Transcriptome Analysis of Muscle Reveals Potential Candidate Genes and Pathways Affecting Intramuscular Fat Content in Duroc Pigs

Xueyan Zhao

Shandong Academy of Agricultural Sciences <https://orcid.org/0000-0001-8683-2569>

Hongmei Hu

Shandong Academy of Agricultural Sciences

Cheng Wang

Shandong Academy of Agricultural Sciences

Yanping Wang

Shandong Academy of Agricultural Sciences

Haichao Lin

Shandong Academy of Agricultural Sciences

Jiying Wang (✉ jnwangjiying@163.com)

<https://orcid.org/0000-0002-6312-1390>

Research article

Keywords: Intramuscular fat, Longissimus dorsi muscle, Duroc, RNA sequencing, Differentially expressed genes, Pathway

Posted Date: May 27th, 2019

DOI: <https://doi.org/10.21203/rs.2.9878/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Frontiers in Genetics on August 11th, 2020. See the published version at <https://doi.org/10.3389/fgene.2020.00877>.

Abstract

Background Intramuscular fat (IMF) content plays a key role in meat quality of pork. Duroc is a common commercial pig breed, which is extensively used as the terminal male parent of DLY (Duroc × Landrace × Yorkshire) commercial pigs. For identification potential candidate genes and pathways, IMF content of 28 purebred Duroc pigs were measured, and transcriptome differences of muscles from individuals with divergent IMF content were analyzed based on RNA sequencing. **Results** As a result, a total of 309 DEGs were detected using edgeR and DESeq2 ($p < 0.05$, $|\log_2(\text{fold change})| \geq 1$) between the IMF-High and -Low groups. Functional analysis of the DEGs revealed 27 Gene Ontology (GO) terms, two pathways ($q < 0.05$) and 23 key DEGs, which closely related to lipid metabolism and fat cell differentiation. Among these key DEGs, LEP, SFRP1, CIDEA, PNPLA3 and MT3, are potential candidate genes affecting IMF content due to their biological functions, the significant correlation between these genes' expression with IMF content values in the 28 Duroc samples and proofs from previous studies. Meanwhile, a weighted gene co-expression network analysis (WGCNA) also revealed seven out of these key DEGs, including ADIPOQ, FABP4, PLIN1, CIDEA, ACAA1, CYP2A19 and ADIRF, were hub genes in turquoise module which may exhibit important functions in influencing IMF content. Furthermore, two pathways, regulation of lipolysis in adipocytes and PPAR signaling pathway, were potential pathways affecting IMF content because of their significant enrichment for the DEGs and co-expression genes of turquoise modules. **Conclusions** Our study demonstrates that some DEGs and pathways detected have essential role in regulating IMF content, and are supposed to be potential candidate genes and pathways affecting the trait in Duroc pigs. The study provides information for understanding molecular mechanism of IMF content, and would be of help for improving pork quality of Duroc pigs.

Background

Meat quality is a main economic trait in pig production, which can be evaluated by multiple indicators, such as intramuscular fat (IMF) content, pH, water holding capacity, color and tenderness. Of these, IMF content is arguably one of the most important one, and has closely correlations with some other meat quality traits, including flavor, juiciness and tenderness [1]. IMF corresponds to the amount of fat accumulates between muscle fibers, chemically, which covers the sum of phospholipids, triglycerides and cholesterol. High level of IMF content or marbling has a positive influence on the eating quality of pork [2]. However, for the past century, thinner backfat is considered as an important parameter for pig production. Although, this selection led to higher muscularity and growth, IMF content, meat juiciness and tenderness obviously decreased [3]. Meanwhile, IMF content is very difficult to be improved by traditional breeding methods. Hence, to satisfy the consumer preference for IMF content, it is worthwhile to understand the molecular basis of IMF content for improving it based on molecular breeding,

IMF content is a complex trait, which is regulated by many genes. Quantitative trait loci (QTL) and genome-wide association studies (GWAS) have been performed to determine the underlying genetic factors influencing it in previous studies. To date, 652 QTLs for IMF content have been deposited in pigQTLdb (https://www.animalgenome.org/cgi-bin/QTLdb/SS/traitmap?trait_ID=16). However, most of

QTLs have been distinguished mainly by low-density microsatellite markers using linkage mapping method, and span a large chromosomal region, which always comprise hundreds of genes. So, it is difficult for the further fine-mapping to identify causative genes [4]. The emergence of high-throughput genotyping platforms, SNP arrays, makes it possible find genetic variants and QTLs associated with the traits of interest in a narrower region. Despite these progresses, compared with other traits, the progress of studying the genetic basis of IMF content trait is still slow.

Next generation sequencing of RNA (RNA-seq), a method based on next-generation sequencing (NGS), has the potential to give unprecedented details about the transcriptional landscape of a certain organism. So, it is a new method for studying complex quantitative traits controlled by many interacting genes. Compared with real-time PCR and microarrays, RNA-seq, with low background noise and high dynamic ranges of gene expression levels quantification, integrates advanced molecular biology techniques and bioinformatics [5]. To date, RNA-seq has been carried out in different livestock species, such as cattle [6], sheep [7], and pig [8], to identify genes involved in IMF content and other meat quality traits. Several studies on transcriptome profiling of porcine longissimus dorsi (LD) between breeds with different IMF content [9], or between individuals with extremely high or low IMF content in Berkshire [10] have been reported. However, studies on transcriptome profiling in Duroc for identify differentially expressed genes (DEGs) and pathways related to IMF content are still not reported.

Duroc pig population is extensively used as the terminal male parent of the DLY (Duroc × Landrace × Yorkshire) commercial pigs due to its high meat quality and large muscle mass. As a coloured breed, Duroc has substantially higher levels of IMF content than white breeds [11]. Moreover, in the study, we found the IMF content of Duroc still varied significantly (1.17% to 4.23%). Accordingly, individuals with extreme IMF content in Duroc population are good materials for transcriptomics study of IMF content. Hence, in the present study, we performed RNA sequencing of LD muscles from Duroc pigs with variant IMF content, analyzed the transcriptome differences between the groups with extremely high and low IMF content, determined key DEGs and biological pathways affecting IMF content. This study provides valuable information for understanding the molecular basis underlying IMF content of pigs.

Results

Phenotypic variation between extreme groups

IMF content of LD muscles from a purebred population of 28 Duroc pigs were measured. The average of IMF content was 2.36% with a range of 1.17% to 4.23%. To select suitable samples used for the transcriptome analysis, we conducted Principal Component Analysis (PCA) using 300 genes with the most variance. The cumulative proportion of the first two principle components (PCs) was 38.19% (PC 1 = 29.90%, PC 2 = 8.29%, Fig.1A). According to the result of PCA, D8, D15, D19, and D21 individuals were assigned to the IMF-High group, and D13, D17, D18, and D26 individuals were assigned to the IMF-Low group. It was consistent with those selected based on IMF content values, suggesting close relationship existed between expressions of these genes and IMF content values. The average of IMF content was

3.70% and 1.47% for IMF-High and -Low group respectively. And the difference between the two groups was of extreme significance ($p < 0.01$, Table S1).

Summary of sequencing data

In this study, the transcriptome of LD muscle of the 28 Duroc pigs were sequenced using paired-end RNA-seq approach. Totally, 52.43 - 86.93 million raw reads per sample were generated. After filtering 2.31% of raw reads, an average of 62.56 million clean reads were used for following analyses (Table S1). of these, 94.36% of clean reads were successfully mapped to Sus scrofa11.1 genome assembly, with 90.34% being uniquely mapped (Table S1). Additionally, out of the reads mapped to the reference genome, 91.58% and 3.86% were located in exon and intron regions, respectively, and the remaining were in intergenic regions (Table S1).

According to alignment result of RNA sequencing data with Sus scrofa genome, a total of 26,918 genes were expressed in eight IMF content extremely individuals. The correlation coefficient of mean gene expression level between IMF-High and -Low groups was 0.96. The high correlation showed that most of the genes had a similar expression behavior in the two groups. Taking into account those genes with fragments per kilobase of transcript per million fragments mapped (FPKM) value > 1 in more than four pigs, 18,411 and 18,950 genes were found in IMF-High and -Low group, respectively.

Identification of differential expression genes

Using edgeR, 457 genes were identified as DEGs when we compared the genes expression in terms of $|\log_2(\text{fold change})| \geq 1$ and $p\text{-value} < 0.05$ between two groups. Of these, 334 and 123 genes were up- and down-regulated, respectively. Whereas, using DESeq2 under the same screening criteria, 316 DEGs were identified. Of these, 241 and 75 genes were up- and down-regulated, respectively. By comparing the DEGs detected by edgeR and DESeq2, a total of 309 overlapped DEGs (Table S2) were discovered by both methods, and were used for the following functional enrichment analyses. Of these, 240 genes had a higher expression and 69 genes had a lower expression in the IMF-High group. The heatmap and hierarchical clustering of the 309 DEGs indicated that the expression patterns were consistent within groups and different between groups (Fig.1B, 1C). Meanwhile, hierarchical clustering also exhibited the relationships of expression patterns between the eight IMF extreme individuals and other 20 individuals in the 28 Duroc pigs (Fig. 1C).

Furthermore, to confirm the DEGs, correlation analysis was also carried out between FPKM values of DEGs and IMF contents of 28 pigs measured in the study (Table S2). Of these 309 DEGs, 182 DEGs had medium correlation ($|r| \geq 0.3$), and 107 were significantly correlated ($p < 0.05$) between expression and IMF contents, suggesting these DEGs have similar expression pattern in the 28 samples as the eight IMF extreme individuals.

Using DESeq2 and edgeR, we also screened DEGs under stricter conditions ($|\log_2(\text{fold change})| \geq 1$ and q -value < 0.05). As a result, 10 and 12 DEGs were identified respectively, and seven genes were overlapped (Table 1). Further gene annotation revealed that two of the seven DEGs (*LEP* and *CIDEA*) directly related with lipid metabolism, three (*SPP1*, *ENSSSCG00000034371* and *SFRP1*) located in the region of QTLs influencing IMF content and marbling. Furthermore, correlation analysis indicated that the expressions of *LEP*, *CIDEA* and *SFRP1* were significantly correlated with IMF content ($r = 0.55$, $p = 2.23 \times 10^{-3}$; $r = 0.47$, $p = 1.20 \times 10^{-2}$; $r = 0.45$, $p = 1.69 \times 10^{-2}$) in the 28 Duroc samples. Of these seven DEGs, two have not been annotated with the specific gene name, and the remaining five ones were regarded as key genes affecting IMF content and used for following protein-protein interaction (PPI) network analysis.

Gene Ontology (GO) enrichment analysis of differentially expressed genes

To have a functional view of our DEGs detected between the two groups, we carried out GO term enrichment analysis for them and determined 58 significantly enriched GO terms ($q < 0.05$, Fig.2A and Table S3). Interestingly, all the significant enriched GO terms were enriched for up-regulated genes. Among the 58 enriched GO terms, most of them belonged to biological process (BP) category, only two terms belonged to molecular function (MF) and cellular component (CC) category, respectively. Importantly, 27 of 58 GO terms were closely associated with lipid metabolism and fat cell differentiation, such as lipid catabolic process, low-density lipoprotein receptor particle metabolic process, white fat cell differentiation, and positive regulation of cholesterol efflux (Fig.2B). Moreover, other significant GO terms related to some physiological and biological events, such as response to cytokine, response to tumor necrosis factor, alcohol metabolic, and detoxification, were also enriched.

Figure 2B illustrates the 18 up-regulated genes involved in the significantly enriched GO terms related to lipid metabolism and fat cell differentiation. Among these genes, *APOE*, *ADIPOQ*, *PPARG*, *CEBPA*, *LIPE*, *MT3*, and *PNPLA3*, were involved in more than five GO terms, including catabolism of triglyceride-rich lipoprotein constituents (*APOE*), control of fat metabolism and insulin sensitivity (*ADIPOQ*), regulation of adipocyte differentiation (*PPARG*), inducing adipocyte terminal differentiation (*CEBPA*), and triacylglycerol hydrolysis (*LIPE* and *PNPLA3*). Furthermore, the other enriched genes, such as *FABP4* [12, 13] and *FASN* [14] are candidate genes affecting IMF content by QTL mapping. By correlation analysis between expression and IMF content, the correlations of *CIDEA* ($r = 0.47$, $p = 1.20 \times 10^{-2}$), *SFRP1* ($r = 0.48$, $p = 1.69 \times 10^{-2}$), *ADIRF* ($r = 0.42$, $p = 2.43 \times 10^{-2}$), *PNPLA3* ($r = 0.42$, $p = 2.72 \times 10^{-2}$) and *MT3* ($r = 0.52$, $p = 4.82 \times 10^{-3}$) were of great significance at 0.05 level or 0.01 level.

Pathway enrichment analysis of differentially expressed genes

In addition, we conducted Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis for the 309 DEGs. Five pathways were significantly enriched ($q < 0.05$, Table 2) for the up-regulated DEGs, while no one was detected for the down-regulated genes. Two of the significant pathways, regulation of lipolysis in adipocytes ($q = 5.05E-03$) and PPAR signaling pathway ($q = 5.05E-03$), were IMF-related. And these two IMF-related pathways covered eight genes, *PLIN1*, *LIPE*, *PTGER3*, *ADRB1*, *FABP4*, *ACAA1*, *PPARG*, and *ADIPOQ*, which were relevant to lipid metabolism and fat cell differentiation. It is noticeable that, except for *PTGER3*, other genes were all found in enriched GO terms for the DEGs.

Together with the five key genes detected under stricter conditions ($|\log_2(\text{fold change})| \geq 1$ and $q\text{-value} < 0.05$), we determined in total number of 23 key genes playing the vital role in regulating IMF content due to one overlapped (Table S4). Among these genes, some (*APOE*, *ADIPOQ*, *PPARG*, *CEBPA*, *LIPE*, *MT3*, and *PNPLA3*) were involved in more than five GO terms, some (*PLIN1*, *LIPE*, *ADRB1*, *FABP4*, *ACAA1*, *PPARG*, and *ADIPOQ*) were simultaneously detected by GO and KEGG analyses. The results implied that there could be possible interactions among these genes. To further analyze the interaction relationships of the 23 key genes, we constructed PPI network analysis using Search Tool for the Retrieval of Interacting Genes (STRING). Except *FABP4*, *CYP2A19* and *ADIPOQ*, the other genes were successful matched to suscrofa database of STRING. By the visualization of Cytoscape software, a PPI network consisted of 12 core proteins were constructed (Fig. 3). The top hub genes interacted with more than six of other genes were *PPARG*, *LEP*, *LIPE* and *PLIN1*.

Co-expressed genes associated with IMF content

The Weighted Gene Co-expression Network Analysis (WGCNA) relies on the assumption that strongly correlated expression levels of gene sets indicate those genes work cooperatively in related pathways and contribute to the resulting phenotype. By WGCNA, we constructed co-expression network with 5175 selected genes and identified 29 modules (Fig. 4A). Six modules, including skyblue module, orange module, turquoise module, pink module, darkgreen module, and darkred, were correlated with IMF content (correlation > 0.4).

For each module, we performed functional enrichment analysis, and found that genes in turquoise module remarkably participated in IMF-related biological processes or pathways. For the co-expressed genes in turquoise module, 82 significant enriched GO terms were identified ($q < 0.05$, Table S5) in GO analysis. Among these, positive regulation of lipid localization and lipid particle were two terms related to lipid metabolism (Fig. 4B). Meanwhile, two pathways were also identified in KEGG analysis ($q < 0.05$), regulation of lipolysis in adipocytes and PPAR signaling pathway, which were canonical pathways regulating fat deposition (Fig. 4B, Table S6). Notably, the two terms and pathways related to lipid metabolism or regulating fat deposition were all found to be significantly enriched ones in the functional analysis of the DEGs.

Above mentioned results revealed that genes in the turquoise module may exhibit important functions in influencing IMF content. To identify hub genes, we calculated the weight referring to connection strength

between two genes in the turquoise module. The weight values ranged from 0.02 to 0.39. Correlation network was constructed with the weight value > 0.36 by Cytoscape (Fig. 4C). The network contained 39 genes, seven of which, *ADIPOQ*, *FABP4*, *PLIN1*, *CIDEA*, *ACAA1*, *CYP2A19*, and *ADIRF*, were also among the key genes identified by GO and pathway enrichment analysis of DEGs. Furthermore, except *CYP2A19*, the expressions of the other six hub genes were all significantly correlated with IMF content (*ADIRF*, $r = 0.42$, $p = 2.43E-02$), or their correlation coefficients were above 0.3, which also proved that these hub genes were potential candidate genes regulating IMF content in Duroc.

Discussion

Duroc, one of the most excellent breeds for its high meat production and meat quality, is usually used as the terminal male parent of the commercial cross-bred pigs such as DLY pigs. Though Duroc is known for its high IMF content [11] compared to the other lean breeds, the IMF content of Duroc population still vary in a wide range. In this study, IMF content of 28 Duroc pigs, which were reared under the same environmental conditions and with the same nutrition levels, varied from 1.17% to 4.23%. Combined with phenotypic data and the result of PCA, four high IMF content individuals and four low IMF content individuals with significant difference ($p < 0.0016$) were selected from the 28 Duroc pigs. High throughput RNA-seq technology was employed to identify DEGs between the two groups, and several functional enrichment analyses were followed to identify plausible candidate genes and pathways affecting IMF content. The outcome of the study might be helpful for explaining genetic variations related to IMF content in Duroc.

DESeq2 and edgeR are both very popular and easy-to-use packages for differential expression analysis of RNA-Seq data [15]. In this study, these two methods were used for detecting DEGs, and overlapped genes detected by them were considered as DEGs, which would be helpful to reduce false-positive DEGs detected. As a consequence, 309 DEGs were detected (Table S2) under the criteria of $|\log_2(\text{fold change})| \geq 1$ and $p\text{-value} < 0.05$. To further identify critical genes from these DEGs, we further screened the GO terms and pathways referring to metabolisms of major chemical composition of IMF (Fig. 2B). The results of GO enrichment analysis showed that the up-regulated DEGs significantly enriched in GO terms associated with lipid metabolic, fatty acids metabolic, cholesterol metabolic, and sterol metabolic. Meanwhile, GO terms related to fat cell differentiation were also enriched for up-regulated DEGs. Interestingly, there were no down-regulated DEGs enriched in any IMF-related GO terms. It may suggest that the increase of IMF content in Duroc may be resulted from the up-regulation of some critical genes relevant to IMF content.

Among the key DEGs detected in the study, *LEP*, *SFRP1*, *CIDEA*, *PNPLA3* and *MT3*, were not only repeatedly discovered as the key genes in the functional enrichment analysis (Table 1, Fig. 2B), but their expression were significantly correlated with IMF content values in the 28 Duroc samples (Table S2). *LEP* gene encodes a protein, leptin, which is secreted by white adipocytes into the circulation and regulates the body's fat storage system through the inhibition of lipogenesis and the stimulation of lipolysis [16]. However, a study on *LEP* gene expression in IMF tissue between growing Erhualian and large white pigs, the expression of *LEP* may contribute to fast sediment of IMF [17]. *SFRP1* is a Wnt antagonist in Wnt/ β -

catenin signaling pathway which plays an important role in regulating adipogenesis and controlling adipocyte differentiation [18]. *CIDEA* was discovered highly expressed in adipose tissues, and its expression level was significantly higher in obese pigs than in their lean counterparts [19]. The protein encoded by *CIDEA*, an adipocyte lipid drop protein, negatively regulates lipolysis and promotes triglyceride accumulation [20]. *PNPLA3* gene, belongs to the patatin-like phospholipase (PNPLA) family, exhibits lipase and transacylase activity [21]. In human, an association was revealed between *PNPLA3* expression and obesity as well as related phenotypes [22]. In pigs, *PNPLA3* is mainly expressed in adipose and liver tissue, participating in the restoration of lipid homeostasis upon aberrant intracellular lipid accumulation. Consistent with our results, in Sodhi's study, *PNPLA3* was also distinguished as one of important candidate genes affecting IMF content by comparative transcriptome analysis in different pig breeds [23]. *MT3* protein, metallothionein 3, is a member of the mammalian metallothionein (MT) family. Wang et al. found that *MT3* was expressed in human adipocytes and played a critical role in protecting adipocytes from hypoxic damage [24]. By the model of *MT3* knockout mice, Byun et al. found *MT3* may be involved in central leptin signaling and the consequent increase in peripheral energy expenditure [25]. However, the biological function of porcine MT3 needs to be further clarified. In summary, the above evidences suggested that they are potential candidate genes affecting IMF content. Except the five genes mentioned above, the other key genes, such as *FABP4*, *CIDEA*, *PPARG*, and *ADIPOQ*, whose correlation coefficients were all above 0.35 (Table S4), should also be paid attentions too.

To identify important pathways affecting IMF content, we conducted KEGG analysis using DEGs ($p < 0.05$ and $|\log_2(\text{fold change})| \geq 1$, Table 2) and co-expression genes in turquoise modules (Table S6). Regulation of lipolysis in adipocytes and PPAR signaling pathway were both identified as enriched pathways for the two gene sets. According to the description of KEGG database (https://www.genome.jp/dbget-bin/www_bget?map04923), regulation of lipolysis in adipocytes is a pathway regulating triacylglycerol to release fatty acids and glycerol for other organs as energy substrates. Some studies have demonstrated that the IMF content in the LD is determined by a balance between fat accumulation and degradation [26]. It is implied that more lipid storage may lead to the increase of lipolysis [27]. Consequently, in the study, regulation of lipolysis in adipocytes pathway enriched in the High-IMF group compared with Low-IMF group. PPAR signaling pathway is a canonical pathway involving in lipid metabolism. The expression pattern of many candidate genes in this pathway were correlated with meat quality traits in LD muscle of pigs [28]. Previous studies showed that the PPAR signaling pathway was significant enriched pathway for DEGs of subcutaneous and intramuscular stromal vascular cells during adipogenic differentiation [29], as well as DEGs of transcriptome analysis between Laiwu and Yorkshire pigs' LD muscles [30]. Therefore, consistent with the previous studies, our results suggested that PPAR signaling pathway was a pivotal pathway affecting IMF content.

Conclusions

In this study, we performed a high throughput RNA sequencing to evaluate the transcriptome profiles differences of eight Duroc pigs with extreme IMF content. A total of 309 DEGs were screened by DESeq2

and edgeR. GO and pathway enrichment analysis, correlation analysis between gene expressions and IMF content in 28 Duroc pigs and WGCNA revealed some potential candidate genes, such as *LEP*, *SFRP1*, *CIDEA*, *PNPLA3* and *MT3*, and pathways, such as regulation of lipolysis in adipocytes and PPAR signaling pathway. These potential candidate genes and pathways play an essential role in affecting IMF content of pigs, and deserved to carry further studies to unravel their specific mechanism on IMF content. This study provides useful information for understanding molecular mechanism of intramuscular fat content in Duroc pigs.

Methods

Animals and sample collection

The pig specimens were from a Duroc breeding farm in Shandong province of China. A total of 28 purebred Duroc pigs at 108.29 ± 6.00 kg (mean $\pm$ SD) of body weight were used in this study to select pigs representing high- (n=4) and low-IMF (n=4) content groups. These pigs were fed *ad libitum* with commercial feed and water under the same environmental, and were slaughtered in one batch after euthanized by electric shock. All the experimental procedures were approved by Institutional Animal Care and Use Committee of Institute of Animal Husbandry and Veterinary Medicine, Shandong Academy of Agricultural Sciences (approval code, IACC20060101, 1 January 2006). About 0.2 g sample of LD muscle closed to the last 4th of the thoracic vertebrae was collected for each pig, put into a tube with RNAlater Stabilization Solution (Thermo Fisher, Waltham, MA, USA), and frozen at -80°C for RNA extraction.

For the determination of IMF content, the LD muscle samples taken from the last 4th and 5th thoracic vertebra, which were removed adipose and connective tissues, were grounded after oven-drying. Then, IMF content were examined using the Soxhlet petroleum-ether method, and expressed as the weight percentage of wet muscle tissue. Each sample was measured in triplicate to ensure its accuracy.

RNA isolation, library construction and sequencing

Total RNA was isolated from 28 porcine LD samples using TRNzol reagent (Invitrogen, Life Technologies). RNA was examined with a NanoDrop 1000 instrument (Thermo Fisher Scientific) and Qubit® RNA Assay Kit in Qubit® 2.0 Fluorometer (Life Technologies, CA, USA), and RNA integrity was confirmed using a 2100 Bioanalyzer (Agilent Technologies, CA, USA). RNA libraries were constructed using NEBNext® Ultra™ RNA Library Prep Kit for Illumina® (NEB, USA) following manufacturer's recommendations. Briefly, the mRNA was isolated from total RNA using poly-dT oligonucleotides attached to magnetic beads. The mRNA was fragmented, and then reverse-transcribed into first-strand cDNA with random hexamers. After generating double-stranded cDNA, end repair, A-tailing, adaptor ligation and cDNA purification and enrichment were then performed. Paired-end raw sequences (150 bp) were generated using an Illumina HiSeq4000 Sequencing System.

Quality assessment of sequencing data and mapping of RNA-seq reads

In order to ensure the quality of bioinformatics analysis, raw reads were firstly filtered to obtain clean reads. We filtered raw reads in FASTQ format by removing reads according to the following criteria: containing adapter sequences, ambiguous base-content > 0.1%, 50% of bases whose Q_{phred} quality scores ≤ 20 . Clean reads were then mapped to *Sus scrofa* reference genome (Sscrofa11.1) and the annotation database Ensembl Genes 95 using HISAT2 software [31]. FeatureCounts tool of subread package was used for calculating the number of reads mapped to each gene for estimating gene expression levels [32]. Gene expression levels were also measured in the expected number of fragments per kilobase of exon per million fragments mapped (FPKM).

Differential gene expression analysis

For analyzing the expression correlation of 28 Duroc pigs and selecting IMF extreme individuals, PCA was performed with R package prcomp using 300 genes with the most variance in 28 Duroc pigs. Combined with phenotype data, eight pigs with extremely high and low IMF content were ultimately determined for the following transcriptome analysis.

Differential expression analysis was carried out by the R package DESeq2 [33] and edgeR [34]. Genes with $|\log_2(\text{fold change})| \geq 1$ and $p\text{-value} < 0.05$ and an adjusted $p\text{-value} < 0.05$ were regarded as DEGs for each method. To reduce false-positive result, the overlapped DEGs detected by the two methods were regarded as the final DEGs and used for the following functional enrichment analyses. Meanwhile, the resulting p -values were adjusted using the Benjamini and Hochberg's approach for controlling the false discovery rate (FDR). And to detect key genes affecting IMF, we also conducted the differential expression analysis using stricter criteria, $|\log_2(\text{fold change})| \geq 1$ and $q\text{-value} < 0.05$. Of particular note is that, in the study, we did not use genders as a sample selection condition. Among the eight individuals, there were only two male pigs which belonged to IMF-High group. Therefore, the DEGs located in Y-chromosome were not taken into account.

To further confirm the DEGs, Person correlation analysis between the expression levels of DEGs and IMF content values of the data set of 28 Duroc pigs were conducted using the R package Hmisc. In addition, to visualize the DEGs between the two IMF groups and among the 28 Duroc pigs, heatmap was generated using the R package pheatmap based on the FPKM values and hierarchical clustering was analyzed by R function hclust using correlation coefficients of DEGs obtained in Person correlation analysis as the distance between a pair of individuals.

Functional enrichment and PPI network analysis of DEGs

The DEGs detected by DESeq2 and edgeR were subjected to GO and KEGG pathway enrichment analysis using R package clusterProfiler [35]. GO terms and pathways with q -value < 0.05 were considered to be significantly enriched. Genes, involved in the significantly enriched GO terms and pathways related to lipid metabolism and fat cell differentiation, were treated as key genes affecting IMF.

To evaluate the interactive relationships of the genes and identify hub genes, the key genes, identified by differential gene expression analysis under stricter criteria (q -value < 0.05) and GO and KEGG pathway enrichment analysis of DEGs, were uploaded into STRING (https://string-db.org/cgi/input.pl?sessionId=WcgfuVcCfC0u&input_page_show_search=on). Active interaction sources of STRING included textmining, experiments, database, co-expression, neighborhood, gene fusion, co-occurrence with a medium confidence score (0.40). The PPI network was visualized by Cytoscape software [36].

Weighted Gene Co-expression Network Analysis (WGCNA)

To construct co-expression network, we further applied WGCNA [37] using FPKM values resulting from RNA-Seq of eight Duroc pigs with extreme IMF content. Genes with FPKM values < 1 in more than four individuals and SD > 1.2 were deleted. Consequently, we built a signed co-expression network using R package WGCNA with 5175 genes by creating a matrix of pairwise Pearson correlation coefficients. The power of 9 based on the scale-free topology criterion was used as the soft-threshold. Next, the Topological Overlap Measure (TOM), representing the overlap in shared neighbors, was calculated using the adjacency matrix. By hierarchical clustering and dynamic tree cutting, genes, clustered into distinct modules, were assigned to a color. The minimum module size was set to 30. The Pearson's correlations between the module eigengene and the IMF content values were calculated for estimating the Module-Trait relationships (MTRs). Modules with a correlation coefficients > 0.4 with IMF content values were selected. In each module, intra-modular connectivity and the gene-trait correlations were also assessed. Genes in the modules were conducted functional annotation by R package clusterProfiler. GO terms and pathways with q -value < 0.05 were considered as significant ones.

List Of Abbreviations

IMF: intramuscular fat; DLY: Duroc $\times$ Landrace $\times$ Yorkshire; KEGG: Kyoto Encyclopedia of Genes and Genomes; DEGs: differentially expressed genes; LD: longissimus dorsi; QTL: quantitative trait loci; GWAS: genome-wide association studies; NGS: Next-generation sequencing; RNA-seq: RNA sequencing; FPKM: fragments per kilobase of transcript per million fragments mapped; GO: gene ontology; BP: biological process; MF: molecular function; CC: cellular component; STRING: Search Tool for the Retrieval of Interacting Genes; PPI: protein-protein interaction; WGCNA: a weighted gene co-expression network analysis; MTRs: Module-Trait relationships; PCA: principal components analysis.

Declarations

Ethics approval and consent to participate

Please refer to Method section for the detailed information.

Availability of data and material

All data generated in this study were included in the main article and its supplementary files. The raw data of RNA sequencing have been deposited in the National Center for Biotechnology Information Sequence Read Archive with accession No. PRJNA527944 (Available online: <https://www.ncbi.nlm.nih.gov/sra/PRJNA527944>).

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Funding

This study was supported by Key Research and Development Project of Shandong Province (2018GNC110004), Shandong Swine Industry Technology System Innovation (SDAIT-08-03), Agricultural scientific and technological innovation project of Shandong Academy of Agricultural Sciences (CXGC2017B02), and the National Natural Science Foundations of China (31372293). The funding agency provided financial support for the research but was not involved in the design of the study nor collection, analysis, and interpretation of data or in writing the manuscript.

Authors' contributions

W.J. conceived and designed the experiments. Z.X. carried out computational analysis and carried out the experimental validations. Z.X. and W.J. wrote the manuscript. W.C., H.H, W.Y., and L.H. contributed to the sample collecting, sequencing, data analysis, and interpretation of data. All authors read and approved the final manuscript.

Acknowledgements

The authors wish to thank the Pig Farm of Jianghai and the Breeding Swine Quality Supervision and Testing Center, Ministry of Agriculture (Wuhan) for their cooperation in this study.

Reference

1. Klont RE, Brocks L, Eikelenboom G. Muscle fibre type and meat quality. *Meat Sci.*1998; 49(S1):S219-29.
2. Laack RL, Van, Stevens SG, Stalder KJ. The influence of ultimate pH and intramuscular fat content on pork tenderness and tenderization. *J Anim Sci.* 2001; 79(2):392-7.

3. Hernández-Sánchez J, Amills M, Pena RN, Mercadé A, Manunza A, Quintanilla R. Genomic architecture of heritability and genetic correlations for intramuscular and back fat contents in Duroc pigs. *J Anim Sci.* 2013; 91(2):623-32.
4. Pearson TA, Manolio TA. How to interpret a genome-wide association study. *JAMA.* 2008; 299(11):1335-44.
5. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet.* 2009; 10(1):57-63.
6. Zhang YY, Wang HB, Wang YN, Wang HC, Zhang S, Hong JY, et al. Transcriptome analysis of mRNA and microRNAs in intramuscular fat tissues of castrated and intact male Chinese Qinchuan cattle. *PLoS One.* 2017; 12(10):e0185961.
7. Miao X, Luo Q, Qin X, Guo Y. Genome-wide analysis of microRNAs identifies the lipid metabolism pathway to be a defining factor in adipose tissue from different sheep. *Sci Rep.* 2015; 5:18470.
8. Cardoso TF, Cánovas A, Canela-Xandri O, González-Prendes R, Amills M, Quintanilla R. RNA-seq based detection of differentially expressed genes in the skeletal muscle of Duroc pigs with distinct lipid profiles. *Sci Rep.* 2017; 7:40005.
9. Li XJ, Zhou J, Liu LQ, Qian K, Wang CL. Identification of genes in longissimus dorsi muscle differentially expressed between Wannanhua and Yorkshire pigs using RNA-sequencing. *Anim Genet.* 2016; 47(3):324-33.
10. Lim KS, Lee KT, Park JE, Chung WH, Jang GW, Choi BH, et al. Identification of differentially expressed genes in longissimus muscle of pigs with high and low intramuscular fat content using RNA sequencing. *Anim Genet.* 2017; 48(2):166-74.
11. Cisneros F, Ellis M, Baker DH, Easter RA, McKeith FK. The influence of short-term feeding of amino acid-deficient diets and high dietary leucine levels on the intramuscular fat content of pig muscle. *J Anim Sci.* 2010; 63(3):517-22.
12. Gao Y, Zhang YH, Zhang S, Li F, Wang S, Dai L, et al. Association of A-FABP gene polymorphism in intron 1 with meat quality traits in Junmu No. 1 white swine. *Gene.* 2011; 487(2):170-3.
13. Bink MC, Te Pas MF, Harders FL, Janss LL. A transmission/disequilibrium test approach to screen for quantitative trait loci in two selected lines of Large White pigs. *Genet Res.* 2000; 75(1):115-21.
14. Choi JS, Jin SK, Jeong YH, Jung YC, Jung JH, Shim KS, et al. Relationships between Single Nucleotide Polymorphism Markers and Meat Quality Traits of Duroc Breeding Stocks in Korea. *Asian-Australas J Anim Sci.* 2016; 29(9):1229-38.

15. Varet H, Brillet-Guéguen L, Coppée J-Y, Dillies M-A. SARTools: A DESeq2- and edgeR-Based R Pipeline for Comprehensive Differential Analysis of RNA-Seq Data. *PLoS One*. 2016; 11(6):e0157022.
16. David GG, Allen SJ, Elias CF. Role of the adipocyte-derived hormone leptin in reproductive control. *Horm Mol Biol Clin Investig*. 2014; 19(3):141-9.
17. Gao QX, Li J, Liu HL, Wang LY, Xu YX. Comparative study on lipogenic and lipolytic gene expression in intramuscular fat tissue between growing Erhualian and large white pigs. *Yi Chuan Xue Bao*. 2004; 31(11):1218-25.
18. Lagathu C, Christodoulides C, Tan CY, Virtue S, Laudes M, Campbell M, et al. Secreted frizzled-related protein 1 regulates adipose tissue expansion and is dysregulated in severe obesity. *Int J Obes (Lond)*. 2010; 34(12):1695-705.
19. Li YH, Lei T, Chen XD, Xia T, Peng Y, Long QQ, et al. Molecular cloning, chromosomal location and expression pattern of porcine CIDEa and CIDEc. *Mol Biol Rep*. 2009; 36(3):575-82.
20. Puri V, Konda S, Ranjit S, Aouadi M, Chawla A, Chouinard M, et al. Fat-specific protein 27, a novel lipid droplet protein that enhances triglyceride storage. *J Biol Chem*. 2007; 282(47):34213-8.
21. Gross RW, Jenkins C. Identification, cloning, expression, and a purification of three novel human calcium-independent phospholipase a2 family members possessing triacylglycerol lipase and acylglycerol transacylase activities. *J Biol Chem*. 2004; 279(47):48968-75.
22. Wieser V, Adolph TE, Enrich B, Moser P, Moschen AR, Tilg H. Weight loss induced by bariatric surgery restores adipose tissue PNPLA3 expression. *Liver Int*. 2017; 37(2):299-306.
23. Sodhi SS, Park WC, Ghosh M, Jin NK, Sharma N, Shin KY, et al. Comparative transcriptomic analysis to identify differentially expressed genes in fat tissue of adult Berkshire and Jeju Native Pig using RNA-seq. *Mol Biol Rep*. 2014; 41(9):6305-15.
24. Wang B, Wood IS, Trayhurn P. PCR arrays identify metallothionein-3 as a highly hypoxia-inducible gene in human adipocytes. *Biochem Biophys Res Commun*. 2008; 368(1):88-93.
25. Byun HR, Kim DK, Koh JY. Obesity and downregulated hypothalamic leptin receptors in male metallothionein-3-null mice. *Neurobiol Dis*. 2011; 44(1):125-32.
26. Jeong J, Kwon EG, Im SK, Seo KS, Baik M. Expression of fat deposition and fat removal genes is associated with intramuscular fat content in longissimus dorsi muscle of Korean cattle steers. *J Anim Sci*. 2012; 90(6):2044-53.
27. Duncan RE, Ahmadian M, Jaworski K, Sarkadi-Nagy E, Sul HS. Regulation of lipolysis in adipocytes. *Annu Rev Nutr*. 2007;27:79-101.

28. Wang W, Xue W, Xu X, Jin B, Zhang X. Correlations of genes expression in PPAR signalling pathway with porcine meat quality traits. Czech J Anim Sci. 2016; 61(7):333-9.

29. Jiang S, Wei H, Song T, Yang Y, Peng J, Jiang S. Transcriptome Comparison between Porcine Subcutaneous and Intramuscular Stromal Vascular Cells during Adipogenic Differentiation. PLoS One. 2013; 8(10):e77094.

30. Chen W, Fang GF, Wang SD, Wang H, Zeng YQ. Longissimus lumborum muscle transcriptome analysis of Laiwu and Yorkshire pigs differing in intramuscular fat content. Genes Genom. 2017; 39(7):759-66.

31. Kim D, Langmead B, Salzberg S. HISAT: a fast spliced aligner with low memory requirements. Nat Methods. 2015; 12(4):357-60.

32. Liao Y, Smyth G, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014; 30(7):923-30.

33. Love MI, Wolfgang H, Simon A. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Bio. 2014; 15(12):550.

34. Robinson MD, Mccarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010; 26(1):139-40.

35. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. OMICS. 2012; 16(5):284-7.

36. Smoot ME, Ono K, Ruscheinski J, Wang PL, Ideker T. Cytoscape 2.8. New features for data integration and network visualization. Bioinformatics. 2011; 27(3):431-2.

37. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008; 9:559.

Tables

Table 1. Description of the differentially expressed genes detected between IMF-High and -Low groups with $|\log_2(\text{fold change})| \geq 1$ and $q\text{-value} < 0.05$

ID	Gene Name	log2(fold change)	q-value	
			DEseq2	edgeR
ENSSSCG00000034371	-	13.31	3.82E-07	4.08E-05
ENSSSCG00000009216	<i>SPP1</i>	5.43	6.11E-09	9.69E-05
ENSSSCG00000040464	<i>LEP</i>	4.04	2.12E-02	3.94E-02
ENSSSCG00000011557	<i>CIDEA</i>	3.58	2.12E-02	5.06E-03
ENSSSCG00000025822	<i>SFRP1</i>	2.61	2.92E-02	8.22E-03
ENSSSCG00000040849	-	-2.63	4.24E-02	1.72E-02
ENSSSCG00000013892	<i>KCNN1</i>	-3.23	1.14E-08	6.84E-04

Table 2. Significant enriched pathways of differentially expressed genes (*q*-value ≤ 0.05)

Description	Ratio	<i>p</i> _value	<i>q</i> _value	GeneName
Regulation of lipolysis in adipocytes	5/47	4.94E-05	5.05E-03	<i>PLIN1, LIPE, PTGER3, ADRB1, FABP4</i>
Retinol metabolism	4/47	1.20E-04	5.05E-03	<i>CYP2A19, SDR16C5, RETSAT, AOX1</i>
PPAR signaling pathway	5/47	1.34E-04	5.05E-03	<i>PLIN1, ACAA1, PPARG, ADIPOQ, FABP4</i>
Vitamin B6 metabolism	2/47	1.77E-03	4.14E-02	<i>PSAT1, AOX1</i>
cAMP signaling pathway	6/47	1.83E-03	4.14E-02	<i>LIPE, PTGER3, HCAR1, ADRB1, PPP1R1B</i>

Supplementary Material Legend

- Additional file 1: Table S1. Summary of sequencing data of 28 Duroc pigs.
- Additional file 2: Table S2. DEGs detected by DESeq2 and edgeR in longissimus dorsi muscle between the IMF-high and -low groups of Duroc pigs ($|\log_2(\text{fold change})| \geq 1, p < 0.05$)
- Additional file 3: Table S3. Significant enriched Gene ontology terms (*q* value < 0.05) of DEG detected in the study.
- Additional file 4: Table S4. Key DEGs used for protein-protein interaction network analysis.
- Additional file 5: Table S5. Significant enriched GO terms (*q* value < 0.05) of co-expressed genes in turquoise module.
- Additional file 6: Table S6. Significant enriched pathways (*q* value < 0.05) of co-expressed genes in turquoise module

Figures

Figure 1

Transcriptomic difference and differentially expressed genes between the IMF-high and -low groups. (A) Principal Component Analysis (PCA) of 300 genes with the most variance from RNA sequencing data of 28 Duroc pigs. (B) Heatmap of DEGs with $|\log_2(\text{fold change})| \geq 1$ and p value < 0.05 . (C) Hierarchical clustering of the 28 Duroc pigs with DEGs. The DEGs used in picture B and C were 309 overlapped genes identified by DEseq2 and edgeR with $|\log_2(\text{fold change})| \geq 1$ and p value < 0.05 . DEG, differentially expressed gene.

Figure 3

Protein-protein interaction (PPI) network of key genes affecting IMF content. The analysis used 23 key genes identified by differential expression analysis under stricter criteria ($|\log_2(\text{fold change})| \geq 1$ and $q\text{-value} < 0.05$) and functional enrichment analysis of DEGs. Genes without interaction were not exhibited in the picture.

Figure 4

WGCNA turquoise module related to IMF content. (A) Cluster dendrogram showing the co-expression modules defined by WGCNA and labeled by colors. (B) Significant enriched GO terms and pathways of the co-expression genes in the turquoise module ($q < 0.05$). (C) Protein-protein interaction (PPI) network of genes with weight value > 0.36 in the turquoise module. The DEGs were represented in red. WGCNA, Weighted Gene Co-expression Network Analysis. GO, Gene Ontology; DEG, differentially expressed gene.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SUPPLYMENT.xlsx](#)
- [ARRIVEGuidelinesChecklist.pdf](#)