

RESEARCH

Open Access



Integration of epigenomic and genomic data to predict residual feed intake and the feed conversion ratio in dairy sheep via machine learning algorithms

Pablo Augusto de Souza Fonseca¹, Aroa Suarez-Vega¹, Cristina Esteban-Blanco¹, Héctor Marina¹, Rocío Pelayo¹, Beatriz Gutiérrez-Gil¹ and Juan-José Arranz^{1*}

Abstract

Background Feed efficiency (FE) is an essential trait in livestock species because of the constant demand to increase the productivity and sustainability of livestock production systems. A better understanding of the biological mechanisms associated with FEs might help improve the estimation and selection of superior animals. In this work, differentially methylated regions (DMRs) were identified via genome-wide bisulfite sequencing (GWBS) by comparing the DNA methylation profiles of milk somatic cells from dairy ewes that were divergent in terms of residual feed intake. The DMRs were identified by comparing divergent groups for residual feed intake (RFI), the feed conversion ratio (FCR), and the consensus between both metrics (Cons). Additionally, the predictive performance of these DMRs and genetic variants mapped within these regions was evaluated via three machine learning (ML) models (xgboost, random forest (RF), and multilayer feedforward artificial neural network (deeplearning)). The average performance of each model was based on the root mean squared error (RMSE) and squared Spearman correlation (ρ^2). Finally, the best model for each scenario was selected on the basis of the highest ratio between ρ^2 and RMSE.

Results In total, 12,257, 9,328, and 6,723 genes were annotated for DMRs detected in the RFI, FCR, and Cons groups, respectively. These genes are associated with important pathways for regulating FE in dairy sheep, such as protein digestion and absorption, hormone synthesis and secretion, control of energy availability, cellular signaling, and feed behavior pathways. With respect to the ML predictions, the smallest mean RMSE (0.17) was obtained using RF, which was used to predict RFI. The highest mean ρ^2 (0.20) was obtained when the RFI was predicted via the mean methylation within the DMRs identified, the consensus groups were compared, and the genetic variants mapped within these DMRs were included. The best overall models were obtained for the prediction of RFI using the DMRs obtained in the comparison of RFI groups (RMSE = 0.10, ρ^2 = 0.86) using xgboost and the DMRs plus the genetic variants identified via the Cons groups (RMSE = 0.07, ρ^2 = 0.62) using RF.

*Correspondence:
Juan-José Arranz
jjarrs@unileon.es

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Conclusions The results provide new insights into the biological mechanisms associated with FE and the control of these processes through epigenetic mechanisms. Additionally, the potential use of epigenetic information as a biomarker for the prediction of FE can be suggested based on the obtained results.

Keywords Omics, Prediction, Artificial intelligence, Livestock, Feed efficiency

Background

Recently, an increasing demand for highly nutritional food has emerged due to the constant growth of the world population, reaching an expected population size of 10.9 billion by 2100 [1]. However, concomitantly with this increase in production, it is necessary to maintain sustainability and animal welfare standards, resulting in a challenge for producers [2]. Among the different factors, nutrition is a major component affecting sheep milk composition. Consequently, appropriate nutritional management is directly associated with improvements in the profile of bioactive substances in milk [3]. Importantly, in animal production systems, management decisions related to feeding strategies can account for up to 75% of all variable costs on dairy farms [4, 5, 6]. An important concept in animal production is feed efficiency (FE), which can be defined as the required amount of feed that an animal needs to produce one product unit. More efficient animals produce the same amount of product, eat less, and have the potential to reduce production costs and the environmental impact while maximizing profitability. Indeed, more feed-efficient sheep produce less methane when fed high-quality pellets [7].

The evaluation of FEs in dairy sheep is still challenging because of the lack of a set model for calculating an efficiency metric to be used in selection programs. Two commonly studied metrics for FE in livestock species are the feed conversion ratio (FCR) and residual feed intake (RFI). The FCR is the ratio of feed intake to meat gain or milk production. However, the high correlations between feed intake and the rate of gain/production might result in an undesirable selection of heavier animals at the mature weight [8]. On the other hand, the RFI is the difference between the actual and the predicted feed intake required to obtain the rate of gain/production [9]. Compared with FCR, RFI seems to be phenotypically and genetically independent of the rate of gain [10, 11, 12, 13, 14]. Therefore, it is suggested that a more interesting use of RFI in selection programs may be considered compared with FCR.

Few studies have evaluated the heritability of RFI and FCR in sheep, especially dairy sheep. The available studies in dairy sheep estimate the heritability between 0.11 and 0.54 for RFI and 0.05–0.77 for FCR [7, 14, 15, 16], which suggests the possibility of improvement in the FE through genetic selection. The observed high variability in the heritability estimates might be explained by the phenotypical variability observed between and

within flocks in major physiological processes contributing to FE, such as feed intake and digestion, anabolism, catabolism, physical activity and thermoregulation [17, 18]. Some studies have applied omics technologies, such as RNA sequencing (RNA-seq) and metagenomics, to better understand the genetic basis of feed efficiency in sheep [19, 20, 21]. To the best of our knowledge, no studies have evaluated the contribution of epigenetic markers over FEs in dairy sheep. Individual nutritional status is directly affected by epigenetic alterations or vice versa because of its action on enzymes related to epigenetic processes or even changes in the availability of substrates for those enzymes [22, 23, 24]. Consequently, epigenetic alterations are interesting targets for the identification of biomarkers associated with FE.

Recently, rumen and plasma metabolomic data were efficiently used to discriminate animals with high and low RFI values through the application of a machine learning (ML) model, partial least squares discriminant analysis [25, 26]. Additionally, recently, FE values for Assaf ewes were predicted using milk somatic cell transcriptomic data (median Spearman correlations of 0.37 and 0.22 for FCR and RFI, respectively) [27]. In other species, different ML models have been used to estimate FE metrics, applying omics data as features in predictions. For example, genomic and metabolomic data were used in supervised and semisupervised methods for RFI prediction in pigs and cattle, achieving interesting predictive accuracy [28, 29, 30]. On the basis of the potential biological contribution of epigenetic markers to FEs, the use of epigenetic data as features in ML models has the potential to generate useful results regarding the predictive accuracy of those models.

Considering this information, the main objectives of the current study are (1) to identify differentially methylated loci (DMLs) and regions (DMRs) between groups of animals divergent for RFI, FCR, and both (consensus) and (2) to evaluate the predictive accuracy of ML fed exclusively with the methylation levels within the detected DMRs and with a combination of the methylation levels and generic variants mapped within these DMRs.

Methods

Samples and feed efficiency estimation

The 21 Spanish Assaf female ewes used in the current study belong to the experimental flock of the Instituto de Ganadería de Montaña (CSIC- Universidad de León). Therefore, these animals were not privately owned by

another institution/individual/farm, and obtaining informed consent from the owner was not necessary. A detailed description of the animals used in the current study and the nutrient protein restriction (NPR) used during the prepubertal stage is explained by Suárez-Vega et al. (2024) [27]. Briefly, 40 Spanish Assaf female lambs were managed under standard commercial conditions and fed a standard diet for replacing ewe lambs, which consisted of 16% crude protein [31]. To evaluate the impact of a feed restriction challenge due to a trade market problem and a shortage of concentrate inputs, the 40 ewes were subsequently divided into two groups: one group of animals received the standard diet mentioned above for 64 days, and the second group received the same diet without soybean meal (44% reduction in protein intake) during the same period [32]. The 64-day NPR period in the prepubertal stage coincided with the allometric growth of the mammary gland [33]. These ewes synchronized in oestrus, so the lambing took place in a short time window to avoid variations caused by the lactation stage. Animals were housed in individual pens and milked twice daily in a 1 × 10 stall milking parlor (DeLaval). No tranquilizers or anesthesia were administered to the animals during the experiment.

After this challenge, in a second lactation, 21 ewes out of the 40 initial ewes (the reduction in animals was caused by animals that did not become pregnant or did not reach the lactation stage) were fed ad libitum with a total mixed ratio (TMR) [34]. The distribution of these animals in the NPR challenge group was as follows: 11 controls and 10 NPRs. Again, before this second lactation, the ewes had synchronized oestrus, and the lambing also took place in a short time window, similar to the first lactation. After three weeks of adaptation to the TMR and for three additional weeks, the dry matter intake (DMI) and milk yield of these ewes were measured daily. Individual feed intake for each animal was estimated daily by weighing the refused dry matter. Additionally, morning and evening milk products were weighed for each animal to calculate the daily milk yield. Protein, fat, and lactose contents were estimated for each animal as described by Barrio et al. (2022) [34]. The changes in body weight (BW) were calculated for each ewe by recording two consecutive days at the beginning and end of the FE evaluation.

The FCR and RFI of each ewe were estimated for the evaluated period. The FCR was estimated as the ratio between the DMI and energy-corrected milk yield (ECM, kg/d), where the ECM was calculated via INRA (2018) [35] equations for sheep.

$$\text{ECM} = \text{kg/d milk yield} \times [(0.0071 \times \text{g/kg milk fat}) + (0.0043 \times \text{g/kg milk protein}) + 0.2224]$$

The RFI was estimated using the GLM procedure of the SAS software package (version 9.4; SAS Institute, Inc.) and expressed as the residuals of the following regression model:

$$\text{DMI} = \mu + a \times \text{ECM} + b \times \text{MBW} + c \times \text{BWC} + \text{RFI}$$

where.

DMI = mean dry matter intake during the experiment (kg/d); μ = intercept; ECM = energy-corrected milk yield (kg/d); MBW = mean metabolic body weight during the experiment ($\text{BW}^{0.75}$; kg); BWC = BW change per day (kg/d); RFI = residuals.

DNA extraction and whole-genome bisulfite sequencing

Milk samples were collected at the end of the period when FE was evaluated, between days 40 and 85 at the second lactation (between March and April 2021). All the ewes were born between 25/10/2018 and 06/11/2018. Therefore, the ewes had between 28 and 30 months of age during the experiment. DNA was isolated from milk somatic cells using a noninvasive sampling procedure. A final volume of 100 ml of milk was collected from the 21 animals during the morning milk, and FE estimation was performed during the second lactation for whole-genome bisulfite sequencing (WGBS). The milking samples were collected at the period when FE was evaluated, between days 40 and 50 at second lactation. The protocol for milk somatic cell isolation and DNA extraction was described by Fonseca et al. (2023) [36]. The sequencing libraries (150 bp paired-end) were constructed on the Novogene platform in Cambridge (UK) and sequenced on an Illumina NovaSeq 6000 system, with a minimum coverage depth of 20X for each sample. Detailed information regarding library preparation and WGBS is available from Fonseca et al. (2022) [68]. The reads obtained from the sequencing data are available in the European Nucleotide Archive (ENA) repository under accession number PRJEB56589.

Differential methylation between divergent feed-efficient groups

The quality control (QC) and alignment steps for the WGBS reads obtained for the 21 ewes are described in detail in Fonseca et al. (2023) [32]. Briefly, FastQC [37] and Trim Galore software (version 0.6.5) [38] were used for QC and read trimming (reads were trimmed based on the quality scores, adapters were removed, and short reads were filtered), respectively. The reads were subsequently aligned to the ovine reference genome (Oar_Ram_v2.0), and the methylation calling procedure was performed via Bsseeker2 [39]. In the present study, methylated sites with fewer than five reads mapped within the region were filtered out.

The effect of prepubertal protein restriction on the RFI and FCR values was evaluated via a linear model in R (v. 4.2.0). The 21 ewes were subsequently divided into high- and low-efficiency groups on the basis of their RFI and FCR values. Additionally, high- and low-efficiency groups were created on the animals with consensus values (Cons) for both metrics. For RFI, the selected groups were composed of seven ewes with a negative RFI (L-RFI, more efficient) and seven with the highest RFI (H-RFI, less efficient). Similarly, for FCR, seven animals in each extreme of the FCR distribution composed the high-FCR (H-FCR, more efficient) and low-FCR (L-FCR, less efficient) groups. The consensus groups (H-Cons and L-Cons) were formed by six animals at each extreme of the RFI and FCR distributions, which did not have contrasting FE values between the metrics (H-RFI and H-FCR or L-RFI and L-FCR).

The DMLs and DMRs associated with FEs were identified by comparing the high and low groups for RFI, FCR, and consensus individually via the R package DSS [40]. For each grouping, the DMLs were identified by applying a multifactorial model where the FE groups, the NPR groups, and the interaction between the FE and NPR groups were included as terms. The DMLs identified for the FE group term were further selected to identify DMRs. Consequently, DMLs were identified on the FE classification via the RFI, FCR, and consensus groups. The DMRs were determined on regions harboring statistically significant methylated sites based on the following criteria: FDR-adjusted p value $< 10^{-5}$ for methylated sites, minimum length of 50 bp, minimum number of 50 methylated sites, significant percentage of methylated sites in the region (0.5), and a methylation difference greater than 0.1. Additionally, the DMRs mapped within regions located less than 50 bp from each other were merged into a single DMR.

Gene ontology and metabolic pathway enrichment analysis of the identified DMRs

The gene annotation for each DMR was performed using the R package “genomation” [41]. The DMRs were assigned to candidate genes and annotated for genomic context (promoter, intron, exon, and intergenic) on the ovine reference genome Oar_Ram_v2.0 (annotation release 104). For DMRs that were not mapped within a gene coordinate (intergenic regions), the closest gene was assigned to the DMR. Gene Ontology (GO) term and KEGG pathway enrichment analyses were performed via the R package gprofiler2. For GO terms, only terms with more than 2 and fewer than 1,000 assigned genes were considered enriched. In addition, a redundancy reduction analysis was performed using the “rutils” package in R. The Wang method was applied through the `go_reduce()` function to estimate semantic similarities between the

GO terms. The terms with a similarity threshold higher than 0.7 were subsequently grouped. Enriched status was defined on a false discovery rate (FDR) < 0.05 . The enrichment analysis was performed individually for DMRs obtained for the RFI, FCR, and consensus groups.

Variant calling within DMRs using whole-genome bisulfite sequencing reads

The variant calling procedure was performed in the EpiDiverse pipeline (<https://github.com/EpiDiverse/SNP>), which is essentially based on SAMtools software [42]. In the first step, the base quality scores from bam files are manipulated via a “double-masking” procedure via Revelio [43]. After this step, a combination of variant calling with FreeBayes (<https://github.com/ekg/freebayes>) and postcalling filtering with BCFtools is performed for the alignments. Both single-nucleotide variants (SNPs) and insertions and deletions (INDELs) were called using this approach. To use the abovementioned pipeline, the WGBS reads were realigned against the ovine reference genome Oar_Ram_v2.0 via bwa-meth software (<https://github.com/brentp/bwa-meth>). This step was required to obtain bam files with quality scores, which was not possible when using Bssseeker2. Importantly, only the regions comprising the detected DMRs were included in the variant calling and that false-positives can be included during the double-masking procedure via Revelio. The variants identified were subjected to a QC using the software Plink v2 [44] where only variants with minor allelic frequency (MAF) > 0.1 and with a linkage disequilibrium (r^2) < 0.2 with other variants were retained.

The genetic relationship among the ewes was assessed using a subset of 21,222 variants obtained after the filtering 55,627 variants available in the 50 K Affymetrix SNP chip [45] by MAF > 0.1 , call rate > 0.9 , and with a linkage disequilibrium (r^2) < 0.2 . The kinship coefficient was estimated using the KING-robust estimator [46] using the software Plink v2 [44].

Prediction of residual feed intake and the feed conversion ratio via machine learning models

A summary of the workflow applied in the ML predictions for RFI and FCR is presented in Fig. 1. Predictive models were built using the methylation means within each detected DMR as features. To estimate the predictive potential of these DMRs identified in the comparison between extreme RFI groups to predict RFI values (mRFI_RFI), DMRs identified in the comparison between extreme FCR groups were used to predict FCR values (mFCR_FCR), and DMRs identified in the comparison between extreme consensus groups were used to predict both RFI (mCons_RFI) and FCR (mCons_FCR) values. The ML models were built on the basis of a multilayer feedforward artificial neural network (deeplearning) and



Fig. 1 Flowchart describing the pipeline applied for machine learning-based predictions for residual feed intake (RFI) and the feed conversion ratio (FCR) using the mean methylation level within differentially methylated regions (DMRs) and genetic variants (single-nucleotide variants and insertions and deletions). The models in the hyperparameter tuning step were selected on the basis of the Spearman correlation (ρ). After this step, the best overall model was selected on the basis of the ratio between the squared Spearman correlation (ρ^2) and the root mean squared error (RMSE)

random forest (RF) using the R package h2o [47] and extreme gradient boosting (xgboost) using the R package xgboost (<https://cran.r-project.org/web/packages/xgboost/index.html>). For the deep learning and RF algorithms, a discrete random search composed of 5-fold cross-validation was performed during the hyperparameter tuning stage. These three methods were compared because of the different characteristics related with the ability to handle large and complex data, computational efficiency and overfitting when small datasets are evaluated.

The following parameters were included in the grid for the tuning stage: hidden (number of layers and neurons per layer) = [(5, 5, 5, 5), (10, 10, 10, 10), (50, 50, 50, 100, 100, 100)]; epochs (number of times to iterate (stream) the dataset) = [5, 10, 50, 100, 200]; L1 regularization =

[0, 1×10^{-5} , 1×10^{-4}]; rate (learning rate) = [0, 0.01, 0.005, 0.001]; ρ (adaptive learning rate time decay factor) = [0, 0.01, 0.005, 0.001]; input dropout ratio = [0, 0.1, 0.2].

The grid used in the hyperparameter tuning for RF included ntree (number of trees) = [5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100]; mtries (number of columns to randomly select at each level) = [5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100]; and the sample rate (row sample rate per tree) = [0.5, 0.8, 1].

The xgboost grid search was performed extensively using 1,000 maximum boosting interactions and 10 interactions as early stops (stopping if no improvement for 10 consecutive trees) in a grid composed of the following parameters: eta (learning rate) = [0.001, 0.01, 0.05];

max_depth (learning rate)= [5, 10]; min_child_weight (minimum sum of instance weights (Hessian) needed in a child)= [1, 5, 10]; subsample (subsample ratio of the training instance)= 1; colsample_bytree (subsample ratio of columns when constructing each tree)= [0.5, 0.8, 1].

In total, 100 models were built for each ML algorithm through the random assignment of the 21 ewes in the training (2/3, 14 animals) and test (1/3, 7 animals) samples. Importantly, there were not completely equal training and test samples generated in all 100 iterations of random assignments. Among the 100 fitted models, the best model was defined as the ratio between the squared Pearson correlation (ρ^2) and the root mean squared error (RMSE) between the predicted and real values of RFI and FCR in the training stage, where higher ratios suggest better models. Finally, the best model was defined, and the predictive accuracy was estimated on the basis of the RMSE and ρ^2 , which were based on 10 rounds of leave-one-out cross-validation.

Similarly, the same steps described above were applied to test the predictive accuracy of ML models using the methylation means within DMRs plus the variants, from here called VARs, identified and mapped within the DMRs. This approach was used to test a putative improvement in predictive accuracy by the inclusion of genomic information.

Results

Comparison of residual feed intake and feed conversion rates between extreme feed efficiency groups and nutritional protein restriction groups

The ewes belonging to each FE extreme group (H-RFI, L-RFI, H-FCR, L-FCR, H-Cons, and L-Cons) and the respective FE metrics and the subdivisions within each group based on the NPR groups (NPR and control) are shown in Additional file 1. Significant differences were observed between the mean values of RFI (p value = 1.03×10^{-06}) and FCR (p value = 0.005) observed for the H and L groups (Table 1). On the other hand, the NPR factor does not seem to significantly affect either the RFI (p value = 0.56) or the FCR (p value = 0.14) metric. Despite this absence of effect over FE metrics, we decided

to include the NPR group as a fixed effect on the model used to identify DMLs between extremely divergent FE groups owing to potential biological processes that could be affected by the NPR challenge.

Additional File 2 shows the distribution of the molecular kinship coefficients assessed across all the samples and within each FE group. A mean below second-degree relationship was observed for all the evaluated groups, and no higher kinship between high- and low-efficient animals has been observed, irrespective of the index considered.

Whole-genome bisulfite sequencing, differentially methylated region calling, and variant calling

The alignment statistics obtained via Bssseeker2 for the samples included in DML and DMR calling are available in Additional file 2. A mean of 192.57 ± 23.24 million paired reads were aligned across Oar_Ram_v2.0, corresponding to an average mapping of 0.66 ± 0.03 . The DMRs detected for each FE group are presented in Additional File 4. In total, 28,255, 17,431, and 9,720 DMRs were identified in the comparison of the divergent groups for RFI, FCR, and Cons, respectively. The distribution of DMRs per chromosome is shown in Fig. 2A, where it is possible to note that chromosomes 1, 2, and 3 had the largest number of DMRs for all the FE groups. The gene annotation process resulted in the assignment of 12,257, 9,328, and 6,723 genes to the DMRs detected for the RFI, FCR, and Cons groups, respectively. In general, most of the genes with assigned DMRs were shared between at least two FE groups (Fig. 2B). In total, 2,921 genes were shared among all three groups. The percentage of DMRs assigned to different genic contexts (intergenic, promoter, exon, and intron) was very similar across the FE groups (Fig. 2C-E). For all the FE groups, the majority of the DMRs were mapped in intergenic and intronic regions. Importantly, the same DMR can be mapped in different contexts because of the presence of different transcripts for each gene.

The variant calling process initially identified 194,326, 129,159, and 78,488 variants for the RFI, FCR and Cons groups, respectively. After the QC, 59,606, 43,492, and 28,103 variants were retained for the RFI, FCR and RFI groups.

Enriched metabolic pathways associated with genes harboring FE-associated DMRs

A total of 79, 56, and 22 enriched KEGG pathways were identified as enriched ($FDR < 0.05$) for the genes harboring DMRs identified in the comparison of the divergent groups for RFI, FCR, and Cons, respectively (Additional file 5). Only genes harboring DMRs in exons, introns, or promoters were used as inputs for the enrichment analyses. Figures of enriched pathways shared between the FE

Table 1 Comparison between residual feed intake (RFI) and feed conversion ratio (FCR) values between high and low groups and between Ewes subjected to nutritional protein restriction (NPR) and control Ewes

Group	RFI		FCR	
	Mean (\pm SD)	P value	Mean (\pm SD)	P value
High	0.29 \pm 0.08	1.03 $\times 10^{-06}$	1.97 \pm 0.25	0.005
Low	-0.05 \pm 0.04		1.38 \pm 0.11	
NPR	0.12 \pm 0.13	0.56	1.74 \pm 0.36	0.14
Control	0.08 \pm 0.18		1.54 \pm 0.21	

SD: standard deviation

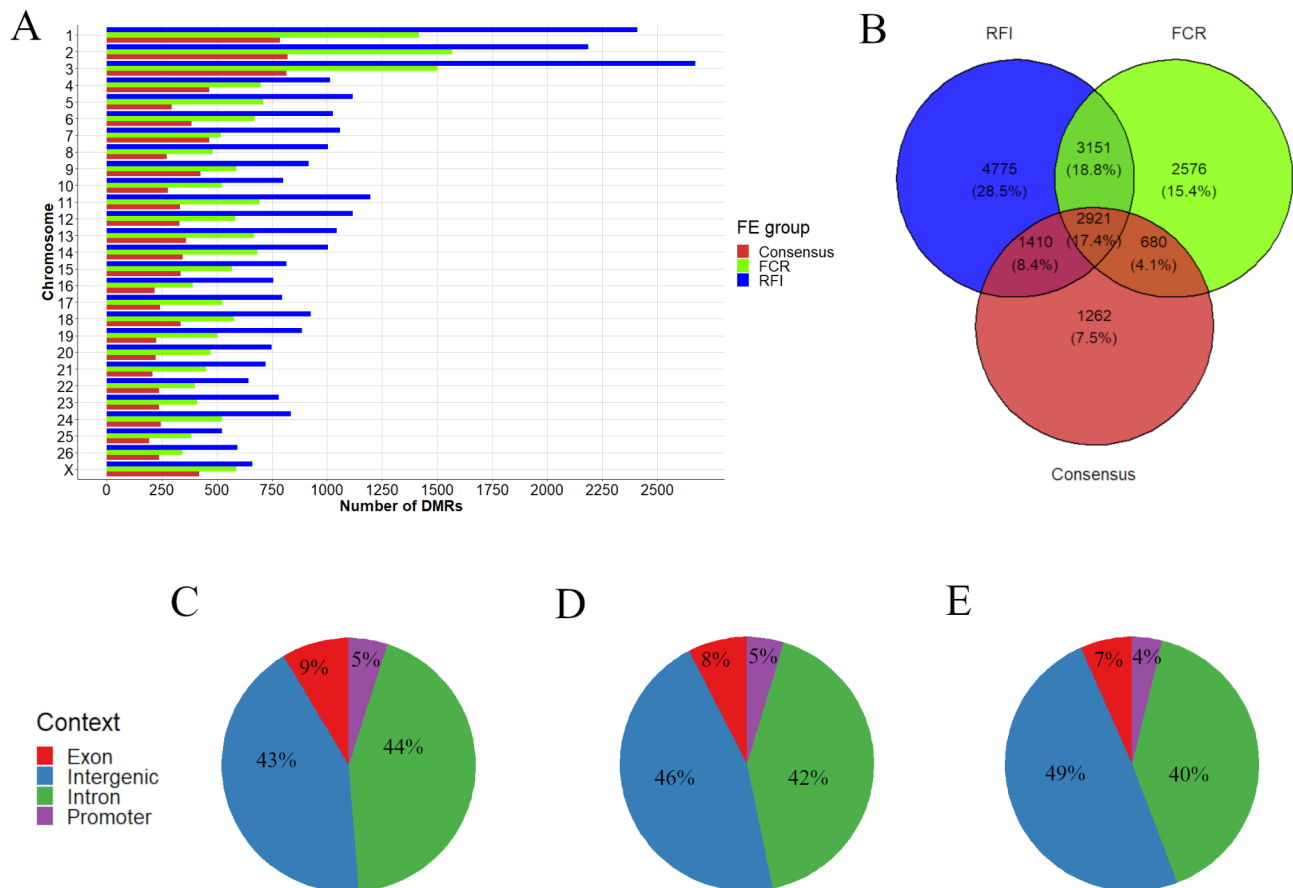


Fig. 2 Distribution of differentially methylated regions (DMRs) per chromosome (**A**) for residual feed intake (RFI, in blue), feed conversion ratio (FCR, in green), and consensus (Cons, in red). Venn diagram showing the number of genes harboring DMRs shared between RFI, FCR, and Cons. Percentages of DMRs mapped in exons (red), intergenic regions (blue), introns (green), and promoters (purple) for RFI (**C**), FCR (**B**), and Cons (**D**)

groups are shown in Fig. 3. Sixteen enriched pathways were shared among RFI, FCR, and Cons, whereas 33 pathways were shared exclusively between RFI and FCR. Additionally, one and two enriched pathways were shared between RFI and FCR and the Cons group, respectively. The genes harboring DMRs identified for the RFI groups presented the greatest number of exclusively enriched terms (29 pathways).

Metabolic pathways for the FE status were observed in the 16 pathways shared among the three FE groups. Among these 16 pathways, the following pathways were interesting to highlight: the calcium signaling pathway, the MAPK signaling pathway, protein digestion and absorption, glycerolipid metabolism, the Wnt signaling pathway, glutamatergic synapses, the PI3K-Akt signaling pathway, the phospholipase D signaling pathway, the ECM-receptor interaction pathway, and the phosphatidylinositol signaling system. Additionally, enriched metabolic pathways involved with relevant biological processes associated with the FE status were observed among those pathways shared between RFI and FCR, and others were exclusively identified for the RFI group.

Among the enriched pathways shared between the RFI and FCR groups, the Apelin signaling pathway, oxytocin signaling pathway, growth hormone synthesis, secretion and action, insulin secretion, parathyroid hormone synthesis, secretion and action, and GnRH secretion, among others, were interesting. The 29 enriched pathways identified exclusively for the RFI group presented biologically relevant pathways, such as the relaxin signaling pathway; GABAergic synapse; insulin signaling pathway; valine, leucine, and isoleucine degradation; insulin resistance; GnRH signaling pathway; murine metabolism; and the thyroid hormone signaling pathway. Importantly, despite some pathways being identified exclusively for the RFI group, biological redundancy is observed with other pathways shared between RFI and FCR, such as the GnRH signaling pathway and GnRH secretion. This redundancy might be explained by genes annotated exclusively for one of the FE metrics acting in similar biological processes compared with the functional profile of those genes shared between both FE metrics.

The comparison of the enriched KEGG pathways associated with the genes harboring DMRs revealed a similar

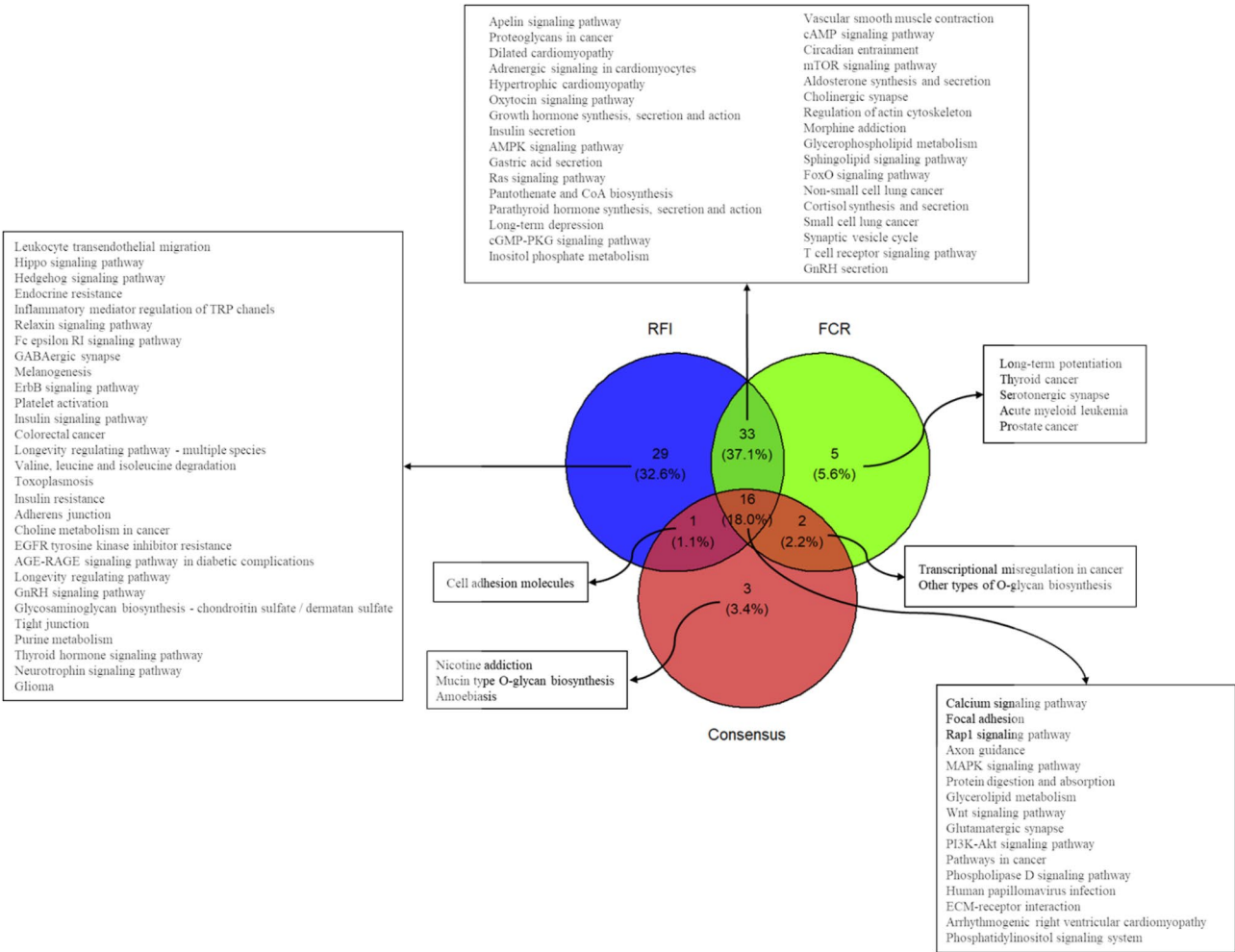


Fig. 3 Venn diagram showing the number of enriched pathways identified for the genes harboring differentially methylated regions (DMRs) shared between the residual feed intake (RFI), feed conversion ratio (FCR), and consensus groups

pattern to that of the number of genes shared between the groups. As mentioned, 16 KEGG pathways were enriched for the genes harboring DMRs identified via the three analysis approaches (RFI, FCR, and Cons). In addition, 33 pathways were shared between RFI and FCR. This might suggest that when the animals are grouped on the basis of a consensus between both metrics, a perfect agreement is not necessarily obtained. If this were the case, we would expect all the enriched terms from the Cons grouping to be shared with the RFI and FCR groups. Nevertheless, it is interesting to highlight that the enriched KEGG pathways identified exclusively for the Cons group (nicotine addiction, mucin-type O-glycan biosynthesis, and amoebiasis) or shared only with RFI (cell adhesion molecules) or FCR (transcriptional misregulation in cancer and other types of O-glycan biosynthesis) are related to broad metabolic processes without any direct association with FE. Similarly, the enriched pathways identified exclusively in the list of genes harboring DMRs identified for the FCR group are also related

to broad metabolic processes without any direct association with FEs (long-term potentiation, thyroid cancer, serotonergic synapses, acute myeloid leukemia, and prostate cancer). On the other hand, the enriched pathways shared between all three groups, between RFI and FCR only and exclusively enriched for RFI, are associated with functionally interesting metabolic processes for FEs. Table 2 shows the functionally relevant metabolic pathways enriched in the genes harboring DMRs in the RFI, FCR, and Cons groups.

Predictive accuracy of machine learning models for the residual feed intake and feed conversion rate

The features included in the datasets used in the ML algorithms for the predictions of FE metrics were composed by the mean methylation levels within DMRs and the genetic variants identified within the DRMs: RFI (28,255 DMRs and 59,606 variants), FCR (17,431 DMRs and 43,492 variants), and Cons (9,720 DMRs and 28,103 variants).

Table 2 Functionally relevant enriched KEGG pathways for feed efficiency (FE) identified for the genes harboring differentially methylated regions (DMRs) for residual feed intake (RFI), feed conversion ratio (FCR), and consensus (Cons)

KEGG pathway	P value	FE metric
Hormones and peptides		
Aldosterone synthesis and secretion	0.0079	RFI
Aldosterone synthesis and secretion	0.0084	FCR
Apelin signaling pathway	0.0004	RFI
Apelin signaling pathway	0.0197	FCR
GnRH secretion	0.0300	RFI
GnRH secretion	0.0023	FCR
GnRH signaling pathway	0.0030	RFI
Growth hormone synthesis, secretion, and action	0.0015	RFI
Growth hormone synthesis, secretion, and action	0.0100	FCR
Insulin resistance	0.0009	RFI
Insulin secretion	0.0015	RFI
Insulin secretion	0.0011	FCR
Insulin signaling pathway	0.0286	RFI
Oxytocin signaling pathway	0.0013	RFI
Oxytocin signaling pathway	0.0003	FCR
Parathyroid hormone synthesis, secretion and action	0.0020	RFI
Parathyroid hormone synthesis, secretion and action	3.17×10^{-5}	FCR
Thyroid hormone signaling pathway	0.0070	RFI
Synapses		
GABAergic synapse	0.0216	RFI
Glutamatergic synapse	0.0005	RFI
Glutamatergic synapse	2.57×10^{-5}	FCR
Glutamatergic synapse	0.0147	Cons
Serotonergic synapse	0.0184	FCR
Signaling pathway		
Calcium signaling pathway	0.0001	RFI
Calcium signaling pathway	5.97×10^{-5}	FCR
Calcium signaling pathway	7.63×10^{-5}	Cons
cAMP signaling pathway	0.0031	RFI
cAMP signaling pathway	0.0032	FCR
Hippo signaling pathway	0.0095	RFI
Phosphatidylinositol signaling system	0.0022	RFI
Phosphatidylinositol signaling system	0.0021	FCR
Phosphatidylinositol signaling system	0.0429	Cons
ErbB signaling pathway	0.0261	RFI
PI3K-Akt signaling pathway	0.0001	RFI
PI3K-Akt signaling pathway	0.0079	FCR
PI3K-Akt signaling pathway	0.0167	Cons
Rap1 signaling pathway	2.58×10^{-6}	RFI
Rap1 signaling pathway	8.16×10^{-5}	FCR
Rap1 signaling pathway	0.0002	Cons
Ras signaling pathway	0.0020	RFI
Ras signaling pathway	0.0378	FCR
Wnt signaling pathway	0.0046	RFI
Wnt signaling pathway	0.0156	FCR
Wnt signaling pathway	0.0147	Cons
Others		
Circadian entrainment	0.0067	RFI
Circadian entrainment	0.0126	FCR
Cortisol synthesis and secretion	0.0164	RFI
Cortisol synthesis and secretion	0.0360	FCR

Table 2 (continued)

KEGG pathway	P value	FE metric
Inflammatory mediator regulation of TRP channels	0.0128	RFI
Protein digestion and absorption	0.0179	RFI
Protein digestion and absorption	0.0053	FCR
Protein digestion and absorption	0.0016	Cons
Valine, leucine and isoleucine degradation	0.0432	RFI

Table 3 Mean (\pm standard deviation) for the root mean squared error (RMSE) and mean squared pearson correlation (ρ^2) obtained in the 100 rounds of random assignment of Ewes in the training and test datasets for each machine learning model to predict residual feed intake (RFI) and the feed conversion ratio (FCR)

Prediction	Deeplearning		RF		xgboost	
	Mean RMSE	Mean ρ^2	Mean RMSE	Mean ρ^2	Mean RMSE	Mean ρ^2
mRFI_RFI	0.20 \pm 0.05	0.18 \pm 0.22	0.17 \pm 0.03	0.14 \pm 0.16	0.18 \pm 0.03	0.16 \pm 0.20
mFCR_FCR	0.35 \pm 0.10	0.14 \pm 0.18	0.31 \pm 0.07	0.15 \pm 0.17	0.35 \pm 0.07	0.14 \pm 0.17
mCons_RFI	0.19 \pm 0.08	0.16 \pm 0.19	0.17 \pm 0.03	0.17 \pm 0.21	0.18 \pm 0.02	0.16 \pm 0.16
mCons_FCR	0.35 \pm 0.11	0.18 \pm 0.20	0.31 \pm 0.08	0.14 \pm 0.17	0.34 \pm 0.08	0.17 \pm 0.21
mRFI_RFI+VARs	0.25 \pm 0.15	0.18 \pm 0.22	0.17 \pm 0.03	0.15 \pm 0.16	0.17 \pm 0.03	0.11 \pm 0.15
mFCR_FCR+VARs	0.59 \pm 0.41	0.19 \pm 0.20	0.31 \pm 0.09	0.14 \pm 0.14	0.35 \pm 0.07	0.18 \pm 0.20
mCons_RFI+VARs	0.24 \pm 0.08	0.20 \pm 0.23	0.18 \pm 0.04	0.17 \pm 0.19	0.18 \pm 0.03	0.18 \pm 0.20
mCons_FCR+VARs	0.44 \pm 0.13	0.20 \pm 0.22	0.30 \pm 0.08	0.13 \pm 0.15	0.33 \pm 0.08	0.15 \pm 0.20

mRFI_RFI: Mean methylation within DRMs identified comparing RFI groups used to predict RFI; mFCR_FCR: Mean methylation within DRMs identified comparing FCR groups used to predict FCR; mCons_RFI: Mean methylation within DRMs identified comparing Consensus groups used to predict RFI; mCons_FCR: Mean methylation within DRMs identified comparing Consensus groups used to predict FCR; mRFI_RFI+SNPs: Mean methylation within DRMs identified comparing RFI groups plus SNPs within the same DMRs used to predict RFI; mFCR_FCR+SNPs: Mean methylation within DRMs identified comparing FCR groups plus SNPs within the same DMRs used to predict FCR; mCons_RFI+SNPs: Mean methylation within DRMs identified comparing Consensus groups plus SNPs within the same DMRs used to predict RFI; mCons_FCR+SNPs: Mean methylation within DRMs identified comparing Consensus groups plus SNPs within the same DMRs used to predict FCR

Table 4 Best overall models, which are based on the highest ratio between the root mean squared error (RMSE) and squared spearman correlation (ρ^2), identified for each predictive scenario and the respective hyperparameters used in the model

Prediction	Model	Hyperparameters	RMSE	ρ^2
mRFI_RFI	xgboost	max_depth = 5; eta = 0.05; nround = 315; min_child_weight = 1; subsample = 1; colsample_bytree = 0.5	0.10	0.86
mFCR_FCR	deeplearning	hidden = 5,5,5,5,5; epochs = 5.087879; l1 = 1e-04; rate = 1	0.19	0.73
mCons_RFI	deeplearning	hidden = 100,100,100; epochs = 51.294697; l1 = 1e-05; rate = 0.001	0.07	0.37
mCons_FCR	xgboost	max_depth = 5; eta = 0.05; nround = 343; min_child_weight = 1; subsample = 1; colsample_bytree = 0.5	0.17	0.93
mRFI_RFI+VARs	xgboost	max_depth = 5; eta = 0.05; nround = 303; min_child_weight = 1; subsample = 1; colsample_bytree = 0.5	0.12	0.73
mFCR_FCR+VARs	RF	mtry = 80; ntrees = 27; sample_rate = 1.0	0.19	0.51
mCons_RFI+VARs	RF	mtry = 45; ntrees = 15; sample_rate = 0.5	0.07	0.62
mCons_FCR+VARs	deeplearning	hidden = 10,10,10,10; epochs = 10.450117; l1 = 1e-04; rate = 1	0.23	0.80

The predictive performance of the ML algorithms was compared for 100 rounds of random assignment of animals in the training and test datasets. Similar performance results were obtained across the ML algorithms and FE datasets (Table 3). The smallest mean RMSE (0.17) was obtained for mRFI_RFI, mCons_RFI, and mRFI_RFI+VARs via RF and mRFI_RFI+VARs the mean ρ^2 value, the largest observed value (0.20) was obtained for mCons_RFI+VARs and mCons_FCR+VARs via deeplearning. An interesting variation among the results obtained in the 100 rounds of random assignments was observed for the RMSE and ρ^2 values (Additional file 6). Therefore, the best model (defined as the model with the highest value for the ratio between ρ^2 and RMSE)

and the respective hyperparameters for each tested scenario were identified and are shown in Table 4. The highest ratios were obtained for mCons_RFI+VARs (RMSE = 0.07, ρ^2 = 0.62) via RF and mRFI_RFI via xgboost (RMSE = 0.10, ρ^2 = 0.86). In addition, the highest correlation between the actual and predicted values was obtained in the mCons_FCR scenario via xgboost (ρ^2 = 0.93). On the other hand, the lowest correlation in the best model was obtained in the mCons_RFI scenario via deeplearning (ρ^2 = 0.37).

Discussion

[48] Different nutritional models have been proposed to estimate the energy requirements of small ruminants [48], but they struggle to account for environmental, feeding, and individual variability. A meta-analysis on resilience and efficiency traits in meat sheep reported pooled heritability of 0.32 ± 0.15 for RFI and 0.12 ± 0.03 for FCR [49], aligning with previous findings where RFI shows higher heritability (0.11–0.54) than FCR (0.05–0.77) across sheep breed [7, 14, 15, 16]. Most estimates focus on meat sheep, while pooled heritability for FE in dairy sheep remains uncalculated due to limited data.

The difficulty in obtaining genetic parameters for FE traits in dairy sheep arises from the challenge of accurately measuring feed intake and energy requirements. Since FE represents an animal's ability to convert feed into metabolically available nutrients, defining an appropriate timeframe for measurement is essential. For dairy animals, selecting a timeframe that reflects the entire lactation period is crucial to ensure FE estimates lead to sustained improvements [50]. Therefore, the development of studies aimed at the estimation of FE in dairy sheep is important for the dairy sheep industry. The investigation of the genetic basis of different FE estimates has the potential to help provide a better understanding of the biological processes captured by each estimate. Consequently, this helps develop more precise nutritional models and FE estimates. Additionally, owing to the difficulty of measuring a straightforward operational system to collect data for FE estimation in dairy breeds, a potential alternative would be the creation of a reference population from which biomarkers could be used to efficiently train predictive models for the estimation of FEs in new animals.

Functional profile of DMRs identified between divergent FE groups

Evaluating epigenetic marks associated with FE offers a valuable approach, as nutritional status influences epigenetic modifications and vice versa through enzyme regulation [22, 23, 24]. This study identified DMRs by comparing high- and low-FE dairy ewes using RFI and FCR, reaching a consensus between both metrics for animal grouping. The RFI-based grouping resulted in the highest number of DMRs (28,255), likely due to the greater difference between L-RFI and H-RFI groups compared to L-FCR and H-FCR groups. This also explains the larger number of genes exclusively harboring DMRs in the RFI comparison. Interestingly, most DMRs identified in the H-Cons vs. L-Cons comparison were mapped within genes shared with at least one of the two FE metrics, with 2,921 genes overlapping across all three contrasting approaches. Functional analysis of genes harboring DMRs provided further insights into the

relationships among identified DMRs. This study focused on understanding the effects of DMRs on FE in Assaf ewes. The NPR group was included as an effect in the model used for DMR detection. However, as Fonseca et al. (2023) [36] provides an in-depth discussion on DMRs directly influenced by NPR, the present study emphasizes DMRs identified from FE group comparisons.

The protein digestion and absorption pathway was enriched in the three groups (RFI, FCR, and Cons). A review of the biological determinants of between-animal variation in FE suggested that feeding and digestive-related terms/pathways could be associated with RFI due to its covariation with feed intake in beef cattle [51]. Interestingly, several collagen-related genes were associated with this enriched pathway (protein digestion and absorption pathway) and harbored DMRs. In sheep, severe dietary protein restriction results in the downregulation of *COL1A1*, *COL1A2*, *COL3A1*, and *COL5A1* in the rumen epithelium [52]. Among these genes, *COL5A1* was identified in the present study as associated with DMRs for RFI (six DMRs), FCR (seven DMRs), and Cons (one DMR). Interestingly, several members of the collagen synthesis gene family (despite not having the same genes identified here) were identified as differentially expressed in the mammary gland transcriptome between High-FE and Low-FE Assaf ewes [27]. These proteins play crucial roles in the transport and absorption of substances across membranes, acting as gatekeepers for a dynamic response to different metabolic states [53].

Among the functionally relevant enriched pathways, a group of hormones and peptide-related pathways were identified as good candidates for the FE status evaluated in the current study. These pathways were shared between the RFI and FCR groups and were associated with the synthesis, secretion, and signaling of apelin, aldosterone, GnRH, growth hormone, insulin, oxytocin, parathyroid, and thyroid hormones [54].

The apelin gene (*APLN*) encodes a bioactive peptide that is associated with adipose tissue metabolism and the regulation of glucose and insulin levels, consequently playing an important role in energy metabolism [54, 55]. Additionally, in fish and mice, apelin has been described as an important appetite regulator [56, 57, 58]. Although the Apelin signaling pathway was enriched for the genes harboring the identified DMRs for both RFI and FCR contrasts, only the analysis comparing H-RFI and L-RFI identified DMRs associated with the *APLN* gene (X:113587436–113587514 and X:113577488–113577543). However, interestingly, two DMRs (6:299653–299738 and 6:368918–369448) identified exclusively for the FCR groups were related to the *APELA* (Apelin receptor early endogenous ligand) gene. The *APELA* gene is suggested to have originated from the same common ancestor of apelin, and in chickens, it

seems to play important roles in hepatic lipid metabolism [59]. Consequently, *APLN* and *APELA* might play similar roles in the control of FE in sheep.

The aldosterone synthesis and secretion pathway is associated with insulin levels through a regulatory mechanism related to potassium depletion [60]. In mice, genetic aldosterone deficiency results in increased insulin secretion without alterations in insulin sensitivity or the prevention of adipocyte dysfunction [61, 62] [64]. Here, the insulin secretion pathway was also enriched for the genes associated with DRMs identified in both the FCR and RFI groups. In contrast, the insulin secretion and resistance pathways were enriched for the genes associated with DRMs identified in the RFI groups. The effects of insulin on the regulation of feed intake and energy expenditure are well understood and characterized through several mechanisms, such as the regulation of leptin release [63, 64]. In total, 28, four, and one genes were associated with the aldosterone synthesis and secretion pathway and insulin secretion, resistance, and signaling pathways, respectively. Seven genes from the adenylate cyclase gene family (*ADCY1*, *ADCY2*, *ADCY3*, *ADCY5*, *ADCY6*, *ADCY7*, and *ADCY9*) are among the genes associated with aldosterone synthesis and secretion and insulin secretion pathways. The *Adcy*/PKA-dependent pathway plays a crucial role in regulating triglyceride storage or mobilization in white adipose tissue [65]. *ADCY1*, *ADCY3*, *ADCY5*, and *ADCY6* were previously related to the regulation of insulin secretion in mice and humans [66, 67, 68, 69, 70]. *ADCY3* haploinsufficiency results in reduced expression of genes related to thermogenesis, fatty acid oxidation, and insulin signaling and increased expression of adipogenesis-related genes [68]. In light of the above findings, the DRMs identified in these genes are interesting biomarkers for FE because of the functional impact of these genes on insulin levels, feed intake, and energy expenditure.

Growth hormone synthesis, secretion, and action and the GnRH secretion pathway were enriched in both the RFI and FCR groups. The leptin protein acts over GnRH, increasing its secretion when the energy reserves are abundant to promote reproductive functions [71]. On the other hand, ghrelin acts as a ligand of growth hormone (GH) secretagogue receptor to stimulate GH release, appetite, and food intake [72]. Interestingly, DRMs were associated with the ghrelin gene (*GHRL*) for both the RFI (19:54284696–54284775 and 19:54289697–54289756) and FCR (19:54303886–54303936 and 19:54289130–54289180) groups. Despite the absence of the leptin gene (*LEP*) among the genes harboring DRMs in both groups, another functionally relevant gene for leptin activity was identified as harboring DRMs in the RFI and FCR groups. Phosphoinositide-3-kinase regulatory subunit 1 (*PIK3RI*) couples the leptin and insulin signaling pathways and is

suggested to be required for the acute effect of leptin [73, 74].

In total, five DRMs were identified in the comparison between RFI groups (16:11522806–11522871, 16:11531077–11531139, 16:11526075–11526144, 16:11488225–11488529, and 16:11683716–11684150), and two were identified in the comparison between FCR (16:11606919–11607131 and 16:11457093–11457253) groups. Additionally, three DRMs were identified in the comparison between the H-Cons and L-Cons groups (16:11759779–11759847, 16:11755506–11755700, and 16:11472257–11472375). Among the genes associated with these DRMs, those related to thyroid hormones were identified. The effects of thyroid hormones on energy expenditure are well described in the literature [75]. The parathyroid hormone synthesis, secretion, and action pathways were enriched in both the RFI and FCR groups, whereas the thyroid hormone signaling pathway was enriched exclusively in the RFI group. The parathyroid hormone 1 receptor (*PTH1R*), which is associated with the parathyroid hormone synthesis pathway, showed DRMs in the RFI (19:52593996–52594053) and FCR (19:52626022–52626090) groups. The parathyroid hormone (PTH) is essential for calcium homeostasis and, consequently, for bone metabolism [76]. The action of PTH is mediated by *PTH1R*, so agonists of *PTH1R* are used to treat hypoparathyroidism. Additionally, treatment with PTH in patients with hypoparathyroidism resulted in a decrease in body weight (potentially by appetite reduction) and hypercalcemic episodes [77] [80]. Three DRMs were identified in the thyroid hormone receptor beta (*THRB*) gene for the RFI group (26:41130304–41130566, 26:41177101–41177241, and 26:41190490–41190562), which is associated with the thyroid hormone signaling pathway. *THRB* acts as a mediator of thyroid hormones, and mutations in *THRB* might result in resistance to thyroid hormones, leading to elevated levels of thyroid hormones, normal or elevated levels of TSH, and goiter [78].

The oxytocin signaling pathway was enriched for the genes associated with DRMs identified in the RFI and FCR groups. Recently, increasing evidence has linked oxytocin action with the regulation of food intake and energy expenditure [79, 80, 81]. Here, four DRMs associated with the oxytocin receptor gene (*OXTR*) for the RFI group were identified (19:17654618–17654782, 19:17674987–17675207, 19:17651048–17651183, 19:17667957–17668049). In humans, the methylation of the *OXTR* is associated with a series of callous-unemotional traits in youth, such as rigid thinking in anorexia nervosa [82]. These results suggest a link between this gene and feeding behavior. Additionally, the overexpression of *OXTRs* is associated with abnormal mammary gland development in mice [83]. Therefore, these results

suggest a potential impact of *OXTR* methylation on FEs in dairy sheep. In addition, transcriptomic analysis of the hypothalamus of pigs with extremely divergent FE statuses revealed that both the oxytocin signaling pathway and serotonergic synapses were enriched with differentially expressed genes [84]. Among the differentially expressed genes associated with the serotonergic synapse mentioned above, *KCNN2*, *KCND2*, and *PRKCG* were also identified as harboring DMRs in the current study. The differentially expressed genes associated with the oxytocin signaling pathway in the hypothalamus of pigs divergent from FEs and harboring DMRs identified in the present study were *ADCY1*, *ADCY7*, *CACNG5*, *CDKN1A*, *CAMKK2*, and *PRKCG*.

Among the enriched pathways identified in the present study, several pathways related to cellular signaling were enriched (calcium, cAMP, Hippo, phosphatidylinositol, ErbB, PI3K-Akt, RAP1, Ras, and Wnt signaling pathways). The calcium signaling pathway was previously identified as enriched for the genes differentially expressed in the comparison between high- and low-FE groups in chickens, pigs, and Nelore cattle [85, 86, 87] [91, 92]. The activation of the Ras/P13-K/Akt pathway is controlled by the connection of the epidermal growth factor receptor (EGFR), which collaborates with ErbB3 [88]. The duration of P13-K signaling is negatively regulated by phosphatase and tensin homolog (*PTEN*) [89]. Additionally, serine/threonine kinase 3 (*AKT3*) acts as one of the main regulators of the P13/Akt/mTOR pathway and modulates cell survival and proliferation in the liver of pigs with increased feed efficiency [90]. Interestingly, both *PTEN* and *AKT3* harbored DMRs identified in the comparison between divergent FEs in the current study. Two DMRs for the RFI groups (22:9249556–9249643 and 22:9295392–9295468) and two for the FCR groups (22:9498044–9498120 and 22:9444188–9444313) were identified in the *PTEN* gene. With respect to *AKT3*, one DMR was identified for the RFI groups in this gene (12:32754054–32754513). In addition, DMRs were also identified in other genes that play crucial roles in these pathways, such as *MTOR*, *MAP2K7*, *MAPK9*, *MAP2K1*, *MAPK10*, *PIK3R1*, and *EGFR*. [96]

The functional profile of the enriched pathways identified here suggests a significant association with the regulation of feed behavior and energy expenditure through hormonal activity and nervous system regulation. In addition to the abovementioned pathways, the GABAergic synapse (RFI), glutamatergic synapse (RFI, FCR, and Cons), and serotonergic synapse (FCR) pathways were enriched. GABAergic synapses act with inhibitory inputs in the regulation of energy and glucose homeostasis via nutrient and adiposity signals, with leptin acting in the regulation of GABAergic synapse strength [91]. The glutamatergic system interacts with the central

oxytocinergic system to regulate feeding behavior [92]. With respect to the serotonergic system, an inverse relationship between serotonin levels in the brain, food intake, and body weight has been reported [93]. Another important component of feed behavior and energy expenditure regulation is the circadian rhythm [94, 95]. Among the genes associated with DMRs and related to the enriched pathway Circadian entrainment, *PER2* (period circadian regulator) stands out with three DMRs for the RFI (1:3049840–3050077, 1:3027343–3027447, and 1:3022247–3022377) and FCR (1:3050049–3050304, 1:3032669–3032785, and 1:3046139–3046254) groups. *PER2* is an important regulator of the circadian clock, generating a circadian rhythm in the central nervous system and peripheral organs [96]. Additionally, *PER2* is responsible for the regulation of *PPAR γ* , a nuclear receptor that plays crucial roles in adipogenesis, the inflammatory response, and insulin sensitivity [97, 98, 99]. In cattle, the silencing of *PER2* was associated with the suppression of lipid synthesis in the mammary gland through the regulation of *SREBF1* and *PPAR γ* [100].

Here, key biological pathways and candidate genes associated with DMRs in divergent FE animals were identified. While these pathways are functionally important, other DMR-related genes may also influence FE traits. Validating DMR functionality through approaches like ATAC-seq or eQTL studies is essential. However, these findings provide valuable insights and potential targets for future functional studies to enhance the understanding of biological mechanisms underlying FE in dairy sheep.

Applicability of machine learning models for RFI and FCR prediction via methylation profiles and genetic variants within DMRs

Recent studies applying ML to predict/classify FE metrics in livestock using omics data have increased. Transcriptomic data classified pigs and beef cattle into high- and low-FE groups [101, 102, 103], while SNPs, rumen bacteria, and serum metabolites were also used for classification in pigs and cattle [104, 105, 106]. In sheep, dry matter intake was predicted via milk spectra in the Sarda dairy breed, with correlations of 0.33 (records), 0.32 (ewes), and 0.23 (days) [107]. To our knowledge, this study is the first to apply ML for FE prediction using epigenomic data in livestock. Notably, we predicted RFI and FCR rather than classifying animals into divergent groups.

Predictions were conducted with a limited sample size (21 animals). To assess the impact of animal selection, 100 rounds of random sampling were performed for each scenario and ML algorithm. Results showed significant variability in the training and test sets, likely due to differences in methylation levels and genetic variants within

DMRs. An important issue here is data leakage [108], which occurs when training data contain information unavailable during testing. Common causes include feature selection before cross-validation [109], biased data splitting [110], and leakage of target information through the input features [111]. In this study, DMRs were identified from a prior comparison of divergent FE groups, and ML predictions included both these animals and new ones. This may have contributed to variability in the results. Further analysis is needed to better understand this issue and improve the use of preselected biomarkers for FE prediction.

Despite these considerations, our findings suggest the potential of epigenetic marks and genetic variants for predicting RFI and FCR in dairy sheep. Due to the scarcity of studies predicting FE metrics rather than classifying animals, direct comparisons are challenging. However, a similar study in pigs using SNPs achieved the highest Spearman correlations of 0.34 for RFI and 0.36 for average daily gain [104]. Another study using ML for RFI prediction in pigs reported Spearman correlations of 0.28–0.27 with support vector machines and gradient boosting with 1,000 SNPs [28]. Our models, on average, yielded comparable results, with the best models surpassing previous predictive accuracy. The application of ML to predict FE metrics using epigenomic data could benefit the sheep industry by enabling more efficient and precise selection of animals with superior feed efficiency. By identifying epigenetic markers and genetic variants associated with RFI and FCR, breeders can make data-driven decisions to improve productivity while reducing feed costs. This approach could enhance genetic selection programs, optimize nutritional strategies, and contribute to sustainable sheep farming by minimizing resource waste. Additionally, early prediction of FE traits could help farmers identify high-efficiency animals without the need for costly and time-consuming individual feed intake measurements, ultimately improving profitability and environmental sustainability.

Conclusion

The results obtained here suggest relevant differences in epigenetic marks between divergently feed-efficient dairy sheep. These epigenetic marks are mapped within genes directly related to relevant metabolic pathways for the FE status. Among these pathways, protein digestion and absorption, hormone synthesis and secretion, control of energy availability, cellular signaling, and feed behavior pathways stand out as potential targets for the control of FE in dairy sheep. These results provide new insights into the biological mechanisms associated with FE and the control of these processes through epigenetic mechanisms. Additionally, the potential use of epigenetic information, individually or in combination with genetic

variants, for the prediction of FE metrics was evaluated. The results suggest the potential use of this epigenetic information as a biomarker for predicting RFI and FCR in dairy sheep. However, these results are preliminary must be validated in larger samples to evaluate the impact of sample composition and feature selection on the predictions.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-025-11520-1>.

Supplementary Material 1
Supplementary Material 2
Supplementary Material 3
Supplementary Material 4
Supplementary Material 5
Supplementary Material 6
Supplementary Material 7

Acknowledgements

Not applicable.

Author contributions

PASF: Formal data analysis, Writing – original draft. ASV: Conceptualization, Methodology, Writing – review & editing. BGG: Conceptualization, Investigation. CEB: Laboratory experiments and analysis; HM: Methodology, Investigation. RP: Laboratory experiments and analysis. JJA: Conceptualization, Validation, Writing – review & editing.

Funding

This work has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No 772787 (SMARTER). PASF is the beneficiary of a Maria Zambrano Grant of the University of Leon funded by the Ministry of Universities (Madrid, Spain) and financed by the European Union-Next Generation EU.

Data availability

All sequence data obtained in the present study were uploaded to the European Nucleotide Archive under the accession number PRJEB56589 (<https://www.ebi.ac.uk/ena/browser/view/PRJEB56589>).

Declarations

Ethics approval and consent to participate

All procedures involving animals in this study were performed following Spanish regulations regarding the protection of animals used for experimental and other scientific proposes (Royal Decree 53/2013) under the supervision of the Ethical and Animal Welfare Committee of the University of León after the approval of the competent body, Junta de Castilla y León. The nutritional challenge described in this work was approved by the Ethics Committee of the Instituto de Ganadería de Montaña (IGM, CSIC-ULE) in León (Spain) (Reference 100102/2018-1). In addition, the experimental design and the analysis performed in the current study were following ARRIVE guidelines.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Departamento de Producción Animal, Facultad de Veterinaria, Universidad de León, Campus de Vegazana, León 24007, Spain

Received: 18 November 2024 / Accepted: 24 March 2025

Published online: 31 March 2025

References

- United Nations. Revision of World Population Prospects. United Nations. 2019.
- Maja MM, Ayano SF. The impact of population growth on natural resources and farmers' capacity to adapt to climate change in Low-Income countries. *Earth Systems and Environment*; 2021.
- Nudda A, Atzori AS, Correddu F, Battaccone G, Lunesu MF, Cannas A et al. Effects of nutrition on main components of sheep milk. *Small Ruminant Res*. 2020;184.
- United Nations Food and Agriculture Organization. The state of food and agriculture 2016 (SOFA): climate change, agriculture and food security. *Livest Balance*. 2016.
- Connor EE. Invited review: improving feed efficiency in dairy production: challenges and possibilities. *Animal*. 2015;9.
- Ho CKM, Malcolm B, Doyle PT. Potential impacts of negative associative effects between concentrate supplements, pasture and conserved forage for milk production and dairy farm profit. *Anim Prod Sci*. 2013;53.
- Paganoni B, Rose G, Macleay C, Jones C, Brown DJ, Kearney G et al. More feed efficient sheep produce less methane and carbon dioxide when eating high-quality pellets. *J Anim Sci*. 2017;95.
- Rauw WM, Kanis E, Noordhuizen-Stassen EN, Grommers FJ. Undesirable side effects of selection for high production efficiency in farm animals: A review. *Livest Prod Sci*. 1998;56.
- Koch RM, Swiger LA, Chambers D, Gregory KE. Efficiency of feed use in beef cattle. *J Anim Sci*. 1963;22.
- Arthur PF, Archer JA, Johnston DJ, Herd RM, Richardson EC, Parnell PF. Genetic and phenotypic variance and covariance components for feed intake, feed efficiency, and other postweaning traits in Angus cattle. *J Anim Sci*. 2001;79.
- Herd RM, Bishop SC. Genetic variation in residual feed intake and its association with other production traits in British Hereford cattle. *Livest Prod Sci*. 2000;63.
- Baker SD, Szasz JL, Klein TA, Kuber PS, Hunt CW, Glaze JB et al. Residual feed intake of purebred Angus steers: effects on meat quality and palatability. *J Anim Sci*. 2006;84.
- Hoque MA, Suzuki K. Genetics of residual feed intake in cattle and pigs: A review. *Asian-Australas J Anim Sci*. 2009;22.
- Tortoreau F, Marie-Etancelin C, Weisbecker JL, Marcon D, Bouvier F, Moreno-Romieux C et al. Genetic parameters for feed efficiency in Romane Rams and responses to single-generation selection. *Animal*. 2020;14.
- Cammack KM, Leymaster KA, Jenkins TG, Nielsen MK. Estimates of genetic parameters for feed intake, feeding behavior, and daily gain in composite Ram lambs. *J Anim Sci*. 2005;83.
- Fogarty NM, Lee GJ, Ingham VM, Gaunt GM, Cummins LJ. Variation in feed intake of grazing crossbred Ewes and genetic correlations with production traits. *Aust J Agric Res*. 2006;57.
- Herd RM, Arthur PF. Physiological basis for residual feed intake. *J Anim Sci*. 2009;87.
- González-garcía E, Santos JP, Dos, Hassoun P. Residual feed intake in dairy Ewes: an evidence of intraflock variability. *Animals*. 2020;10.
- McLoughlin S, Spillane C, Claffey N, Smith PE, O'Rourke T, Diskin MG et al. Rumen Microbiome composition is altered in sheep divergent in feed efficiency. *Front Microbiol*. 2020;11.
- Zhang YK, Zhang XX, Li FD, Li C, Li GZ, Zhang DY et al. Characterization of the rumen microbiota and its relationship with residual feed intake in sheep. *Animal*. 2021;15.
- Zhang D, Zhang X, Li F, Li C, La Y, Mo F et al. Transcriptome analysis identifies candidate genes and pathways associated with feed efficiency in Hu sheep. *Front Genet*. 2019;10.
- Lillicrop KA, Hoile SP, Grenfell L, Burdige GC. DNA methylation, ageing and the influence of early life nutrition. *Proceedings of the Nutrition Society*. 2014.
- Kim Kchol, Friso S, Choi SW. DNA methylation, an epigenetic mechanism connecting folate to healthy embryonic development and aging. *J Nutr Biochem*. 2009.
- Ford D, Ions LJ, Alatawi F, Wakeling LA. The potential role of epigenetic responses to diet in ageing. *Proceedings of the Nutrition Society*. 2011.
- Touitou F, Tortoreau F, Bret L, Marty-Gasset N, Marcon D, Meynadier A. Evaluation of the links between lamb feed efficiency and rumen and plasma metabolomic data. *Metabolites*. 2022;12.
- Marina H, Arranz JJ, Suárez-Vega A, Pelayo R, Gutiérrez-Gil B, Toral PG et al. Assessment of milk metabolites as biomarkers for predicting feed efficiency in dairy sheep. *J Dairy Sci*. 2024;107.
- Suárez-Vega A, Gutiérrez-Gil B, Fonseca PAS, Hervás G, Pelayo R, Toral PG, et al. Milk transcriptome biomarker identification to enhance feed efficiency and reduce nutritional costs in dairy Ewes. *Animal*. 2024;18:101250.
- Piles M, Bergsma R, Gianola D, Gilbert H, Tusell L. Feature selection stability and accuracy of prediction models for genomic prediction of residual feed intake in pigs using machine learning. *Front Genet*. 2021;12.
- Yao C, Zhu X, Weigel KA. Semi-supervised learning for genomic prediction of novel traits with small reference populations: an application to residual feed intake in dairy cattle. *Genet Selection Evol*. 2016;48.
- Martin MJ, Dórea JRR, Borchers MR, Wallace RL, Bertics SJ, DeNise SK et al. Comparison of methods to predict feed intake and residual feed intake using behavioral and metabolite data in addition to classical performance variables. *J Dairy Sci*. 2021;104.
- INRA. Alimentation des ruminants: Apports nutritionnels-Besoins et réponses des animaux-Rationnement-Tables des valeurs des aliments. Éditions Quae 728p ISBN: 978-2-7592-2867-6. 2018.
- Fonseca PAS, Suárez-Vega A, Pelayo R, Marina H, Alonso-García M, Gutiérrez-Gil B et al. Intergenerational impact of dietary protein restriction in dairy ewes on epigenetic marks in the perirenal fat of their suckling lambs. *Scientific Reports* 2023 13:1 [Internet]. 2023 [cited 2023 Mar 20];13:1–15. Available from: <https://www.nature.com/articles/s41598-023-31546-3>
- Hughes K. Comparative mammary gland postnatal development and tumorigenesis in the sheep, cow, Cat and Rabbit: exploring the menagerie. *Semin Cell Dev Biol*. 2021.
- Barrio E, Hervás G, Gindri M, Friggens NC, Toral PG, Frutos P. Relationship between feed efficiency and resilience in dairy Ewes subjected to acute underfeeding. *J Dairy Sci*. 2023;1:14.
- Nozière P. INRA feeding system for ruminants. INRA feeding system for ruminants. 2018.
- Fonseca PAS, Suárez-Vega A, Esteban-Blanco C, Pelayo R, Marina H, Gutiérrez-Gil B et al. Epigenetic regulation of functional candidate genes for milk production traits in dairy sheep subjected to protein restriction in the prepubertal stage. *BMC Genomics*. 2023;24.
- Andrews S. FASTQC A Quality Control tool for High Throughput Sequence Data. Babraham Institute. 2015.
- Krueger F. Trim Galore! A wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files. Babraham Institute. 2015.
- Guo W, Fiziev P, Yan W, Cokus S, Sun X, Zhang MQ et al. BS-Seeker2: A versatile aligning pipeline for bisulfite sequencing data. *BMC Genomics*. 2013.
- Feng H, Conneely KN, Wu H. A bayesian hierarchical model to detect differentially methylated loci from single nucleotide resolution sequencing data. *Nucleic Acids Res*. 2014;42.
- Akalin A, Franke V, Vlahoviček K, Mason CE, Schübeler D. Genomation: A toolkit to summarize, annotate and visualize genomic intervals. *Bioinformatics*. 2015.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N et al. The sequence alignment/map format and samtools. *Bioinformatics*. 2009.
- Nunn A, Otto C, Fasold M, Stadler PF, Langenberger D. Manipulating base quality scores enables variant calling from bisulfite sequencing alignments using conventional bayesian approaches. *BMC Genomics*. 2022;23.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*. 2007.
- Marina H, Pelayo R, Gutiérrez-Gil B, Suárez-Vega A, Esteban-Blanco C, Reverter A et al. Low-density SNP panel for efficient imputation and genomic selection of milk production and technological traits in dairy sheep. *J Dairy Sci*. 2022;105.
- Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. Robust relationship inference in genome-wide association studies. *Bioinformatics*. 2010;26.
- Bash E. Machine Learning with R and H2O. Packt. 2018;1.

48. Cannas A, Tedeschi LO, Atzori AS, Lunesu MF. How can nutrition models increase the production efficiency of sheep and goat operations? *Anim Front*. 2019;9.
49. Mucha S, Tortoreau F, Doeschl-Wilson A, Rupp R, Conington J. Animal board invited review: Meta-analysis of genetic parameters for resilience and efficiency traits in goats and sheep. *Animal*. 2022;16:100456.
50. Lovendahl P, Difford GF, Li B, Chagunda MGG, Huhtanen P, Lidauer MH et al. Review: selecting for improved feed efficiency and reduced methane emissions in dairy cattle. *Animal*. 2018.
51. Cantalapiedra-Hijar G, Abo-Ismael M, Carstens GE, Guan LL, Hegarty R, Kenny DA et al. Review: biological determinants of between-animal variation in feed efficiency of growing beef cattle. *Animal*. 2018.
52. Xue Y, Lin L, Hu F, Zhu W, Mao S. Disruption of ruminal homeostasis by malnutrition involved in systemic ruminal microbiota-host interactions in a pregnant sheep model. *Microbiome*. 2020;8.
53. Pizzagalli MD, Bensimon A, Superti-Furga G. A guide to plasma membrane solute carrier proteins. *FEBS J*. 2021.
54. Bertrand C, Valet P, Castan-Laurell I. Apelin and energy metabolism. *Front Physiol*. 2015;6.
55. Ravussin E, Smith SR. Role of the Adipocyte in Metabolism and Endocrine Function. *Endocrinology: Adult and Pediatric*. 2015.
56. Assan D, Huang Y, Mustapha UF, Addah MN, Li G, Chen H. Fish feed intake, feeding behavior, and the physiological response of Apelin to fasting and refeeding. *Front Endocrinol (Lausanne)*. 2021.
57. O'Shea M, Hansen MJ, Tatemoto K, Morris MJ. Inhibitory effect of apelin-12 on nocturnal food intake in the rat. *Nutr Neurosci*. 2003;6.
58. Valle A, Hoggard N, Adams AC, Roca P, Speakman JR. Chronic central administration of apelin-13 over 10 days increases food intake, body weight, locomotor activity and body temperature in C57BL/6 mice. *J Neuroendocrinol*. 2008;20.
59. Tan W, Zheng H, Wang D, Tian F, Li H, Liu X. Expression characteristics and regulatory mechanism of Apela gene in liver of chicken (*Gallus gallus*). *PLoS ONE*. 2020;15.
60. Conn JW. Hypertension, the potassium ion and impaired carbohydrate tolerance. *N Engl J Med*. 1965;273.
61. Luther JM, Luo P, Kreger MT, Brissova M, Dai C, Whitfield TT et al. Aldosterone decreases glucose-stimulated insulin secretion in vivo in mice and in murine Islets. *Diabetologia*. 2011;54.
62. Luo P, Dematteo A, Wang Z, Zhu L, Wang A, Kim HS et al. Aldosterone deficiency prevents high-fat-feeding-induced hyperglycaemia and adipocyte dysfunction in mice. *Diabetologia*. 2013;56.
63. Mitchell CS, Begg DP. The regulation of food intake by insulin in the central nervous system. *J Neuroendocrinol*. 2021.
64. Lee MJ, Fried SK. Multilevel regulation of leptin storage, turnover, and secretion by feeding and insulin in rat adipose tissue. *J Lipid Res*. 2006;47.
65. Enns LC, Ladiges W. Protein kinase A signaling as an anti-aging target. *Ageing Res Rev*. 2010.
66. Kitaguchi T, Oya M, Wada Y, Tsuboi T, Miyawaki A. Extracellular calcium influx activates adenylate cyclase 1 and potentiates insulin secretion in MIN6 cells. *Biochem J*. 2013;450.
67. Hodson DJ, Mitchell RK, Marselli L, Pullen TJ, Brias SG, Semplici F et al. ADCY5 couples glucose to insulin secretion in human Islets. *Diabetes*. 2014;63.
68. Tong T, Shen Y, Lee HW, Yu R, Park T. Adenyl cyclase 3 haploinsufficiency confers susceptibility to diet-induced obesity and insulin resistance in mice. *Sci Rep*. 2016;6.
69. Jia L, Jiang Y, Li X, Chen Z. Purβ promotes hepatic glucose production by increasing Adcy6 transcription. *Mol Metab*. 2020;31.
70. Alhaidan Y, Christesen HT, Lundberg E, Balwi MAA, Brusgaard K. CRISPR/Cas9 ADCY7 knockout stimulates the insulin secretion pathway leading to excessive insulin secretion. *Front Endocrinol (Lausanne)*. 2021;12.
71. Guzmán A, Hernández-Coronado CG, Rosales-Torres AM, Hernández-Medrano JH. Leptin regulates neuropeptides associated with food intake and GnRH secretion. *Ann Endocrinol (Paris)*. 2019;80.
72. De Vriese C, Delporte C, Ghrelin. A new peptide regulating growth hormone release and food intake. *Int J Biochem Cell Biology*. 2008.
73. Niswender KD, Morrison CD, Clegg DJ, Olson R, Baskin DG, Myers MG et al. Insulin activation of phosphatidylinositol 3-kinase in the hypothalamic arcuate nucleus: A key mediator of insulin-induced anorexia. *Diabetes*. 2003;52.
74. Hill JW, Williams KW, Ye C, Luo J, Balthasar N, Coppari R et al. Acute effects of leptin require PI3K signaling in hypothalamic Proopiomelanocortin neurons in mice. *J Clin Invest*. 2008;118.
75. Yavuz S, Del Prado SSN, Celi FS. Thyroid hormone action and energy expenditure. *J Endocr Soc*. 2019;3.
76. Potts JT, Kronenberg HM, Rosenblatt M. Parathyroid hormone: chemistry, biosynthesis, and mode of action. *Adv Protein Chem*. 1982;35.
77. Harsløf T, Sikkær T, Sørensen L, Pedersen SB, Mosekilde L, Langdahl BL et al. The effect of treatment with PTH on undercarboxylated osteocalcin and energy metabolism in hypoparathyroidism. *J Clin Endocrinol Metab*. 2015;100.
78. Ortiga-Carvalho TM, Sidhaye AR, Wondisford FE. Thyroid hormone receptors and resistance to thyroid hormone disorders. *Nat Rev Endocrinol*. 2014.
79. Spetter MS, Hallschmid M. Current findings on the role of Oxytocin in the regulation of food intake. *Physiol Behav*. 2017.
80. Ott V, Finlayson G, Lehnert H, Heitmann B, Heinrichs M, Born J et al. Oxytocin reduces reward-driven food intake in humans. *Diabetes*. 2013;62.
81. Kerem L, Lawson EA. The effects of Oxytocin on appetite regulation, food intake and metabolism in humans. *Int J Mol Sci*. 2021.
82. Maud C, Ryan J, McIntosh JE, Olsson CA. The role of Oxytocin receptor gene (OXTR) DNA methylation (DNAm) in human social and emotional functioning: A systematic narrative review. *BMC Psychiatry*. 2018;18.
83. Li D, Ji Y, Zhao C, Yao Y, Yang A, Jin H et al. OXTR overexpression leads to abnormal mammary gland development in mice. *J Endocrinol*. 2018;239.
84. Hou Y, Hu M, Zhou H, Li C, Li X, Liu X et al. Neuronal signal transduction-involved genes in pig hypothalamus affect feed efficiency as revealed by transcriptome analysis. *Biomed Res Int*. 2018;2018.
85. de Oliveira PSN, Cesar ASM, do Nascimento ML, Chaves AS, Tizioto PC, Tullio RR, et al. Identification of genomic regions associated with feed efficiency in Nelore cattle. *BMC Genet*. 2014;15:100.
86. Xiao C, Deng J, Zeng L, Sun T, Yang Z, Yang X. Transcriptome analysis identifies candidate genes and signaling pathways associated with feed efficiency in Xiayan chicken. *Front Genet*. 2021;12.
87. Xu C, Wang X, Zhou S, Wu J, Geng Q, Ruan D et al. Brain transcriptome analysis reveals potential transcription factors and biological pathways associated with feed efficiency in commercial Dly pigs. *DNA Cell Biol*. 2021;40.
88. Schuurbijs OCJ, Kaanders JHAM, Van Der Heijden HFM, Dekhuijzen RPN, Oyen WJG, Bussink J. The PI3-K/AKT-pathway and radiation resistance mechanisms in non-small cell lung cancer. *J Thorac Oncol*. 2009.
89. Carracedo A, Pandolfi PP. The PTEN-PI3K pathway: of feedbacks and cross-talks. *Oncogene*. 2008.
90. Horodynska J, Hamill RM, Reyer H, Trakooljul N, Lawlor PG, McCormack UM et al. RNA-seq of liver from pigs divergent in feed efficiency highlights shifts in macronutrient metabolism, hepatic growth and immune response. *Front Genet*. 2019;10.
91. Lee DK, Jeong JH, Chun SK, Chua S, Jo YH. Interplay between glucose and leptin signalling determines the strength of GABAergic synapses at POMC neurons. *Nat Commun*. 2015;6.
92. Jalali M, Zendehdel M, Babapour V, Gilanpour H. Interaction between central oxytocinergic and glutamatergic systems on food intake in neonatal chicks: role of NMDA and AMPA receptors. *Int J Pept Res Ther*. 2019;25.
93. Lam DD, Garfield AS, Marston OJ, Shaw J, Heisler LK. Brain serotonin system in the coordination of food intake and body weight. *Pharmacol Biochem Behav*. 2010.
94. McHill AW, Wright KP. Role of sleep and circadian disruption on energy expenditure and in metabolic predisposition to human obesity and metabolic disease. *Obes Rev*. 2017.
95. Boumans IJMM, de Boer IJM, Hofstede GJ, la Fleur SE, Bokkers EAM. The importance of hormonal circadian rhythms in daily feeding patterns: an illustration with simulated pigs. *Horm Behav*. 2017;93.
96. Nishide SY, Hashimoto K, Nishio T, Honma KI, Honma S. Organ-specific development characterizes circadian clock gene Per2 expression in rats. *Am J Physiol Regul Integr Comp Physiol*. 2014;306.
97. Wafer R, Tandon P, Minchin JEN. The role of peroxisome proliferator-activated receptor gamma (PPARG) in adipogenesis: applying knowledge from the fish aquaculture industry to biomedical research. *Front Endocrinol (Lausanne)*. 2017.
98. Picard FAJ. PPAR(gamma) and glucose homeostasis. *Annu Rev Nutr*. 2002;22.
99. Janani C, Ranjitha Kumari BD. PPAR gamma gene - A review. *Diabetes and Metabolic Syndrome: Clinical Research and Reviews*. 2015.
100. Jing Y, Chen Y, Wang S, Ouyang J, Hu L, Yang Q et al. Circadian gene per2 silencing downregulates Pparg and srebf1 and suppresses lipid synthesis in bovine mammary epithelial cells. *Biology (Basel)*. 2021;10.
101. Piles M, Fernandez-Lozano C, Velasco-Galilea M, González-Rodríguez O, Sánchez JP, Torralardona D et al. Machine learning applied to transcriptomic

- data to identify genes associated with feed efficiency in pigs. *Genet Selection Evol.* 2019;51.
102. Chen W, Alexandre PA, Ribeiro G, Fukumasu H, Sun W, Reverter A et al. Identification of predictor genes for feed efficiency in beef cattle by applying machine learning methods to Multi-Tissue transcriptome data. *Front Genet.* 2021;12.
 103. Messad F, Louveau I, Koffi B, Gilbert H, Gondret F. Investigation of muscle transcriptomes using gradient boosting machine learning identifies molecular predictors of feed efficiency in growing pigs. *BMC Genomics.* 2019;20.
 104. Tusell L, Bergsma R, Gilbert H, Gianola D, Piles M. Machine learning prediction of crossbred pig feed efficiency and growth rate from single nucleotide polymorphisms. *Front Genet.* 2020;11.
 105. Foroutan A, Fitzsimmons C, Mandal R, Berjanskii MV, Wishart DS. Serum metabolite biomarkers for predicting residual feed intake (RFI) of young Angus bulls. *Metabolites.* 2020;10.
 106. Clemmons BA, Martino C, Powers JB, Campagna SR, Voy BH, Donohoe DR et al. Rumen bacteria and serum metabolites predictive of feed efficiency phenotypes in beef cattle. *Sci Rep.* 2019;9.
 107. Ledda A, Carta S, Correddu F, Cesarani A, Atzori AS, Battacone G et al. Dry matter intake prediction from milk spectra in Sarda dairy sheep. *Animals.* 2023;13.
 108. Kaufman S, Rosset S, Perlich C. Leakage in data mining: Formulation, detection, and avoidance. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* 2011.
 109. Varma S, Simon R. Bias in error Estimation when using cross-validation for model selection. *BMC Bioinformatics.* 2006;7.
 110. Wen J, Thibeau-Sutre E, Diaz-Melo M, Samper-González J, Routier A, Bottani S et al. Convolutional neural networks for classification of Alzheimer's disease: overview and reproducible evaluation. *Med Image Anal.* 2020;63.
 111. Oakden-Rayner L, Dunnmon J, Carneiro G, Re C. Hidden stratification causes clinically meaningful failures in machine learning for medical imaging. *ACM CHIL 2020 - Proceedings of the 2020 ACM Conference on Health, Inference, and Learning.* 2020.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.