Chen, Y., Atashi, H., Fang, L., Grelet, C., Gengler, N., Li, B. (2026) Integrative analysis of genomic, transcriptomic, and mid-infrared spectra reveals the genetic basis of nitrogen use efficiency in dairy cattle, *Proceedings 13th World Congress on Genetics Applied to Livestock Production* 23551: https://doi.org/10.31274/wcgalp.23551 ©2026 Chen, et al. CC BY-NC-ND

Integrative analysis of genomic, transcriptomic, and mid-infrared spectra reveals the genetic basis of nitrogen use efficiency in dairy cattle

Y. Chen[1,2*], H. Atashi[2], L. Fang[3], C. Grelet[4], N. Gengler[2], and B. Li[3,5]

*[1]Hunan Agricultural University. Changsha, China; [2]University of Liège, Gembloux, Belgium; [3]Aarhus University, Aarhus, Denmark; [4]Walloon Agricultural Research Center, Gembloux, Belgium; [5]Scotland's Rural College (SRUC), Edinburgh, UK; *Corresponding author: yansen.chen@hunau.edu.cn*

Introduction

Nitrogen use efficiency (**NUE**) is an essential trait for enhancing both the economic and environmental sustainability of dairy production (Chen *et al*., 2021). However, NUE in dairy cattle ranges from only 15% to 80%~~%~~ (Cantalapiedra-Hijar *et al*., 2018; Grelet *et al*., 2020), meaning that more than half of the nitrogen intake is excreted rather than converted into milk protein. The recent development of milk mid-infrared (**MIR**) spectra-based prediction models for NUE (Grelet *et al*., 2020) has enabled large-scale phenotyping and opened new opportunities for genetic improvement. Although preliminary genetic analyses of predicted NUE have been conducted (Chen *et al*., 2021), its underlying genetic mechanisms remain poorly understood. Single-step GWAS (**ssGWAS**) can help elucidate the genetic architecture of complex traits, including NUE. Transcriptome-wide association study (**TWAS**), colocalization (**COLOC**), and Mendelian randomization (**MR**) can further integrate genomic and transcriptomic evidence to identify causal genes.

This study aimed to (1) characterize the genetic architecture of NUE across early, mid, and late lactation stages using ssGWAS; (2) identify genes putatively associated with NUE across the three stages through TWAS, COLOC, and MR; and (3) prioritize candidate genes using a newly developed global multi-omics score (**gMOS**) that integrates multi-source evidence from GWAS, TWAS, COLOC, and MR.

Materials & Methods

This study used two datasets. The **first dataset**: 1,918,884 test-day records collected between 2019 and 2023 as part of the official milk recording scheme in the Walloon region of Belgium. It included MIR spectra, milk yield, parity, and 50K SNP genotypes of cows. The genotypes were imputed and subjected to quality control, resulting in 564,060 SNPs distributed across 29 chromosomes (CHR), following the procedures described in Chen *et al*. (2023). The MIR spectra were first standardized following Grelet *et al*. (2015) and combined with milk yield and parity to predict NUE. Daily NUE records were then categorized into three lactation stages based on days in milk (**DIM**) of each parity: early (**NUE1**, 5–50 DIM), mid (**NUE2**, 51–200 DIM), and late (**NUE3**, 201–365 DIM). For each cow, NUE values within each stage were averaged, and only cows with NUE available for all three stages were retained for each parity. In total, 97,510 records from 59,909 cows across 466 herds, derived from the first five parities, were kept for analysis. The pedigree file contained 170,550 animals, among which 7,199 cows were genotyped for 564,060 SNPs. The **second dataset:** all transcriptomic data from the cattle Genotype-Tissue Expression Project (CattleGTEx) resource were applied to this study including all *cis*-QTL (**eQTL**) of cattle genome across 23 cattle tissues derived earlier using > 8,000 RNA-seq data (Liu *et al*., 2022).

ssGWAS. Using the combined phenotypic, pedigree, and genotypic information, three single-trait repeatability models were applied to estimate variance components and genomic breeding values. The model was as follows:

$$\mathbf{y} = \mathbf{Hh} + \mathbf{Xb} + \mathbf{Za} + \mathbf{Wp} + \mathbf{e} \quad (1)$$

where **y** was the vector of NUE (NUE1 or NUE2 or NUE3), **h** was the herd × year-season of calving classes, **b** was the parity × age of calving, **a** is the additive genetic effects, **p** is the permanent environment, **e** is the residual. **H**, **X**, **Z**, **W** are the incidence matrices assigning observations to effects.

The SNP effects for the three traits were obtained by back-solving the animals' GEBV following Wang *et al*. (2012). SNP P-values were computed according to the procedure described in Aguilar *et al*. (2019). A significance threshold of 8.86×10^{-8} (0.05/564,060) was applied, based on the Bonferroni correction for multiple testing. Significant SNPs located within 500 kb of each other were grouped into a single genomic region for gene annotation and subsequent COLOC analyses. All computations were performed using the BLUPF90 software family (Misztal *et al*., 2014).

TWAS, COLOC, and MR. TWAS was conducted using the S-PrediXcan framework by integrating the GWAS summary statistics for NUE with pre-trained Elastic Net expression models from CattleGTEx for 23 tissues/cell types. All TWAS computations were performed via the FarmGTEx TWAS Server (https://twas.farmgtex.org/). Colocalization analysis was conducted using the COLOC R package to identify genes for which ssGWAS and *cis*-eQTL signals were likely driven by the same causal variant (PP.H4 $\geq$ 0.80). For these colocalized genes, MR was performed using the TwoSampleMR R package, with *cis*-eQTL SNPs selected as instruments through a stepwise *P*-value threshold (5×10^{-6} to 10^{-3}) requiring at least three variants reaching the significance threshold in the *cis*-eQTL region. Exposure and outcome datasets were harmonized, and MR was conducted using weighted mode, with significant causal effects declared at *P*-value < 0.05.

Global multi-omics score. To rank and prioritize candidate genes incorporating results from multiple metrics, we developed a global multi-omics score (**gMOS**) that quantifies the importance of each gene over time by synthesizing evidence from GWAS, TWAS, COLOC, and MR. Method-specific weights were assigned to time and tissue dimensions, and the overall evidence score for each gene was defined as:

$$\mathrm{gMOS} = \mathrm{TissueScore} + \mathrm{TimeScore} + \mathrm{MR_consistent} \quad (2)$$

where the tissue-level component is $\mathrm{TissueScore} = 0.5 \times \sum \mathrm{TWASorg} + 1.0 \times \sum \mathrm{COLOCorg} + 2.0 \times \sum \mathrm{MRorg}$, and the time-level component is $\mathrm{TimeScore} = 1.0 \times \mathrm{GWAStime} + 0.5 \times \sum \mathrm{TWAStime} + 1.0 \times \sum \mathrm{COLOCtime} + 2.0 \times \sum \mathrm{MRtime}$. One additional modifier was included: MR_consistent = +0.5 for genes showing consistent MR effect directions across $\geqslant$2 lactation stages. The top eight candidate genes, as determined by the gMOS ranking, were selected for visualization.

Results & Discussion

Across lactation, NUE showed a declining trend from early to late stages, with mean values (± SD) of 30.10 ± 5.77% for NUE1, 25.64 ± 3.94% for NUE2, and 24.68 ± 3.61% for NUE3.These values are consistent with previous reports (Chen *et al*., 2021). The ssGWAS results showed similar genetic architectures across lactation stages (Fig. 1A), with major peaks on CHR14 (0.14–2.19 Mb) and CHR25 (25.91–27.68 Mb). Notably, the CHR14 region was also associated with the nitrogen efficiency index (Chen *et al*., 2023). In total, 334 significant SNPs (P < 0.05/564,060) were detected. Clustering significant SNPs within 500

kb yielded 17 genomic regions, annotated to 243 unique genes. Among these, 63 genes were shared across all three traits and were located on CHR14 and CHR25. TWAS detected 47 unique NUE-associated genes with significant effects across 21 tissues, with the strongest enrichment in blood, muscle, and liver/uterus (Fig. 1B). However, the number of significant genes differed among the three traits (15 vs. 64 vs. 45), as did the number of enriched tissues (9 vs. 20 vs. 19). COLOC detected 34 unique genes with significant effects across 17 tissues, with enrichment primarily in blood, muscle, and monocytes (Fig. 1D). COLOC identified a relatively balanced number of genes across the three traits (42 vs. 44 vs. 38), and the number of enriched tissues was also comparable (14 vs. 16 vs. 14). Mammary and liver tissues also showed stable but moderate levels of colocalization. For instance, in mammary tissue, *DGAT1* showed strong eQTL signals and a clear association with NUE1 (Fig. 1C). MR detected 8 unique genes with significant effects across 9 tissues (40 genes in total), with the strongest enrichment in blood and monocytes (12 each). As monocytes are immune cells circulating in blood, the concordant enrichment suggests a causal contribution of immuno-metabolic regulation to NUE.

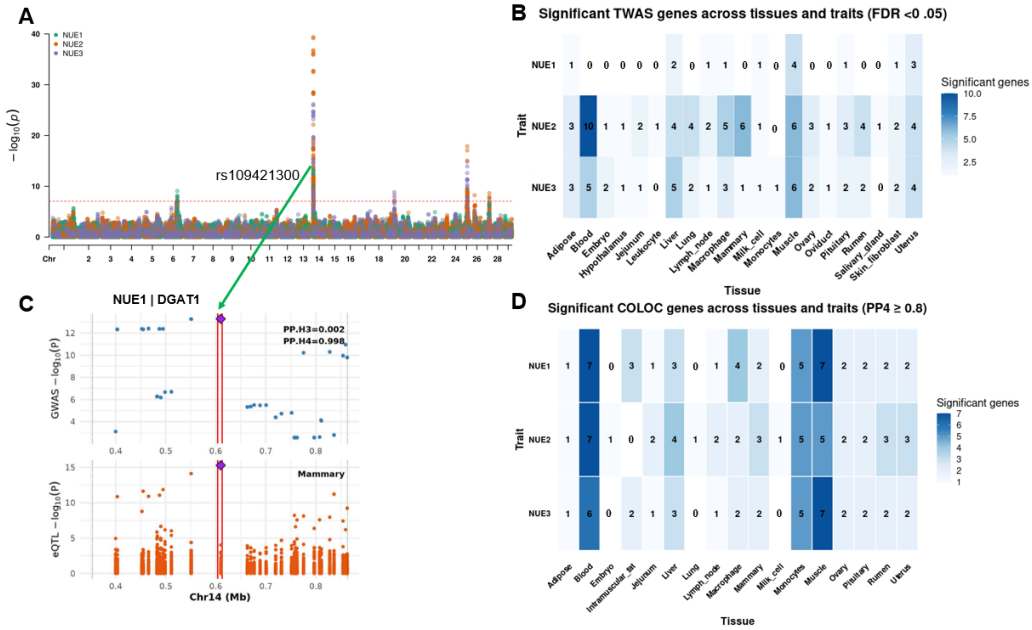

Figure 1. Genetic analyses of nitrogen use efficiency (NUE) across early (NUE1), mid (NUE2), and late (NUE3) lactation. A: Manhattan plots from ssGWAS for NUE1–NUE3. B: Number of significant TWAS genes (FDR < 0.05) detected across 21 tissues. C: Local ssGWAS–eQTL colocalization around *DGAT1* for NUE1 in mammary tissue. D: Number of significant COLOC genes (PP.H4 ⩾ 0.80) detected across 17 tissues.

Fig. 2A summarizes the 258 significant genes detected across the four analytical approaches, with *DGAT1* being the only gene consistently identified by all four analyses. Using our gMOS framework to integrate evidence across methods, tissues, and lactation stages, *DGAT1* also received by far the highest aggregated score among all candidates (Fig. 2B), demonstrating strong cross-method convergence. Compared to Fonseca *et al*. (2025) study, gMOS jointly incorporates association (ssGWAS), predicted expression (TWAS), colocalization (COLOC), causal inference (MR), and recurrence across tissues and stages. This multi-layered integration provides stronger evidence for gene prioritization and explains the molecular mechanism why *DGAT1* emerges as the most consistently supported candidate for NUE. In future work, the gene-level scores generated by gMOS will be incorporated into weighted genomic selection models to further refine breeding strategies aimed at NUE.

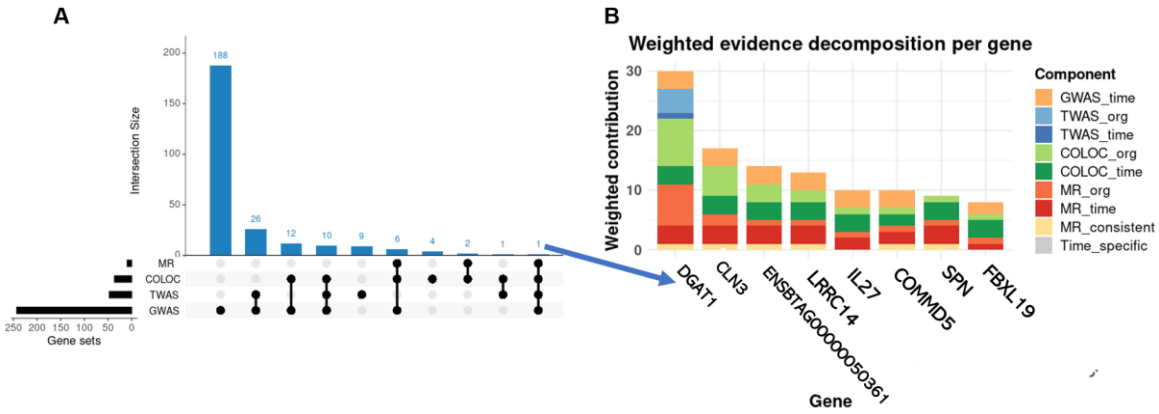

Figure 2. Integration of multi-omics evidence to prioritize candidate genes for nitrogen use efficiency. A: Overlap of genes detected by ssGWAS, TWAS, COLOC, and MR. B: Weighted multi-omics contributions for the eight top-ranked genes.

Conclusions

Using large-scale GWAS across three lactation stages and multi-omics post-GWAS analyses including TWAS, COLOC, and MR, our gMOS framework prioritized eight core genes, with *DGAT1* identified as the top candidate, for NUE.

Acknowledgements

This work was supported by COST Action CA22112 (EU-LI-PHE). Y. Chen acknowledges funding from the Luxembourg National Research Fund (FNR, grant no. 17939764, NITMETFUN). B. Li acknowledges UK Biotechnology and Biological Sciences Research Council (BBSRC) grant BB/X009505/1. Computational resources were provided by CÉCI, funded by FNRS (grant no. 2.5020.11). The authors also thank CattleGTEx project for providing eQTL data.

References

Aguilar I., Legarra A., Cardoso F., *et al*. (2019) Genet. Sel. Evol. 51:1–8. https://doi:10.1186/s12711-019-0469-3

Chen Y., Vanderick S., Mota R.R., *et al*. (2021) J. Dairy Sci. 104:4413–4423. https://doi:10.3168/jds.2020-18849

Chen Y., Atashi H., Grelet C., *et al*. (2023) J. Dairy Sci. 106:3397–3410. https://doi:10.3168/jds.2022-22351

Dewhurst R.J., Cheng L., Cabrita A.R.J., *et al*. (2018) Anim. 12:1827–1837. https://doi:10.1017/S1751731117003391

Fonseca P.A.S., Suárez-Vega A., Casas L., *et al*. (2025) PNAS Nexus:pgaf361. https://doi:10.1093/pnasnexus/pgaf361

Grelet C., Froidmont E., Foldager L., *et al*. (2020) J. Dairy Sci. 103:4435–4445. https://doi:10.3168/jds.2019-17910

Grelet C., Pierna J.A.F., Dardenne P., *et al*. (2015) J. Dairy Sci. 98:2150–2160. https://doi:10.3168/jds.2014-8764

Liu S., Gao Y., Canela-Xandri O., *et al*. (2022). A multi-tissue atlas of regulatory variants in cattle. Nat Genet. 54:1438-1447

Misztal I., Tsuruta S., Lourenco D., *et al*. Manual for BLUPF90 family of programs http://nce.ads.uga.edu/wiki/lib/exe/fetch.php?media=blupf90_all8.pdf, visited on: 13-11-2025

Wang H., Misztal I., Aguilar I., *et al*. (2012) Genet. Res. 94:73–83. https://doi:10.1017/S0016672312000274