東北大学機関リポジトリ
TOUR
Tohoku University Repository

Identification of Predictor Genes of Feed Efficiency in Beef Cattle by Applying Machine Learning (ML) Methods to Multi-tissue Transcriptome Data

Identification of Predictor Genes of Feed Efficiency in Beef Cattle by Applying Machine Learning (ML) Methods to Multi-tissue Transcriptome Data

Weihao CHEN[1,2], Antonio REVERTER[2], Yutao LI[2] and Wei SUN[1]

[1]College of Animal Science and Technology, Yangzhou University
[2]CSIRO Agriculture & Food, Queensland Bioscience Precinct

Machine learning (ML) methods have shown promising results in identifying candidate genes when applied to large transcriptome datasets. However, no attempt has been made to compare the performance of combining different ML methods together in the prediction of high and low feed efficiency (HFE and LFE) animals. In this study, using RNA-seq data of five tissues from 18 Nellore bulls, we evaluated the prediction accuracies of five analytical methods in classifying animals according to their feed efficiency potential. Of five methods, the two-step ML method combining RF and XGBoost (RX), identified the smallest subsets of potential predictor genes across all tissues with the highest classification accuracy for 9 HFE and 9 LFE animals. Besides, genes identified by the RX, there was a correlation between the gene's prediction ranking ("Gain" values) and its relevance to the networks ("Betweenness"), reflecting a key biological role to the phenotype. When comparing co-expression gene network differences between LFE and HFE groups from the RX, the number of connections between genes with maximum expression in skeletal muscle represented the biggest change between HFE and LFE networks. This indicates more FE related pathways activated in HFE. The results demonstrate a great potential for applying a combination of ML methods to large transcriptome datasets to identify biologically important genes for accurately classifying FE samples.