in the present study that are also involved in phosphorylation processes and kinases include BMP2K, CCNT1, EPHA5, FBP1, GK2, LRRK2, MAPK10, NPFFR2, and TBK1. Attention should be given to the potential roles these genes may also play on casein phosphorylation. In addition, the results from Moioli et al.[4] showed that the most represented categories of genes associated with protein content had functions related to the phosphorus/phosphate metabolic processes. To explore this further, future research could focus on quantifying the phosphorylation levels of caseins using LC/ESI-MS methods.

Studies in dairy cows have reported the protein profile of milk to be associated with several regions

across twenty chromosomes, particularly those coding for protein variants of β-CN, κ-CN, or β-LG, and DGAT1.[41] In addition, it has been found that the phosphorylation level of isoforms of $\alpha_{s1}$-CN was associated with different genes (DGAT1 vs β-LG), suggesting that phosphorylation of isoforms is likely to be mediated by different enzymes.[42] Furthermore, up- and down-regulatory systems have been proposed for phosphorylation levels of $\alpha_s$-CNs.[17]

The present study aligns with previous research where the gene SLC4A4 has been associated with $\alpha_{s2}$-CN content in dairy cows.[14] This gene has also been associated with other milk traits in dairy cattle, including milk yield,[43] and milk curd-firming characteristics.[44] The SLC4A4 gene is involved in the regulation of intracellular pH, secretion and absorption of bicarbonate, and active transport of glucose which is uptaken by the mammary epithelial cells for milk synthesis.[43] The bicarbonate transport facilitated by SLC4A4 is also linked to calcium ion concentration and homeostasis.[45] Since $\alpha_s$-CNs interact with calcium via phosphoserine and phosphothreonine, any changes in calcium availability could influence the micellisation of caseins in milk.

Despite the genes DCK and GC not being previously associated with milk protein profile, they have also been considered top candidate genes for milk yield[43] and mastitis in dairy cows,[46,47] and GC has been associated with fertility traits.[48]

No previous association has been reported between protein composition of milk and ADAMTS3 and NPFFR2. However, these genes have been associated with milk yield and protein yield in a large GWAS study with Holstein dairy cattle in the US.[48] Additionally, ADAMTS3 has been associated with milk yield of dairy goats in China.[49] Also, the expression of ADAMTS3 was found to significantly increase during inflammatory response in bovine mammary epithelial cells.[50] Its roles in extracellular matrix remodeling, cell signaling, proteolytic activity, and tissue remodeling suggest that it could influence the mammary gland environment in ways that impact protein secretion.

The gene NPFFR2 was previously considered a strong candidate gene affecting mastitis occurrence in cattle.[47] NPFFR2 has been suggested to regulate feeding behavior and energy expenditure in mammals and it has been linked to heat tolerance,[51] so it could indirectly affect the mammary gland's ability to synthesize and secrete milk proteins.

The gene USO1, associated with both $\alpha_{s1}$- ($-\log_{10} p = 4.86$) and $\alpha_{s2}$-CN ($-\log_{10} p = 8.47$), is involved in protein amino acid binding and one of its processes is intracellular protein transport. It has been associated with milk yield in buffaloes.[52] The gene CXCL13 was one of the important candidate genes for $\alpha_{s2}$-CN ($-\log_{10} p = 8.19$) and has been associated by others with milk traits and lactation persistency,[43] and with clinical mastitis.[46]

### Candidate genes for the content of α-lactalbumin

For α-lactalbumin, OAR3 was important. The candidate gene CCNT1 was located close to LALBA (80 Kb distance), and a small LD was observed in the region (detailed in Supplemental Figure S8). Interestingly, CCNT1 has been previously associated with protein percentage in Assaf and Churra breeds.[11]

Some of the relevant genes for α-LA, CNTN1, PDZRN4, and TBK1, were also candidate genes for milk production and composition in the present study. It is important to highlight that α-LA is the regulatory protein of the lactose synthase enzyme system that catalyzes and regulates the synthesis of lactose in the lactating mammary gland, which increases the volume of secreted fluids through osmotic effects and serves as a primer for the synthesis of oligosaccharides.[53] Therefore, common genes regulating α-LA content and milk production were expected.

The other strong candidate gene for α-LA, ADAMTS20, is involved in procollagen processing, extracellular matrix remodeling, inflammation, cell migration, and angiogenesis cell differentiation, lactogenic activity of mammary epithelial cells, and stimulation of synthesis of milk proteins. Interestingly, ADAMTS20 has been associated with milk traits in dairy goats,[54] and other ADAMTS genes have been associated with mastitis in cattle.[50]

A further candidate gene for α-LA, MUC19 (mucin protein), has been correlated with growth traits of Chinese cattle,[55] and is considered as artificial selection signature for milk production in a Brazilian local cattle breed.[56] Other candidate genes for α-LA, CNTN1, LRRK2, and CPNE8 have been previously associated with udder scores in Canadian Angus cows.[57]

### Manhattan plots and candidate genes for MCP and pH

For RCT and milk pH, OAR16 was important, with SNPs significantly associated at the suggestive significance threshold. Clustering of SNPs in OAR16 was also evidenced in the Manhattan plots for lactose, ratio of casein to protein, and ratio of casein to calcium, K20, and A30, despite the SNPs not being significant. The clustering regions observed in OAR16 indicate that the genetic control of MCP, pH, lactose,

ratio of casein-to-protein, and ratio of casein-to-calcium is likely related.

In OAR16 (38.969–39.028 Mb) is located PRL. However, this was not a candidate gene in the present study, and it was located over 200 Kb from the candidate genes EMB (for pH) and SLC38A9 (for RCT). Previous studies have identified a SNP in the PRL gene that was significantly associated with MCP in Sarda sheep. This effect was suggested to be mediated by certain PRL isoforms that may have negatively impacted the activation of milk protein gene transcription.[58]

The candidate gene EMB (embigin), associated with milk pH, is involved in cell adhesion, which is consistent with the fact that milk pH may be influenced by the influx of minerals when the junction between mammary epithelial cells is disrupted, as seen in cases of infection (mastitis) or accumulated mechanical tissue damage in late lactation.[59] Embigin's role in cell adhesion could affect the integrity of mammary epithelial cells and affect the balance of ions and pH in milk. Interestingly, the gene EMB has been previously associated with milk yield in dairy sheep.[2]

In addition, OAR25 was important for RCT, K20, and A30, with notable clustering of SNPs. This finding aligns with the work of Marina et al.[39] who identified SNPs in OAR25 significantly associated with A30, particularly in the 9.56 Mb region. However, in our study, the 21.54 Mb region of OAR25 was more important.

Clustering in OAR25 was also observed for κ-CN, despite the SNPs not being significant, and for $\alpha_{s1}$-CN with a significant intragenic SNP in WAPL. This suggests that the genetic control of milk coagulation properties could also be related to the genetic control of contents of κ-CN and $\alpha_{s1}$-CN. Further investigation into the genomic correlations between RCT (the MCP with the strongest SNP signal) and the other milk traits (as shown in Supplementary File 2, Table S7), revealed strong correlations with ratio of casein to calcium ($r_G$ = 0.92, $SE$ = 0.25), milk pH ($r_G$ = 0.88, $SE$ = 0.13), lactose ($r_G$ = −0.86, $SE$ = 0.22), and ratio of casein to protein ($r_G$ = −0.77, SE = 0.15). These findings further support our previous study on pedigree-based genetic correlations.[60]

Surprisingly, neither OAR3, where LALBA and BLG are located, nor OAR6, where the casein genes are located, were identified as important hot zones for milk coagulation properties in the present study. Gaspa et al.[10] also did not find significant associations with milk coagulation properties in OAR6; instead, these were observed in OAR12. In contrast, Bertelsen et al.[61] found that a SNP close to SLC4A4 (BTA 6) had an additive genetic effect on curd firming rate in bovine milk, and the association was suggested to be due to linkage with the casein gene cluster. These differing GWAS results highlight how the selection strategies of different breeds or populations can influence milk quality for processing.

## Conclusion

The present GWAS study on milk production, milk composition, and milk coagulation traits in dairy sheep identified candidate genes previously associated with milk production, composition, mastitis, immune response, and udder conformation in dairy farm animals. These genes are likely influencing the studied milk traits indirectly, and their exact mechanisms could be further investigated. Several genes reported here are involved in phosphorylation processes and kinases. The gene PDZRN4 appears to play an important role in milk production in New Zealand dairy sheep, while the gene SLC4A4 may be influencing $\alpha_{s2}$-CN content.

Another finding of this study is the distinct genetic background for protein composition and milk coagulation properties. The genetic background for milk coagulation properties was found to be more similar to that of the ratio of casein to calcium, pH, lactose, and the ratio of casein to protein, rather than being solely attributed to the casein genes. This suggests that MCPs can be influenced by factors other than milk protein composition.

The content of casein fractions in total protein was associated with genes located close to the casein gene cluster as well as with genes involved in the phosphorylation process. Little is known about the genes regulating the phosphorylation of caseins in both bovine and ovine milks and this could be an area of further investigation in dairy sheep. To the best of our knowledge, this study represents the first genome-wide association study focused on protein composition in dairy sheep, specifically examining the quantities of individual casein and whey protein fractions in total protein.

The findings of the present study lack robustness for large-scale applications due to the small sample size, and further genomic studies must be performed with larger populations of dairy sheep. It is recommended that, beyond GWAS, linkage disequilibrium is investigated to search for QTLs for marker-assisted selection. Whole genome sequencing could help pinpoint the exact causative DNA polymorphisms, followed by studies on gene expression to validate the results, before implementing gene-assisted selection in dairy sheep.

## Acknowledgments

The authors would like to thank the commercial farm, and the scientific communities from the NZ3M group, Massey University, and AgResearch.

## Authors' contributions

Conceptualization, ACM, NLV, and WM; methodology, ACM and NLV; software, ACM, NLV; validation, NLV; formal analysis, ACM; investigation, ACM; resources, WM, NLV; data curation, ACM; writing, ACM; review NLV, SML, MW, and WM; editing ACM, visualization, ACM, NLV; supervision, NLV, SML, MW, and WM; project administration, ACM; funding acquisition, WM.

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Funding

This study was funded by the MBIE New Zealand Milks Mean More (NZ3M) Endeavour Programme [Contract MAUX1803]. This work was also supported by the Riddet Institute through PhD scholarship to ACM.

## References

1. García-Gámez E, Gutierrez-Gil B, Sahana G, Sanchez JP, Bayón Y, Arranz JJ. GWA analysis for milk production traits in dairy sheep and genetic support for a QTN influencing milk protein percentage in the LALBA gene. *PLoS One*. 2012;7:e47782.
2. Li H, Wu XL, Tait R, Jr, et al. Genome-wide association study of milk production traits in a crossbred dairy sheep population using three statistical models. *Anim Genet*. 2020;51(4):624–628.
3. Mi B-N, Zhang L-G, Bai U, et al. Genome-wide association study of milk production traits in Dairy Meade sheep. *Acta Veterinaria Zootechn Sin*. 2021;52:3294–3303.
4. Moioli B, Steri R, Napolitano F, Catillo G. A genome wide scan highlights differences in the genetic architecture of fat and protein contents in dairy sheep. *Small Ruminant Res*. 2015;131:21–28.
5. Sutera A, Riggio V, Mastrangelo S, et al. Genome-wide association studies for milk production traits in Valle del Belice sheep using repeated measures. *Anim Genet*. 2019;50:311–314.
6. Sutera AM, Moscarelli A, Mastrangelo S, et al. Genome-wide association study identifies new candidate markers for somatic cells score in a local dairy sheep. *Front Genet*. 2021;12:643531.
7. Diez-Tascón C, Bayón Y, Arranz JJ, De La Fuente F, San Primitivo F. Mapping quantitative trait loci for milk production traits on ovine chromosome 6. *J Dairy Res*. 2001;68(3):389–397.
8. Gutiérrez-Gil B, El-Zarei MF, Alvarez L, et al. Quantitative trait loci underlying milk production traits in sheep. *Anim Genet*. 2009;40(4):423–434.
9. Costilla R, Chan P, McRae K, et al. Genome-wide association studies for milk traits in New Zealand dairy sheep. *N Z J Anim Sci Prod*. 2023;82:64–68.
10. Gaspa G, Serdino J, Manca M, et al. Genome-wide association study of milk coagulation properties in dairy sheep. *J Anim Sci*. 2016;94(suppl_5):159–159.
11. Marina H, Pelayo R, Suárez-Vega A, Gutiérrez-Gil B, Esteban-Blanco C, Arranz J. Genome-wide association studies (GWAS) and post-GWAS analyses for technological traits in Assaf and Churra dairy breeds. *J Dairy Sci*. 2021;104:11850–11866.
12. Noce A, Pazzola M, Dettori ML, et al. Variations at regulatory regions of the milk protein genes are associated with milk traits and coagulation properties in the Sarda sheep. *Anim Genet*. 2016;47(6):717–726.
13. Pazzola M, Vacca GM, Paschino P, Bittante G, Dettori ML. Novel genes associated with dairy traits in sarda sheep. *Animals*. 2021;11(8):2207.
14. Dadousis C, Pegolo S, Rosa GJ, Bittante G, Cecchinato A. Genome-wide association and pathway-based analysis using latent variables related to milk protein composition and cheesemaking traits in dairy cattle. *J Dairy Sci*. 2017;100:9085–9102.
15. Amalfitano N, Mota LFM, Rosa GM, Cecchinato A, Bittante G. Role of CSN2, CSN3, and BLG genes and the polygenic background in the cattle milk protein profile. *J Dairy Sci*. 2022;105(7):6001–6020.
16. Bobe G, Beitz D, Freeman A, Lindberg G. Effect of milk protein genotypes on milk protein composition and its genetic parameter estimates. *J Dairy Sci*. 1999;82(12):2797–2804.
17. Fang Z-H, Bovenhuis H, van Valenberg H, et al. Genome-wide association study for αS1-and αS2-casein phosphorylation in Dutch Holstein Friesian. *J Dairy Sci*. 2019;102(2):1374–1385.
18. McMahon DJ, Brown RJ. Evaluation of Formagraph for comparing rennet solutions. *J Dairy Sci*. 1982;65(8):1639–1642.
19. Marshall AC. Animal factors affecting the cheese-making properties and the heat coagulation time of milk from dairy sheep in a New Zealand flock. *New Zealand Journal of Agricultural Research*. 2024;68(3):1–18.
20. Bobe G, Beitz DC, Freeman AE, Lindberg GL. Separation and quantification of bovine milk proteins by reversed-phase high-performance liquid chromatography. *J Agric Food Chem*. 1998;46(2):458–463.
21. Bonfatti V, Grigoletto L, Cecchinato A, Gallo L, Carnier P. Validation of a new reversed-phase high-performance liquid chromatography method for separation and quantification of bovine milk protein genetic variants. *J Chromatogr A*. 2008;1195:101–106.
22. Vigolo V, Franzoi M, Cendron F, et al. Characterization of the genetic polymorphism linked to the β-casein A1/A2 alleles using different molecular and biochemical methods. *J Dairy Sci*. 2022;105:8946–8955.
23. Gilmour A, Gogel B, Cullis B, Welham S, Thompson R. 2015. *ASReml User Guide Release 4.1 Structural Specification*. Hemel Hempstead: VSN International Ltd.

24. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet*. 2011;88(1):76–82.
25. Armstrong RA. When to use the B onferroni correction. *Ophthalmic Physiol Opt*. 2014;34(5):502–508.
26. Lander E, Kruglyak L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet*. 1995;11(3):241–247.
27. Turner SD. 2014. qqman: An R package for visualizing GWAS results using QQ and Manhattan plots. *J Open Source Softw*. 3(25):731.
28. Martin FJ, Amode MR, Aneja A, et al. Ensembl 2023. *Nucleic Acids Res*. 2023;51(D1):D933–D941.
29. Davenport KM, Bickhart DM, Worley K, et al. An improved ovine reference genome assembly to facilitate in-depth functional annotation of the sheep genome. *Gigascience*. 2022;11:giab096.
30. Marshall AC, Lopez-Villalobos N, Loveday SM, Ellis A, McNabb W. Modelling lactation curves for dairy sheep in a New Zealand flock. *Animals*. 2023;13(3):349.
31. Marshall AC, Vigolo V, Marchi MD, et al. Effect of protein polymorphisms on milk composition, coagulation properties, and protein profile in dairy sheep. *Int Dairy J*. 2025;160:106102.
32. Carta A, Usai MG, Sechi T, et al. 2014. Exploring the genetic variation between Sarda and Lacaune dairy sheep breeds by genome wide association study on economic traits *Proceedings, 10th Word Congress of Genetics Applied to Livestock Production, Paper 339*, Vancouver, Canada.
33. Czech B, Wang Y, Szyda J. 2023. Genome-wide association study of heat stress response in *Bos taurus*. bioRxiv [Preprint]: 2023.06. 05.543663.
34. Shin D-H, Lee H-J, Cho S, et al. Deleted copy number variation of Hanwoo and Holstein using next generation sequencing at the population level. *BMC Genomics*. 2014;15(1):240.
35. Devani K, Plastow G, Orsel K, Valente TS. Genome-wide association study for mammary structure in Canadian Angus cows. *Plos Oone*. 2020;15:e0237818.
36. Compton AG, Albrecht DE, Seto JT, et al. Mutations in contactin-1, a neural adhesion and neuromuscular junction protein, cause a familial form of lethal congenital myopathy. *Am J Hum Genet*. 2008;83(6):714–724.
37. Jales Neto LH, Wicik Z, Torres GH, et al. Overexpression of SNTG2, TRAF3IP2, and ITGA6 transcripts is associated with osteoporotic vertebral fracture in elderly women from community. *Mol Genet Genomic Med*. 2020;8(9):e1391.
38. Sanchez M-P, Rocha D, Charles M, et al. Sequence-based GWAS and post-GWAS analyses reveal a key role of SLC37A1, ANKH, and regulatory regions on bovine milk mineral content. *Sci Rep*. 2021;11(1):7537.
39. Marina H, Reverter A, Gutiérrez-Gil B, et al. Gene networks driving genetic variation in milk and cheese-making traits of Spanish Assaf sheep. *Genes (Basel)*. 2020;11(7):715.
40. Bijl E, Holland JW, Boland M. Posttranslational modifications of caseins. In: *Milk proteins*. Amsterdam, The Netherlands: Elsevier; 2020:173–211.
41. Schopen G, Visker M, Koks P, Mullaart E, Van Arendonk J, Bovenhuis H. Whole-genome association study for milk protein composition in dairy cattle. *J Dairy Sci*. 2011;94(6):3148–3158.
42. Bijl E, van Valenberg H, Huppertz T, van Hooijdonk A, Bovenhuis H. Phosphorylation of αS1-casein is regulated by different genes. *J Dairy Sci*. 2014;97(11):7240–7246.
43. Pedrosa VB, Schenkel FS, Chen S-Y, et al. Genomewide association analyses of lactation persistency and milk production traits in Holstein cattle based on imputed whole-genome sequence data. *Genes (Basel)*. 2021;12(11):1830.
44. Dadousis C, Cipolat-Gotet C, Schiavon S, Bittante G, Cecchinato A. Inferring individual cow effects, dairy system effects and feeding effects on latent variables underlying milk protein composition and cheese-making traits in dairy cattle. *J Dairy Res*. 2017a;85(1):87–97.
45. Thornell IM, Bevensee MO. Regulators of Slc4 bicarbonate transporter activity. *Front Physiol*. 2015;6:166.
46. Sodeland M, Grove H, Kent M, et al. Molecular characterization of a long range haplotype affecting protein yield and mastitis susceptibility in Norwegian Red cattle. *BMC Genet*. 2011;12(1):70.
47. Wu X, Lund MS, Sahana G, et al. Association analysis for udder health based on SNP-panel and sequence data in Danish Holsteins. *Genet Sel Evol*. 2015;47(1):1–14.
48. Jiang J, Ma L, Prakapenka D, VanRaden PM, Cole JB, Da Y. A large-scale genome-wide association study in US Holstein cattle. *Front Genet*. 2019;10:442321.
49. Ni J, Xian M, Ren Y, et al. Whole-genome resequencing reveals candidate genes associated with milk production trait in Guanzhong dairy goats. *Anim Genet*. 2024;55(1):168–172.
50. Sheng H, Zhang J, Pan C, et al. Genome-wide identification of bovine ADAMTS gene family and analysis of its expression profile in the inflammatory process of mammary epithelial cells. *Int J Biol Macromol*. 2023;244:125304.
51. Cheruiyot EK, Haile-Mariam M, Cocks BG, MacLeod IM, Xiang R, Pryce JE. New loci and neuronal pathways for resilience to heat stress in cattle. *Sci Rep*. 2021;11:16619.
52. El-Halawany N, Abdel-Shafy H, Abd-El-Monsif AS, Abdel-Latif MA, Al-Tohamy AF, Abd El-Moneim OM. Genome-wide association study for milk production in Egyptian buffalo. *Livestock Science*. 2017;198:10–16.
53. Brew K. *α-Lactalbumin. Advanced Dairy Chemistry: Volume 1A: Proteins: Basic Aspects*. 4th ed. New York, USA: Springer; 2012:261–273.
54. Kang X, Li M, Liu M, et al. Copy number variation analysis reveals variants associated with milk production traits in dairy goats. *Genomics*. 2020;112(6):4934–4937.
55. Chen Y, Peng W, Zhang Z, et al. The relationship between MUC19 copy number variation and growth traits of Chinese cattle. *Gene*. 2023;851:147010.
56. Campos BM, do Carmo AS, da Silva TBR, et al. Identification of artificial selection signatures in Caracu breed lines selected for milk production and meat production. *Livestock Science*. 2017;206:82–87.
57. Devani K, Crowley JJ, Plastow G, Orsel K, Valente TS. Genetic parameter estimations and genomic insights for teat and udder structure in young and mature Canadian Angus cows. *J Anim Sci*. 2021;99(5):skab087.

58. Dettori ML, Pazzola M, Petretto E, Vacca GM. Association analysis between SPP1, POFUT1 and PRLR gene variation and milk yield, composition and coagulation traits in Sarda sheep. *Animals*. 2020;10(7):1216.
59. Hettinga KA. Lactose in the dairy production chain. In: *Lactose*. Cambridge, USA: Elsevier. 2019:231–266.
60. Marshall AC, Lopez-Villalobos N, Loveday SM, Weeks M, McNabb W. Estimation of genetic parameters for production, composition and processability of milk from dairy sheep in a New Zealand flock. *N Z J Agric Res*. 2024;68(5):1–18.
61. Bertelsen H, Gregersen V, Poulsen N, et al. Detection of genetic variation affecting milk coagulation properties in Danish Holstein dairy cattle by analyses of pooled whole-genome sequences from phenotypically extreme samples (pool-seq). *J Anim Sci*. 2016;94: 1365–1376.