

RESEARCH

Open Access



# Using nontargeted LC-MS metabolomics to identify the Association of Biomarkers in pig feces with feed efficiency

Jie Wu<sup>1,2†</sup>, Yong Ye<sup>1,2†</sup>, Jianping Quan<sup>1,2</sup>, Rongrong Ding<sup>1,2</sup>, Xingwang Wang<sup>1,2</sup>, Zhanwei Zhuang<sup>1,2</sup>, Shenping Zhou<sup>1,2</sup>, Qian Geng<sup>1,2</sup>, Cineng Xu<sup>1,2</sup>, Linjun Hong<sup>1,2</sup>, Zheng Xu<sup>1,2</sup>, Enqin Zheng<sup>1,2</sup>, Gengyuan Cai<sup>1,2</sup>, Zhenfang Wu<sup>1,2,3,4\*</sup> and Jie Yang<sup>1,2\*</sup>

## Abstract

**Background:** Improving feed efficiency is economically and environmentally beneficial in the pig industry. A deeper understanding of feed efficiency is essential on many levels for its highly complex nature. The aim of this project is to explore the relationship between fecal metabolites and feed efficiency-related traits, thereby identifying metabolites that may assist in the screening of the feed efficiency of pigs.

**Results:** We performed fecal metabolomics analysis on 50 individuals selected from 225 Duroc x (Landrace x Yorkshire) (DLY) commercial pigs, 25 with an extremely high feed efficiency and 25 with an extremely low feed efficiency. A total of 6749 and 5644 m/z features were detected in positive and negative ionization modes by liquid chromatography-mass spectrometry (LC/MS). Regrettably, the PCA could not classify the samples accurately. To improve the classification, OPLS-DA was introduced. However, the predictive ability of the OPLS-DA model did not perform well. Then, through weighted coexpression network analysis (WGCNA), we found that one module in each positive and negative mode was related to residual feed intake (RFI), and six and three metabolites were further identified. The nine metabolites were found to be involved in multiple metabolic pathways, including lipid metabolism (primary bile acid synthesis, linoleic acid metabolism), vitamin D, glucose metabolism, and others. Then, Lasso regression analysis was used to evaluate the importance of nine metabolites obtained by the annotation process.

**Conclusions:** Altogether, this study provides new insights for the subsequent evaluation of commercial pig feed efficiency through small molecule metabolites, but also provide a reference for the development of new feed additives.

**Keywords:** Feed efficiency, Pig, LC-MS, WGCNA

\* Correspondence: [wzfe@mail@163.com](mailto:wzfe@mail@163.com); [jieyang2012@hotmail.com](mailto:jieyang2012@hotmail.com)

†Jie Wu and Yong Ye contributed equally to this work.

<sup>1</sup>College of Animal Science and National Engineering Research Center for Breeding Swine Industry, South China Agricultural University, Guangzhou 510642, China

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



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, 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 changes were made. 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/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Background

Feed remains the main input cost in all animal production. Additionally, pork is one of the most important sources of meat for human beings. To cope with increasing market demand and breeding costs, improving the feed efficiency (FE) of pigs has always been a concern of breeders. Since FE cannot be measured directly for its complex pleiotropy, the feed conversion ratio (FCR) and residual feed intake (RFI) are often used as alternative selection conditions for FE [1–4]. However, although RFI traits have a strong genetic correlation with FCR (0.76 to 0.99) [5], a unified indicator for evaluating feed efficiency has not yet been determined.

At present, there are various omics methods to explore the molecular mechanisms that affect the feed efficiency of pigs, including transcriptomics [6, 7], genomics [5] and 16S rRNA gene sequencing [8]. However, metabolomics is rarely used to study pig feed efficiency phenotypes. Metabolites produced by the intestinal microbiota are increasingly recognized as an important part of human physiology [9]. As the downstream of the gene regulation network and protein interaction network, the metabolite can provide more detailed biological terminal information. By analyzing the changes in the expression of metabolites, it can help researchers find novel biomarkers and further understand the currently known metabolic pathways so that they can be applied to the study of various apparent traits. Furthermore, fecal metabolites are the final products of the metabolism of cells and intestinal microbiota, which can help to reflect the absorption and digestion of nutrients by the intestinal flora and digestive tract more comprehensively. Finally, the analysis of fecal metabolomics provides a noninvasive way to study the correlation between biological traits and metabolites. Therefore, the fecal metabolome can not only partially explain the composition of the gut microbiota but also be used as biomarkers to investigate the relationship between gut microbial metabolism and host phenotypes [10].

Duroc × (Landrace × Yorkshire) (DLY) commercial pigs has the advantages of high lean meat rate, low backfat thickness, high water holding capacity (WHC), suitable pH value (pH 1 > 5.9; 6.2 > pH 2 > 5.5) and IMF content (about 2.5%) for pork, etc. [11]. Desired carcass and meat quality traits makes DLY pigs currently account for the largest sales share in the Chinese pork market. Therefore, exploring the factors that affect the feed efficiency of DLY pigs has great significance for improving the economic benefits of the swine industry.

Liquid chromatography-mass spectrometry (LC/MS) is a relatively high-resolution separation analysis technique. Because of its high sensitivity, wide dynamic range, and lack of derivatization, LC-MS analysis has become a common technique for metabolomics research [12, 13].

However, few studies have used LC/MS technology to analyze the relationship between the fecal metabolome and feed efficiency. Hence, in this project, we analyzed the metabolomics of feces based on high and low feed efficiency groupings, aiming to unearth some small molecular metabolites that can evaluate feed efficiency and provide a new perspective for feed management in the pig industry.

## Results

### Difference analysis of FE-related traits

The summary of FE-related phenotypic data, such as average daily feed intake (ADFI), body weight (BW) and average daily gain (ADG), are shown in Table 1 and Supplementary Table S1. The calculated FCR and RFI data were used to compare the differences between the high- and low-FE groups. There were significant differences in the FCR phenotype between these two groups ( $p < 0.0001$ , ANOVA) (Fig. 1A). Similarly, the RFI of the low-FE group was significantly higher than that of the high-FE group ( $p = 0.0019$ , unpaired t-test) (Fig. 1B).

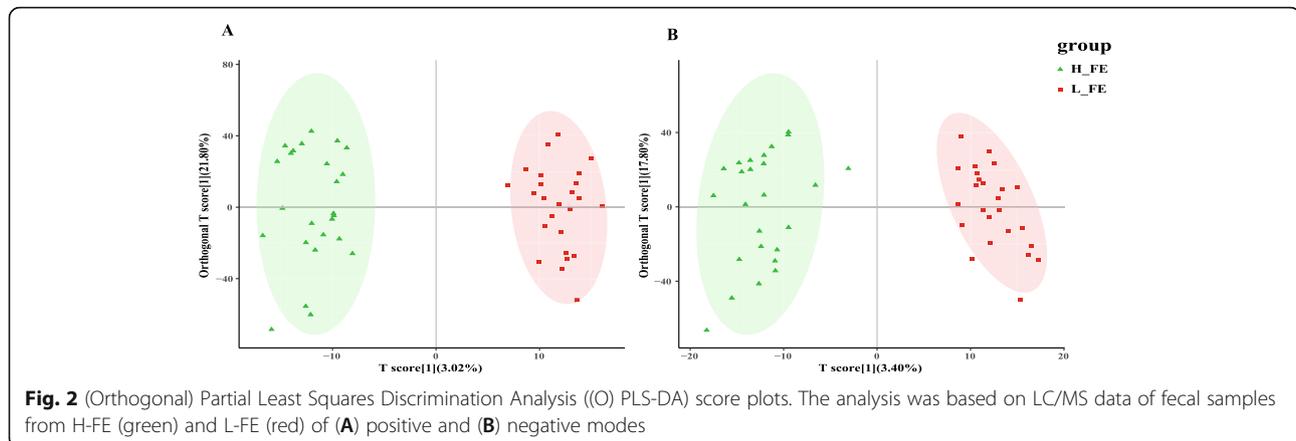
### Statistics and analysis of different metabolites

By liquid chromatography-mass spectrometry, a total of 5644 and 6749 m/z features were detected in negative and positive ionization modes, respectively. Our results showed that there was no clear separation between the high-FE and low-FE groups by PCA analysis (Fig. S1). Furthermore, we used the OPLS-DA model, which is a powerful statistical modeling tool, to identify differential metabolites between the two groups (Fig. 2). The ellipses in score plots for OPLS-DA were defined as the 95% critical limit of the Hotelling T<sup>2</sup>. For OPLS-DA analysis in both positive and negative ionization modes, the results showed that there was a complete separation between the high-FE group and the low-FE group ( $R^2X = 0.357$ ,  $R^2Y = 0.966$  and  $Q^2Y = 0.292$  in positive mode;  $R^2X = 0.330$ ,  $R^2Y = 0.948$  and  $Q^2Y = 0.178$  in negative mode) (Fig. 2A, B). In positive and negative ion modes, the first two components of the OPLS-DA explained 24.8 and 21.2% of the variance, respectively. From the results of OPLS-DA, it can be found that although there was a clear differentiation between the two groups in the positive and negative modes, a lower  $Q^2Y$  value indicated poor predictability and low quality of the model.

### Metabolism network analysis

Referring to the results of the above difference analysis, we tried to use another approach to explore the association between metabolome networks and phenotypes. WGCNA is a method of grouping genes with similar gene expression patterns into a module and finding the hub gene in the module. This analysis was then performed in the positive and negative ion modes. By





of the training set and the test set of 0.8817 and 0.8542, respectively (Fig. 4A). Subsequently, to evaluate the most valuable potential biomarkers, we ranked the importance of each metabolite according to the absolute value of the regression coefficient in the regression model (Fig. 4B).

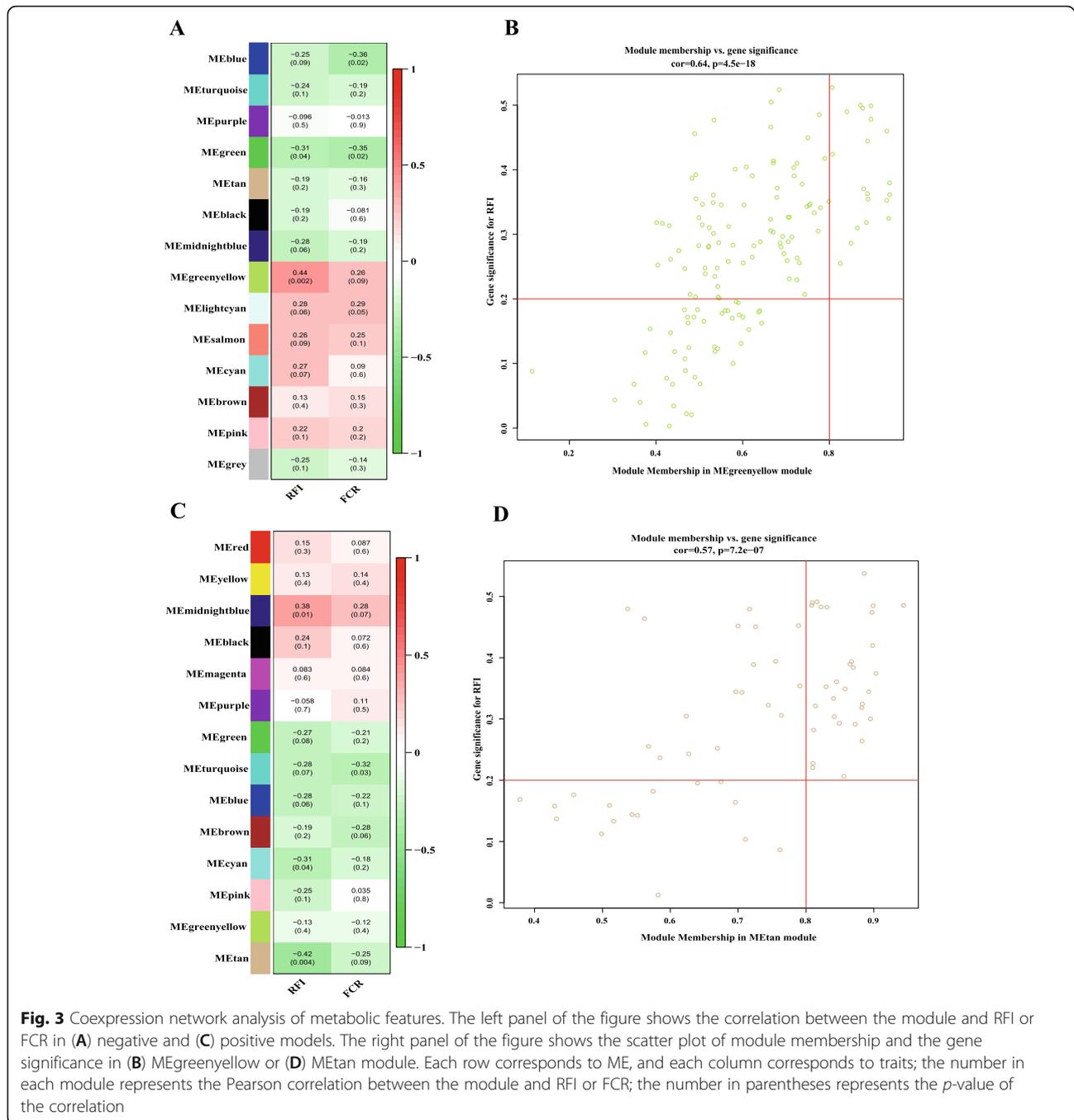
## Discussion

Improving the feed efficiency traits of livestock is of great significance, but it is not easy to estimate. Therefore, any measure that can effectively predict feed efficiency is meaningful for production. Although there is currently much work to study FE at the genetic level, few studies have linked metabolites to feed efficiency phenotypic traits. In this study, we analyzed and compared the metabolites in the feces of pigs in the high-FE and low-FE groups by LC-MC technology and interpretation tools, including WGCNA and Lasso regression. To the best of our knowledge, this is the first report combining these methods to study the metabolomic profile related to feed efficiency and related traits in DLY pigs.

At present, FCR and RFI are commonly used to evaluate FE traits, and it is believed that RFI can better represent feed efficiency [2, 3, 14], which is consistent with our WGCNA analysis results. The RFI and FCR are continuously varying quantitative traits, and the factors affecting quantitative traits are diverse and have different weights. There are two strategies to analyze and study quantitative traits: (i) one is to group quantitative traits according to thresholds, our PCA and OPLS-DA analysis was to directly determine the experimental animals into two groups of high or low feed efficiency and then analyze them. This analysis method can identify the influencing factors that affect the phenotype with greater weight as soon as possible; (ii) another strategy is to correlate the values of quantitative traits directly with the influencing factors. The WGCNA correlation analysis we performed can more comprehensively take into account the continuity effect of metabolite changes on the phenotype. The two methods can play a complementary

role, facilitating a more rapid and comprehensive search for factors influencing traits. In short, the two methods can play a complementary role, facilitating a quick and comprehensive search for factors affecting the trait.

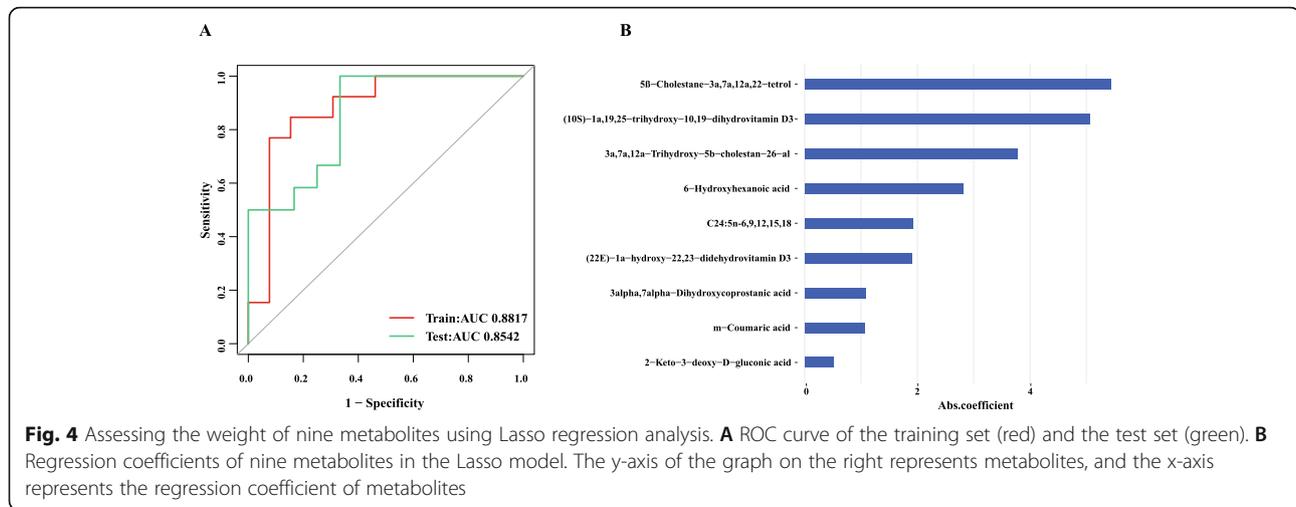
In our data, we found that the use of powerful tools such as PCA and OPLS-DA were not sufficient to distinguish the different features between the high- and low-FE animals. There are many possible explanations for the unsatisfactory results of PCA and OPLS-DA, including but not limited to (1) the sampling process was carried out after the individual growth indicators were measured. When the pig reaches the weight (approximately 100 kg), its metabolic activity is often not as active as before, and the increase in weight has little effect on the growth performance of pigs after 100 kg [15]. Notably, collected fecal samples should be immediately stored at  $-80^{\circ}\text{C}$  to  $-20^{\circ}\text{C}$  temperature until processed to avoid microbial fermentation. Sample storage is a critical and sensitive step, and freeze-thaw cycles need to be minimized to prevent possible metabolite degradation [16]. Additionally, to maximize avoidance of additional variability, although difficult to achieve, we recommend collecting fecal samples from multiple time points per individual and analyze an aliquot of the homogenized and mixed samples, or by metabolic characterization of multiple samples from each animal to minimize this variability [17]; (2) throughout the experiment, all test subjects were clinically healthy. In contrast, liver metabolism and skeletal muscle metabolism are greatly affected in infected or inflamed piglets and a significant decrease in growth performance will be observed in growing pigs [18]. Therefore, there is no physiological interference between the FE groups that could cause large metabolome differences; (3) the number of animal individuals in our study (25 individuals per group) may not have high statistical power, so more animal groups and more targeted experimental designs may be needed to evaluate feed efficiency in the future. Because the results of the PCA and OPLS-DA models were not ideal,



**Fig. 3** Coexpression network analysis of metabolic features. The left panel of the figure shows the correlation between the module and RFI or FCR in (A) negative and (C) positive models. The right panel of the figure shows the scatter plot of module membership and the gene significance in (B) MEgreenyellow or (D) METan module. Each row corresponds to ME, and each column corresponds to traits; the number in each module represents the Pearson correlation between the module and RFI or FCR; the number in parentheses represents the *p*-value of the correlation

we then adopted WGCNA analysis to select the modules and metabolites most closely related to RFI and FCR. After screening and annotation, we obtained nine metabolites in these models. Based on these metabolites, we identified four pathways from the KEGG database that were also significantly related to feed efficiency, including lipid metabolism (primary bile acid synthesis, linoleic acid metabolism), vitamin D, and glucose metabolism. Moreover, the Lasso regression model showed that all nine annotated metabolites contribute to feed efficiency.

The metabolite 22-OH-THC is a kind of bile alcohol, which is the end product of catabolism of cholestanic acids [19–21]. Bile alcohol may be regarded as an intermediate and side product from the normal pathways in bile acid biosynthesis [22]. Notably, THC26 and DHCA were mainly involved in the biosynthesis of primary bile acids. The specific synthesis process is that cholesterol 7 $\alpha$ -hydroxylase (CYP27A1) catalyzes the oxidation of steroid side chains to form THC26 or DHCA in the mitochondria of liver cells and then obtains the primary



bile acid cholic acid (CD) or chenodeoxycholic acid (CDCA) under the catalysis of various enzymes [23–27]. Interestingly, although the synthesis of bile acids is determined by a variety of cytochrome P450 enzymes (CYPs), both THC26 and DHCA are intermediate products catalyzed by CYP27A1 [28]. Bile acids start from the catabolism of cholesterol and are the final product of cholesterol catabolism; they play a critical role in food digestion and nutrient absorption, helping the absorption of lipids and fat-soluble vitamins in the intestine [27, 29–31]. After passing down the intestine with bile, approximately 95% of bile acids are reabsorbed in the terminal ileum and circulate back to the liver through the portal vein [23, 30, 32]. The performance of these functions of bile acid mainly depends on its enterohepatic circulation process, which is of great significance for nutrient absorption and distribution, metabolic regulation and homeostasis [23, 30, 32–34]. The results of metabolite network analysis showed that three metabolites related to bile acid synthesis were significantly negatively correlated with RFI traits, which means that they were positively correlated with feed efficiency. This result was consistent with our finding that these metabolites had higher levels in the high feed efficiency group. At present, little is known about the transport of bile acids and intermediates between different chambers, which may provide some references for understanding the important factors in the synthesis of bile acids.

In addition, (10S)-1 $\alpha$ ,19,25-trihydroxy-10,19-dihydrovitamin D3 and (22E)-1 $\alpha$ -hydroxy-22,23-didehydrovitamin D3 belong to vitamin D, which is a steroid derivative [35, 36]. Vitamin D has various effects on lipid metabolism and immune system function through its effects on nuclear hormone receptors (such as vitamin D receptor and PPAR $\gamma$ ) [37, 38]. Similarly, our results were consistent with the effect of vitamin D on lipid metabolism mediated through these receptors. Previous studies

have shown that the CYP27A1 enzyme can catalyze the hydroxylation of compounds both in the biosynthesis of bile acids and the bioactivation of vitamin D3 [39–41]. The acidic pathway (or alternative pathway) of bile acid synthesis is initiated by CYP27A1, which is a mitochondrial cytochrome P450 enzyme widely distributed in most tissues and macrophages [23, 26]. CYP27A1 cannot only catalyze the 25-hydroxylation of vitamin D3, which is required for the conversion of vitamin D3 into a functionally active form, but may also regulate cholesterol homeostasis by promoting the synthesis of bile acids or producing active oxysterols [26, 41–43]. Although there are no current reports on the effect of adding this enzyme, this warrants further research. Additionally, metabolite C24:5n-6 was involved in the alpha linolenic acid and linoleic acid metabolism pathways [44]. Linoleic acid is the main dietary n-6 polyunsaturated fatty acid (PUFA), and livestock mainly obtain it from diets such as vegetable oil, soybeans, and corn [45]. Previous studies reported that higher n-6 PUFA intake can reduce liver fat in overweight individuals, improve liver metabolism, and regulate the balance between fatty acid oxidation and lipid synthesis [46, 47]. In the process of linoleic acid metabolism, linoleic acid is catalyzed by the rate-limiting enzyme fatty acid desaturase-2 (FADS2), and after a series of extensions, C24:5n-6 is produced by FADS2 catalyzed C24:4n-6 [48]. Our results showed that C24:5n-6 correlated negatively and significantly with RFI traits and was significantly higher in the high-FE group than in the low-FE group ( $p = 0.002$ ). Notably, in the process of linoleic acid metabolism (elongation and desaturation), there was no significant difference between the upstream metabolite linoleic acid in the high and low FE groups, while the downstream C24:5n-6 was extremely different in the two groups. Besides, little research has been conducted on these three metabolites (m-coumaric acids, 6-hydroxyhexanoic acid and 2-keto-

3-deoxy-d-gluconic acid), and there is still insufficient evidence to show that they are closely related to feed efficiency. We hope that with the continuous updating and refining of metabonomics database, we will have enough information to intensively elaborate them in the future. Based on this, these findings can also provide some references for further analysis of linoleic acid metabolism. Improving feed efficiency is a concern, but these enzyme and potential metabolic markers deserve further evaluation and research to clarify their biological significance.

## Conclusions

In this study, the PCA and OPLS-DA models have low explanatory variances for pig feed efficiency phenotypes. Then, through WGCNA coexpression analysis, we found two important modules and 210 potential metabolic features related to feed efficiency in the positive and negative modes. The nine potential metabolic markers obtained from screening and annotation were assessed by the Lasso model, which offered possibilities of collecting metabolic hallmarks in feces under minimizing stress reactions in the future. In short, this study identified important pathways and candidate biomarkers closely related to RFI, that contribute to understanding the molecular mechanism of feed efficiency more comprehensively and provides an important reference for further verifying the application of metabolite signatures to identify pig feed efficiency traits.

## Methods

### Animals and sample collection

A total of 225 female DLY pigs in this study were provided by Guangdong Wen's Foodstuffs Group Co., Ltd. (Yunfu, China). The experimental animals were randomly divided into 30 pens, and each pen had 6 to 8 pigs. All pigs were arranged in a controlled environment with free access to water and food. The pig farm has strict epidemic prevention measures, the living conditions of the pigs such as humidity and temperature are controlled by sophisticated equipment, and veterinarians also regularly check the health status of the pigs to ensure that the pigs we randomly selected in the future are healthy throughout the entire cycle. Phenotypic data such as feed intake and weight per meal were recorded by the Osborne Feed Intake Recording Equipment (FIRE) Pig Performance Testing System (Osborne, KS, United States). The entire experimental process lasted approximately 12 weeks from an initial weight of approximately 30 kg to 100 kg. The calculation methods for FCR and RFI are based on previously described studies [49]. In this study, FE was defined as the weight gain from 30 kg to 100 kg of body weight divided by the total feed intake, which is the inverse of the commonly used

feed conversion ratio (FCR). It means low FCR and RFI ratios correspond to high FE. After ranking the FCR values of 225 pigs, the top 25 highest FCR and 25 lowest FCR were selected as the Low-FE and High-FE groups, respectively. Fecal samples of 50 female piglets were collected after rectal stimulation. From each pig 2–3 tubes (2-ml sterile tubes) of fecal samples were collected, and the tubes were immediately transferred to liquid nitrogen for temporary storage. Then, to minimize the effects of microbial fermentation, the samples were stored at  $-80^{\circ}\text{C}$  until analysis. In the entire monitored experiment, except for the sampling required by the research, no behaviors that caused animal stress reactions were carried out. The experimental protocol used in this study is in accordance with the Animal Protection and Use Committee of South China Agricultural University (SCAU, Guangzhou, China) (Approval number SCAU#0017).

### Fecal sample pretreatment

Fifty milligrams of fecal sample were accurately weighed, and 400  $\mu\text{l}$  of extraction solution (methanol: water = 4:1) was added to the samples. Then, the high-throughput tissue grinder was used to crush at low temperature (60 Hz,  $-20^{\circ}\text{C}$ ). After vortex mixing and ultrasound at 40 kHz for 30 min at  $5^{\circ}\text{C}$ , the extracted samples were placed at  $-20^{\circ}\text{C}$  for 30 min. The solution was then centrifuged at 13,000 g for 15 min ( $4^{\circ}\text{C}$ ), and the supernatant was extracted and injected into the LC-MS system for analysis.

### LC-MS analysis and quality control

The instrument platform for LC-MS analysis in this study was the AB SCIEX UPLC-TripleTOF system. The chromatographic conditions were as follows: the column was a BEH C18 column (100 mm  $\times$  2.1 mm i.d., 1.7  $\mu\text{m}$ ; Waters, Milford, USA); mobile phase A was water (containing 0.1% formic acid), and mobile phase B was acetonitrile/isopropanol (1/1) (containing 0.1% formic acid). The solvent gradient changed according to the following conditions: from 0 to 3 min, 95% (A): 5% (B) to 80% (A): 20% (B); from 3 to 9 min, 80% (A): 20% (B) to 5% (A): 95% (B); from 9 to 13 min, 5% (A): 95% (B) to 5% (A): 95% (B); from 13 to 13.1 min, 5% (A): 95% (B) to 95% (A): 5% (B), from 13.1 to 16 min, 95% (A): 5% (B) to 95% (A): 5% (B) for equilibrating the systems. The sample injection volume was 20  $\mu\text{L}$  and the flow rate was set to 0.4 mL/min. The column temperature was maintained at  $40^{\circ}\text{C}$ . During the period of analysis, all these samples were stored at  $4^{\circ}\text{C}$ . The flow rate was 0.40 mL/min, the injection volume was 20  $\mu\text{L}$ , and the column temperature was  $40^{\circ}\text{C}$ . In addition, the sample mass spectrometer signals were collected using positive and negative ion scanning modes. The mass spectrometry conditions were as follows: electrospray capillary voltage,

injection voltage and collision voltage: 1.0 kV, 40 V and 6 eV; ion source temperature and desolvation temperature: 120 °C and 500 °C; carrier gas flow rate: 900 L/h; mass spectrometry scan range: 50–1000 m/z; resolution: 30,000.

To evaluate the stability of the analysis system and find the variables with large variations in the analysis system during analysis, all test samples were mixed as quality control (QC) samples. In the process of instrument analysis, a QC sample was inserted every 8–10 samples.

#### Data analysis

Before conducting statistical analysis, the raw data were imported into the metabolomics software ProgenesisQI (Waters Corporation, Milford, USA) to generate the matrix of retention time, mass-to-charge ratio, and peak intensity for baseline filtering, peak identification, integration, retention time correction, and peak alignment. Furthermore, to obtain the final data matrix for subsequent analysis, the preprocessing process was as follows: (i) only variables with nonzero values above 80% in all samples were retained; (ii) missing values were filled using the k-nearest neighbors (KNN) approach in the R DMwR package; (iii) standardized values were obtained via the Z-score method, and the variables with relative standard deviation (RSD)  $\geq 30\%$  of the QC samples were deleted.

The *p* values for statistical differences in these phenotypes were based on ANOVA, Wilcoxon rank-sum test or unpaired Student *t*-tests, depending on the distribution of the data. If the data were normally distributed and homogenous, ANOVA was used to evaluate whether these traits were statistically significant, if the data were normally distributed but not homogeneous, unpaired Student *t*-tests were used, while a Wilcoxon rank-sum test was used otherwise. Principal component analysis (PCA) and (Orthogonal) partial least squares discrimination analysis (OPLS-DA) models were built using the ropls package in R [50]. PCA was used to observe the overall distribution between samples and the dispersion degree between groups. OPLS-DA was used to distinguish the different metabolites between groups. The goodness of fit ( $R^2$ ) and goodness of prediction ( $Q^2$ ) in cross-validation were used to evaluate the performance of the OPLS-DA model, and 500 permutation tests were performed.

#### Weighted gene correlation network analysis

Network and clustering analyses were performed using the R package Weighted Gene Coexpression Network Analysis (WGCNA) [51]. The Pearson correlation coefficient was calculated to obtain a coexpression similarity measure and used to subsequently construct an

adjacency matrix using soft threshold combined with topological overlap matrix (TOM). Then, hierarchical clustering was performed based on the TOM. Briefly, the soft thresholds of the positive and negative ion modes were set to 3 and 8, respectively, to achieve the approximate scale-free topology of the signed network ( $R^2 > 0.9$ ) (Fig. S3). In the dynamic tree cutting algorithm, deepSplit was set to 2 and minModuleSize was set to 50. The first principal component of the metabolite module was used as the feature vector of the module (including most of the variation information of all metabolites in the module), used to calculate the correlation coefficient between the metabolite module and feed efficiency, and then the most relevant module for subsequent analysis was selected. Subsequently, the gene significance (GS) and module membership (MM) of the most relevant module were calculated. Among these, GS can represent the correlation between metabolic characteristics and phenotype, and MM can represent the correlation between metabolic characteristics and module feature vectors.  $GS > 0.2$  and  $MM > 0.8$  were set as the threshold to screen the hub genes. Since WGCNA was first used for transcriptome data, we followed the term hub gene to represent the important metabolites identified. Subsequently, hub genes were identified by using the online Human Metabolome Database (HMDB) [52] and the METLIN public database [53]. The *p*-values of the hub genes were computed using the Wilcoxon test. The pathways in which hub genes participated were identified in the Kyoto Encyclopedia of Genes and Genomes (KEGG) database [54].

#### Lasso-penalized linear regression

We performed the Lasso regression in R using the glmnet [55] and caret packages. The sample data were randomly divided into a training set and a test set at a 1:1 ratio. Ten cross-validations were performed to calculate the lambda value ( $\lambda = 0.08678594$ ). Receiver operating characteristic (ROC) curves were generated using the pROC curve, predictions were made on the training set and the test set, and the importance of the variables was evaluated by the varimp function of the caret package.

#### Abbreviations

ADFI: Average daily feed intake; BW: Body weight; C24:5n-6: C24:5n-6,9,12,15,18; CA: Cholic acid; CDCA: Chenodeoxycholic acid; CYP27A1: Cholesterol 7 $\alpha$ -hydroxylase; DHCA: 3 $\alpha$ ,7 $\alpha$ -Dihydroxycoprostanic acid; FADS2: Fatty acid desaturase-2; FCR: Feed conversion ratio; FE: Feed efficiency; GS: Gene significance; H-FE: High feed efficiency; KDG: 2-Keto-3-deoxy-D-gluconic acid; L-FE: Low feed efficiency; MM: Module membership; OPLS-DA: Orthogonal partial least squares discriminant analysis; PCA: Principal component analysis; PUFA: Polyunsaturated fatty acid; RFI: Residual feed intake; THC26: 3 $\alpha$ ,7 $\alpha$ ,12 $\alpha$ -Trihydroxy-5 $\beta$ -cholestan-26- $\alpha$ l; WGCNA: Weighted gene co-expression network analysis; 22-OH-THC: 5 $\beta$ -Cholestane-3 $\alpha$ ,7 $\alpha$ ,12 $\alpha$ ,22-tetrol

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40813-021-00219-w>.

**Additional file 1: Table S1.** The detailed information of experimental animals.

**Additional file 2: Table S2.** The nine metabolites in the positive and negative modes obtained by WGCNA.

**Additional file 3: Figure S1.** Principal Component Analysis (PCA) scores plots. The analysis was based on LC/MS data of fecal samples from H-FE (green) and L-FE (red) of (A) positive and (B) negative model. **Figure S2.** Clustering dendrogram and module-trait correlation plots. Each coloured row represents a colour-coded module which contains a group of highly connected metabolic features. A total of 14 and 14 modules was identified in (A) positive and (B) negative model, respectively. **Figure S3.** Soft-thresholding values estimation. Scale independence and mean connectivity of various soft-thresholding values ( $\beta$ ) in (A) negative and (B) positive model.

### Acknowledgements

We thank Wen's Foodstuffs Group for constant support.

### Authors' contributions

Conceptualization, JY, ZW and YY; methodology, YY; software, JW, RD; formal analysis, JW, YY, JY, ZW; investigation, JQ, RD, XW, ZZ, SZ, QG, CX, LH, ZX, EZ, GC; resources, ZW; data curation, JY, ZW; writing—original draft preparation, JW, YY; writing—review and editing, JW, JQ, JY; visualization, JW, YY; supervision, ZW; project administration, JY, ZW; funding acquisition, GC, JY, ZW. All authors have read and agreed to the published version of the manuscript. The author(s) read and approved the final manuscript.

### Funding

This work was supported by the National Natural Science Foundation of China (31972540), the National High-quality lean-type Pig Breeding United Research Program of China, the Guangdong YangFan Innovative and Entrepreneurial Research Team Program (2016YT03H062), and the Key-Area Research and Development Program of Guangdong Province (2018B020203002).

### Availability of data and materials

All data generated or analysed during this study are included in this published article and its supplementary files.

### Declarations

#### Ethics approval and consent to participate

The experimental protocol used in this study was approved by the Animal Protection and Use Committee of South China Agricultural University (SCAU, Guangzhou, China) (Approval number SCAU#0017).

#### Consent for publication

Not applicable.

#### Competing interests

The authors have declared that no competing interest exists.

#### Author details

<sup>1</sup>College of Animal Science and National Engineering Research Center for Breeding Swine Industry, South China Agricultural University, Guangzhou 510642, China. <sup>2</sup>Guangdong Provincial Laboratory of Lingnan Modern Agricultural Science and Technology, Guangzhou 510642, China. <sup>3</sup>State Key Laboratory for Conservation and Utilization of Subtropical Agro-bioresources, Guangzhou 510642, China. <sup>4</sup>Guangdong Provincial Key Laboratory of Agro-animal Genomics and Molecular Breeding, Guangzhou 510642, China.

Received: 5 February 2021 Accepted: 11 May 2021

Published online: 02 June 2021

### References

- Carmelo VAO, Banerjee P, da Silva Diniz WJ, Kadarmideen HN. Metabolic networks and pathways associated with feed efficiency and related-traits in Duroc and Landrace pigs. *Sci Rep.* 2020;10(1):1–14.
- Do DN, Strathe AB, Ostersen T, Pant SD, Kadarmideen HN. Genome-wide association and pathway analysis of feed efficiency in pigs reveal candidate genes and pathways for residual feed intake. *Front Genet.* 2014;5:307.
- McCormack UM, Curião T, Metzler-Zebeli BU, Magowan E, Berry DP, Reyer H, et al. Porcine feed efficiency-associated intestinal microbiota and physiological traits: Finding consistent cross-locational biomarkers for residual feed intake. *Msystems.* 2019;4(4):e00324–18.
- Saintilan R, Mérour I, Schwob S, Sellier P, Bidanel J, Gilbert H. Genetic parameters and halothane genotype effect for residual feed intake in Piétrain growing pigs. *Livest Sci.* 2011;142(1–3):203–9. <https://doi.org/10.1016/j.livsci.2011.07.013>.
- Do DN, Strathe AB, Jensen J, Mark T, Kadarmideen HN. Genetic parameters for different measures of feed efficiency and related traits in boars of three pig breeds. *J Anim Sci.* 2013;91(9):4069–79. <https://doi.org/10.2527/jas.2012-6197>.
- Vigors S, O'Doherty JV, Bryan K, Sweeney T. A comparative analysis of the transcriptome profiles of liver and muscle tissue in pigs divergent for feed efficiency. *BMC Genomics.* 2019;20(1):461. <https://doi.org/10.1186/s12864-019-5740-z>.
- Ramayo-Caldas Y, Ballester M, Sánchez JP, González-Rodríguez O, Revilla M, Reyer H, et al. Integrative approach using liver and duodenum RNA-Seq data identifies candidate genes and pathways associated with feed efficiency in pigs. *Sie Rep.* 2018;8(1):558. <https://doi.org/10.1038/s41598-017-19072-5>.
- Quan J, Cai G, Ye J, Yang M, Ding R, Wang X, et al. A global comparison of the microbiome compositions of three gut locations in commercial pigs with extreme feed conversion ratios. *Sci Rep.* 2018;8(1):1–10.
- De Vadder F, Kovatcheva-Datchary P, Goncalves D, Vinera J, Zitoun C, Duchamp A, et al. Microbiota-generated metabolites promote metabolic benefits via gut-brain neural circuits. *Cell.* 2014;156(1–2):84–96. <https://doi.org/10.1016/j.cell.2013.12.016>.
- Zierer J, Jackson MA, Kastenmüller G, Mangino M, Long T, Telenti A, et al. The fecal metabolome as a functional readout of the gut microbiome. *Nat Genet.* 2018;50(6):790–5. <https://doi.org/10.1038/s41588-018-0135-7>.
- Jiang Y, Zhu L, Tang G, Li M, Jiang A, Cen W, et al. Carcass and meat quality traits of four commercial pig crossbreeds in China. *Genet Mol Res.* 2012; 11(4):4447–55. <https://doi.org/10.4238/2012.September.19.6>.
- Novais FJ, Pires PRL, Alexandre PA, Dromms RA, Iglesias AH, Ferraz JBS, et al. Identification of a metabolomic signature associated with feed efficiency in beef cattle. *BMC Genomics.* 2019;20(1):1–10.
- Gika HG, Theodoridis GA, Plumb RS, Wilson ID. Current practice of liquid chromatography–mass spectrometry in metabolomics and metabonomics. *J Pharm Biomed Anal.* 2014;87:12–25. <https://doi.org/10.1016/j.jpba.2013.06.032>.
- Ding R, Yang M, Wang X, Quan J, Zhuang Z, Zhou S, et al. Genetic architecture of feeding behavior and feed efficiency in a Duroc pig population. *Front Genet.* 2018;9:220. <https://doi.org/10.3389/fgene.2018.00220>.
- Oliveira E, Bertol T, Coldebela A, Santos Filho J, Scandolera A, Warpechowski M. Live performance, carcass quality, and economic assessment of over 100kg slaughtered pigs. *Arq Bras Med Vet Zootecnia.* 2015;67(6):1743–50. <https://doi.org/10.1590/1678-4162-7632>.
- Smirnov KS, Maier TV, Walker A, Heinzmann SS, Forcisi S, Martinez I, et al. Challenges of metabolomics in human gut microbiota research. *Int J Med Microbiol.* 2016;306(5):266–79. <https://doi.org/10.1016/j.ijmm.2016.03.006>.
- Karu N, Deng L, Slae M, Guo AC, Sajed T, Huynh H, et al. A review on human fecal metabolomics: methods, applications and the human fecal metabolome database. *Anal Chim Acta.* 2018;1030:1–24. <https://doi.org/10.1016/j.aca.2018.05.031>.
- Helm ET, Curry SM, De Mille CM, Schweer WP, Burrough ER, Zuber EA, et al. Impact of porcine reproductive and respiratory syndrome virus on muscle metabolism of growing pigs. *J Anim Sci.* 2019;97(8):3213–27. <https://doi.org/10.1093/jas/skz168>.

19. Iguchi Y, Yamaguchi M, Sato H, Kihira K, Nishimaki-Mogami T, Une M. Bile alcohols function as the ligands of membrane-type bile acid-activated G protein-coupled receptor. *J Lipid Res.* 2010;51(6):1432–41. <https://doi.org/10.1194/jlr.M004051>.
20. Iguchi Y, Kihira K, Nishimaki-Mogami T, Une M. Structure-activity relationship of bile alcohols as human farnesoid X receptor agonist. *Steroids.* 2010;75(1):95–100. <https://doi.org/10.1016/j.steroids.2009.11.002>.
21. Kuramoto T, Matsumoto N, Hoshita T. Syntheses of 22- and 23-hydroxylated bile alcohols. *Chem Pharm Bull.* 1978;26(9):2788–92. <https://doi.org/10.1248/cpb.26.2788>.
22. Kuroki S, Shimazu K, Kuwabara M, Une M, Hoshita T. Identification of bile alcohols in human bile. *J Lipid Res.* 1985;26(2):230–40. [https://doi.org/10.1016/S0022-2275\(20\)34393-5](https://doi.org/10.1016/S0022-2275(20)34393-5).
23. Chiang JYL. Bile acid metabolism and signaling. *Compr Physiol.* 2013;3(3):1191–212. <https://doi.org/10.1002/cphy.c120023>.
24. Li T, Apte U. Bile acid metabolism and signaling in cholestasis, inflammation, and cancer. *Adv Pharmacol.* 2015;74:263. <https://doi.org/10.1016/bs.apha.2015.04.003>.
25. Cali JJ, Russell DW. Characterization of human sterol 27-hydroxylase. A mitochondrial cytochrome P-450 that catalyzes multiple oxidation reaction in bile acid biosynthesis. *J Biol Chem.* 1991;266(12):7774.
26. Pikuleva IA, Babiker A, Waterman MR, Bjorkhem I. Activities of recombinant human cytochrome P450c27 (CYP27) which produce intermediates of alternative bile acid biosynthetic pathways. *J Biol Chem.* 1998;273(29):18153–60. <https://doi.org/10.1074/jbc.273.29.18153>.
27. Shapiro H, Kolodziejczyk AA, Halstuch D, Elinav E. Bile acids in glucose metabolism in health and disease. *J Exp Med.* 2018;215(2):383–96. <https://doi.org/10.1084/jem.20171965>.
28. Gómez C, Stücheli S, Kratschmar DV, Bouitbir J, Odermatt A. Development and validation of a highly sensitive LC-MS/MS method for the analysis of bile acids in serum, plasma, and liver tissue samples. *Metabolites.* 2020;10(7):282.
29. Halilbasic E, Claudel T, Trauner M. Bile acid transporters and regulatory nuclear receptors in the liver and beyond. *J Hepatol.* 2013;58(1):155–68. <https://doi.org/10.1016/j.jhep.2012.08.002>.
30. Schneider KM, Albers S, Trautwein C. Role of bile acids in the gut-liver axis. *J Hepatol.* 2018;68(5):1083–5. <https://doi.org/10.1016/j.jhep.2017.11.025>.
31. de Aguiar Vallim TQ, Tarling EJ, Edwards PA. Pleiotropic roles of bile acids in metabolism. *Cell Metab.* 2013;17(5):657–69. <https://doi.org/10.1016/j.cmet.2013.03.013>.
32. García M, Thirouard L, Sedès L, Monrose M, Holota H, Caira F, et al. Nuclear receptor metabolism of bile acids and xenobiotics: a coordinated detoxification system with impact on health and diseases. *Int J Mol Sci.* 2018;19(11):3630.
33. Roberts DMS, Magnusson BM, Burczynski FJ, Weiss M. Enterohepatic circulation. *Clin Pharmacokinet.* 2002;41(10):751–90. <https://doi.org/10.2165/00003088-200241100-00005>.
34. Hofmann AF. The enterohepatic circulation of bile acids in mammals: form and functions. *Front Bioence.* 2009;14(7):2584–98.
35. Sicinski RR, Prah J, Smith CM, DeLuca HF. New 1 $\alpha$ ,25-dihydroxy-19-norvitamin D3 compounds of high biological activity: synthesis and biological evaluation of 2-hydroxymethyl, 2-methyl, and 2-methylene analogues. *J Med Chem.* 1998;41(23):4662–74. <https://doi.org/10.1021/jm9802618>.
36. Mullard G. Metabolomic investigations into human apocrine sweat secretions. University of Nottingham; 2012.
37. Salamon H, Bruiners N, Lakehal K, Shi L, Ravi J, Yamaguchi KD, et al. Cutting edge: vitamin D regulates lipid metabolism in mycobacterium tuberculosis infection. *J Immunol.* 2014;193(1):30–4. <https://doi.org/10.4049/jimmunol.1400736>.
38. White JH. Vitamin D metabolism and signaling in the immune system. *Rev Endocr Metab Disord.* 2012;13(1):21–9. <https://doi.org/10.1007/s11154-011-9195-z>.
39. Su P, Rennert H, Shayiq RM, Yamamoto R, Zheng Y-M, Addya S, et al. A cDNA encoding a rat mitochondrial cytochrome P450 catalyzing both the 26-hydroxylation of cholesterol and 25-hydroxylation of vitamin D3: gonadotropic regulation of the cognate mRNA in ovaries. *DNA Cell Biol.* 1990;9(9):657–65. <https://doi.org/10.1089/dna.1990.9.657>.
40. Usui E, Noshiro M, Ohyama Y, Okuda K. Unique property of liver mitochondrial P450 to catalyze the two physiologically important reactions involved in both cholesterol catabolism and vitamin D activation. *FEBS Lett.* 1990;274(1–2):175–7. [https://doi.org/10.1016/0014-5793\(90\)81357-t](https://doi.org/10.1016/0014-5793(90)81357-t).
41. Jones G, Prosser DE, Kaufmann M. Cytochrome P450-mediated metabolism of vitamin D. *J Lipid Res.* 2014;55(1):13–31. <https://doi.org/10.1194/jlr.R031534>.
42. Shinkyo R, Sakaki T, Kamakura M, Ohta M, Inouye K. Metabolism of vitamin D by human microsomal CYP2R1. *Biochem Biophys Res Commun.* 2004;324(1):451–7. <https://doi.org/10.1016/j.bbrc.2004.09.073>.
43. Bodin K, Andersson U, Rystedt E, Ellis E, Norlin M, Pikuleva I, et al. Metabolism of 4 beta -hydroxycholesterol in humans. *J Biol Chem.* 2002;277(35):31534–40. <https://doi.org/10.1074/jbc.M201712200>.
44. Martin J, Stapleton RD. Parenteral and Enteral Nutrition with Omega-3 Fatty Acids. In: Rajendram R, Preedy VR, Patel VB, editors. *Diet and Nutrition in Critical Care.* New York, NY: Springer New York; 2021. p. 1–18.
45. Pérez MM, Gonçalves ECS, Salgado JCS, Rocha MdS, Almeida PZd, Vici AC, et al. Production of omegas-6 and 9 from the hydrolysis of açai and Buriti oils by lipase immobilized on a hydrophobic support. *Molecules.* 2018;23(11):3015, DOI: <https://doi.org/10.3390/molecules23113015>.
46. Bjermo H, Iggman D, Kullberg J, Dahlman I, Johansson L, Persson L, et al. Effects of n-6 PUFAs compared with SFAs on liver fat, lipoproteins, and inflammation in abdominal obesity: a randomized controlled trial. *Am J Clin Nutr.* 2012;95(5):1003–12. <https://doi.org/10.3945/ajcn.111.030114>.
47. Clarke SD. The multi-dimensional regulation of gene expression by fatty acids: polyunsaturated fats as nutrient sensors. *Curr Opin Lipidol.* 2004;15(1):13–8. <https://doi.org/10.1097/00041433-200402000-00004>.
48. Glaser C, Heinrich J, Koletzko B. Role of FADS1 and FADS2 polymorphisms in polyunsaturated fatty acid metabolism. *Metab Clin Exp.* 2010;59(7):993–9. <https://doi.org/10.1016/j.metabol.2009.10.022>.
49. Wu J, Wang X, Ding R, Quan J, Ye Y, Gu T, et al. Identification of important proteins and pathways affecting feed efficiency in DLY pigs by iTRAQ-based proteomic analysis. *Animals.* 2020;10(2):189. <https://doi.org/10.3390/ani10020189>.
50. Thévenot EA, Roux A, Xu Y, Ezan E, Junot C. Analysis of the human adult urinary metabolome variations with age, body mass index, and gender by implementing a comprehensive workflow for univariate and OPLS statistical analyses. *J Proteome Res.* 2015;14(8):3322–35. <https://doi.org/10.1021/acs.jproteome.5b00354>.
51. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics.* 2008;9(1):559. <https://doi.org/10.1186/1471-2105-9-559>.
52. Wishart DS, Feunang YD, Marcu A, Guo AC, Liang K, Vázquez-Fresno R, et al. HMDB 4.0: the human metabolome database for 2018. *Nucleic Acids Res.* 2018;46(D1):D608–17. <http://www.hmdb.ca/>. <https://doi.org/10.1093/nar/gkx1089>.
53. Sana TR, Roark JC, Li X, Waddell K, Fischer SM. Molecular formula and METL IN personal metabolite database matching applied to the identification of compounds generated by LC/TOF-MS. *J Biomol Tech.* 2008;19(4):258–66. <https://metlin.scripps.edu/>.
54. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* 2017;45(D1):D353–61. <http://www.genome.jp/kegg/>. <https://doi.org/10.1093/nar/gkw1092>.
55. Tibshirani R. Regression shrinkage and selection via the lasso. *J R Stat Soc Ser B Methodol.* 1996;58(1):267–88.

## Publisher's Note

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