Proc. Assoc. Advmt. Anim. Breed. Genet. 26: 233-236

TISSUE-SPECIFIC RNA EXPRESSION DATA ENHANCES GENOMIC PREDICTION OF METHANE EMISSIONS IN SHEEP

S.W. Alemu, T.P. Bilton, B.J. Perry, H.M. Henry, K.M. McRae, K.G. Dodds, J.C. McEwan, S.M. Clarke and S.J. Rowe

AgResearch, Invermay Agricultural Centre, Private Bag 50034, Mosgiel 9053, New Zealand

SUMMARY

Reducing methane emissions in sheep through selective breeding requires accurate genomic prediction methods. This study investigated whether incorporating RNA tissue expression data could enhance prediction accuracy for methane emissions. Analysis was conducted on 1,366 sheep, including RNA sequencing data from three tissues (longissimus muscle, rumen wall, and rumen papillae) in a subset of 48 animals. To address the high dimensionality of RNA expression data (>20,000 genes), we employed principal component analysis (PCA) and selected components based on their correlation with methane emissions. These selected components were integrated as intermediate traits in a Neural Network Genomic Best Linear Unbiased Prediction (NNGBLUP) model, evaluating up to nine principal components (PCs) sequentially for each tissue. While the traditional GBLUP approach demonstrated robust performance within our flock sheep population, the integration of tissue-specific RNA expression data through selected PCs provided modest improvements in prediction accuracy. These findings suggest that RNA tissue expression data can complement existing genomic prediction methods, despite constraints from limited sample size and genetic diversity. Our results highlight the potential of incorporating molecular phenotypes in breeding programs aimed at reducing methane emissions in sheep.

INTRODUCTION

Addressing methane emissions from sheep is key for enhancing agricultural sustainability, especially in regions where sheep farming plays a significant economic role. Current genomic prediction methods for methane emissions rely primarily on genetic markers, but these approaches may not fully capture the biological complexity of methane production. Our study focuses specifically on the potential of RNA tissue expression data to enhance prediction accuracy for methane emissions in sheep. Previous research has demonstrated the value of genomic prediction methods using rumen microbiome information (Hess *et al.* 2023), but the role of tissue-specific gene expression in methane production remains poorly understood. This study aims to evaluate whether RNA expression profiles from three tissues - rumen wall, rumen papillae, and longissimus muscle - can improve the accuracy of genomic prediction for methane emission beyond traditional genomic approaches. While rumen wall and papillae tissues are directly involved in digestive processes and methane production, the inclusion of longissimus muscle allows us to explore diverse biological information that might enhance genomic prediction accuracy. By incorporating RNA expression data from these tissues into prediction models, we aim to develop more accurate genomic prediction strategies for methane emissions in sheep breeding programs.

MATERIALS AND METHOD

The animal experiments conducted adhered strictly to the guidelines of the 1999 New Zealand Animal Welfare Act and the AgResearch Code of Ethical Conduct. The trials were approved by the AgResearch Animal Ethics Committee (approval number 15601).

Genotype. All animals were genotyped using a 15K SNP chip, which directly focused on the available genetic markers without any imputation to higher densities. Filtering was performed such that only autosomal SNPs were retained, and SNPs with a minor allele frequency greater than 0.01 and a call rate of 70% were included. This resulted in a total of 13,916 SNPs, with a mean call rate of 0.96. A genomic relationship matrix (GRM) was constructed using the KGD software (Dodds *et al.* 2015). For each pair of individuals, only SNPs with non-missing genotypes were used to compute the genomic relationship, assuming missingness at random.

RNA-seq. Tissue samples (rumen wall, rumen papillae, and longissimus muscle) were collected from 48 males born in 2021. Females were excluded to maintain the breeding population, ensuring herd sustainability. Total RNA from the snap-frozen tissue was extracted using the NucleoSpin RNA extraction kit (Macherey-Nagel), according to the manufacturer's guidelines, following the protocol for RNA purification from cultured cells and tissue. For the longissimus sample, the protocol for difficult-to-lyse tissues was followed. Sequencing libraries were made with the Illumina Stranded Total RNA with Ribo-Zero Plus and sequenced on the Illumina Novaseq 6000. Analysis began with a filtering step, performed independently for each tissue type, where genes with zero counts in more than 10% of samples were removed. The filtered data were transformed using a centred log-ratio transformation using the proper package and standardized by scaling to zero mean and unit variance. Principal component analysis (PCA) was then conducted on each individual tissue.

Prediction Set-Up. The dataset was divided into training and validation sets to test the models' predictive power. The training set included data from 2019 to 2021 (804 animals), while data from 2022 and 2023 (562 animals) served as the validation set. A subset of 48 animals, born in 2021 but recorded in 2022, was enriched with RNA data for analysis.

Table 1. Distribution of experimental sheep by birth year, gender, and RNA sequencing

Year of birth	Males	Females	Total animals	RNA sequenced
2019	130	145	275	-
2020	126	135	261	-
2021	128	140	268	48(males)
2022	124	136	260	-
2023	142	160	302	-
Total	650	716	1,366	48

Statistical analysis. The models included fixed effects of birth and rearing rank, dam age (1, 2, 3+ years), and contemporary group (flock-birth year combination). Birthday deviation from the flock-year mean was used as a covariate to adjust phenotypes. Enteric methane emissions were measured using portable accumulation chambers (PAC). Animals were off feed for 1–4 hours after 3 days of ad-lib grazing on pasture with sufficient dry matter. Using the adjusted methane phenotype, this study employed two genomic prediction approaches. The first method was Genomic Best Linear Unbiased Prediction (GBLUP), which served as our baseline model. The second approach was Neural Network GBLUP (NN-GBLUP), implemented within a Bayesian framework which extended traditional GBLUP by incorporating RNA expression data through a neural network structure (Zhoa *et al.* 2022). This model processed information through three layers: an input layer containing genomic data, an intermediate layer representing RNA expression information through principal components (PCs), and an output layer predicting methane emissions. The analysis involved 60,000 chains with a burn-in phase of 10,000, sampled every 10 steps and implemented in JWAS software (Cheng *et al*. 2018).

To identify the most informative RNA expression patterns, we first performed principal component analysis and selected PCs based on their correlation with methane emissions in the subset

of 48 animals with RNA data. We systematically evaluated a series of models incorporating different numbers of principal components (PCs). For each tissue, we created nine distinct models by incrementally including PCs that showed the strongest correlations with methane emissions. Starting with the two most strongly correlated PCs, we sequentially added components to create models with 3 to 10 PCs. The selected PC based on correlation for the three tissues is shown in Table 2.

Table 2. Selected principal components by tissue type based on correlation with methane emissions

Tissue	Selected PC based on correlation with methane
Longissimus	44, 10, 43, 2, 24, 37, 1, 25, 3, 12
Rumen Papillae	35, 8, 24, 30, 25, 16, 41, 45, 32, 16
Rumen wall	30, 10, 40, 37, 42, 43, 23, 19, 39, 33

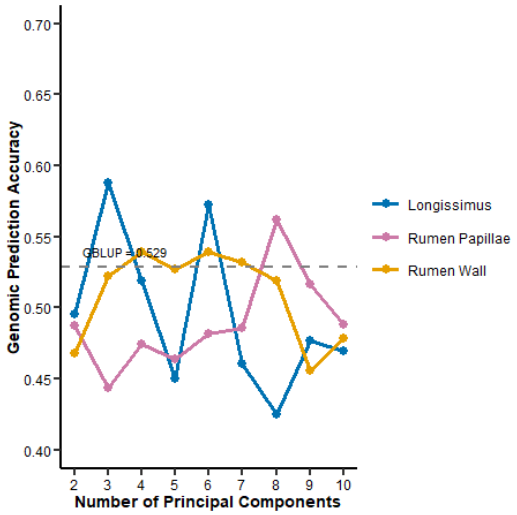

Figure 1. Accuracy of methane emission prediction using tissue-specific principal components

RESULTS AND DISCUSSION

The genomic prediction accuracy for the baseline GBLUP was 0.529 (Figure 1), providing a reference point for evaluating the incorporation of RNA tissue expression data. For rumen wall tissue, while the addition of some PCs resulted in prediction accuracies lower than the GBLUP baseline, specific combinations showed improved performance. The highest prediction accuracy of 0.539 was achieved using PCs: PC1, PC2, PC10, PC24, PC37, and PC43. These specific combinations are detailed in Table 2, where principal components were selected based on their correlation with methane emissions in the subset of 48 animals with RNA data.

In the case of rumen papillae tissue, the integration of 8 PCs resulted in a prediction accuracy of 0.562, representing a substantial improvement over the GBLUP baseline. This improvement suggests that rumen papillae expression patterns may capture important biological signals related to methane production.

Longissimus tissue demonstrated the highest prediction accuracy of 0.588 with 3 PCs, surpassing the GBLUP baseline, suggesting insights into methane production mechanisms despite its indirect role in ruminal processes. While specific studies on muscle gene expression's direct correlation with methane emissions are limited, microbial gene abundance in the gastrointestinal tract has been

shown to influence metabolic processes and traits like methane production (Roehe *et al.*2016). This underscores the potential for muscle tissue gene expression patterns to indirectly impact methane emissions through metabolic pathways.

These results are particularly noteworthy given the constraints of the study design. The data consisted of animals from a single flock, and RNA tissue information was available for only 48 individuals. Under these limitations, the fact that the traditional GBLUP model maintained comparable performance highlights its robustness in scenarios with high genetic relationships within the study population. However, the observed improvements in prediction accuracy when incorporating RNA tissue expression data, albeit modest, suggest potential value in integrating molecular phenotypes for enhancing genomic prediction.

The tissue-specific variations in prediction accuracy improvements indicate that different tissues may capture distinct aspects of the biological mechanisms underlying methane production. Future studies with larger RNA-sequenced populations and diverse genetic backgrounds could further validate and potentially amplify these improvements in genomic prediction accuracy.

CONCLUSION

In conclusion, our study demonstrates a proof-of-concept that integrating RNA tissue expression data can modestly improve genomic prediction accuracy for methane emissions in sheep. While our results show promising improvements, it is essential to acknowledge that these findings are preliminary given the limited sample size of 48 individuals with RNA data and the constraints of a single-flock study design. Further validation across larger and more diverse populations is necessary to solidify and potentially amplify these results.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the AgResearch Animal Genomics Research laboratory and field team farm for the RNA extractions and sequencing, and detailed tissue collection, respectively.

REFERENCES

Cheng H, Rohan F.R. and Garrick D (2018) *Proc. World Congr. Genet. Appl. Livest. Prod.* **11**: 859.

Dodds K.G., McEwan J.C., Brauning R., Anderson R.M., van Stijn T.C., Kristjansson T. and Clarke S.M. (2015) *BMC Genomics* **16**: 1047.

Hess M.K., Zetouni L., Hess A.S., Budel J., Dodds K.G., Henry H.M., Brauning R., McCulloch A.F., Hickey S.M., Johnson P.L., Elmes S., Wing J., Bryson B., Knowler K., Hyndman D., Baird H., McRae K.M., Jonker A., Janssen P.H., McEwan J.C. and Rowe S.J (2023) *Genet. Sel. Evol.* **55**: 53.

Roehe R., Dewhurst R.J., Duthie C.A., Rooke J.A., McKain N., Ross D.W., Hyslop J.J., Waterhouse A., Freeman T.C., Watson M., Wallace R.J. and Yu P. (2016) *PLoS Genet*. **12**: e1005846.

Zhao T., Zeng J. and Cheng H. (2022) *Genetics.* **221**: iyac034.