SMARTER

SMAll RuminanTs breeding for Efficiency and Resilience

Research and Innovation action: H2020 – 772787
Call: H2020-SFS-2017-2
Type of action: Research and Innovation Action (RIA)
Work programme topic: SFS-15-2016-2017
Duration of the project: 01 November 2018 – 31 October 2022

GUIDELINES
Complete parts that are highlighted in yellow
Guidelines are highlighted in green

Peer-rewiewed paper on the genetic basis of feed resource use utilisation

Publication of at least one scientific paper with the general objective of assessing the genetic architecture of thecomplex trait of resource use efficiency.

Juan-José Arranz

Universidad de León

* Deliverable leader – Contact: jjarrs@unileon.es

DELIVERABLE D1.3

Workpackage N°WP1

Due date: M50

Actual date: 15/12/2022

Dissemination level: Public

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 772787

About the SMARTER research project

SMARTER will develop and deploy innovative strategies to improve Resilience and Efficiency (R&E) related traits in sheep and goats. SMARTER will find these strategies by: i) generating and validating novel R&E related traits at a phenotypic and genetic level ii) improving and developing new genome-based solutions and tools relevant for the data structure and size of small ruminant populations, iii) establishing new breeding and selection strategies for various breeds and environments that consider R&E traits.

SMARTER with help from stakeholders chose several key R&E traits including feed efficiency, health (resistance to disease, survival) and welfare. Experimental populations will be used to identify and dissect new predictors of these R&E traits and the trade-off between animal ability to overcome external challenges. SMARTER will estimate the underlying genetic and genomic variability governing these R&E related traits. This variability will be related to performance in different environments including genotype-by-environment interactions (conventional, agro-ecological and organic systems) in commercial populations. The outcome will be accurate genomic predictions for R&E traits in different environments across different breeds and populations. SMARTER will also create a new cooperative European and international initiative that will use genomic selection across countries. This initiative will make selection for R&E traits faster and more efficient. SMARTER will also characterize the phenotype and genome of traditional and underutilized breeds. Finally, SMARTER will propose new breeding strategies that utilise R&E traits and trade-offs and balance economic, social and environmental challenges.

The overall impact of the multi-actor SMARTER project will be ready-to-use effective and efficient tools to make small ruminant production resilient through improved profitability and efficiency.

Table of contents

1 Summary

This deliverable presents two manuscripts that compile the results of part of the work done in WP1. These results help to understand the genetic basis of the trait *resource use efficiency* in small ruminants. The study of this trait is the main objective of WP1.

The first manuscript [INIA-UY] analyses different ways of estimating Residual feed intake, which is the main indicator of animal feed efficiency. This study shows different models to estimate RFI in Autralian Merino sheep with meat/wool aptitude. The main result of this experiment indicates that a reduction in the duration of the experimental period from 42 to 35 days of measurement does not yield a reduction in the reliability of the estimates. This aspect is very important since it allows the optimisation of testing station facilities and the evaluation of a more significant number of animals, increasing the predictive capacity of the breeding scheme and reducing its costs.

In the second study, RNAseq-type methodology is used in somatic cells from the milk of high and low-feed-efficiency dairy ewes [UNILEON]. This method allows the sampling of animals without biopsies and to analyse of the gene expression profile of the milk somatic cells, which are an excellent proxy for the secretory cells of the mammary epithelium. These animals were compared through differential expression analysis (DEA) and sparse Partial Least Square-Discriminant analysis (sPLS-DA). In the DEA, 79 genes were identified as differentially expressed between both conditions, while the sPLS-DA identified 261 predictive genes [variable importance in projection (VIP) > 2]. The DEA allowed the identification of genes associated with the immune system and stress in Low-FE animals. In addition, the sPLS-DA approach revealed the importance of genes involved in cell division and cellular lipid metabolic process for the High-FE sheep in the lactating mammary gland transcriptome. A set of discriminant genes, commonly identified by the two statistical approaches, was also detected, including some involved in cell proliferation or encoding heat-shock proteins.

2 Introduction

There are several reasons why studying the genetic basis of feed efficiency in small ruminants is important. One reason is that feed efficiency is a key factor in the profitability of a farm. Small ruminants that are more efficient at converting feed into product (meat, wool and milk) are more profitable for the farmer and make the production system have a lower environmental impact when evaluating the emission of greenhouse gases per kilogram of product. This can lead to cost savings for the farmer and improved sustainability for the livestock sector as a whole.

Another reason is that feed efficiency is an important trait for the health and welfare of small ruminants. For example, dairy animals that are less efficient at converting feed into milk are more likely to suffer from health problems such as obesity, which can negatively impact their overall health and wellbeing. Another reason why feed efficiency is important in small ruminants is because these animals are often raised in environments where feed is scarce or of poor quality. For example, sheep and goats are often raised on marginal lands or in arid regions with limited access to high-quality feed. In these environments, feed efficiency can be a key factor in the survival and productivity of the animals. By understanding the genetic basis of feed efficiency, we can develop breeding programs that aim to improve the health and welfare of small ruminants. Finally, studying the genetic basis of feed efficiency can also help us to understand the underlying biological mechanisms that contribute to this trait. This

knowledge can be used to develop new technologies and approaches that can help farmers to improve the feed efficiency of their animals. Overall, the study of the genetic basis of feed efficiency in small ruminants is important for improving the profitability and sustainability of the livestock sector and promoting animal health and welfare.

One of the objectives of WP1 of the SMARTER project is the knowledge of the genetic basis of the trait "Feed efficiency" in small ruminants. For this purpose, tools of classical quantitative genetics have been used, and it has been possible to determine the genetic parameters of the main indicators of feed efficiency. As presented in deliverable 1.2, FE traits have moderate heritability and variable genetic correlations with the different production traits. In any scenario analysed, there is a clear consensus that the feed efficiency trait can be used in the different breeding programs for meat/wool and dairy animals. In any case, it must be taken into account that, to estimate feed efficiency, it is necessary to precisely analyse feed intake, animal weight, and the quantity and quality of the product. This makes it complicated to implement this character in small ruminant breeding programs. For this reason, alternatives have been investigated throughout the project to simplify feed efficiency estimation processes without losing precision. Therefore, the first chapter of this deliverable refers to an experiment, the results of which have been published in the journal Livestock Science (Amarilho-Silveira et al., 2022) and which proposes a one-week reduction in the classical estimation procedures and which would allow the analysis of 20% more animals in the same evaluation period.

On the other hand, the SMARTER project also proposes using approaches with genomic tools to analyse the genetic architecture of feed efficiency, such as genome-wide association analysis (GWAS). For this purpose, SNP genotyping in the different commercial populations of sheep and goats measured for feed efficiency is undergoing. The project also proposes using certain molecular markers analysed in task 1.1 derived from the analysis of the transcriptome or epigenome of the somatic cells of the milk to unravel the biological basis of feed efficiency.

In the second chapter of this deliverable, the results of the GWAS analyses are not yet available and therefore, the results of the study conducted in the experimental population of Assaf breed where feed efficiency was measured, and RNAseq analysed mammary transcriptome expression are presented. This study (Suarez Vega et al., under review) is currently under review in the Journal of Dairy Science and is expected to be accepted for publication in the coming weeks.

3 Manuscript 1: *Residual feed intake for Australian Merino sheep estimated in less than 42 days of trial* (Amarilho-Silveira., 2022)

This manuscript is published in open access at the link at the end of the reference.

Amarilho-Silveira, F., de Barbieri, I., Cobuci, J. A., Balconi, C. M., de Ferreira, G. F., & Ciappesoni, G. (2022). Residual feed intake for Australian Merino sheep estimated in less than 42 days of trial. *Livestock Science*, *258*, 104889. https://doi.org/10.1016/J.LIVSCI.2022.104889

Livestock Science 258 (2022) 104889

Contents lists available at ScienceDirect

Livestock Science

journal homepage: www.elsevier.com/locate/livsci

Residual feed intake for Australian Merino sheep estimated in less than 42 days of trial

Fernando Amarilho-Silveira [a,*], Ignacio de Barbieri [c], Jaime Araujo Cobuci [a], Camila Marques Balconi [b], Gracialda Ferreira de Ferreira [c], Gabriel Ciappesoni [b]

[a] Universidade Federal do Rio Grande do Sul, Avenida Bento Gonçalves 7712, RS 91509 900, Brazil
[b] Instituto Nacional de Investigación Agropecuaria, Ruta 48 km 10, Canelones 90200, Uruguay
[c] Instituto Nacional de Investigación Agropecuaria, Ruta 5 km 386, Tacuarembó 45000, Uruguay

HIGHLIGHTS

- There is an increase in body weight and feed intake along the trials.
- The best correlations for average daily gain within of models was greater than 0.94.
- The feed intake correlations presented values greater than 0.93.
- The greater R-square for residual feed estimate was equal at 0.753.
- The best reduced model is the linear with 35 days.

ARTICLE INFO

Keywords:
Feed efficiency
Average daily gain
Repeated measures
Genetic model

ABSTRACT

The evaluation of sheep feed intake (FI) in feed efficiency tests is expensive. Decreasing the test period could be a resource-saving tool by reducing the cost of evaluating each animal and allowing to test a greater number of animals per year. For this reason, the objective of this research was to explore residual feed intake (RFI) models and to decreasing the test duration. Data was collected from 286 Australian Merino sheep of three performed trials, the test period consisted of 56 days (14 days of feed and facilities adaptation and 42 days of FI and average daily gain (ADG) evaluation). Two models were used to calculate RFI, Model 1 (based on Koch et al. (1963) linear model) and Model 2 (repeated measures, weekly model). Model 1 included ADG and FI estimates in a linear regression. The second model included weekly average FI as repeated measure and the weekly ADG. The increase in body weight during the test period was not perfectly linear, presenting a marked variance increase in two of the three tests while FI presented a tendency to increase throughout of the evaluation period, however presenting a high variance per day. In the 42-days tests, Pearson and Spearman correlations between models for ADG were of 0.89 and 0.87, respectively. The best correlations were detected for FI between 42 and 35-days models, presenting Pearson and Spearman correlations of 0.95 and 0.94 in the linear model, and 0.96 and 0.95 in the weekly model. When considering RFI, the correlations between linear and weekly 42-days models were from 0.93 to 0.92, respectively. The 35-days RFI length models (linear and weekly) presented a Pearson and Spearman correlations greater than 0.98 with the 42-days models. Therefore, the RFI models 35-days of duration allowed to decrease seven days of the FI test while maintaining accuracy and explaining 75.3% of the FI in the linear model, and 63.6% of the weekly model. Reducing seven days of testing would provide a greater data collection into a year of phenotypic evaluation.

1. Introduction

Understanding the role of feed intake (FI) and growth rate in feed efficiency is indispensable to select economically efficient sheep. Moreover, it is necessary to optimize FI tests without losing precision in order to increase feed efficiency selection intensity. Increasing the

* Corresponding author.
E-mail address: fernando@proagovina.com.br (F. Amarilho-Silveira).

https://doi.org/10.1016/j.livsci.2022.104889
Received 16 August 2021; Received in revised form 18 February 2022; Accepted 28 February 2022
Available online 3 March 2022

4 Manuscript 2: *Feed efficiency in dairy sheep: An insight from the milk transcriptome* (Suárez Vega et al., under review)

Suarez-Vega, A., Frutos, P., Gutiérrez-Gil, B., Esteban-Blanco, C., Toral P.G., Arranz, J.J., Hervás, G. Feed efficiency in dairy sheep: An insight from the milk transcriptome. *Journal of Dairy Science* (under review).

RUNNING HEAD: FEED EFFICIENCY & MILK TRANSCRIPTOME IN SHEEP

Feed efficiency in dairy sheep: An insight from the milk transcriptome

Aroa Suárez-Vega[1], Pilar Frutos[2], Beatriz Gutiérrez-Gil[1], Cristina Esteban Blanco [1], Pablo G. Toral[2], Juan-José Arranz[1*] and Gonzalo Hervás[2]

[1]Departamento de Producción Animal, Facultad de Veterinaria, Universidad de León, Campus de Vegazana s/n, 24007 León, Spain

[2]Instituto de Ganadería de Montaña (CSIC-University of León), Finca Marzanas s/n, 24346 Grulleros, León, Spain

*Corresponding author: Juan-José Arranz. E-mail: jjarrs@unileon.es

Interpretative Summary

Feed efficiency in dairy sheep: An insight from the milk transcriptome. By *Suárez-Vega et al.* The knowledge of the genetic architecture behind feed efficiency would allow the breeding of more efficient animals maximizing farm profitability and reducing the environmental impact of animal production. This study analyzes high throughput gene expression data from milk samples to determine key genes and biological mechanisms associated with feeding efficiency in dairy sheep. This research demonstrates the informative potential of the milk transcriptome as a target tissue to determine the biological basis of feed efficiency in dairy sheep, highlighting the importance of genes involved in cell division and lipid metabolism in the lactating mammary gland of high-feed efficiency sheep.

ABSTRACT

As higher feed efficiency in dairy ruminants means a higher capability to transform feed nutrients into milk or milk components, differences in feed efficiency are expected to be partly linked to changes in the physiology of the mammary glands. Therefore, this study aimed to determine the biological functions and key regulatory genes associated with feed efficiency in dairy sheep using the milk somatic cell transcriptome. RNA-Seq data from high (**H-FE**, n = 8) and low (**L-FE**, n = 8) feed efficiency ewes were compared through differential expression analysis (**DEA**) and sparse Partial Least Square-Discriminant analysis (sPLS-DA). In the DEA, 79 genes were identified as differentially expressed between both conditions, while the sPLS-DA identified 261 predictive genes [variable importance in projection (**VIP**) > 2] that discriminated H-FE and L-FE sheep. The DEA between sheep with divergent feed efficiency allowed the identification of genes associated with the immune system and stress in L-FE animals. In addition, the sPLS-DA approach revealed the importance of genes involved in cell division (e.g., *KIF4A* and *PRC1*) and cellular lipid metabolic process (e.g., *LPL*, *SCD*, *GPAM*, and *ACOX3*) for the H-FE sheep in the lactating mammary gland transcriptome. A set of discriminant genes, commonly identified by the two statistical approaches, was also detected, including some involved in cell proliferation (e.g., *SESN2*, *KIF20A*, or *TOP2A*) or encoding heat-shock proteins (*CRYAB* or *HSPB1*). These results provide novel insights into the biological basis of feed efficiency in dairy sheep, highlighting the informative potential of the mammary

gland transcriptome as a target tissue and revealing the usefulness of combining univariate and multivariate analysis approaches to elucidate the molecular mechanisms controlling complex traits.

Keywords: dairy sheep, feed efficiency, mammary gland, RNA-Seq, sPLS-DA

INTRODUCTION

Feed costs represent a high proportion of total costs associated with the livestock industry (up to 65–70%; Zhang et al., 2017). Breeding more efficient animals would maximize farm profitability and also reduce the environmental impact of animal production (Lovendahll et al., 2018). However, the challenges and costs of estimating feed efficiency make the implementation of this phenotype in animal breeding schemes difficult. Therefore, elucidating the complex genetic architecture behind feed efficiency is one of the major research goals in this field.

RNA sequencing is widely used in animal breeding to determine genes influencing complex traits, such as milk production, reproductive performance, and quality of carcasses (Wickramasinghe et al., 2014). For feed efficiency, most studies using RNA-Seq data have been performed in pigs, chicken, and beef cattle (e.g., Chen et al., 2021; Piles et al., 2019; Xiao et al., 2021). In dairy ruminants, we have only found the following studies comparing divergent feed efficiency cows using liver and white blood cell transcriptomes: (Khansefid et al., 2017; Salleh et al., 2017, 2018; Dorji et al., 2021)). However, we are not aware of any RNA-Seq approach linking the lactating mammary gland transcriptome and feed efficiency. In dairy sheep, hardly any transcriptomic studies have been conducted to characterize feed efficiency. To our knowledge, the only research aiming at identifying differentially expressed genes in sheep with extreme residual feed intake (**RFI**) values was published by Zhang et al. (2019) using liver samples.

As higher feed efficiency means a higher capability of the animal to transform feed nutrients into milk or milk components, we hypothesize that high and low feed efficiency sheep would show transcriptomic differences in the mammary cells. Therefore, this study aims to characterize the transcriptome of the mammary gland in lactating sheep with divergent feed efficiency values by identifying genes, metabolic pathways, and biological processes potentially involved in this phenotype. At a practical level, concerning the interest in the dairy sheep industry, our objective is to obtain feed efficiency-related data

at the molecular level that can be used in the future in commercial population selection programs.

MATERIAL AND METHODS

Ethics statements

All experimental procedures were approved by the Research Ethics Committee of the Instituto de Ganadería de Montaña, the Spanish National Research Council (**CSIC**), and the Junta de Castilla y León (Spain), following procedures described in Spanish and European Union legislation (R. D. 53/2013 and Council Directive 2010/63/EU).

Animals and sampling

This study constitutes a part of a larger research project aiming at providing new insights into the physiological mechanisms contributing to feed efficiency variation in dairy ruminants. A detailed description of the sheep management practices and calculations for the feed efficiency index (**FEI**) are detailed in Toral et al. (2021).

For these analyses, we selected animals with divergent FEI values from a group of 40 lactating Assaf ewes. Briefly, FEI was computed as the difference between the recorded (**DMI_R**) and the predicted (**DMI_P**) dry matter intake over a three-week experimental period.

$$FEI = DMI_R - DMI_P$$

DMI_P was calculated as follows:

$$DMI_P = ME_{mp}/ME_{TMR}$$

Where **ME_{mp}** are the metabolizable energy requirements for maintenance, production, and BW change (MJ/d), and ME_{TMR} is the metabolizable energy of the total mixed ration (**TMR**; MJ/kg of DM). Both values were estimated using equations for metabolizable energy requirements for nonpregnant lactating sheep and TMR formulation and tables of the nutritional value of feed materials from the Agricultural and Food Research Council (AFRC, 1993).

Sheep with extreme FEI values (8 high feed efficiency sheep (**H-FE**), FEI = −0.29 (SD = 0.23), RFI = −0.16 (SD = 0.25), and 8 low feed efficiency sheep (**L-FE**), FEI = 0.81 (SD = 0.24), RFI = 0.19 (SD = 0.24)) were sampled for RNA-Seq. Milk samples were obtained

as described previously (Toral et al., 2016). To summarize, 50 mL of fresh milk was collected from each animal one hour after milking and 10 min after injection of 5 IU of oxytocin/animal (Facilpart, Laboratorios SYVA, León, Spain) to maximize milk somatic cell concentration. To prevent RNA degradation, udders were cleaned with soap and water and disinfected with povidone-iodine, and the nipples were also flushed with RNAseZap (Ambion, Austin, TX, USA). A sterile gauze was used to cover the collection tube to avoid contamination.

For RNA extraction, the milk somatic cells (**MSC**) were pelleted by centrifugation at 650×g for 10 min at 4 °C in the presence of a final concentration of 0.5 mM of EDTA. Then, the pellet was washed twice with 10, and 2 mL of PBS (pH 7.2 and 0.5 mM of EDTA) followed by centrifugation at 650×g for 10 min at 4 °C.

The last pellet was kept in RNAlater (Sigma-Aldrich, Madrid, Spain) and stored at −80 ºC until RNA extraction using 500 µL of TRIzol according to the manufacturer's instructions (Invitrogen, Carlsbad, CA, USA). The RNA quality was evaluated using an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA), obtaining a mean RNA integrity number of 8.2 (SD = 0.6). The RNA sequencing was conducted at CNAG (Centro Nacional de Análisis Genómico, Barcelona, Spain), where the TrueSeq Stranded Total RNA Library Prep Kit (Illumina, San Diego, CA, USA) was used for library preparation. A HiSeq™ 3000/4000 sequencing system (Illumina) was used to generate stranded paired-end reads of 75 bp. The datasets generated for this study can be found in the ArrayExpress - EMBL-EBI database under the accession E-MTAB-12355.

Alignment and quantification

The alignment to the ovine reference genome (assembly ARS-UI_Ramb_v2.0) available at NCBI was performed using STAR v. 2.7.0 (Dobin et al., 2013). The quantification of the gene expression for the different samples was carried out using RSEM v.1.3.3 software (Li and Dewey, 2011). The options applied for the quantification were "–paired-end" to indicate our data were paired-end, "–estimate-rspd" to estimate the start position of the distribution, "–calc-ci" to calculate 95% credibility intervals and posterior mean estimates, "–seed 12345" to set the seed for the random number of generators used in calculating posterior mean estimates and credibility intervals, and "–p 8" to fix the parallel environment.

Differential expression analysis

To perform the differential expression analysis (**DEA**), we first imported the samples into the R environment with the Tximport package (Soneson et al., 2016). Once the matrix of counts per gene and sample was created, we kept genes with more than 10 counts in at least three samples. Then, technical replicates from the same sample were collapsed using the "collapseReplicates" function on DESeq2 (Love et al., 2014). The DEA between L-FE and H-FE animals was performed using DESeq2 (Love et al., 2014). We selected differentially expressed genes with a False Discovery Rate (**FDR**) <0.05, and log2FoldChange> |1.5|.

Sparse Partial Least Square-Discriminant analysis

The matrix of counts, normalized with DESeq2, was used to perform a sparse Partial Least Square-Discriminant analysis (**sPLS-DA**) in order to identify the key genes driving discrimination of our samples into H-FE and L-FE classes. To perform the sPLS-DA we used the R package mixOmics (Rohart et al., 2017). We used the function "tune.splsda" to assess the optimal number of components and variables to select in each component. For this step, the function "tune" implements repeated (N = 15) and stratified (10-fold) the cross-validation to obtain the best predictive performance for the model. Then, the function "splsda" was used to classify the samples and select the variables. The function "vip" allowed us to obtain the variable importance in projection (**VIP**) coefficients, which reflect the relative importance of variables to explain each component. Genes with a VIP higher than two were selected for further analysis.

Functional enrichment analyses

Gene ontology (**GO**) terms and pathways were analyzed to explore the biological relevance of genes associated with feed efficiency in the DEA and sPLS-DA analyses. For the functional enrichment analyses, we used the ToppGene Suite (Chen et al., 2009). GO terms and pathways were selected as functionally enriched when the q-value (Benjamini and Hochberg correction) was < 0.05, and at least two genes were clustered in the term. To visualize the results, the GOplot R package was used (Walter et al., 2015).

RESULTS

Mapping statistics summary

An average of 31.4 million reads per library (n = 16) was generated. Overall, 96.65% of the reads aligned to the ovine genome; among them, 78.61% were uniquely mapped reads. A total of 15,116 genes were expressed (i.e., those detected in at least 3 samples and with 10 counts or more).

Differentially expressed genes between H-FE and L-FE sheep and functional enrichment analyses

In the DEA, 79 genes were identified as differentially expressed (**DEG**) between H-FE and L-FE animals (FDR < 0.05 and log2FoldChange > |1.5|; Table S1; doi:10.17632/3mw4hnpvr6.1), 10 genes had higher expression in the lactating mammary gland of H-FE animals, and 69 had higher expression in the L-FE.).

Functional enrichment analysis was performed to determine which GO terms were enriched among the DEGs. We found enriched (FDR < 0.05) 27 GO terms in the biological process category (**BP**)-GO terms and two pathways (Source = MSigDB C2 BIOCARTA (v7.5.1)) (Table S2; doi:10.17632/3mw4hnpvr6.1). There was non-enrichment in the molecular function (**MF**) and cellular component (**CC**) categories. A total of 11 BP-GO terms remained after reducing the terms with a gene overlap greater than 80% (Figure 1). The highest enriched BP-GO terms were "response to lipid" (16 genes; FDR = 1.637E-2), "regulation of protein modification process" (16 genes; FDR = 3.134E-2), and "positive regulation of DNA-templated transcription" (16 genes; FDR = 3.134E-2). All 11 BP-GO terms had a negative z-score, meaning they were downregulated in the H-FE animals, but the BP-GO term "mitotic spindle midzone assembly" had a positive z-score (1.41) and was enriched with two genes (*KIF4A* and *PRC1*) (FDR = 4.976E-2). The pathways found enriched were "Overview of proinflammatory and profibrotic mediators" (5 genes; FDR = 2.991E-2) and "p53 transcriptional gene network" (4 genes; FDR = 2.991E-2).

Discriminant genes between H-FE and L-FE

The supervised analysis with the sPLS-DA method was applied to discriminate between H-FE and L-FE animals (Figure 2A). The tune function led to an sPLS-DA model with one component and 380 predictive genes that could help to classify sheep as H-FE and L-FE; 261 out of those 380 genes had a VIP > 2 (Table S3; doi:10.17632/3mw4hnpvr6.1). The prediction obtained with the first component was **AUC** (area under the curve) = 1 (*P*-

value = 0.0007775), with a significantly balanced error rate of 0.32. The 20 genes with the highest loading are represented in Figure 2B. The loading weights were positive for the H-FE group and negative for the L-FE group. Interestingly, among the set of discriminant genes in the sPLS-DA (VIP > 2), 18 genes were also detected with the DEGs: *CCNA2*, *CRYAB*, *GALNT17*, *HS3ST1*, *HSPB1*, *IQCF1*, *KIF20A*, *KIF4A*, *LOC101111669*, *LOC101115355*, *LOC101117955*, *MYO7A*, *NKX3-1*, *PDE4C*, *PRC1*, *PRDM5*, *SESN2*, and *TOP2A* (Table 1).

The functional enrichment analysis was performed with the discriminant genes between high and L-FE sheep with a VIP > 2. We identified 40 GO terms enriched in the BP category, five GO terms in the MF category, and 14 terms enriched in the CC category (Table S4; doi:10.17632/3mw4hnpvr6.1). The highest enriched terms in each GO category were "nuclear division" for the BP (26 genes, FDR = 1.154E-6), "ATP binding" for the MF (34 genes, FDR = 1.575E-2), and "chromosome, centromeric region" for the CC (15 genes, FDR = 7.342E-5). With the enrichment analysis using the pathway databases, 21 pathways were found to be significantly enriched (Table S4; doi:10.17632/3mw4hnpvr6.1), the highest enriched one "Cell Cycle, Mitotic" (22 genes, FDR = 1.454E-4, BioSystems: REACTOME). In Figure 2C, GO terms (2, 7, and 4 GO terms from the MF, BP, and CC categories) and pathways (2 pathways from the REACTOME database) remaining after eliminating those with a gene overlap greater than 80% are represented. All non-redundant terms and pathways had a positive z-score, meaning that, in general, the genes clustered in each term/pathway had higher expression in the H-FE condition.

DISCUSSION

The characterization of the genetic basis of economically relevant breeding traits is crucial to understanding the biology underlying these phenotypes and selecting animals with higher genetic merit. Regarding feed efficiency, RNA-Seq has been applied over the last decade to determine genes and markers related to this trait in several livestock species mainly intended for meat production (e.g., Ramayo-Caldas et al., 2018; Higgins et al., 2019; Zhang et al., 2019; Xiao et al., 2021). Nevertheless, less research has been performed in dairy cattle (Khansefid et al., 2017; Salleh et al., 2017, 2018), and we are not aware of studies on dairy sheep. The liver has been the most commonly used organ to study gene expression differences in relation to feed efficiency (Khansefid et al., 2017;

Salleh et al., 2017, 2018). However, the collection of biopsies from an internal organ would not be feasible in practice. In this study, we analyzed RNA-Seq data from milk, which offers a novel perspective for the genetic characterization of feed efficiency. The milk transcriptome has successfully been used to examine differences in mammary metabolism due to breed (Suárez-Vega et al., 2016; Michailidou et al., 2021), lactation stage (Wickramasinghe et al., 2012; Suárez-Vega et al., 2015; Arora et al., 2019), dietary lipid supplementation (Suárez-Vega et al., 2017, 2019), and mastitis (Asselstine et al., 2019).

We used two bioinformatic approaches, DEA and sPLS-DA. DEA independently tests the expression level of each gene between conditions allowing the determination of the DEGs. Our study identified 79 DEGs between high- and L-FE sheep. A similar number of DEG was identified by other authors when the liver was used as the target tissue (55 DEGs when studying FE for daily gain and body weight in Hu sheep (Zhang et al., 2019a) (FDR < 0.05 and log2FoldChange > |1.5|) and 70 and 19 DEGs (FDR < 0.05) for Holstein and Jersey dairy cattle breeds, respectively (Salleh et al., 2017)). Most of the genes found to be differentially expressed had a higher expression in the L-FE condition. Thus, most GO terms found enriched were related to the biological processes upregulated in the L-FE sheep. The highest enriched term among the DEGs was "response to lipid", suggesting that dietary lipids can induce a different response in more or less efficient animals. Several studies in nutrigenomics in ruminants have demonstrated that lipid supplementation affects the lactating mammary gland transcriptome (Mach et al., 2011; Suárez-Vega et al., 2017, 2019). Moreover, it has been shown that there is variability in the individual response to dietary unsaturated fatty acids (Frutos et al., 2017). Although L-FE and H-FE ewes received the same TMR, a detailed fatty acid analysis of their ruminal digesta suggested a lower biohydrogenation extent of dietary fatty acids in the less efficient group (Toral et al., 2021), which might contribute to explain the present findings.

In addition, it is worth mentioning that almost all the genes in the GO term "lipid response" are clustered between the terms "response to cytokine" and "response to oxidative stress". Metabolic adaptations to high energy demands, such as lactation, lead to lipid mobilization, which might be higher in less efficient sheep. This would be supported by their lower BW gain over the trial and the lower milk de novo fatty acids/*cis*-9 18:1 ratio, a potential proxy of energy deficiency and body fat mobilization (Toral et

al., 2021). Lipid mobilization has been demonstrated to favor inflammatory responses and oxidative stress (Contreras and Sordillo, 2011). In addition, several RNA-Seq studies with different target tissues and species, such as beef (Lindholm-Perry et al., 2017; de Lima et al., 2020), dairy cows (Salleh et al., 2017), pigs (Ramayo-Caldas et al., 2018; Horodyska et al., 2019), sheep (Zhang et al., 2019a), and poultry (Yang et al., 2020) support our results and highlight the association between feed efficiency and the immune system and stress, indicating that the latter processes may increase maintenance requirements and so reduce production in L-FE animals (Patience et al., 2015).

Regarding upregulated genes in H-FE sheep, we found the enriched term "mitotic spindle midzone assembly", which is associated with anaphase and cell division (Wadsworth, 2021). The vast majority of cell proliferation in the mammary gland occurs during its allometric growth before puberty and during pregnancy, with the number of secretory cells in the mature udder correlating with milk yield (Tucker, 1987). It has been demonstrated in dairy cattle and mice that there is a constant but low proportion of cell division during lactation (Knight, 2000; Sorensen et al., 2006). The genes involved in "mitotic spindle midzone assembly", *KIF4A* and *PRC1*, showed low abundance in our transcriptomic data (< 10 fragments per kilobase per million mapped reads (**FPKM**)), which agrees with the results reported in cattle postulating that cell division during lactation is low. However, the higher expression of these two genes in H-FE animals suggests that cell division might also be higher than in L-FE, consistent with the observed differences in milk yield between groups (Toral et al., 2021).

Systems biology is particularly interesting when determining the genetic basis of complex phenotypes, such as feed efficiency. Co-expression network analysis has been successfully used to analyze the genetic architecture of feed efficiency by finding modules of highly co-expressed genes (Alexandre et al., 2015; Ramayo-Caldas et al., 2018; Salleh et al., 2018; de Lima et al., 2020). However, to date, no studies have exploited sPLS-DA to study transcriptomics behind feed efficiency. The sPLS-DA approach aims to determine the most discriminant set of genes between the sample groups (Rohart et al., 2017). As a multivariate methodology, modeling transcripts as a set, sPLS-DA provides a more accurate picture of the context of the biological system and complements the findings obtained from univariate approaches. In our analysis, the number of discriminant genes between H-FE and L-FE sheep was higher than the number

of genes detected by the DEA approach (261 vs 79 genes, respectively). These results allowed us to determine which biological processes were more relevant for the H-FE condition (Figure 2), complementing the results obtained by the DEA approach in which most enriched GO terms were associated with L-FE. The most enriched MF was "ATP binding". Some studies in livestock species have related energy homeostasis production with feed efficiency (Alexandre et al., 2015; Ramayo-Caldas et al., 2018; Salleh et al., 2018; Yang et al., 2020). However, mutations in the ATP-binding domain have also been demonstrated to affect anaphase chromosome segregation in cultured cells (Wagenbach and Maney, 1999). Thereunder, the majority of the terms enriched were linked to "nuclear division", "chromosome organization", and "cell division", among others, with more than 20 genes clustered within these terms. This result highlights the importance of the findings previously discussed from the DEA analysis, suggesting a greater cell division in the lactating udder of H-FE sheep. Thus, we hypothesize that the higher milk yield of more efficient animals (Toral et al., 2021) could be due to a higher number of secretory cells. Moreover, the sPLS-DA methodology allowed the identification of a higher number of genes linked to cell division than DEA. This reinforces the use of a systems biology approach to understanding the complexity of the biological processes behind feed efficiency, which may be underestimated using univariate analyses (Brito et al., 2020).

Another GO term enriched and associated with H-FE was "cellular lipid metabolic process", consistent with the greater milk fat yield in more efficient animals (Toral et al., 2021). Some well-known genes involved in mammary lipid metabolism were found as discriminant genes between H-FE and L-FE sheep: *LPL*, *SCD*, *GPAM*, and *ACOX3*. The *LPL* gene product, the lipoprotein lipase, is involved in the mammary uptake of plasma fatty acids (Bionaz and Loor, 2008; McManaman, 2009) and, in cattle, an association between *LPL* abundance and maintenance of milk synthesis through lactation has been suggested (Bionaz and Loor, 2008). Regarding the stearoyl-CoA desaturase (*SCD*), an enzyme implicated in the desaturation of fatty acids, polymorphisms in the *SCD* gene have been associated with milk and protein yields in dairy cattle (Macciotta et al., 2008; Alim et al., 2012). Another gene associated with lipid metabolisms was *GPAM*, which encodes for mitochondrial glycerol-3-phosphate acetyltransferase, a protein involved in triglyceride synthesis (Roy et al., 2006). In dairy cows, mutations in the *GPAM* gene were significantly correlated with changes in milk fat and protein or milk yield (Yu et al., 2021). The most discriminant gene between H-FE and L-FE sheep was *ACOX3*. This

gene encodes for acyl-Coenzyme A Oxidase 3, which is involved in peroxisomal β-oxidation. The majority of cellular energy is supplied by the oxidation of carbohydrates, fats, or protein. Although *ACOX3* was upregulated in the H-FE group, in general, genes involved in fatty acid oxidation have been related to low feed efficiency in livestock (Alexandre et al., 2015; Tizioto et al., 2015; Zhang et al., 2019b) and negative energy balance (Swartz et al., 2021), when using the liver as target tissue. However, oxidation in mammary tissue has received less attention, and its relationship with feed efficiency is thus less clear. In any event, genome-wide association studies in dairy cows suggested associations between *ACOX3* and fat percentage and some fatty acid concentrations in dairy cows (Ibeagha-Awemu et al., 2016; Bahithige et al., 2021).

Finally, we conducted a literature review on those genes found in common by the two methodological approaches (DEA and sPLS-DA) to corroborate the association with feed efficiency. Some genes, such as *CRYAB*, *HSPB1*, or *PRC1* have been related to feed efficiency in livestock using other target tissues. The *CRYAB* gene, which encodes for Crystallin Alpha B protein, is upregulated in the liver, duodenum, and adipose tissue of L-FE pigs (Gondret et al., 2017; Ramayo-Caldas et al., 2018), in the jejunum in cattle (Lindholm-Perry et al., 2016), and in breast muscle in poultry (Bottje et al., 2014). This gene, and *HSPB1* (also known as *HSP27*), are members of the heat-shock protein family. The expression of heat-shock proteins increases as a cellular response mechanism to a stressor (Archana et al., 2017). *HSPB1* codifies heat-shock protein Family B (Small) Member 1, and contradictory findings have been found for this gene regarding feed efficiency. In beef cattle, higher expression of *HSPB1* has been associated with H-FE animals (Jung et al., 2017; Carvalho et al., 2019), whereas in broilers, this gene is upregulated in L-FE animals (Bottje et al., 2012). In our study, *CRYAB* and *HSPB1* were upregulated in L-FE sheep, which agrees with the results reported in pigs by Ramayo-Caldas et al. (2018). Findings in dairy cows demonstrated that less efficient animals have higher heat production than efficient ones (Kennedy et al., 2021). Previous studies have revealed associations between SNPs in heat-shock proteins and traits such as respiration rate and body temperature (Charoensook et al., 2012; Deb et al., 2013; Singh et al., 2017). Thus, it could be speculated that animals with a higher expression of heat-shock proteins are less efficient due to energy losses in greater heat production. Lastly, we would like to highlight the role of gene *SESN2* in lactation and its potential impact on feed efficiency, which could be a novel finding. *SESN2*, which codifies for sestrin2, belongs to a family

of conserved, stress-inducible regulators of metabolism. A study on the influence of the expression of this gene on lactation suggested that *SESN2* negatively regulates cell proliferation and casein synthesis in cow mammary epithelial cells (Luo et al., 2018). Thus, the upregulation of *SESN2* in L-FE animals would lead to a decrease in milk yield, confirmed in the companion paper by Toral et al. (2021). The fact that *SESN2* is involved in cell proliferation, such as other genes found in common by DEA and sPLS-DA approaches (specifically *KIF20A*, *KIF4A*, *TOP2A*, *NKX3*-1, *CCNA2*, *PRC1*), emphasizes the complementarity of the different methodologies applied and supports the potential relevance of mammary cell division for feed efficiency in dairy sheep.

CONCLUSION

The results from this study provide novel insights into the biological basis of feed efficiency in dairy sheep, highlighting the informative potential of the mammary gland as a target tissue and revealing the usefulness of combining univariate and multivariate analysis approaches to elucidate the molecular mechanisms controlling complex phenotypes. The DEA between sheep with divergent feed efficiency allowed the identification of genes associated with the immune system and stress in L-FE animals. In addition, the sPLS-DA approach revealed the importance of genes involved in cell division (e.g., *KIF4A* and *PRC1*) and cellular lipid metabolic process (e.g., *LPL*, *SCD*, *GPAM*, and *ACOX3*) for the H-FE sheep in the lactating mammary gland transcriptome. We also detected a set of genes commonly identified by the two statistical approaches, including some involved in cell proliferation (e.g., *SESN2*, *KIF20A*, or *TOP2A*) or encoding heat-shock proteins (*CRYAB* or *HSPB1*). Further research would be needed to elucidate the potential role of these genes as candidate biomarkers of feed efficiency in dairy sheep.

ACKNOWLEDGMENTS

This research was funded by the Junta de Castilla y León (JCyL, Spain; project CSI276P18) and the Spanish Research State Agency (Agencia Estatal de Investigación, Spain; project PID2020-113441RB-I00, MCIN/AEI/10.13039/501100011033). Cofunding by the European Regional Development Fund (ERDF/FEDER) is also acknowledged. Research activities for the University of Leon were funded by the

SMARTER project through the H2020 research and innovation program of the European Commission (Grant Agreement n°772787).

REFERENCES

AFRC (Agricultural and Food Research Council). 1993. Energy and Protein Requirements of Ruminants. CAB International.

Alexandre, P.A., L.J.A. Kogelman, M.H.A. Santana, D. Passarelli, L.H. Pulz, P. Fantinato-Neto, P.L. Silva, P.R. Leme, R.F. Strefezzi, L.L. Coutinho, J.B.S. Ferraz, J.P. Eler, H.N. Kadarmideen, and H. Fukumasu. 2015. Liver transcriptomic networks reveal main biological processes associated with feed efficiency in beef cattle. BMC Genomics 16:1–13. doi:10.1186/S12864-015-2292-8/FIGURES/2.

Alim, M.A., Y.P. Fan, X.P. Wu, Y. Xie, Y. Zhang, S.L. Zhang, D.X. Sun, Y. Zhang, Q. Zhang, L. Liu, and G. Guo. 2012. Genetic effects of stearoyl-coenzyme A desaturase (SCD) polymorphism on milk production traits in the Chinese dairy population. Mol. Biol. Rep. 39:8733–8740. doi:10.1007/S11033-012-1733-6.

Archana, P., J. Aleena, P. Pragna, M. Vidya, P. Abdul Niyas, M. Bagath, G. Krishnan, A. Manimaran, V. Beena, E. Kurien, V. Sejian, and R. Bhatta. 2017. Role of heat shock proteins in livestock adaptation to heat stress. J. Dairy, Vet. Anim. Res. Volume 5. doi:10.15406/JDVAR.2017.05.00127.

Arora, R., A. Sharma, U. Sharma, Y. Girdhar, M. Kaur, P. Kapoor, S. Ahlawat, and R.K. Vijh. 2019. Buffalo milk transcriptome: A comparative analysis of early, mid and late lactation. Sci. Reports 2019 91 9:1–9. doi:10.1038/s41598-019-42513-2.

Asselstine, V., F. Miglior, A. Suárez-Vega, P.A.S. Fonseca, B. Mallard, N. Karrow, A. Islas-Trejo, J.F. Medrano, and A. Cánovas. 2019. Genetic mechanisms regulating the host response during mastitis. J. Dairy Sci. 102. doi:10.3168/jds.2019-16504.

Bahithige, H., P. Chathurangi Ariyarathne, M. Correa-Luna, H.T. Blair, D.J. Garrick, N. Lopez-Villalobos, M.B.P.C.; ; Correa-Luna, H.T.; Blair, D.J.; Garrick, and N. Lopez-Villalobos. 2021. Identification of Genomic Regions Associated with Concentrations of Milk Fat, Protein, Urea and Efficiency of Crude Protein Utilization in Grazing Dairy Cows. Genes 2021, Vol. 12, Page 456 12:456. doi:10.3390/GENES12030456.

Bionaz, M., and J.J. Loor. 2008. Gene networks driving bovine milk fat synthesis during the lactation cycle. BMC Genomics 9:366.

Bottje, W.G., B.-W. Kong, J.Y. Lee, T. Washington, J. Baum, S. Dridi, T. Wing, and J. Hardiman. 2014. Potential Roles of mTOR and Protein Degradation Pathways in the Phenotypic Expression of Feed Efficiency in Broilers. Biochem. Physiol. Open Access 2014 31 3:1–8. doi:10.4172/2168-9652.1000125.

Bottje, W.G., B.W. Kong, J.J. Song, J.Y. Lee, B.M. Hargis, K. Lassiter, T. Wing, and J. Hardiman. 2012. Gene expression in breast muscle associated with feed efficiency in a single male broiler line using a chicken 44K microarray. II. Differentially expressed focus genes. Poult. Sci. 91:2576–2587. doi:10.3382/PS.2012-02204.

Brito, L.F., H.R. Oliveira, K. Houlahan, P.A.S. Fonseca, S. Lam, A.M. Butty, D.J. Seymour, G. Vargas, T.C.S. Chud, F.F. Silva, C.F. Baes, A. Cánovas, F. Miglior, and F.S. Schenkel. 2020. Genetic mechanisms underlying feed utilization and implementation of genomic selection for improved feed efficiency in dairy cattle. Can. J. Anim. Sci. 100:587–604. doi:10.1139/CJAS-2019-0193/ASSET/IMAGES/CJAS-2019-0193TAB5.GIF.

Carvalho, E.B., M.P. Gionbelli, R.T.S. Rodrigues, S.F.M. Bonilha, C.J. Newbold, S.E.F. Guimarães, W. Silva, L.L. Verardo, F.F. Silva, E. Detmann, and M.S. Duarte. 2019. Differentially expressed mRNAs, proteins and miRNAs associated to energy metabolism in skeletal muscle of beef cattle identified for low and high residual feed intake. BMC Genomics 20:1–12. doi:10.1186/S12864-019-5890-Z/FIGURES/3.

Charoensook, R., K. Gatphayak, A.R. Sharifi, C. Chaisongkram, B. Brenig, and C. Knorr. 2012. Polymorphisms in the bovine HSP90AB1 gene are associated with heat tolerance in Thai indigenous cattle. Trop. Anim. Health Prod. 44:921–928. doi:10.1007/S11250-011-9989-8.

Chen, J., E.E. Bardes, B.J. Aronow, and A.G. Jegga. 2009. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization.. Nucleic Acids Res. 37:W305-11. doi:10.1093/nar/gkp427.

Chen, W., P.A. Alexandre, G. Ribeiro, H. Fukumasu, W. Sun, A. Reverter, and Y. Li. 2021. Identification of Predictor Genes for Feed Efficiency in Beef Cattle by Applying Machine Learning Methods to Multi-Tissue Transcriptome Data. Front. Genet. 12. doi:10.3389/FGENE.2021.619857.

Contreras, G.A., and L.M. Sordillo. 2011. Lipid mobilization and inflammatory responses during the transition period of dairy cows. Comp. Immunol. Microbiol. Infect. Dis. 34:281–289. doi:10.1016/J.CIMID.2011.01.004.

Deb, R., B. Sajjanar, U. Singh, S. Kumar, M.P. Brahmane, R. Singh, G. Sengar, and A. Sharma. 2013. Promoter variants at AP2 box region of Hsp70.1 affect thermal stress response and milk production traits in Frieswal cross bred cattle. Gene 532:230–235. doi:10.1016/J.GENE.2013.09.037.

Dobin, A., C.A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, and T.R. Gingeras. 2013. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29:15–21. doi:10.1093/bioinformatics/bts635 [doi].

Dorji, J., I.M. MacLeod, A.J. Chamberlain, C.J. Vander Jagt, P.N. Ho, M. Khansefid, B.A. Mason, C.P. Prowse-Wilkins, L.C. Marett, W.J. Wales, B.G. Cocks, J.E. Pryce, and H.D. Daetwyler. 2021. Mitochondrial protein gene expression and the oxidative phosphorylation pathway associated with feed efficiency and energy balance in dairy cattle. J. Dairy Sci. 104:575–587. doi:10.3168/JDS.2020-18503.

Frutos, P., P.G. Toral, and G. Hervás. 2017. Individual variation of the extent of milk fat depression in dairy ewes fed fish oil: Milk fatty acid profile and mRNA abundance of candidate genes involved in mammary lipogenesis. J. Dairy Sci. 100:9611–9622. doi:10.3168/JDS.2017-13354.

Gondret, F., A. Vincent, M. Houée-Bigot, A. Siegel, S. Lagarrigue, D. Causeur, H. Gilbert, and I. Louveau. 2017. A transcriptome multi-tissue analysis identifies

biological pathways and genes associated with variations in feed efficiency of growing pigs. BMC Genomics 18. doi:10.1186/S12864-017-3639-0.

Higgins, M.G., D.A. Kenny, C. Fitzsimons, G. Blackshields, S. Coyle, C. McKenna, M. McGee, D.W. Morris, and S.M. Waters. 2019. The effect of breed and diet type on the global transcriptome of hepatic tissue in beef cattle divergent for feed efficiency. BMC Genomics 20:1–11. doi:10.1186/S12864-019-5906-8/FIGURES/1.

Horodyska, J., R.M. Hamill, H. Reyer, N. Trakooljul, P.G. Lawlor, U.M. McCormack, and K. Wimmers. 2019. RNA-seq of liver from pigs divergent in feed efficiency highlights shifts in macronutrient metabolism, hepatic growth and immune response. Front. Genet. 10:117. doi:10.3389/FGENE.2019.00117/BIBTEX.

Ibeagha-Awemu, E.M., S.O. Peters, K.A. Akwanji, I.G. Imumorin, and X. Zhao. 2016. High density genome wide genotyping-by-sequencing and association identifies common and low frequency SNPs, and novel candidate genes influencing cow milk traits. Sci. Reports 2016 61 6:1–18. doi:10.1038/srep31109.

Jung, U.S., M.J. Kim, T. Wang, J.S. Lee, S.W. Jeon, N.C. Jo, W.S. Kim, M. Baik, and H.G. Lee. 2017. Upregulated heat shock protein beta-1 associated with caloric restriction and high feed efficiency in longissimus dorsi muscle of steer. Livest. Sci. 202:109–114. doi:10.1016/J.LIVSCI.2017.05.009.

Kennedy, K.M., F. Becker, H.M. Hammon, and B. Kuhla. 2021. Differences in net fat oxidation, heat production, and liver mitochondrial DNA copy numbers between high and low feed-efficient dairy cows. J. Dairy Sci. 104:9287–9303. doi:10.3168/JDS.2020-20031.

Khansefid, M., C.A. Millen, Y. Chen, J.E. Pryce, A.J. Chamberlain, C.J. Vander Jagt, C. Gondro, and M.E. Goddard. 2017. Gene expression analysis of blood, liver, and muscle in cattle divergently selected for high and low residual feed intake. J. Anim. Sci. 95:4764–4775. doi:10.2527/JAS2016.1320.

Knight, C.H. 2000. The importance of cell division in udder development and lactation. Livest. Prod. Sci. 66:169–176. doi:10.1016/S0301-6226(00)00224-4.

Li, B., and C.N. Dewey. 2011. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12:323. doi:10.1186/1471-2105-12-323.

de Lima, A.O., J.E. Koltes, W.J.S. Diniz, P.S.N. de Oliveira, A.S.M. Cesar, P.C. Tizioto, J. Afonso, M.M. de Souza, J. Petrini, M.I.P. Rocha, T.F. Cardoso, A.Z. Neto, L.L. Coutinho, G.B. Mourão, and L.C.A. Regitano. 2020. Potential Biomarkers for Feed Efficiency-Related Traits in Nelore Cattle Identified by Co-expression Network and Integrative Genomics Analyses. Front. Genet. 11:189. doi:10.3389/FGENE.2020.00189/BIBTEX.

Lindholm-Perry, A.K., V.M. Artegoitia, J.R. Miles, and A.P. Foote. 2017. Expression of cytokine genes and receptors in white blood cells associated with divergent body weight gain in beef steers. Agri Gene 6:37–39. doi:10.1016/J.AGGENE.2017.09.005.

Lindholm-Perry, A.K., A.R. Butler, R.J. Kern, R. Hill, L.A. Kuehn, J.E. Wells, W.T. Oliver, K.E. Hales, A.P. Foote, and H.C. Freetly. 2016. Differential gene expression

in the duodenum, jejunum and ileum among crossbred beef steers with divergent gain and feed intake phenotypes. Anim. Genet. 47:408–427. doi:10.1111/AGE.12440.

Love, M.I., W. Huber, and S. Anders. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15:550. doi:s13059-014-0550-8 [pii].

Lovendahll, P., G.F. Difford, B. Li, M.G.G. Chagunda, P. Huhtanen, M.H. Lidauer, J. Lassen, and P. Lund. 2018. Review: Selecting for improved feed efficiency and reduced methane emissions in dairy cattle. Animal 12:s336–s349. doi:10.1017/S1751731118002276.

Luo, C., S. Zhao, M. Zhang, Y. Gao, J. Wang, M.D. Hanigan, and N. Zheng. 2018. SESN2 negatively regulates cell proliferation and casein synthesis by inhibition the amino acid-mediated mTORC1 pathway in cow mammary epithelial cells. Sci. Reports 2018 81 8:1–10. doi:10.1038/s41598-018-22208-w.

Macciotta, N.P.P., M. Mele, G. Conte, A. Serra, M. Cassandro, R. Dal Zotto, A.C. Borlino, G. Pagnacco, and P. Secchiari. 2008. Association Between a Polymorphism at the Stearoyl CoA Desaturase Locus and Milk Production Traits in Italian Holsteins. J. Dairy Sci. 91:3184–3189. doi:10.3168/JDS.2007-0947.

MacH, N., A.A.A. Jacobs, L. Kruijt, J. Van Baal, and M.A. Smits. 2011. Alteration of gene expression in mammary gland tissue of dairy cows in response to dietary unsaturated fatty acids. Animal 5:1217–1230. doi:10.1017/S1751731111000103.

McManaman, J.L. 2009. Formation of milk lipids: a molecular perspective. Clin. Lipidol. 4:391. doi:10.2217/CLP.09.15.

Michailidou, S., A. Gelasakis, G. Banos, G. Arsenos, and A. Argiriou. 2021. Comparative Transcriptome Analysis of Milk Somatic Cells During Lactation Between Two Intensively Reared Dairy Sheep Breeds. Front. Genet. 12:1293. doi:10.3389/FGENE.2021.700489/BIBTEX.

Patience, J.F., M.C. Rossoni-Serão, and N.A. Gutiérrez. 2015. A review of feed efficiency in swine: Biology and application. J. Anim. Sci. Biotechnol. 6:1–9. doi:10.1186/S40104-015-0031-2/TABLES/4.

Piles, M., C. Fernandez-Lozano, M. Velasco-Galilea, O. González-Rodríguez, J.P. Sánchez, D. Torrallardona, M. Ballester, and R. Quintanilla. 2019. Machine learning applied to transcriptomic data to identify genes associated with feed efficiency in pigs. Genet. Sel. Evol. 51:1–15. doi:10.1186/S12711-019-0453-Y/TABLES/1.

Ramayo-Caldas, Y., M. Ballester, J.P. Sánchez, O. González-Rodríguez, M. Revilla, H. Reyer, K. Wimmers, D. Torrallardona, and R. Quintanilla. 2018. Integrative approach using liver and duodenum RNA-Seq data identifies candidate genes and pathways associated with feed efficiency in pigs. Sci. Reports 2018 81 8:1–11. doi:10.1038/s41598-017-19072-5.

Rohart, F., B. Gautier, A. Singh, and K.A. Lê Cao. 2017. mixOmics: An R package for 'omics feature selection and multiple data integration. PLOS Comput. Biol. 13:e1005752. doi:10.1371/JOURNAL.PCBI.1005752.

Roy, R., L. Ordovas, S. Taourit, P. Zaragoza, A. Eggen, and C. Rodellar. 2006. Genomic structure and an alternative transcript of bovine mitochondrial glycerol-3-phosphate acyltransferase gene (GPAM). Cytogenet. Genome Res. 112:82–89. doi:10.1159/000087517.

Salleh, M.S., G. Mazzoni, J.K. Höglund, D.W. Olijhoek, P. Lund, P. Løvendahl, and H.N. Kadarmideen. 2017. RNA-Seq transcriptomics and pathway analyses reveal potential regulatory genes and molecular mechanisms in high- and low-residual feed intake in Nordic dairy cattle. BMC Genomics 18. doi:10.1186/S12864-017-3622-9.

Salleh, S.M., G. Mazzoni, P. Løvendahl, and H.N. Kadarmideen. 2018. Gene co-expression networks from RNA sequencing of dairy cattle identifies genes and pathways affecting feed efficiency. BMC Bioinformatics 19. doi:10.1186/s12859-018-2553-z.

Singh, K.M., S. Singh, I. Ganguly, R.K. Nachiappan, A. Ganguly, R. Venkataramanan, A. Chopra, and H.K. Narula. 2017. Association of heat stress protein 90 and 70 gene polymorphism with adaptability traits in Indian sheep (Ovis aries). Cell Stress Chaperones 22:675–684. doi:10.1007/S12192-017-0770-4/FIGURES/4.

Soneson, C., M.I. Love, and M.D. Robinson. 2016. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000Research 4:1521. doi:10.12688/f1000research.7563.2.

Sorensen, M.T., J. V. Nørgaard, P.K. Theil, M. Vestergaard, and K. Sejrsen. 2006. Cell Turnover and Activity in Mammary Tissue During Lactation and the Dry Period in Dairy Cows. J. Dairy Sci. 89:4632–4639. doi:10.3168/JDS.S0022-0302(06)72513-9.

Suárez-Vega, A., B. Gutiérrez-Gil, and J.J. Arranz. 2016. Transcriptome expression analysis of candidate milk genes affecting cheese-related traits in 2 sheep breeds.. J. Dairy Sci.. doi:10.3168/jds.2016-11048.

Suárez-Vega, A., B. Gutiérrez-Gil, C. Klopp, C. Robert-Granie, G. Tosser-Klopp, and J.J. Arranz. 2015. Characterization and comparative analysis of the milk transcriptome in two dairy sheep breeds using RNA sequencing. Sci. Rep. 5. doi:10.1038/srep18399.

Suárez-Vega, A., B. Gutiérrez-Gil, P.G. Toral, G. Hervás, J.J. Arranz, and P. Frutos. 2019. Conjugated linoleic acid (CLA)-induced milk fat depression: application of RNA-Seq technology to elucidate mammary gene regulation in dairy ewes. Sci. Rep. 9. doi:10.1038/s41598-019-40881-3.

Suárez-Vega, A., P.G. Toral, B. Gutiérrez-Gil, G. Hervás, J.J. Arranz, and P. Frutos. 2017. Elucidating fish oil-induced milk fat depression in dairy sheep: Milk somatic cell transcriptome analysis. Sci. Reports 2017 71 7:1–9. doi:10.1038/srep45905.

Swartz, T.H., U. Moallem, H. Kamer, G. Kra, Y. Levin, L.K. Mamedova, B.J. Bradford, and M. Zachut. 2021. Characterization of the liver proteome in dairy cows experiencing negative energy balance at early lactation. J. Proteomics 246:104308. doi:10.1016/J.JPROT.2021.104308.

Tizioto, P.C., L.L. Coutinho, J.E. Decker, R.D. Schnabel, K.O. Rosa, P.S.N. Oliveira, M.M. Souza, G.B. Mourão, R.R. Tullio, A.S. Chaves, D.P.D. Lanna, A. Zerlotini-

Neto, M.A. Mudadu, J.F. Taylor, and L.C.A. Regitano. 2015. Global liver gene expression differences in Nelore steers with divergent residual feed intake phenotypes. BMC Genomics 16:1–14. doi:10.1186/S12864-015-1464-X/FIGURES/4.

Toral, P.G., G. Hervás, C. Fernández-Díez, A. Belenguer, and P. Frutos. 2021. Rumen biohydrogenation and milk fatty acid profile in dairy ewes divergent for feed efficiency. J. Dairy Sci. 104:5569–5582. doi:10.3168/JDS.2020-19061.

Toral, P.G., G. Hervás, A. Suárez-Vega, J.J. Arranz, and P. Frutos. 2016. Isolation of RNA from milk somatic cells as an alternative to biopsies of mammary tissue for nutrigenomic studies in dairy ewes. J. Dairy Sci. 99. doi:10.3168/jds.2016-11184.

Tucker, H.A. 1987. Quantitative Estimates of Mammary Growth During Various Physiological States: A Review. J. Dairy Sci. 70:1958–1966. doi:10.3168/JDS.S0022-0302(87)80238-2.

Wadsworth, P. 2021. The multifunctional spindle midzone in vertebrate cells at a glance. J. Cell Sci. 134. doi:10.1242/JCS.250001/268391.

Wagenbach, M., and T. Maney. 1999. Mutations in the ATP-binding domain affect the subcellular distribution of mitotic centromere-associated kinesin (MCAK). Cell Biol. Int. 23:275–286. doi:10.1006/CBIR.1999.0359.

Walter, W., F. Sánchez-Cabo, and M. Ricote. 2015. GOplot: an R package for visually combining expression data with functional analysis. Bioinformatics 31:2912–2914. doi:10.1093/BIOINFORMATICS/BTV300.

Wickramasinghe, S., A. Cánovas, G. Rincón, and J.F. Medrano. 2014. RNA-Sequencing: A tool to explore new frontiers in animal genetics. Livest. Sci. 166:206–216. doi:10.1016/j.livsci.2014.06.015.

Wickramasinghe, S., G. Rincon, A. Islas-Trejo, and J.F. Medrano. 2012. Transcriptional profiling of bovine milk using RNA sequencing. BMC Genomics 13:45. doi:10.1186/1471-2164-13-45 [doi].

Xiao, C., J. Deng, L. Zeng, T. Sun, Z. Yang, and X. Yang. 2021. Transcriptome Analysis Identifies Candidate Genes and Signaling Pathways Associated With Feed Efficiency in Xiayan Chicken. Front. Genet. 12:368. doi:10.3389/FGENE.2021.607719/BIBTEX.

Yang, L., T. He, F. Xiong, X. Chen, X. Fan, S. Jin, and Z. Geng. 2020. Identification of key genes and pathways associated with feed efficiency of native chickens based on transcriptome data via bioinformatics analysis. BMC Genomics 21:1–18. doi:10.1186/S12864-020-6713-Y/FIGURES/7.

Yu, H., Y. Zhao, A. Iqbal, L. Xia, Z. Bai, H. Sun, X. Fang, R. Yang, and Z. Zhao. 2021. Effects of polymorphism of the GPAM gene on milk quality traits and its relation to triglyceride metabolism in bovine mammary epithelial cells of dairy cattle. Arch. Anim. Breed. 64:35. doi:10.5194/AAB-64-35-2021.

Zhang, D., X. Zhang, F. Li, C. Li, Y. La, F. Mo, G. Li, Y. Zhang, X. Li, Q. Song, Y. Zhao, and W. Wang. 2019a. Transcriptome Analysis Identifies Candidate Genes and Pathways Associated With Feed Efficiency in Hu Sheep. Front. Genet. 10:1183.

doi:10.3389/FGENE.2019.01183/BIBTEX.

Zhang, J., H. Shi, S. Li, Z. Cao, H. Yang, and Y. Wang. 2019b. Integrative hepatic metabolomics and proteomics reveal insights into the mechanism of different feed efficiency with high or low dietary forage levels in Holstein heifers. J. Proteomics 194:1–13. doi:10.1016/J.JPROT.2018.12.026.

Zhang, X., W. Wang, F. Mo, Y. La, C. Li, and F. Li. 2017. Association of residual feed intake with growth and slaughtering performance, blood metabolism, and body composition in growing lambs. Sci. Rep. 7. doi:10.1038/S41598-017-13042-7.

Table 1. Genes found in common by the differential gene expression analysis (DEA) and the sparse Partial Least Square-Discriminant analysis (sPLS-DA).

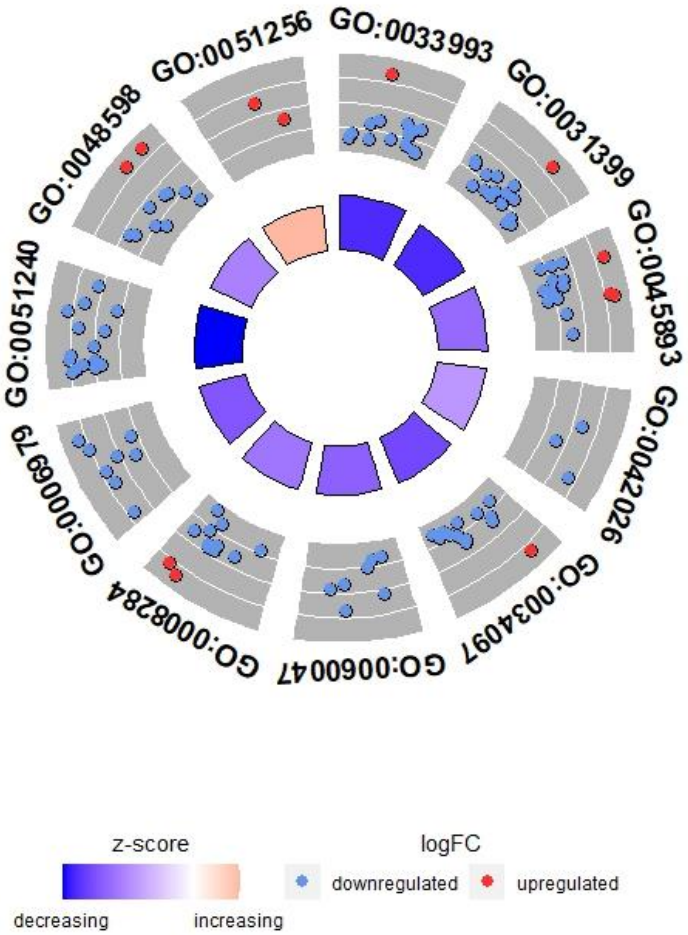
Gene Name	DEA			sPLS-DA	
	FC[1] (log2)	*P*-value	*P*-adj[2]	VIP[3]	Loadings[4]
HS3ST1	-3.22	4.89E-06	1.06E-02	2.32	-0.02
PDE4C	-2.70	2.47E-04	4.79E-02	7.63	-0.06
IQCF1	-2.51	2.05E-04	4.43E-02	5.27	-0.04
HSPB1	-2.20	8.32E-05	3.31E-02	2.78	-0.02
NKX3-1	-2.02	9.85E-05	3.70E-02	2.09	-0.02
SESN2	-1.67	1.12E-04	3.70E-02	7.08	-0.06
LOC101111669	-1.47	1.96E-04	4.28E-02	3.63	-0.03
LOC101115355	-1.06	1.18E-04	3.70E-02	11.83	-0.10
PRC1	1.19	1.93E-04	4.28E-02	15.71	0.13
TOP2A	1.49	1.85E-06	4.66E-03	18.29	0.15
MYO7A	1.67	2.90E-05	2.29E-02	17.56	0.14
CCNA2	1.75	7.10E-05	3.12E-02	10.10	0.08
PRDM5	1.76	1.64E-04	4.08E-02	5.93	0.05
KIF20A	1.85	1.43E-05	1.45E-02	13.10	0.11
KIF4A	1.91	1.30E-05	1.40E-02	13.04	0.11

[1] FC = fold change. Negative values correspond to higher expression in low feed-efficiency animals. Positive values correspond to higher expression in high feed-efficiency animals

[2] *P-adj*= False Discovery Rate (FDR) multiple test correction performed by DESeq2.

[3] VIP= variable importance in projection.

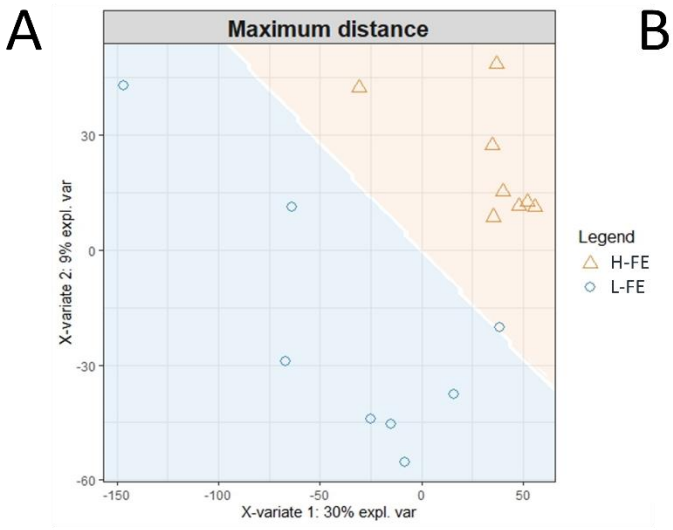
[4] Loadings= value of the gene's loading weight (importance) on the first component of the sPLS-DA.

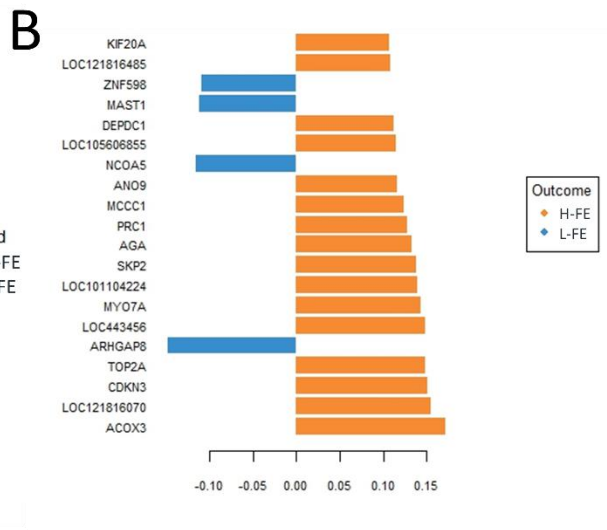


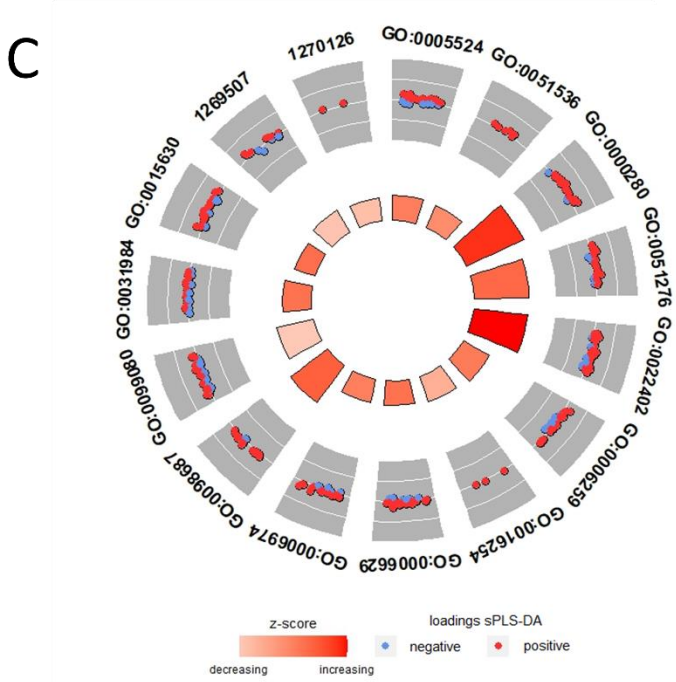
ID	Description
GO:0033993	response to lipid
GO:0031399	regulation of protein modification process
GO:0045893	positive regulation of DNA-templated transcription
GO:0042026	protein refolding
GO:0034097	response to cytokine
GO:0060047	heart contraction
GO:0008284	positive regulation of cell population proliferation
GO:0006979	response to oxidative stress
GO:0051240	positive regulation of multicellular organismal process
GO:0048598	embryonic morphogenesis
GO:0051256	mitotic spindle midzone assembly

671

Figure 1. Functional enrichment results from the differential gene expression analysis between high (H-FE) and low (L-FE) feed efficiency animals. In the GOCircle plot, the significant GO terms enriched (FDR < 0.05) after a reduction of the terms with a gene overlap greater than 80% are represented. The outer circle shows a scatter plot for each GO term of the logFC of the genes clustered in the term. The blue circles are genes downregulated in H-FE, while red circles are upregulated genes in H-FE sheep. The inner circle shows a bar plot representing the z-score for each GO term. The red bar means that the GO term is upregulated for H-FE, while the blue bar indicates the GO term is upregulated for L-FE.







ID	Description
GO:0005524	ATP binding
GO:0051536	iron-sulfur cluster binding
GO:0000280	nuclear division
GO:0051276	chromosome organization
GO:0022402	cell cycle process
GO:0006259	DNA metabolic process
GO:0016254	preassembly of GPI anchor in ER membrane
GO:0006629	lipid metabolic process
GO:0006974	cellular response to DNA damage stimulus
GO:0098687	chromosomal region
GO:0099080	supramolecular complex
GO:0031984	organelle subcompartment
GO:0015630	microtubule cytoskeleton
1269507	Signaling by Rho GTPases
1270126	Interconversion of 2-oxoglutarate and 2-hydroxyglutarate

Figure 2. Results from the sparse Partial Least Square-Discriminant analysis (sPLS-DA). A) Sample prediction area plot from the sPLS-DA model applied on the RNA-Seq data set from high (H-FE; orange triangles) and low (L-FE; blue circles) samples using as the distance for prediction "maximum distance". B) Loading plot of the top 20 discriminating genes on the first component between high and low feed efficiency animals, colors indicate the group in which the mean expression is maximal for each gene (H-FE: orange and L-FE: blue). C) GOCircle plot showing the significant GO terms and pathways enriched (FDR < 0.05) after a reduction of the terms with a gene overlap greater than 80% are represented. The outer circle shows a scatter plot for each term of the logFC of the genes clustered in the term. The blue circles are genes downregulated in H-FE, while the red circles are upregulated genes in H-FE sheep. The inner circle shows a bar plot representing the z-score for each term. The red color means that the GO term is upregulated for the H-FE group; the red color intensity is associated with the value of the z-score.

SUPPLEMENTARY MATERIAL

Supplementary files have been submitted to Mendeley Data with the provisional url https://data.mendeley.com/datasets/3mw4hnpvr6/draft?a=1cbfea49-0082-44bf-8a0e-873aeefa32f4 .

5 Conclusion

The milk somatic cell transcriptome analysis in lactating ewes monitored for estimating feed efficiency combining univariate (DEA) and multivariate (sPLS-DA) approaches has revealed some of the molecular mechanisms involved in feed efficiency in dairy sheep. The DEA allowed the identification of a higher expression of genes associated with the immune system and stress in less efficient animals. In addition, the sPLS-DA approach revealed the importance of genes involved in cell division and cellular lipid metabolic process for more efficient ewes. We also detected a set of genes commonly identified by the two statistical approaches, including some involved in cell proliferation or encoding heat-shock proteins. These results allow us to postulate what happens at the level of the mammary gland in sheep. On the one hand, it is observed that in less efficient animals, part of the cellular energy in the mammary gland is dedicated to the maintenance of the physiology of the organ (immune system and stress control). In contrast, the more efficient animals present a greater number of secretory cells in the acini of their mammary glands and a higher metabolic activity of synthesis of milk components.

6 Deviations or delays

none