BMC Genomics

Trait correlated expression combined with eQTL and ASE analyses identified novel candidate genes affecting intramuscular fat

Yan Liu[1,2,3], Huan Long[1,2], Simin Feng[1,2], Tingting Ma[1,2], Mufeng Wang[1,2], Lizhu Niu[1,2], Xinyi Zhang[1,2], Lianni Wang[1,2], Yu Lei[1,2], Yilong Chen[1,2], Qiankun Wang[1,2] and Xuewen Xu[1,2,3]*

Abstract

Background: Intramuscular fat (IMF) content is a determining factor for meat taste. The Luchuan pig is a fat-type local breed in southern China that is famous for its desirable meat quality due to high IMF, however, the crossbred offspring of Luchuan sows and Duroc boars displayed within-population variation on meat quality, and the reason remains unknown.

Results: In the present study, we identified 212 IMF-correlated genes (FDR ≤ 0.01) using correlation analysis between gene expression level and the value of IMF content. The IMF-correlated genes were significantly enriched in the processes of lipid metabolism and mitochondrial energy metabolism, as well as the AMPK/PPAR signaling pathway. From the IMF-correlated genes, we identified 99 genes associated with expression quantitative trait locus (eQTL) or allele-specific expression (ASE) signals, including 21 genes identified by both cis-eQTL and ASE analyses and 12 genes identified by trans-eQTL analysis. Genome-wide association study (GWAS) of IMF identified a significant QTL on SSC14 (p-value = $2.51E^{-7}$), and the nearest IMF-correlated gene *SFXN4* ($r = 0.28$, FDR = $4.00E^{-4}$) was proposed as the candidate gene. Furthermore, we highlighted another three novel IMF candidate genes, namely *AGT*, *EMG1*, and *PCTP*, by integrated analysis of GWAS, eQTL, and IMF-gene correlation analysis.

Conclusions: The AMPK/PPAR signaling pathway together with the processes of lipid and mitochondrial energy metabolism plays a vital role in regulating porcine IMF content. Trait correlated expression combined with eQTL and ASE analysis highlighted a priority list of genes, which compensated for the shortcoming of GWAS, thereby accelerating the mining of causal genes of IMF.

Keywords: Pig, IMF, eQTL, ASE, GWAS, AMPK/PPAR

* Correspondence: xuewen_xu@mail.hzau.edu.cn
[1]Key Laboratory of Agricultural Animal Genetics, Breeding and Reproduction, Ministry of Education & College of Animal Science and Technology, Huazhong Agricultural University, Wuhan 430070, China
[2]The Cooperative Innovation Center for Sustainable Pig Production, Wuhan 430070, China
Full list of author information is available at the end of the article

Background

Intramuscular fat (IMF) is the adipose tissue deposited between skeletal muscle fibers, and its content is positively correlated with meat juiciness, flavor intensity, and tenderness [1]. A higher IMF content (> 3.5%) of pork is generally accepted because of its positive sensory experiences. However, too much visible fat in the meat will reduce its consumer acceptability due to health concerns [1]. An optimal range of IMF content of 2.0 to 4.0% was recommended as the industry target in the United States according to previous studies [2], and the IMF content has moderate to high heritability: the average heritability is 0.4, ranging from 0.24 to 0.67 in different populations [3–5]. Despite its high heritability, direct selection for IMF content during pig breeding is difficult to put into practice since it is measured post-slaughter. Therefore, seeking the regulatory genes and potential molecular markers of IMF content is an important task for genetic research and pig breeding.

To specifically reveal the genes and pathways involved in intramuscular fat deposition, muscle transcriptomes based on cDNA microarray were compared between extreme animals with the divergent selection of IMF content, such as 2 × 8 Pietrain x Duroc F2 animals [6], 2 × 35 Duroc pigs [7] and 2 × 20 Iberian (25%) x Landrace (75%) back-crossed animals [8]. Recently, the RNA sequencing (RNA-seq) approach was used to investigate muscle transcriptome differences in Duroc pigs with distinct lipid profiles [9] and Berkshire pigs with divergent IMF content [10], which highlighted the vital effect of lipid metabolism-related genes, such as fatty-acid synthase (*FASN*) and stearoyl-CoA desaturase (*SCD*), in IMF content determination. Although the comparative transcriptome profiling approach identified some differential expressed genes (DEGs) related to IMF content, it was hard to prioritize candidate genes that genetically contributed to phenotypic variation. Instead, the approach combining high-throughput genotyping and gene expression analysis, namely expression quantitative trait loci (eQTL) and allele specific expression (ASE) analyses, could discover potential genomic variants that exerted their effects on phenotypic variation by regulating gene expression level. Thus, eQTL and ASE analyses were often used to prioritize causal genes in the genome-wide association studies (GWASs) of complex traits, such as porcine muscle glycogen content [11] and intramuscular fat [12].

The Luchuan pig is a typical Chinese local variety with high fat deposition and good meat quality, widely distributed in southern China, including Guangdong and Guangxi provinces. In recent years, Luchuan sows were crossed with Duroc boars to produce high-quality meat, whereas large within-population variations of fat deposition and meat quality were observed. To identify potential genes controlling meat quality traits in this population, we collected skeletal muscle samples for trait measurement and RNA-seq. Recently, we reported the results of genome-wide eQTL and ASE analyses based on RNA-seq data, in which we confirmed some known candidate genes and revealed some novel candidate genes for meat quality traits [13]. In the present study, we performed the correlation analysis between gene expression level and IMF content, and GWAS of IMF. Further, we conducted an integrated analysis of IMF-correlated genes, eQTL, and GWAS, through which we hope to reveal important pathways and candidate genes affecting IMF content.
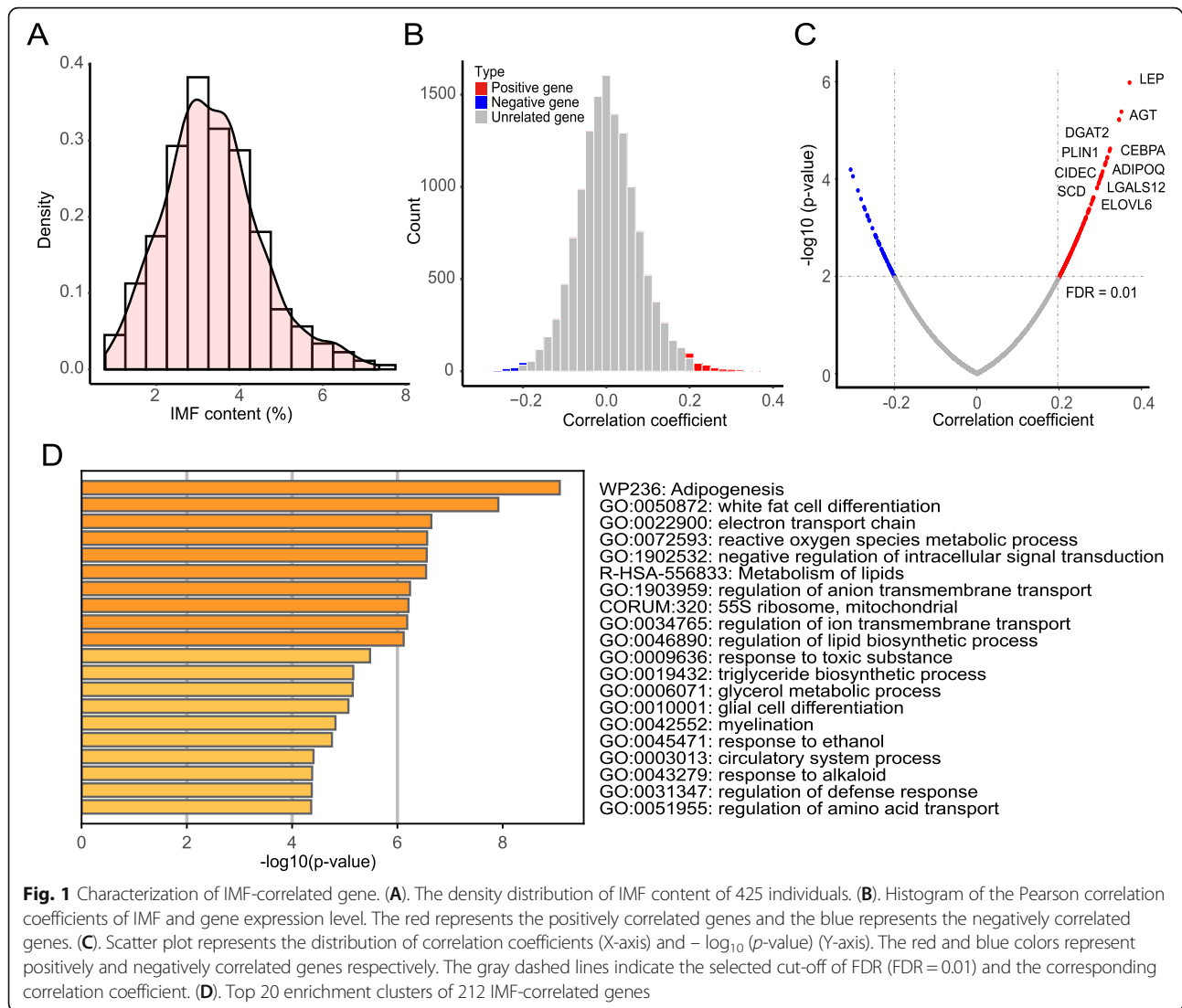
Results

Correlation analysis with trait and gene expression levels identified 212 IMF-correlated genes

As described in our recent paper, the F1 animals of Duroc boars and Luchuan sows displayed great phenotypic variation on meat quality [13]. For IMF content of longissimus dorsi muscle, the 425 F1 animals had an average IMF of 3.35% ± 1.16%, ranging from 0.77 to 7.36% (Fig. 1A). We randomly selected 189 individuals from the population for RNA sequencing with longissimus dorsi muscle, by which the expression levels of all reference genes (25,880 genes) were estimated and a total of 13,450 genes were kept for eQTL and ASE analysis after the QC procedure [13]. Here, we conducted the correlation analysis between IMF and gene expression levels. To do this, the $\log_2$-transformed gene expression level and the phenotypic values of IMF were pre-adjusted for systematic effects using a fixed linear model. Within the 13,450 input genes, correlation analysis identified 212 genes that were significantly correlated with IMF (FDR ≤ 0.01) (Supplementary Table S1). The 212 genes contained 154 negatively correlated genes and 58 positively correlated genes, and the absolute value of the correlation coefficient ranged from 0.20 to 0.37 (Fig. 1B & 1C). These genes were significantly enriched in 20 clusters, which included the processes or GO terms related to adipogenesis, lipid biosynthetic and metabolism process, triglyceride biosynthetic process, electron transport chain, mitochondrial translation and so on (Fig. 1D, Supplementary Table S2).

The IMF-correlated genes were significantly enriched in the processes of lipid and energy metabolism and the AMPK/PPAR signaling pathway

To explore potential signaling pathways affecting IMF, we carried out KEGG, GSEA, and protein-protein interaction (PPI) network analysis using the 212 IMF-correlated genes. The KEGG pathway analysis revealed that five IMF-correlated genes including *PCK1*, *PLIN1*, *PPARG*, *SCD*, and *ADIPOQ* were significantly enriched in the PPAR signaling pathway (q-value = $2.12E^{-2}$), six

Fig. 1 Characterization of IMF-correlated gene. (**A**). The density distribution of IMF content of 425 individuals. (**B**). Histogram of the Pearson correlation coefficients of IMF and gene expression level. The red represents the positively correlated genes and the blue represents the negatively correlated genes. (**C**). Scatter plot represents the distribution of correlation coefficients (X-axis) and $-\log_{10}$ (*p*-value) (Y-axis). The red and blue colors represent positively and negatively correlated genes respectively. The gray dashed lines indicate the selected cut-off of FDR (FDR = 0.01) and the corresponding correlation coefficient. (**D**). Top 20 enrichment clusters of 212 IMF-correlated genes

genes including *NDUFB8, NDUFS3, ATP6V1F, UQCRQ, UQCR10,* and *NDUFB11* were enriched in oxidative phosphorylation (q-value = $3.61E^{-2}$), and seven genes including *FASN, LEP, PCK1, PPARG, SCD, ADIPOQ,* and *STRADB* were enriched in the AMPK signaling pathway (q-value = $6.98E^{-3}$) (Supplementary Table S2). The GSEA analysis with the KEGG gene sets as background revealed that 22 enriched gene sets were positively related with IMF and 24 gene sets were negatively related with IMF (Supplementary Table S3). Oxidative phosphorylation was the most significantly enriched pathway with the highest positive enrichment score (NES = 2.52), followed by fatty acid metabolism (NES = 2.08), PPAR signaling pathway (NES = 2.00) and glycolipid metabolism (NES = 1.92), peroxisome (NES = 1.78) and tricarboxylic acid cycle (NES = 1.74) (Fig. 2).

Furthermore, we used the STRING database to explore the co-expression or interaction of 212 IMF-correlated genes. Among these genes, only 178 genes could be discerned by the STRING database and kept for PPI enrichment analysis. Finally, 124 nodes (genes) and 253 edges were shown in the network (Fig. 3A), whereas another 54 genes were hidden since they did not connect with any nodes. The network contained a significantly higher frequency of interactions than the expected by evaluating the expected edges for a random set of proteins of similar size (expected edges = 123, PPI enrichment *p*-value $< 1.00E^{-16}$). In the PPI network of IMF-correlated genes, the genes related to lipid biosynthesis or metabolic processes displayed an extremely high frequency of interactions, which was followed by the genes related to mitochondrial translation and the oxidative

Liu *et al. BMC Genomics*   (2021) 22:805

Page 4 of 14

Fig. 2 GSEA analysis for all IMF-correlated genes. Six significantly enriched KEGG pathways were output as follows: (**A**). Oxidative phosphorylation, (**B**). Fatty acid metabolism, (**C**). PPAR signaling pathway, (**D**). Glycolipid metabolism, (**E**). Peroxisome and (**F**). Tricarboxylic acid cycle. NES means normalized enrichment score. Permission has been obtained from Kanehisa laboratories to use the KEGG pathway database [48]

phosphorylation pathway (Fig. 3A). We counted the degree of connectivity (edges) for each gene, and *PPARG* had the highest frequency of interactions (Fig. 3B). The top 20 core genes, such as *PPARG, LEP, ADIPOQ, FASN, PCK1* and so on, had the minimum edge of 8 (Fig. 3B), with which a sub-core network was built (Fig. 3C). Interestingly, the 20 core genes were mainly enriched in lipid metabolic process (*n* = 15), PPAR

signaling pathway (*n* = 5), AMPK signaling pathway (*n* = 6) and mitochondrial translation (*n* = 4) (Fig. 3C).

Intersection of IMF-correlated genes and target genes of cis-eQTL, ASE and trans-eQTL

Our previous study identified 2098 genes with cis-eQTL signals, 441 genes with specific trans-eQTL signals (without cis-eQTL signals) and 2253 genes with allele-specific
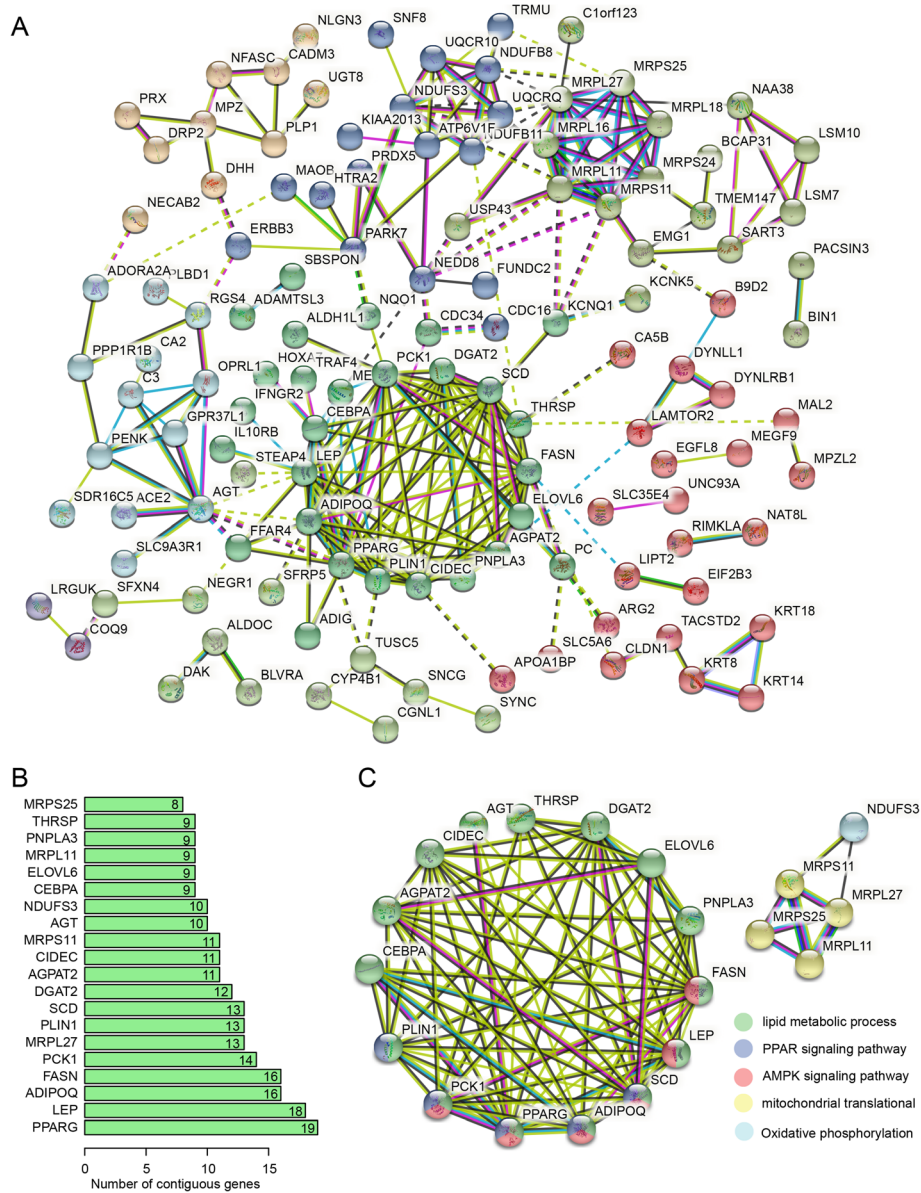
Fig. 3 PPI network analysis of IMF-correlated genes. (**A**). PPI network of 124 IMF-correlated genes. Circles represent genes (nodes) and lines represent the interaction between genes (edges). Line colors represent the different extent of evidence of PPI and circle colors represent different clusters generated by k-means clustering. (**B**). Barplot of edges number of top 20 core genes in the PPI network. (**C**). PPI sub-network of top 20 core genes. Different circle colors represent distinct functional processes or pathways

expression (ASE) signals [13]. Among the 212 IMF-correlated genes, there were 60 cis-eQTL-associated genes and 50 ASE genes (Supplementary Table S1), in which 21 genes were identified by both cis-eQTL and ASE analyses (Table 1, Fig. 4A). In addition, the 212 IMF -correlated genes contained 12 trans-eQTL-associated genes (Fig. 4B, Supplementary Table S1). Taking the *NAXE* gene as an example, whose expression level were positively correlated with IMF ($r = 0.24$, FDR $= 1.90E^{-3}$), and a strong cis-eQTL was identified at SSC4:93673165 (p-value $= 2.76E^{-11}$) (Fig. 4C). ASE analysis revealed that the

reference allele of the SNP (rs326817713) in the first exon of *NAXE* displayed a relatively high expression level for 56 individuals of 69 heterozygotes (Fig. 4D). Meanwhile, the polymorphism of rs326817713 was significantly associated with the expression level of *NAXE* (p-value $= 1.77E^{-11}$) (Fig. 4E). Among the 12 trans-eQTL-associated genes (Supplementary Table S1), the expression of *PARK7* had the most significant correlation with IMF ($r = 0.26$, FDR $= 9.00E^{-4}$). A significant trans-eQTL on chromosome 9 (rs325298336, p-value $= 3.25E^{-7}$) was identified for *PARK7* that located on chromosome 6 (Fig. 4F).

Table 1 The overlap between eQTL analysis, ASE analysis and correlation analysis

gene_name	eQTL analysis			ASE analysis				Correlation analysis		
	SNP_id	*p*value	FDR	SNP_id	Het	ASE	ratio	cor	*p*value	FDR
EMG1	5:64350471	2.28E-29	1.02E-24	5:63751365	103	103	1.00	0.215	5.76E-03	6.20E-03
PCTP	12:32427760	2.80E-12	2.74E-09	12:32062640	43	11	0.26	0.212	6.43E-03	7.30E-03
NAXE	4:93673165	2.76E-11	2.03E-08	4:93495840	69	56	0.81	0.243	1.72E-03	1.90E-03
EIF2B3	6:167199552	2.00E-10	1.17E-07	6:166438671	49	10	0.20	−0.211	6.75E-03	7.10E-03
SLC9A3R1	12:6284901	1.04E-08	3.55E-06	12:6468404	160	98	0.61	0.208	7.47E-03	9.00E-03
ZFAND2B	15:120963216	1.48E-08	4.85E-06	15:121243305	74	73	0.99	0.242	1.78E-03	2.00E-03
MRPS11	1:191048963	1.49E-07	3.63E-05	1:191321445	132	61	0.46	0.242	1.82E-03	1.90E-03
C1orf123	6:158874827	6.78E-07	1.34E-04	6:159011592	43	21	0.49	0.206	8.11E-03	8.90E-03
PRDX5	2:7551456	1.01E-06	1.86E-04	2:7803041	88	10	0.11	0.245	1.60E-03	1.60E-03
MRPL16	2:11716316	1.01E-06	1.86E-04	2:11694100	62	29	0.47	0.209	7.25E-03	7.40E-03
MGST2	8:86943096	4.96E-06	6.83E-04	8:87328624	149	86	0.58	0.240	1.94E-03	1.90E-03
IFNGR2	13:196334437	6.58E-06	8.67E-04	13:197020918	170	16	0.09	0.197	1.15E-02	9.90E-03
IL10RB	13:196629902	1.00E-05	1.21E-03	13:196860215	39	19	0.49	0.202	9.40E-03	8.00E-03
DGAT2	9:9831161	1.02E-05	1.24E-03	9:10063315	152	117	0.77	0.322	2.68E-05	1.00E-05
SMIM20	8:19726922	2.81E-05	2.77E-03	8:19714368	175	143	0.82	0.269	4.93E-04	1.10E-03
AGT	14:59704514	3.81E-05	3.54E-03	14:59648664	177	85	0.48	0.345	6.09E-06	1.00E-05
ATP6V1F	18:20146282	5.54E-05	4.72E-03	18:19756485	166	76	0.46	0.316	3.78E-05	1.00E-04
PLBD1	5:57889765	7.38E-05	5.94E-03	5:57951366	82	16	0.20	−0.265	5.97E-04	3.00E-04
CDC34	2:77362555	9.82E-05	7.42E-03	2:77768209	188	96	0.51	0.249	1.32E-03	1.80E-03
FRMD3	10:31523619	2.91E-04	1.68E-02	10:31714269	115	25	0.22	0.228	3.32E-03	2.40E-03
ENSSSCG00000032003	2:5854819	5.83E-04	2.82E-02	2:4978950	183	135	0.74	0.251	1.18E-03	1.20E-03

"Het" represents the total number of observed heterozygotes, and "ASE" means the number of heterozygotes that displayed ASE. "cor" represents the Pearson correlation coefficient

Integration of GWAS, eQTL analysis, and IMF-gene correlation analysis highlighted four novel candidate genes of IMF

Integration of eQTL (or ASE) analysis and trait correlated expression analysis established the relationship of SNP-gene-trait, whereas the association between SNP and trait remains unknown. To establish the potential triple relationship of SNP-gene-trait, we conducted genome-wide association study for IMF using a fixed linear model with the 189 animals that were genotyped using the Illumina porcine 50 K + SNP iSelect™ BeadChip [13]. It revealed that only one SNP on chromosome 14 (WU_10.2_14_139488835, *p*-value = $2.51E^{-7}$) was associated with IMF on the level of genome-wide significance (*p*-value = $1.39E^{-6}$) (Fig. 5A & 5B). From the list of IMF-correlated genes, the *SFXN4* gene was highlighted as the unique candidate gene of the QTL because it was approximately 1 Mb away from the leading SNP (WU_10.2_14_139488835) (Fig. 5A). Moreover, the expression level of *SFXN4* was significantly correlated with IMF (*r* = 0.28, FDR = $4.00E^{-4}$). Although no cis or trans-eQTL signal was identified for *SFXN4*, six SNPs located in its 3'UTR displayed significant allele-specific expression (Fig. 5C).

In addition, 117 SNPs were associated with IMF on the significant level of *p*-value ≤ $1.00E^{-3}$ (Fig. 5A & 5B), which represented 1.94% of all tested SNPs. By integrated analysis of GWAS, eQTL and IMF-gene correlation analysis, we finally highlighted 9 SNPs that satisfied the triple relationship of SNP-gene-trait, involved in three genes, namely *AGT*, *EMG1* and *PCTP* (Table 2). Taking the *AGT* gene as an example, its expression level was significantly correlated with IMF (*r* = 0.34, FDR = $1.00E^{-5}$), and the polymorphism of ALGA0078039 was associated with IMF (*p*-value = $1.06E^{-4}$) and the expression level of *AGT* (*p*-value = $1.02E^{-3}$).

Discussion

In the present study, the IMF content of the F1 animals of Luchuan sows and Duroc boars displayed a great range of variation, from 0.77 to 7.36%, which was higher than that of Berkshire pigs (from 0.42 to 3.83%) [10]. Too much intrapopulation variation was a disadvantage for commercial production, and therefore, breeders were seeking genetic improvement of this trait. However, to our knowledge, the genetic basis of the IMF content of this population has not been well studied. Thus, identification of IMF-correlated genes could help guide the
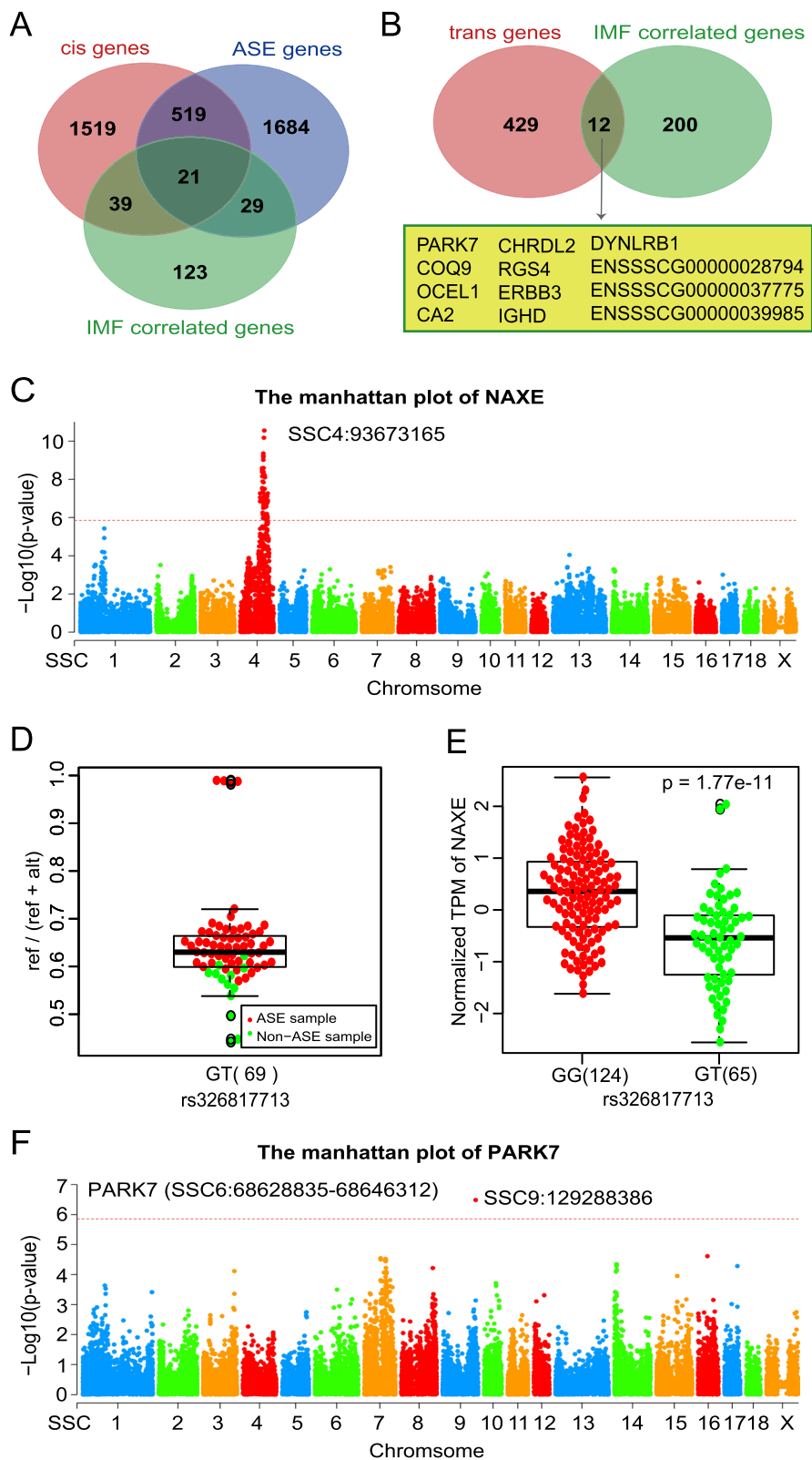
Fig. 4 (See legend on next page.)

(See figure on previous page.)

Fig. 4 eQTL and ASE analyses of IMF-correlated genes. (**A**). Veen diagram of significantly cis-eQTL target genes, ASE genes and IMF-correlated genes. (**B**). Veen diagram of significantly specific trans-eQTL target genes and IMF-correlated genes. (**C**). The Manhattan plot of *NAXE*. The red dashed line represents the cutoff of statistical significance at the genome-wide level ($p = 0.05/36045$). (**D**). Boxplot of the allele-specific expression ratio of *NAXE* for the SNP of rs326817713 (G > T). Allele-specific expression ratio means the number of reads aligned to the reference allele divided by the total number of reads aligned to this site. "ASE sample" (red dot) indicates the heterozygotes with ASE signal at this site, and "Non-ASE sample" (green dot) indicates the heterozygotes without ASE signal at this site. (E). Boxplot of the expression level of *NAXE* with different genotypes of rs326817713. (**F**). The Manhattan plot of *PARK7*. The red dashed line represents the cutoff of statistical significance at the genome-wide level ($p = 0.05/36,045$)
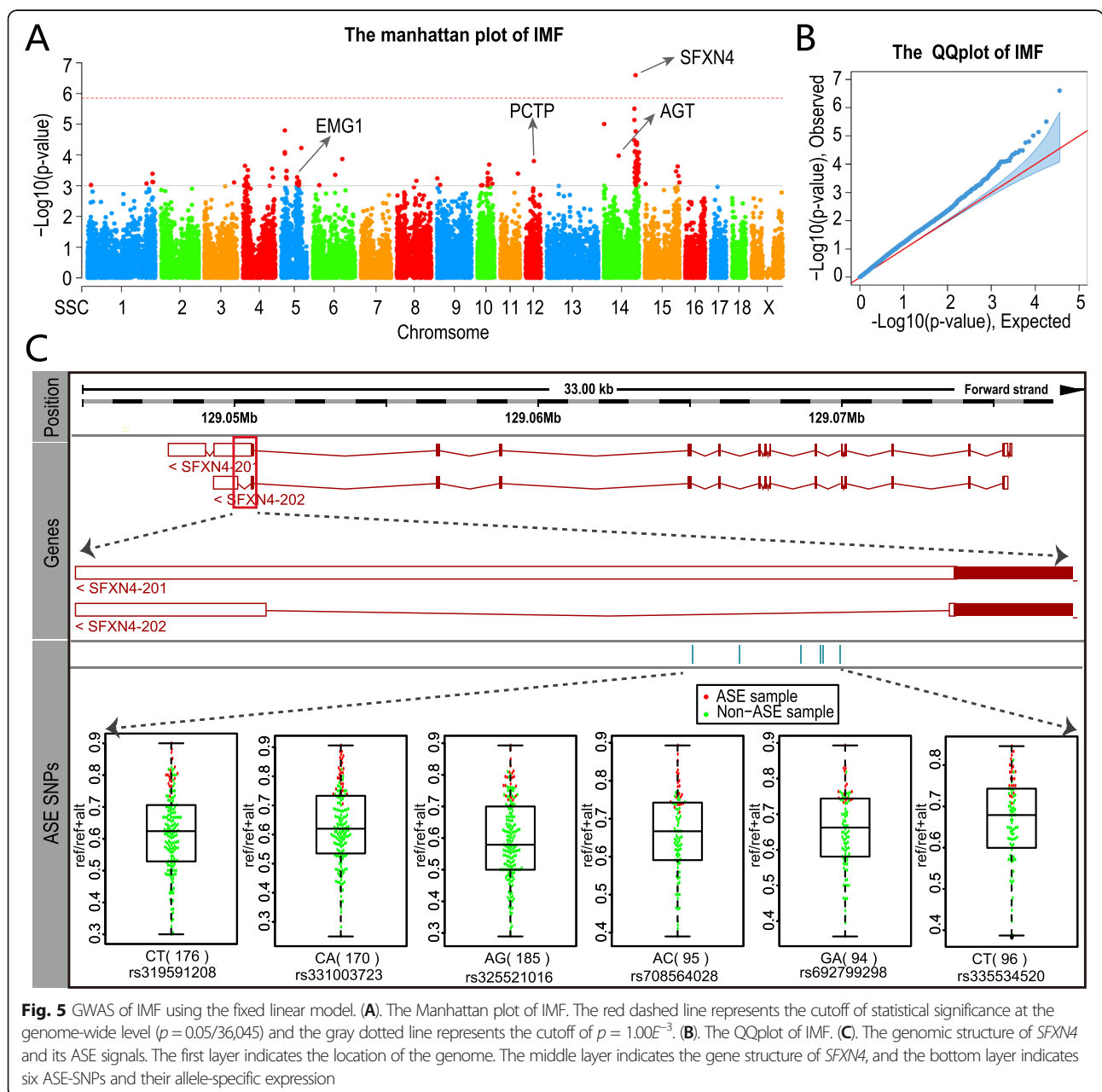

Fig. 5 GWAS of IMF using the fixed linear model. (**A**). The Manhattan plot of IMF. The red dashed line represents the cutoff of statistical significance at the genome-wide level ($p = 0.05/36,045$) and the gray dotted line represents the cutoff of $p = 1.00E^{-3}$. (**B**). The QQplot of IMF. (**C**). The genomic structure of *SFXN4* and its ASE signals. The first layer indicates the location of the genome. The middle layer indicates the gene structure of *SFXN4*, and the bottom layer indicates six ASE-SNPs and their allele-specific expression

Liu *et al. BMC Genomics*   (2021) 22:805

Page 9 of 14

Table 2 The overlap between eQTL analysis, GWAS analysis and correlation analysis

SNP_id	Gene_Name	Correlation analysis			eQTL analysis				GWAS analysis	
		cor	*p*value	FDR	r	*p*value	FDR	type	r	*p*value
ALGA0078039	AGT	0.34	6.09E-06	1.00E-05	0.26	1.02E-03	4.17E-02	cis	0.31	1.06E-04
WU_10.2_5_66693369	EMG1	0.21	5.76E-03	6.20E-03	0.69	7.32E-23	2.10E-18	cis	0.28	5.29E-04
M1GA0007834	EMG1	0.21	5.76E-03	6.20E-03	0.66	2.85E-20	3.58E-16	cis	0.27	7.90E-04
ALGA0032380	EMG1	0.21	5.76E-03	6.20E-03	0.52	7.12E-12	6.03E-09	cis	0.27	8.70E-04
ALGA0032795	EMG1	0.21	5.76E-03	6.20E-03	−0.52	7.46E-12	2.90E-06	trans	−0.27	7.06E-04
ASGA0026250	EMG1	0.21	5.76E-03	6.20E-03	−0.52	7.46E-12	2.90E-06	trans	−0.27	7.06E-04
WU_10.2_5_73855132	EMG1	0.21	5.76E-03	6.20E-03	−0.52	7.46E-12	2.90E-06	trans	−0.27	7.06E-04
WU_10.2_5_73348031	EMG1	0.21	5.76E-03	6.20E-03	−0.50	3.61E-11	1.14E-05	trans	−0.26	9.92E-04
ALGA0065989	PCTP	0.21	6.43E-03	7.30E-03	−0.35	6.72E-06	8.81E-04	cis	−0.30	1.58E-04

"cor" represents the Pearson correlation coefficient. "r" represents the correlation coefficient evaluated by matrixEQTL

discovery of candidate genes and genetic markers and could benefit the research on genetic improvement.

By correlation analysis, we identified 212 IMF-correlated genes. Within the 212 IMF-correlated genes, *FASN, PLIN1, PIN1, CEBPA, CIDEC, SFRP5, SDR16C5,* and *SLC9A3R1* were identified to be DEGs in the muscle of Duroc pigs with extreme lipid profiles [9]. And *SCD, FASN, PLIN1, THRSP, CIDEC, ALDOC, AGT, PC,* and *CA5B* were identified to be DEGs in the muscle of Berkshire pigs with extreme IMF content [10]. In addition, *LEP, SCD, MT3, PENK, B9D2, CYP4B1, KIAA2013* and *ADAMTSL3* were also identified to be DEGs in the muscle of Iberian × Landrace back-crossed pigs with extreme IMF content [14]. Therefore, trait correlated expression analysis confirmed some of the DEGs related to IMF, which supported the reliability of our results. And compared with the strategy of comparative transcriptome study, trait correlated expression analysis depended on a large population that ensured relatively higher statistical power and reliability of results.

The comparative results mentioned above revealed some common functional genes, such as *SCD, PLIN1,* and *FASN,* all of which came from the AMPK/PPAR signaling pathway, indicating the commonality of the signaling pathway involved in the regulation of IMF deposition in different pig breeds. In agreement with the result of GO analysis, both KEGG and GSEA revealed that the IMF-correlated genes were mainly enriched in the processes of fatty acid and energy metabolism, both of which were regulated by the AMPK/PPAR signaling pathway [15]. As a kinase activated by AMP, AMPK plays a key role in the regulation of cellular and overall energy balance, and its activation accelerates ATP production by promoting fatty acid oxidation and glucose transport [15–18]. In addition, AMPK could affect lipolysis by regulating the activity of adipose triglyceride lipase (ATGL) and hormone-sensitive lipase (HSL) [15, 17], reduce fatty acid synthesis through phosphorylation

and inactivation of acetyl CoA carboxylase (ACC) [15, 18], and inhibit fatty acid oxidation by inhibiting fatty acid translocase (CD36) that played a vital role in the homeostasis of lipid metabolism [15, 19, 20]. As a downstream pathway of AMPK, PPARs act as fatty acid sensors and play a crucial role in lipid metabolism, in which PPAR gamma (PPARγ) acts as a major activator of adipogenesis and fatty acid storage [21, 22]. Together, the AMPK/PPAR signaling pathway plays a vital role in regulating IMF content, which was further supported by in vitro [23, 24] and in vivo experiments [25, 26] as well as with transgenic models [27].

Trait correlated expression analysis highlights functional genes that may affect the phenotype by changing the expression levels, eQTL and ASE analyses provide direct clues for characterization of genes carrying genetic variant modulating its expression level, so integrated analysis of both strategies will be an efficient way to prioritize functional candidate genes for complex traits. In the present study, we highlighted 99 IMF-correlated genes that were the eQTL or ASE associated genes identified in our previous study [13]. Of these 99 genes, *SCD, ADIPOQ, FASN,* and *PCK1* were in the AMPK/PPAR signaling pathway, and the genetic polymorphisms of *SCD* [28–31], and *FASN* [32] were verified to be associated with the IMF content and fatty acid composition in pigs. However, so far, no causative mutation controlling IMF content has been identified from these genes, indicating further studies are necessary. Besides, the 99 genes contained some of novel functional candidate genes of IMF. For example, the *NAXE* gene encodes the NAD(P) HX epimerase, also known as ApoA-I binding protein (AIBP), that interacts with apolipoprotein A-I of high-density lipoproteins (HDL), which can regulate lipid rafts via cholesterol and HDL trafficking [33, 34]. And lipid raft microdomains can modulate the surface availability of CD36, which in turn affects fatty acid oxidation and uptake [15, 35, 36]. Another example is

PARK7, also known as *DJ-1*, which regulates lipid raft-dependent endocytosis [37]. The deletion of *PARK7* could promote fatty acid oxidation and prevent hepatic steatosis in mice [38]. Therefore, the identification of strong cis-eQTL, ASE or trans-eQTL signals associated with those genes provided the directions for characterizing the causal variants regulating IMF content.

Trait correlated expression analysis together with eQTL or ASE analysis would benefit the characterization of functional candidate genes underlying the QTL identified by GWAS. By GWAS of IMF, we identified a QTL on chromosome 14. The *SFXN4* gene was regarded as the unique candidate gene for the QTL because it was approximately 1 Mb away from the leading SNP and its expression level was strongly correlated with IMF. In addition, *SFXN4* encodes a mitochondrial transmembrane protein that plays a key role in mitochondrial respiratory homeostasis [39], so we speculate that it may influence lipid metabolism by affecting energy metabolism, which still needs further investigation. Integration analysis of GWAS, eQTL analysis and trait-gene correlation analysis provided support for the candidacy of another three novel genes affecting IMF, namely *AGT*, *EMG1*, and *PCTP*. The *AGT* gene encodes angiotensinogen, a member of the process of lipid synthesis and metabolism, and its silencing attenuated lipid accumulation in adipocytes [40, 41]. *PCTP* encodes phosphatidylcholine transfer protein that has an important effect on lipid rafts and lipid transport by accelerating apolipoprotein A-I-mediated phospholipid and cholesterol efflux [42, 43]. *EMG1* encodes N1-specific pseudouridine methyltransferase involving in ribosome biogenesis [44], but little is known about its effect on lipid metabolism. Notably, the candidacy of these four novel genes affecting IMF needs to be investigated further.

In the present study, we did not find many SNP-gene-trait links. One obvious reason is the limited sample size, which was fine for eQTL, but small for GWAS. Another important reason is that the F1 animals is not suitable for GWAS due to its limited genetic segregation and power. In addition, the low resolution of SNPs in chips may result in the loss of a large amount of genetic information. Nevertheless, we believe the analysis approach establishing the triple relationship of SNP-gene-trait is interesting. Firstly, any relationship supports the rationality of the other two connections, which could justify the result. Secondly, the triple relationship characterized the causality relationship between gene (or quantitative trait gene, QTG) and phenotypic variation, and provide clues for identifying the quantitative trait nucleotide (QTN). To verify the advantage of this approach, a suitable population, a large sample size, and the use of high-density genotyping method such as sequencing are necessary for future study.

Conclusions

Analysis of trait correlated expression revealed the vital role of the processes of lipid biosynthesis and energy metabolism and the AMPK/PPAR signaling pathway in regulating IMF content. Characterization of IMF-correlated genes that associated with eQTL or ASE signals will facilitate cloning causal genes of IMF. We further highlighted four candidate genes of IMF, but their effects on IMF need further study.

Methods

Animal sampling and phenotyping

The experimental population was described in our recent paper [13]. Briefly, it was composed of 425 F1 animals from the crossing of 8 Duroc boars and 158 Luchuan sows (DL pigs), which were fed with the same diets at the same farm. All animals were slaughtered by a standardized procedure at the age of 210 days ±6 days at the same abattoir. All animal procedures followed the guidelines of regulations for the administration of affairs concerning experimental animals, issued by the State Council of the People's Republic of China. The current study was approved by the Scientific Ethics Committee of Huazhong Agricultural University (the approval number was HZAUSW-2016-010), Wuhan, China.

Twenty-five minutes postmortem, samples of the longissimus dorsi muscle at the thoracolumbar junction were taken for IMF content measurements and RNA extraction (stored in liquid nitrogen). IMF content was measured with the Soxhlet extraction method for each animal, with the following steps: 1) prepare ∼60 g minced meat after removing the fascia and fat on the surface of the muscle; 2) dry the minced meat in a 100 °C oven, record the weight loss and calculate the content of water and dry matter; 3) package the powdered dried meat into three separate filter-paper bags for Soxhlet extraction with ether; and 4) calculate the fat content of each sample contained in a filter-paper bag and multiply it by the dry matter content, which yields the IMF content in the muscle. The final IMF content for one individual was the average value of three samples.

RNA preparation and sequencing

The RNA preparation and RNAseq procedure were introduced in our recent paper [13]. Briefly, total RNA was extracted with Trizol reagent (Invitrogen, USA) from the *Longissimus dorsi* muscle of 189 DL pigs (a randomly selected subset of the DL population mentioned above). RNA concentration and integrity were accessed with NanoDrop 2000 spectrophotometers (Thermo Scientific, USA) and Agilent 2100 Bioanalyzer (Agilent Technologies, USA), respectively. The 150-bp, paired-end, nonstrand-specific libraries were prepared from one microgram of total RNA for each

sample using NEBNext, Ultra RNA Library Prep kit for Illumina (NEB, USA) according to the manufacturer's protocols. The cDNA yield of each library was quantified with Qubit (Invitrogen, USA), and the quality of the libraries was assessed using the Agilent 2100 Bioanalyzer (Agilent Technologies, USA). The prepared libraries were sequenced on the Illumina HiSeq 4000 platform in Shanghai Majorbio Biopharm Technology Co., Ltd. For each sample, at least 6.0 Gigabase pairs (Gb) of sequencing data were obtained. After the sequencing reaction was completed, clean reads were obtained from the raw reads through the following quality control steps: 1) removed adapter; 2) reads containing more than 5% N were removed; and 3) low-quality reads were removed, which were performed using SOAPnuke [45] software with the following parameters "-n 0.01 -l 20 -q 0.4 -A 0.25 --cutAdaptor -Q 2 -G --polyX 50 --minLen 150". For the 189 animals, an average of 48.86 million reads was obtained, which was presented detailly in our recent paper [13].

RNA-seq data analysis

The pipeline of RNA-seq data analysis was introduced in our previous study [13]. Briefly, the obtained RNA-seq data was aligned to the genome assembly Sscrofa11.1 (Ensembl Release 90) using HISAT2 V2.0.5 with a transcript annotation index [46]. The aligned reads for each sample were then assembled into genes with StringTie v1.3.3 [47], from which the expression level represented as TPM (transcripts per million mapped reads) value were estimated for each gene. Moreover, those genes whose TPM was greater than 0.01 in more than 90% of samples were used for correlation and eQTL analyses.

Correlation analysis between IMF content and gene expression levels

The phenotypic data and gene expression levels were corrected using the following fixed linear model, $Y = G + Ba + Bo + A + e$, where Y represents the original IMF content or $\log_2$-transformed expression levels, G is gender, Ba is slaughter batch, Bo is the effect of boar, A represents age as a covariate, and e is a random residual. For all expressed genes, the Pearson correlation coefficients were calculated between the residuals of $\log_2$-transformed expression levels and IMF content using cor.test function in R. The statistical significance of correlation analysis was further evaluated by the false discovery rate (FDR) through 10,000 times random permutation, and the FDR by permutation smaller than 0.01 as the significance threshold of correlation. Of note, the most extreme FDR by permutation was set as $1.00E^{-5}$.

Gene ontology and KEGG pathway analysis

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) [48] analysis were carried out using the online software Metascape [49] (https://metascape.org/gp/index.html) with default parameter. In short, the default ontology sources including KEGG Pathway, GO Biological Processes, Reactome Gene Sets, Canonical Pathways and CORUM and so on, were used for gene enrichment analysis. Then, the statistical test based on cumulative hypergeometric distribution was carried out, and the Benjamini & Hochberg false-discovery rate approach was used for adjusting the p-value (q-value). Meanwhile, Metascape performed a hierarchical cluster analysis based on the similarity of terms. Furthermore, the significant level of the KEGG pathway and GO biological processes were filtered by setting a q-value less than 0.05.

Gene set enrichment analysis (GSEA)

GSEA was performed by using the curated gene sets (C2, KEGG gene sets) of the Molecular Signature Database (MSigDB) v.5.2 [50]. All the genes ranked by their corresponding Pearson correlation coefficients were used for GSEA with default parameters following the standard procedure (http://www.broadinstitute.org/gsea/doc/GSEAUserGuideFrame.html). The GSEA software estimated the probability of the normalized enrichment score (NES) for each gene set, and adjusted the p value by FDR. We set the default FDR value (FDR < 0.25) as the level of statistical significance for GSEA.

PPI network generation

The protein-protein interaction (PPI) network was constructed by STRING (https://string-db.org/) [51] using the multiple proteins process and choosing the "*Homo sapiens*" as background. To improve accuracy and intuitiveness, we set the minimum interaction score as 0.4, the number of k-means clusters as 8, and hid disconnected nodes. In addition, the core genes were evaluated by the degree of connectivity, and a sub-core network was built with the same condition.

Overlapping analysis of IMF-correlated genes and eQTL/ASE genes

The characterized IMF-correlated genes were overlapped with the eQTL and ASE associated genes that were identified previously [13]. The pipeline of eQTL and ASE analysis was introduced in our previous study [13]. Briefly, the eQTL analysis was conducted with Matrix-EQTL [52] using a fixed linear model, with sex, slaughter batch and boar as fixed effects, and age and top five principal components based on marker genotypes as covariates accounting for systematic variation. ASE analysis was carried out using the GATK ASEReadCounter tool [53] with the N-masking genome.

Liu *et al. BMC Genomics* (2021) 22:805

Page 12 of 14

GWAS of IMF

As described in our recent paper, the 189 animals were genotyped using the Illumina porcine 50 K + SNP iSelect™ BeadChip and 36,045 SNPs were kept after the step of quality control [13]. We noted that the mixed linear models fitting random effects of relationship matrix could lead to over-fitting of the model. Therefore, GWAS of IMF was performed using MatrixEQTL [52] with the same model as that used for the eQTL analysis [13], which was based on a linear regression analysis without fitting random effects. Unlike eQTL, the dependent variable was normalized IMF instead of normalized gene expression level. To highlight the IMF-associated SNPs, we set two different significance levels at $p = 1.00E^{-3}$ and $p = 1.39E^{-6}$, respectively.

Abbreviations

ACC: acetyl CoA carboxylase; ASE: allele-specific expression; ATGL: adipose triglyceride lipase; CD36: fatty acid translocase; DEGs: differential expressed genes; eQTL: expression quantitative trait locus; QTG: quantitative trait gene; QTL: quantitative trait locus; QTN: quantitative trait nucleotide; FDR: false discovery rate; GO: Gene Ontology; GSEA: Gene Set Enrichment Analysis; GWAS: genome-wide association study; HDL: high-density lipoproteins; HSL: hormone-sensitive lipase; IMF: Intramuscular fat; KEGG: Kyoto Encyclopedia of Genes and Genomes; q-value: adjusting *p*-value; PPI: protein-protein interaction; RNA-seq: RNA sequencing; TPM: transcripts per million mapped reads

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-021-08141-9.

Additional file 1: Supplementary Table S1. The result of correlation analysis between gene expression level and IMF content

Additional file 2: Supplementary Table S2. Enrichment analysis result of 212 IMF-correlated genes with Metascape online tool

Additional file 3: Supplementary Table S3. The result of Gene set enrichment analysis (GSEA) analysis for all the 13, 450 expressed genes

---

Availability of data and materials

The RNAseq data used in the current study are available in the GEO database of NCBI with accession ID GSE124315 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE124315). The datasets supporting the conclusions of this article are included in the article and its Additional files.

Declarations

Ethics approval and consent to participate

All analyses involving animal samples were conducted according to the care and use guidelines of experimental animals established by the Ministry of Agriculture of China. The current study was approved by the Scientific Ethics Committee of Huazhong Agricultural University, Wuhan, China (the approval number is HZAUSW-2016-010). In addition, we confirmed the statement that the study was conducted in accordance with the ARRIVE guidelines.

Consent for publication

Not applicable.

Competing interests

All authors declare that they have no competing interests.

Author details

¹Key Laboratory of Agricultural Animal Genetics, Breeding and Reproduction, Ministry of Education & College of Animal Science and Technology, Huazhong Agricultural University, Wuhan 430070, China. ²The Cooperative Innovation Center for Sustainable Pig Production, Wuhan 430070, China. ³Key Lab of Swine Genetics and Breeding of Ministry of Agriculture and Rural Affairs, Wuhan 430070, China.

References

1. Fernandez X, Monin G, Talmant A, Mourot J, Lebret B. Influence of intramuscular fat content on the quality of pig meat - 2. Consumer acceptability of m. longissimus lumborum. Meat Sci. 1999;53(1):67–72. https://doi.org/10.1016/S0309-1740(99)00038-8.
2. Fortin A, Robertson WM, Tong AK. The eating quality of Canadian pork and its relationship with intramuscular fat. Meat Sci. 2005;69(2):297–305. https://doi.org/10.1016/j.meatsci.2004.07.011.
3. Larzul C, Lefaucheur L, Ecolan P, Gogue J, Talmant A, Sellier P, et al. Phenotypic and genetic parameters for longissimus muscle fiber characteristics in relation to growth, carcass, and meat quality traits in large white pigs. J Anim Sci. 1997;75(12):3126–37. https://doi.org/10.2527/1997.75123126x.
4. Fernandez A, de Pedro E, Nunez N, Silio L, Garcia-Casco J, Rodriguez C. Genetic parameters for meat and fat quality and carcass composition traits in Iberian pigs. Meat Sci. 2003;64(4):405–10. https://doi.org/10.1016/S0309-1740(02)00207-3.
5. Suzuki K, Irie M, Kadowaki H, Shibata T, Kumagai M, Nishida A. Genetic parameter estimates of meat quality traits in Duroc pigs selected for average daily gain, longissimus muscle area, backfat thickness, and intramuscular fat content. J Anim Sci. 2005;83(9):2058–65. https://doi.org/10.2527/2005.8392058x.
6. Liu J, Damon M, Guitton N, Guisle I, Ecolan P, Vincent A, et al. Differentially-expressed genes in pig longissimus muscles with contrasting levels of fat, as identified by combined transcriptomic, reverse transcription PCR, and proteomic analyses. J Agric Food Chem. 2009;57(9):3808–17. https://doi.org/10.1021/jf8033144.
7. Canovas A, Quintanilla R, Amills M, Pena RN. Muscle transcriptomic profiles in pigs with divergent phenotypes for fatness traits. BMC Genomics. 2010;11(1):372. https://doi.org/10.1186/1471-2164-11-372.
8. Pena RN, Noguera JL, Casellas J, Diaz I, Fernandez AI, Folch JM, et al. Transcriptional analysis of intramuscular fatty acid composition in the longissimus thoracis muscle of Iberian x landrace back-crossed pigs. Anim Genet. 2013;44(6):648–60. https://doi.org/10.1111/age.12066.
9. Cardoso TF, Canovas A, Canela-Xandri O, Gonzalez-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(1):40005. https://doi.org/10.1038/srep40005.
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. https://doi.org/10.1111/age.12518.
11. Ma J, Yang J, Zhou L, Ren J, Liu X, Zhang H, et al. A splice mutation in the PHKG1 gene causes high glycogen content and low meat quality in pig

skeletal muscle. PLoS Genet. 2014;10(10):e1004710. https://doi.org/10.1371/journal.pgen.1004710.

12. Gonzalez-Prendes R, Quintanilla R, Marmol-Sanchez E, Pena RN, Ballester M, Cardoso TF, et al. Comparing the mRNA expression profile and the genetic determinism of intramuscular fat traits in the porcine gluteus medius and longissimus dorsi muscles. BMC Genomics. 2019;20(1):170. https://doi.org/10.1186/s12864-019-5557-9.

13. Liu Y, Liu X, Zheng Z, Ma T, Liu Y, Long H, et al. Genome-wide analysis of expression QTL (eQTL) and allele-specific expression (ASE) in pig muscle identifies candidate genes for meat quality traits. Genet Sel Evol. 2020;52(1):59. https://doi.org/10.1186/s12711-020-00579-x.

14. Hamill RM, Aslan O, Mullen AM, O'Doherty JV, McBryan J, Morris DG, et al. Transcriptome analysis of porcine M semimembranosus divergent in intramuscular fat as a consequence of dietary protein restriction. BMC Genomics. 2013;14(1):453. https://doi.org/10.1186/1471-2164-14-453.

15. Bijland S, Mancini SJ, Salt IP. Role of AMP-activated protein kinase in adipose tissue metabolism and inflammation. Clin Sci (Lond). 2013;124(8):491–507. https://doi.org/10.1042/CS20120536.

16. Bolsoni-Lopes A, Alonso-Vale MI. Lipolysis and lipases in white adipose tissue - an update. Arch Endocrinol Metab. 2015;59(4):335–42. https://doi.org/10.1590/2359-3997000000067.

17. Gaidhu MP, Fediuc S, Anthony NM, So M, Mirpourian M, Perry RL, et al. Prolonged AICAR-induced AMP-kinase activation promotes energy dissipation in white adipocytes: novel mechanisms integrating HSL and ATGL. J Lipid Res. 2009;50(4):704–15. https://doi.org/10.1194/jlr.M800480-JLR200.

18. Marcinko K, Steinberg GR. The role of AMPK in controlling metabolism and mitochondrial biogenesis during exercise. Exp Physiol. 2014;99(12):1581–5. https://doi.org/10.1113/expphysiol.2014.082255.

19. Li Y, Yang P, Zhao L, Chen Y, Zhang X, Zeng S, et al. CD36 plays a negative role in the regulation of lipophagy in hepatocytes through an AMPK-dependent pathway. J Lipid Res. 2019;60(4):844–55. https://doi.org/10.1194/jlr.M090969.

20. Samovski D, Sun J, Pietka T, Gross RW, Eckel RH, Su X, et al. Regulation of AMPK activation by CD36 links fatty acid uptake to beta-oxidation. Diabetes. 2015;64(2):353–9. https://doi.org/10.2337/db14-0582.

21. Poulsen L, Siersbaek M, Mandrup S. PPARs: fatty acid sensors controlling metabolism. Semin Cell Dev Biol. 2012;23(6):631–9. https://doi.org/10.1016/j.semcdb.2012.01.003.

22. Yoon MJ, Lee GY, Chung JJ, Ahn YH, Hong SH, Kim JB. Adiponectin increases fatty acid oxidation in skeletal muscle cells by sequential activation of AMP-activated protein kinase, p38 mitogen-activated protein kinase, and peroxisome proliferator-activated receptor alpha. Diabetes. 2006;55(9):2562–70. https://doi.org/10.2337/db05-1322.

23. Chen J, Dodson MV, Jiang Z. Cellular and molecular comparison of redifferentiation of intramuscular- and visceral-adipocyte derived progeny cells. Int J Biol Sci. 2010;6(1):80–8. https://doi.org/10.7150/ijbs.6.80.

24. Habinowski SA, Witters LA. The effects of AICAR on adipocyte differentiation of 3T3-L1 cells. Biochem Biophys Res Commun. 2001;286(5):852–6. https://susceptibility to cholesterol.2001.5484.

25. Chabowski A, Zendzian-Piotrowska M, Nawrocki A, Gorski J. Not only accumulation, but also saturation status of intramuscular lipids is significantly affected by PPARgamma activation. Acta Physiol (Oxf). 2012;205(1):145–58. https://doi.org/10.1111/j.1748-1716.2011.02380.x.

26. Zhao L, Zou T, Gomez NA, Wang B, Zhu MJ, Du M. Raspberry alleviates obesity-induced inflammation and insulin resistance in skeletal muscle through activation of AMP-activated protein kinase (AMPK) alpha1. Nutr Diabetes. 2018;8(1):39. https://doi.org/10.1038/s41387-018-0049-6.

27. Huang J, Xiong Y, Li T, Zhang L, Zhang Z, Zuo B, et al. Ectopic overexpression of swine PPARgamma2 upregulated adipocyte genes expression and triacylglycerol in skeletal muscle of mice. Transgenic Res. 2012;21(6):1311–8. https://doi.org/10.1007/s11248-012-9615-1.

28. Ros-Freixedes R, Gol S, Pena RN, Tor M, Ibanez-Escriche N, Dekkers JC, et al. Genome-wide association study singles out SCD and LEPR as the two Main loci influencing intramuscular fat content and fatty acid composition in Duroc pigs. PLoS One. 2016;11(3):e0152496. https://doi.org/10.1371/journal.pone.0152496.

29. Uemoto Y, Nakano H, Kikuchi T, Sato S, Ishida M, Shibata T, et al. Fine mapping of porcine SSC14 QTL and SCD gene effects on fatty acid composition and melting point of fat in a Duroc purebred population. Anim Genet. 2012;43(2):225–8. https://doi.org/10.1111/j.1365-2052.2011.02236.x.

30. Wang W, Xue W, Jin B, Zhang X, Ma F, Xu X. Candidate gene expression affects intramuscular fat content and fatty acid composition in pigs. J Appl Genet. 2013;54(1):113–8. https://doi.org/10.1007/s13353-012-0131-z.

31. Henriquez-Rodriguez E, Bosch L, Tor M, Pena RN, Estany J. The effect of SCD and LEPR genetic polymorphisms on fat content and composition is maintained throughout fattening in Duroc pigs. Meat Sci. 2016;121:33–9. https://doi.org/10.1016/j.meatsci.2016.05.012.

32. 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. https://doi.org/10.5713/ajas.16.0158.

33. Fang L, Miller YI. Regulation of lipid rafts, angiogenesis and inflammation by AIBP. Curr Opin Lipidol. 2019;30(3):218–23. https://doi.org/10.1097/MOL.0000000000000596.

34. Fang L, Choi SH, Baek JS, Liu C, Almazan F, Ulrich F, et al. Control of angiogenesis by AIBP-mediated cholesterol efflux. Nature. 2013;498(7452):118–22. https://doi.org/10.1038/nature12166.

35. Ehehalt R, Sparla R, Kulaksiz H, Herrmann T, Fullekrug J, Stremmel W. Uptake of long chain fatty acids is regulated by dynamic interaction of FAT/CD36 with cholesterol/sphingolipid enriched microdomains (lipid rafts). BMC Cell Biol. 2008;9(1):45. https://doi.org/10.1186/1471-2121-9-45.

36. Pohl J, Ring A, Korkmaz U, Ehehalt R, Stremmel W. FAT/CD36-mediated long-chain fatty acid uptake in adipocytes requires plasma membrane rafts. Mol Biol Cell. 2005;16(1):24–31. https://doi.org/10.1091/mbc.e04-07-0616.

37. Kim KS, Kim JS, Park JY, Suh YH, Jou I, Joe EH, et al. DJ-1 associates with lipid rafts by palmitoylation and regulates lipid rafts-dependent endocytosis in astrocytes. Hum Mol Genet. 2013;22(23):4805–17. https://doi.org/10.1093/hmg/ddt332.

38. Xu M, Wu H, Li M, Wen Y, Yu C, Xia L, et al. DJ-1 deficiency protects hepatic steatosis by enhancing fatty acid oxidation in mice. Int J Biol Sci. 2018;14(13):1892–900. https://doi.org/10.7150/ijbs.28620.

39. Paul BT, Tesfay L, Winkler CR, Torti FM, Torti SV. Sideroflexin 4 affects Fe-S cluster biogenesis, iron metabolism, mitochondrial respiration and heme biosynthetic enzymes. Sci Rep. 2019;9(1):19634. https://doi.org/10.1038/s41598-019-55907-z.

40. Carroll WX, Kalupahana NS, Booker SL, Siriwardhana N, Lemieux M, Saxton AM, et al. Angiotensinogen gene silencing reduces markers of lipid accumulation and inflammation in cultured adipocytes. Front Endocrinol (Lausanne). 2013;4:10. https://doi.org/10.3389/fendo.2013.00010.

41. Massiera F, Seydoux J, Geloen A, Quignard-Boulange A, Turban S, Saint-Marc P, et al. Angiotensinogen-deficient mice exhibit impairment of diet-induced weight gain with alteration in adipose tissue development and increased locomotor activity. Endocrinology. 2001;142(12):5220–5. https://doi.org/10.1210/endo.142.12.8556.

42. Baez JM, Barbour SE, Cohen DE. Phosphatidylcholine transfer protein promotes apolipoprotein A-I-mediated lipid efflux in Chinese hamster ovary cells. J Biol Chem. 2002;277(8):6198–206. https://doi.org/10.1074/jbc.M106799200.

43. Baez JM, Tabas I, Cohen DE. Decreased lipid efflux and increased susceptibility to cholesterol-induced apoptosis in macrophages lacking phosphatidylcholine transfer protein. Biochem J. 2005;388(Pt 1):57–63. https://doi.org/10.1042/BJ20041899.

44. Thomas SR, Keller CA, Szyk A, Cannon JR, Laronde-Leblanc NA. Structural insight into the functional mechanism of Nep1/Emg1 N1-specific pseudouridine methyltransferase in ribosome biogenesis. Nucleic Acids Res. 2011;39(6):2445–57. https://doi.org/10.1093/nar/gkq1131.

45. Chen Y, Chen Y, Shi C, Huang Z, Zhang Y, Li S, et al. SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. Gigascience. 2018;7(1):1–6. https://doi.org/10.1093/gigascience/gix120.

46. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nat Methods. 2015;12(4):357–60. https://doi.org/10.1038/nmeth.3317.

47. Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol. 2015;33(3):290–5. https://doi.org/10.1038/nbt.3122.

48. Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000;28(1):27–30. https://doi.org/10.1093/nar/28.1.27.

49. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, et al. Metascape provides a biologist-oriented resource for the analysis of

systems-level datasets. Nat Commun. 2019;10(1):1523. https://doi.org/10.1038/s41467-019-09234-6.

50. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005;102(43):15545–50. https://doi.org/10.1073/pnas.0506580102.

51. von Mering C, Huynen M, Jaeggi D, Schmidt S, Bork P, Snel B. STRING: a database of predicted functional associations between proteins. Nucleic Acids Res. 2003;31(1):258–61. https://doi.org/10.1093/nar/gkg034.

52. Shabalin AA. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. Bioinformatics. 2012;28(10):1353–8. https://doi.org/10.1093/bioinformatics/bts163.

53. Castel SE, Levy-Moonshine A, Mohammadi P, Banks E, Lappalainen T. Tools and best practices for data processing in allelic expression analysis. Genome Biol. 2015;16(1):195. https://doi.org/10.1186/s13059-015-0762-6.

Publisher's Note