Kudriashova, A., Chamberlain, A., Tahir, S., Liu, Z., Rochfort, S., Mason, B., Elkins, A., Ram, D., Hemsworth, J., Reich, C., Goddard, M. (2026) Multi-omics identification of blood biomarkers for transition cow disease and associated quantitative trait loci, *Proceedings 13th World Congress on Genetics Applied to Livestock Production* 23897: https://doi.org/10.31274/wcgalp.23897
©2026 Kudriashova, et al. CC BY-NC-ND

Multi-omics identification of blood biomarkers for transition cow disease and associated quantitative trait loci

A. Kudriashova[1,2*], A. Chamberlain[1,3], S. Tahir[1], Z. Liu[1], S. Rochfort[1], B. Mason[1], A. Elkins[1], D. Ram[1], J. Hemsworth[1], C. Reich[1], and M. Goddard[1,2]

[1]*Centre for AgriBioscience, Melbourne, Australia;* [2]*University of Melbourne, Melbourne, Australia;* [3]*La Trobe University, Melbourne, Australia;*
[*]*Corresponding author:* anastasiia.kudriashova@agriculture.vic.gov.au

Introduction

The transition period, approximately three weeks before and three weeks after calving, is the most demanding phase of a dairy cow's lactation cycle. During this time, cows undergo significant metabolic and physiological adaptation that increases their vulnerability to health disorders including mastitis, metritis, retained foetal membranes, milk fever, and ketosis. These conditions affect animal welfare, milk production, increase veterinary costs, and contribute to premature culling, placing substantial burden on dairy farm profitability (Leblanc, 2010).

Despite ongoing genetic improvement, current breeding strategies in Australia do not fully capture the biological complexity of transition health. Transition diseases exhibit low heritability, and incomplete clinical records in Australian herds, limiting the accuracy of estimated breeding values (Pryce *et al*., 2016). Intermediate phenotypes, or endophenotypes, offer a promising alternative. Because they more directly reflect underlying biological processes, molecular endophenotypes such as serum blood biomarkers tend to show higher heritability and can serve as informative proxies for complex health traits (Luke *et al*., 2019).

Recent advances in omics technologies have enabled characterisation of metabolic changes during the transition period. Studies using single omics approaches have identified candidate biomarkers associated with disease risk (Hailemariam *et al*., 2014). However, they capture only partial biological information, and few studies have integrated multiple omics layers or established the genetic basis of candidate biomarkers. Furthermore, most investigations have been conducted in confined housing, with limited data from pasture-based dairy systems (common in Australia) where nutritional and metabolic challenges differ substantially.

This study integrates lipidomics, metabolomics, and transcriptomics with genome-wide association analysis (GWAS) to: (1) identify serum biomarkers differentiating diseased from healthy transition cows, (2) evaluate their heritability, and (3) assess predictive performance for disease classification. This multi-omics approach aims to identify heritable molecular biomarkers for incorporation into genomic selection to improve transition cow health.

Methodology

Animals and sampling. A case-control study was conducted on 15 pasture-based dairy farms in Victoria, Australia. Blood samples were collected transition cows within 70 days post-calving, with the number of animals included in each analysis reported in Table 1. Diseases included mastitis (n=51), milk fever (n=50), endometritis (n=38), retained foetal membrane (n=34), metritis (n=18), and clinical ketosis (n=13); some animals had multiple conditions. Blood was collected from the coccygeal vein into serum and EDTA vacutainers. Serum was separated by centrifugation (1,500 × g, 15 min), whole blood for transcriptomics was preserved in RNAprotect Animal Blood Tubes and stored at −80°C, aliquots for genotyping stored at −20°C.

Omics profiling. Serum lipidomic profiling was conducted using reversed-phase liquid chromatography (LC–MS) following Liu et al. (2025). Untargeted metabolomics employed LC–MS/MS in positive and negative ionization modes. Raw metabolomics data were filtered to remove metabolites with missingness ≥30%, corrected for within-batch drift using QC-based GAM smoothing and between-batch median scaling, missing data were imputed by kNN, and $\log_2$-transformed before statistical analysis. RNA-seq libraries were prepared with the NEXTFLEX Rapid Directional RNA-Seq Kit 2.0 (Revvity) and sequenced (2×150 bp) on a DNBSEQ-G400, yielding ~50 M reads per library. Adapters were removed with QuadTrim (Robinson et al., 2019). Reads were aligned to a customised ARS-UCD2.0 bovine reference using STAR v2 (Dobin *et al*., 2013) in dual-pass mode with Ensembl annotation, duplicates removed with Picard MarkDuplicates, and gene counts generated using featureCounts (Liao *et al*., 2014). All animals were genotyped with the AgVic custom 50K Illumina BeadChip. Genotypes were filtered (GenCall < 0.6 set to missing; >10% missing SNPs/animals removed) and sporadic missing genotypes imputed. Data were imputed to 74K variants using FImpute v3 (Sargolzaei *et al*., 2014), then to the Bovine HD panel using a 2,814-animal reference. HD genotypes (714,452 SNP) and the Run9 1000 Bull Genomes reference (5,036 taurine genomes; Beagle Rsq > 0.9; MAC > 3; high-quality VQSR; excessive heterozygosity removed; Eagle-phased) were imputed to whole-genome sequence using Beagle v5.2 (Ne = 1000). Final retained variants had Beagle imputation accuracy > 0.9 and MAF > 0.05.

Data analysis. GWAS for all samples and differential abundance analysis (DA) between diseased and control groups were performed for each omics layer. QTL mapping identified phenotypes associated with whole genome sequence genotypes using MLMA in GCTA with heritability estimated via GREML (Yang *et al*., 2011). DA (up and downregulated phenotypes) used limma for lipidomics and metabolomics, and edgeR/limma-voom for gene expression (Ritchie *et al*., 2015; Robinson *et al*., 2009). Class and pathway enrichment analyses were performed for differentially abundant phenotypes in each omics. Metabolomics pathway enrichment used the mummichog algorithm (MetaboAnalyst 6.0, www.metaboanalyst.ca). Significantly altered lipids were tested for class representation using Fisher's exact test, comparing the frequency of each category within the significant DA lipids to its frequency in the full lipidomics dataset. Gene expression pathway enrichment used KEGG-based over-representation analysis to identify biological pathways significantly impacted by DA genes. Integration analysis identified common QTLs across omics layers and overlap between GWAS hits and DA phenotypes.

Biomarker discovery. Machine-learning (ML) algorithms were used to predict disease status for all features and GWAS phenotypes separately using Elastic Net and Random Forest (Uddin *et al*., 2019) predictive models through 5×10 repeated cross-validation, with a 70/30 train-test split. Model performance was assessed using area under the ROC curve (AUC) with 95% confidence intervals. Feature importance ranking in predictive models identified biomarkers significantly associated with disease state. Biomarkers selected by both ML methods and DA were designated as consensus biomarkers. All GWAS, DA, and ML models included farm, breed, parity, days in milk, sampling date, haemolysis as covariates, and batch effect included in metabolomics analyses. Statistical significance was defined as FDR < 0.05.

Results & Discussion

Integrating multi-omics data with QTL mapping, DA, and ML identifies heritable, disease-associated biomarkers for early diagnosis and genomic selection in transition cows. QTL mapping revealed lipid associated loci on chromosomes 11 and 19, while metabolite QTL were distributed across chromosomes 10, 11, 16, 18, 23, and 25. Gene expression QTL were detected

in the same regions as lipid QTL on chromosome 19 and with metabolite QTLs on chromosomes 10 and 16, suggesting that transcriptional regulation may play a mediating role in the observed metabolic dysregulation.

DA identified predominantly downregulated features in diseased cows across all omics layers (Table 1). Lipid class enrichment showed reduced lysophosphatidylcholines (LPC) and elevated carnitines, reflecting changes in lipid mobilisation, immune responses and energy metabolism. Tryptophan metabolism and the citrate (TCA) cycle enriched in metabolomics indicated alterations in amino acid and energy metabolism pathways. Gene expression pathways were related to compromised immune function (downregulated cell adhesion molecule interactions, primary immunodeficiency, and Th17 cell differentiation pathways) and increased oxidative stress (oxygen species pathway).

Table 1 Multi-omics summary of differential abundance analysis, genome wide association studies, QTL counts, and their overlap (FDR < 0.05).

Omics	Sample size	Total[1]	GWAS features[2]	QTLs	DA Total (Up/Down)	DA/GWAS overlap
Lipidomics	524	238	36	3	163 (21/142)	11
Metabolomics (neg)	518	1173	7	6	512 (119/393)	1
Metabolomics (pos)	518	1228	7	4	357 (106/251)	1
Gene expression	508	13902	1340	188	8042 (3616/4426)	756

[1] Total number of features identified in omics dataset; [2] Number of GWAS associated feature linked to at least one QTL; [3] Total QTLs detected per omics layer; [4]DA – differential abundance analysis showing the total number of significantly up- and down-regulated features; [5] Number of features overlapping between DA and GWAS results.

Table 2. Predictive performance and heritability of differentially abundant and GWAS biomarkers across multi-omics layers

Omics	Feature Set[1]	Best AUC[2]	Top consensus biomarker[3]	Direction[4]	h^2 (SE)[5]
Gene expression	All	0.811	ENSBTAG00000006221	Up	0.30 (0.19)
Gene expression	GWAS	0.802	ENSBTAG00000076194	Up	0.29 (0.15)
Lipidomics	All	0.775	LPC14.0	Down	0.35 (0.16)
Lipidomics	GWAS	0.786	PI33.1	Down	0.30 (0.16)
Metabolomics (neg)	All	0.824	mz_191.02_rt_1.83	Down	0.32 (0.17)
Metabolomics (neg)	GWAS	0.680	mz_267.07_rt_2.36	-	0.51 (0.16)
Metabolomics (pos)	All	0.823	mz_139.00_rt_1.83	Down	0.31 (0.16)
Metabolomics (pos)	GWAS	0.618	mz_959.48_rt_4.74	Down	0.62 (0.16)

[1] Feature set used for model training (DA = differentially abundant features; GWAS = features with significant genetic associations); [2] Best model AUC obtained from two ML algorithms (Random Forest, Elastic Net); [3] Top consensus biomarker selected by ML models and DA; [4] Direction of biomarker change in diseased cows; [5] Heritability estimate (h^2) with standard error in parentheses. For metabolomics, unidentified metabolites are reported using mass-to-charge ratio (mz) and retention time (rt).

The performance gap between ML models built using all features versus GWAS-associated features was small for gene expression and lipidomics. The small number of GWAS metabolites (Table 1) limited the model's ability to predict disease status, which explains the performance gap between models using all features and those using only GWAS phenotypes (Table 2). We identified 40 consensus biomarkers from lipids, 18 from metabolites, and 39 from genes. Heritability estimates for these biomarkers ranged from 0 to 0.79, indicating low to high genetic control. This suggests that for these omics layers, genetic variants capture

substantial disease-related signal, supporting the potential utility of these biomarkers for genomic selection applications.

Conclusion

The integration of molecular biomarkers with genomic markers provides a deeper biological understanding of transition cow disease processes and enables the identification of both disease indicators and putative causal genetic variants. By combining multi-omics phenotypes with genotype data, this approach supports the development of more accurate genomic breeding values for transition-health traits. Ultimately, these advancements offer a pathway to improving disease resilience, enhancing animal welfare, and increasing farm profitability through more informed and biologically grounded selection strategies.

Acknowledgement

This study was undertaken as part of the DairyBio program, which is jointly funded by Dairy Australia, the Gardiner Foundation, and Agriculture Victoria and is a PhD research project at the University of Melbourne.

Reference

Dobin A., Davis C. A., Schlesinger F., *et al*. (2013). Bioinformatics 29, 15-21. https://doi.org/10.1093/bioinformatics/bts635

Hailemariam, D., Mandal, R., Saleem, F., *et al*. (2014). J Dairy Sci 97, 2680-2693. https://doi.org/10.3168/jds.2013-6803

Leblanc, S. (2010). Journal of Reproduction and Development 56, 29-35. https://doi.org/ 10.1262/jrd.1056S29

Liao Y., Smyth G.K., and Shi W. (2014). Bioinformatics 30, 923-930. https://doi.org/ 10.1093/bioinformatics/btt656

Liu Z., Wang W., Hemsworth J.E., *et al*. (2025). Metabolites 15(4). https://doi.org/10.3390/metabo15040274

Luke T.D.W., Nguyen T.T.T., Rochfort S., *et al*. (2019). J Dairy Sci 102, 11142-11152. https://doi.org/10.3168/jds.2019-17127

Pryce J.E., Parker Gaddis K.L., Koeck A., *et al*. (2016). J Dairy Sci 99, 6855-6873. https://doi.org/10.3168/jds.2016-10854

Ritchie M.E., Phipson B., Wu D., *et al*. (2015). Nucleic acids research 43, 47-47. https://doi.org/10.1093/nar/gkv007

Robinson M.D., McCarthy D.J., and Smyth G.K. (2009). Bioinformatics 26, 139-140. https://doi.org/10.1093/bioinformatics/btp616

Sargolzaei M., Chesnais J.P., and Schenkel F.S. (2014). BMC genomics 15, 478. https://doi:10.1186/1471-2164-15478

Uddin S., Khan A., Hossain M.E., *et al*. (2019).. BMC Med Inform Decis 19, 1-16. https://doi.org/10.1186/s12911-019-1004-8

Yang J., Lee S.H., Goddard M.E., *et al*. (2011). Am J Hum Genet 88, 76-82. https://doi.org/10.1016/j.ajhg.2010.11.011