Blending multivariate models to predict feed efficiency and explore multiple omics in meat sheep

Quentin Le Graverand, Flavie Tortereau, Christel Marie-Etancelin, Annabelle Troegeler-Meynadier, Jean-Louis Weisbecker, Kim-Anh Lê Cao

Blending multivariate models to predict feed efficiency and explore multiple omics in meat sheep

Le Graverand, Q. [1], Tortereau, F. [1], Marie-Etancelin, C. [1], Meynadier, A. [1], Weisbecker, J.L. [1], Lê Cao, K.A.[2]

[1] GenPhySE, Université de Toulouse, INRAE, ENVT, 31326, Castanet-Tolosan, France; [2] School of Mathematics and Statistics, University of Melbourne, VIC 3010, Australia

Selecting sheep for feed efficiency would improve the sustainability of sheep farming by decreasing feeding needs. However, due to the costs of recording feed intake, feed efficiency is rarely selected in sheep. Identifying feed efficiency biomarkers could help resolve this issue. A total of 258 Romane male lambs were phenotyped in the growing period for Residual Feed Intake (RFI) - in three different batches. Rumen fluid and blood were sampled as potential sources of biomarkers for feed efficiency. Multivariate analyses were performed with six distinct 'blocks' of predictors: fixed effects and covariates (FC), genotypes (SNPs), plasma NMR spectra (NMR), ruminal volatile fatty acids (VFAs), long-chain fatty acids (LFAs), bacteria and archaea abundances (16S amplicon sequencing). We modified a Partial Least Square regression approach (PLS) to account for the three batches while selecting biomarkers of feed efficiency (Rohart *et al.*, 2017). Cross-validation was repeated to fit one model per block on our training data (60% of the samples). Then, predictions for the validation set (30% of the samples) were obtained by using a weighted aggregation – based on the performance on each validation set. Testing data (10%) were independently used to assess the overall prediction accuracy based on Pearson correlations. When RFI was predicted from separate blocks, the average accuracy was low to moderate: 0.08 (standard deviation: 0.17) from VFAs to 0.44 (0.13) from SNPs. When RFI was predicted with our approach combining different omics, accuracy increased and reached an average of 0.55 (0.11). Based on weights attributed to blocks of predictors, we were able to rank the most predictive blocks to explain RFI: SNPs, FC, NMR, 16S, LFA and VFA. Furthermore, within each block we identified variables that were highly associated with feed efficiency RFI, including β-hydroxyisovaleric acid and a SNP located on the chromosome 3. To conclude, blending models is useful to integrate heterogeneous omics data: from predicting efficiency, to identifying associations between multi-omics predictors.