Copy Number Variations Contribute to Intramuscular Fat Content Differences By Affecting The Expression of *PELP1* Alternative Splices in Pig

Ligang Wang

Institute of Animal Sciences, Chinese Academy of Agricultural Sciences <https://orcid.org/0000-0002-4376-8373>

Tian Zhang

China Academy of Chinese Medical Sciences

Li Na

Jilin Academy of Agricultural Sciences

Longchao Zhang

Chinese Academy of Agricultural Sciences Institute of Animal Science

Xinhua Hou

Chinese Academy of Agricultural Sciences Institute of Animal Science

Hua Yan

Chinese Academy of Agricultural Sciences Institute of Animal Science

Lixian Wang (✉ iaswlx@263.net)

Chinese Academy of Agricultural Sciences Institute of Animal Science

Research

Keywords: Intramuscular fat, copy number variations, alternative splicing, PELP1, pig

Posted Date: October 18th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-968559/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background

Intramuscular fat (IMF) is a key meat quality trait. Research on the genetic mechanisms of IMF decomposition is valuable for both pork quality improvement and the treatment of obesity and type 2 diabetes. Copy number variations (CNVs) are a type of variant that may influence meat quality.

Results

In this study, a total of 1185 CNV regions (CNVRs) including 393 duplicated CNVRs, 432 deleted CNVRs, and 361 CNVRs with both duplicated and deleted status were identified in a pig F2 resource population using next-generation sequencing data. A genome-wide association study (GWAS) was then performed between CNVs and IMF, and a total of 19 CNVRs were found to be significantly associated with IMF. QTL colocation analysis indicated that 3 of the 19 CNVRs overlapped with known QTLs. RNA-seq and qPCR validation results indicated that CNV150, which is located on the 3'UTR end of the proline, glutamate, and leucine rich protein 1 (*PELP1*) gene, may affect the expression of *PELP1* alternative splices. Sequence alignment and Alphafold2 structure prediction results indicated that the two alternative splices of *PELP1* have a 23 AA sequence variation and a helix-fold structure variation. This region is located in the region of interaction between *PELP1* and other proteins which have been reported to be significantly associated with fat deposition or insulin resistance. We infer that the CNVR may influence IMF content by regulating the alternative splicing of the *PELP1* gene and ultimately affecting the structure of the *PELP1* protein.

Conclusions

In conclusion, we found some CNVRs, especially CNV150, located in *PELP1* that affect IMF. These findings suggest a novel mechanistic approach for meat quality improvement in animals and potential treatment of insulin resistance in human beings.

1. Background

Intramuscular fat (IMF) refers to the total amount of fat located between muscle fibers and as lipid droplets in the muscle cells. In humans, extra IMF deposition has been reported to be associated with type 2 diabetes and insulin resistance [1]. In animals, IMF content directly influences the flavor and juiciness of meat and indirectly influences the tenderness and color of meat [2]. Because pork IMF contains many long-chain polyunsaturated fatty acids, it can also directly affect human health[3].

IMF content is a quantitative trait with low to moderate heritability (0.2–0.4) [4, 5] and is influenced by multiple genes or quantitative trait loci (QTLs). As stated in the PigQTLdb (<http://www.animalgenome.org/cgi-bin/QTLdb/SS/index>) a total of 842 IMF-related QTLs have been reported prior to 23 August 2021 [6]. Although thousands of single-nucleotide polymorphisms (SNPs) have been reported to be associated with IMF in genome-wide association studies (GWAS), most of them only explain a small part of the total genetic variance [4, 7, 8].

Copy number variations (CNVs) are the mutation type with the widest coverage of the genome and are one of the variation types most likely to explain "missing inheritance" beyond the SNP effect. In pigs, there have been several studies revealing CNVs or CNV regions (CNVRs) associated with economic traits. In the research of Rubin et al., CNVs of *KIT* were reported to be associated with coat color [9]. Fowler et al. (2013) found a CNV region related to backfat thickness on chromosome 7 using a pig 60K Beadchip [10]. In the research of Revilla et al. (2017), Chen et al. (2018), Stafuzza et al. (2019), and Krüger et al. (2020), CNVs were reported to be associated with fatty acid composition and growth traits, ear size, number of piglets born alive, and prevalence of porcine endogenous retroviruses [11–14]. In our previous research,

CNVRs have reported associated with IMF content [15]. The previous research used 60K Beadchip data with only 48 CNVs detected.

In this study, we used genome-wide CNV markers identified and genotyped via next-generation sequencing to perform genome-wide association analysis on IMF. Moreover, function analyses of these CNVs were also performed to analyze the potential effects and mechanisms of the CNVs on IMF. We aimed to identify the candidate genes and causal mutation sites of IMF and provide a research basis for pig meat quality breeding.

2. Results

Phenotypic distribution of pig IMF content

For the 592 F2-generation individuals of the Large white pigs × Min pig resource population, the average slaughter weight (240 ± 7 d) was 109.13 kg and the average carcass weight (left side) was 40.38 kg. The average IMF content of these pigs was 2.87%, the maximum was 12.70%, the minimum was 0.73%, the standard deviation was 1.85%, and the coefficient of variation was 0.65.

Porcine genome copy number variation segmentation

After calling and genotyping, all the CNVRs were screened with the silhouette_score > 0.6 in this study. A total of 1185 CNV regions (CNVRs) were identified, including 393 duplicated CNVRs, 432 deleted CNVRs, and 361 CNVRs with both duplicated and deleted status (Figure 1). The length of CNVRs ranged from 2000 bp to 4,196,500 bp with an average length of 30,818 bp. The numbers of copies was relatively evenly distributed on autosomes and basically corresponded to the lengths of the chromosomes.

CNV association analysis

All 592 pigs were used in a genome-wide association study (GWAS) between CNVs and IMF content. The GWAS results are shown in Table 1, and the Manhattan plot of GWAS analysis is shown in Figure 2. Only 1 CNVR significantly associated with IMF could be identified using the general linear model (GLM) method, and 19 CNVRs significantly associated with IMF were identified using the fixed and random model circulating probability unification (FarmCPU) method. The significant CNVR detected by the overlapped two methods was located on chromosome 12 :52194501-52220000. The other CNVRs were located on chromosome 1 (2 CNVRs), chromosome 2 (3 CNVRs), chromosome 3 (1 CNVR), chromosome 4 (1 CNVR), chromosome 7 (2 CNVRs), chromosome 8 (1 CNVR), chromosome 12 (3 CNVRs), chromosome 13 (1 CNVR), chromosome 14 (1 CNVR), chromosome X (2 CNVRs), and chromosome Y (1 CNVR). Among the total 19 CNVRs, 8 CNVRs were deletion CNVRs and the others were duplication CNVRs.

Table 1
Descriptions of the significant CNVRs associated with IMF.

CNVRs	Chromosome	Start	End	Effect	P value	Type	Overlapped gene
CNV150	12	52,194,501	52,220,000	-0.4986684	9.47E-06	Dup	<i>PELP1</i>
CNV11	1	43,146,501	43,151,500	-0.5284976	1.85E-26	Del	-
CNV150	12	52,194,501	52,220,000	-0.4553467	7.08E-25	Dup	<i>PELP1</i>
CNV657	7	79,216,001	79,273,500	0.66226525	1.29E-24	Dup	<i>ENSSSCG00000035754</i>
CNV223	14	2,0385,001	20,388,500	0.77672465	2.35E-24	Dup	-
CNV466	3	25,786,001	25,790,000	0.60166595	1.32E-23	Dup	-
CNV698	8	91,666,001	91,675,500	0.30709026	1.30E-21	Del	-
CNV846	X	75,184,001	75,198,000	0.39222977	3.61E-20	Dup	<i>ENSSSCG00000046526</i>
CNV149	12	49,461,001	49,498,500	-0.2625203	2.53E-17	Dup	<i>SPATA22</i>
CNV385	2	17,955,001	17,967,500	0.34965771	9.45E-16	Dup	-
CNV653	7	78,876,501	78,952,500	-0.3832179	2.58E-15	Dup	<i>SH3BP4</i> , <i>LOC100524322</i> , <i>LOC100524156</i> , <i>ENSSSCG00000044162</i> , <i>R-SSC-381753</i>
CNV771	X	8,710,001	88,350,00	0.43394353	8.49E-13	Dup	-
CNV49	1	236,168,001	236,172,000	-0.2828418	5.89E-12	Del	<i>ENSSSCG00000049310</i>
CNV35	1	174,451,501	174,455,500	-0.3229734	7.50E-12	Del	-
CNV450	2	142,722,001	142,725,000	0.26468096	8.33E-09	Del	-
Dup: duplicate; Del: deletion; <i>PELP1</i> : Proline, Glutamate, and Leucine Rich Protein 1; <i>SPATA22</i> : Spermatogenesis Associated 22; <i>SH3BP4</i> : SH3 Domain Binding Protein 4; <i>FCHO2</i> : FCH And Mu Domain Containing Endocytic Adaptor 2; and <i>ULK4</i> : Unc-51 Like Kinase 4.							

Quality assessment of CNVs using Quantitative real-time PCR (qPCR)

In this study, DNA qPCR validation was performed to confirm the existence of the identified CNVRs. As shown in Figure 3A–3D, all of the selected four CNVRs were validated by DNA qPCR .

Annotation and QTLs co-location of CNVRs

In order to study the function of the CNVRs, we first annotated the CNV coverage area. As shown in Table 1, there were nine known genes (including four olfactory receptor genes) overlapped with six CNVRs, and four novel genes overlapped with four CNVRs. Ten of the 19 CNVRs were located in the intergenic region, without covering any genes. Considering the CNVRs and 842 known IMF-associated QTLs together, 3 of the 19 significant CNVs could be overlapped with at least one of the known QTLs for IMF (Supplementary Table S1).

Significantly related CNV internal RNA-seq

Six individuals with different copy numbers were selected for RNA sequencing to predict the potential effect mechanisms of the CNVRs. The relationships between DNA dosage and RNA expression were then analyzed. The results of the transcript expression levels for the 13 overlapped genes showed that related transcript expression differences could be detected only in the Proline, Glutamate, and Leucine Rich Protein 1 (*PELP1*) genes. The expression of one of the *PELP1* alternative splices (ASs), ENSSSCT00000019597, was significantly different between different individual CNVRs (Table 2).

Table 2
Reads counts of the *PELP1* alternative splices.

Individual	ENSSSCT00000075280 reads count	ENSSSCT00000019507 reads count	Copy numbers
H1	2090	12	Normal
H2	1986	240	Normal
H3	2121	58	Normal
L1	2501	858	Duplicated
L2	2298	436	Duplicated
L3	1984	867	Duplicated

Functional analysis of CNV150

As our research population was constructed using Large white pigs and Min pig F0 individuals, which have different IMF contents (the average IMF content of the Large white is less than 2%, and the average IMF content of Min pigs is more than 4%), we also detected the copy numbers of F0 individuals to confirm the relationship between IMF and CNVR150. From Figure 4, we can see that in Large white pigs, the copy number of CNV150 was multiple, and in Min pigs, the copy number of CNV150 was around 1.

Beyond the RNA-seq data, we also detected the expression of the two *PELP1* alternative splices (ENSSSCT00000075280 and ENSSSCT00000019507) using the qPCR method. As shown in Figure 3E and Figure 3F, the expression of ENSSSCT00000019507 was significantly different between normal-copy-number individuals and duplicated-copy-number individuals. ENSSSCT00000075280 and ENSSSCT00000019507 were coding modulators of the nongenomic activity of the estrogen receptor proteins A0A5G2R420 and F1RFT3.

In order to mine the mechanism of CNV150 in depth, we analyzed the interaction network of *PELP1*. From the protein – protein interaction (PPI) networks (Figure 5), a total of 23 proteins were experimentally determined to interact with *PELP1* (Supplementary Table S2). Among the 23 proteins, the androgen receptor (AR), estrogen receptor (ESR1), glucocorticoid receptor (GR, NR3C1), nuclear receptor subfamily 4 group A member 1(NR4A1), retinoblastoma-associated protein (RB1),

60S ribosomal protein L11 (RPL11), proto-oncogene tyrosine-protein kinase Src (SRC), and signal transducer and activator of transcription 3 (STAT3) have been reported to be related to fat deposition, metabolism, or insulin resistance.

In order to study whether the protein structure variation affected the interaction between PELP1 and its interactive proteins, we aligned the AA of A0A5G2R420 and F1RFT3 and found there was a 23 AA (83–105) difference between these two proteins (Figure 6A). Alpha fold 2 was used to predict the potential structure of these two sequences. In F1RFT3, a helix (about 30 AA) is unfolded as compared to A0A5G2R420 (Figure 6B and Figure 6C). This helix is located between two LXXLL motifs of PELP1.

3. Discussion

The Min is a native Chinese pig breed with an average IMF of > 4.0%, and the Large white is a commercial pig breed with an average IMF of <=2.00%. The Large white × Min F2 separated population is an ideal population in which to investigate the candidate genes or QTLs for IMF. In this study, we first used NGS data of the F2 resource population for CNV calling and genotyping, and then performed CNV-based GWAS for candidate CNV identification. CNV calling using NGS data mainly uses four approaches: paired-end mapping, split-reads, sequence assembly, and read depth (RD) [16]. In this study, we chose to use CNVcaller software [17], which uses the RD method. CNVcaller mitigates the influence of high proportions of gaps and misassembled duplications in the nonhuman reference genome assembly for CNV calling and genotyping and was suitable for our population. In our research, a total of 1185 CNVRs were identified, and this number was smaller than some other pig CNVRs detection research. For example, in the research of Zheng et al., a total of 12,668 CNVRs were detected in 32 Meishan pigs and 29 Duroc pigs [18], and in the research of Wu et al., a total of 18,687 CNVs were identified in Tongcheng and Large white pigs [19]. This may have been caused by the strict standard of CNV definition (silhouette coefficient > 0.6) we used. The qPCR validation results indicated that the selected CNVRs were all real.

In the GWAS analysis, a total of 19 genomic significant CNVRs were identified as being related to IMF. Among the known genes in which significant CNVRs were located, *PELP1* is an ESR coregulator protein [20] and has been reported to be associated with sperm morphology abnormalities in pigs [21]. In the human and great ape *PELP1* gene, duplicated CNV also exists [22, 23]. Homozygous spermatogenesis associated 22 (*SPATA22*) is a sex-related gene associated with infertility and related traits [24]. SH3 domain-binding protein 4 (SH3BP4) is a negative regulator of amino acid-Rag GTPase-mTORC1 signaling and is related to diabetic retinopathy [25, 26]. FCH And Mu Domain Containing Endocytic Adaptor 2 (FCHO2) protein can participate in the early step of clathrin-mediated endocytosis and has lipid-binding activity [27]. Other known genes, namely LOC100524322, LOC100524156, and R-SSC-381753, are olfactory receptors (ORs), and ORs have been reported to be related to IMF or insulin resistance in previous research [28–31].

In order to study the function of the CNVRs located in the intergenic region or in novel genes, we performed QTL co-location analysis, and ultimately found that 3 of the 19 CNVRs were located in the regions of reported IMF-associated QTLs. We infer that these QTLs may affect IMF through structure variation.

As CNVRs usually work through regulation effects or dose effects, we analyzed the RNA expression profiles of some individuals with different CNVR dosages. Interestingly, we found that one of the *PELP1* ASs, named ENSSSCT00000019597, was significantly differently expressed in CNV150-variant individuals. We then validated the differential expression using qPCR and the results were positive. Hence, we inferred that this CNV150 may affect *PELP1* alternative splicing.

In order to confirm the function of CNV150, we analyzed the read depth of CNV 150 in F0-generation individuals and found that the copy number of CNV150 was normal in Min pigs and duplicate in Large white pigs. As shown in Table 1,

this CNVR has a negative effect for IMF, and Min pigs and Large white pigs are high- and low-IMF pigs, respectively. These results were consistent. We then further studied PELP1 and its ASs.

First, we studied whether *PELP1* directly or indirectly affects IMF. In the PPI networks, about half of the proteins had been reported as related to IMF or insulin resistance. Among these genes or proteins, AR and ESR1 can regulate leptin transcript accumulation and protein secretion in adipocytes [32]. NR3C1 transcription factor has been identified as a potential regulator co-localizing within QTLs for fatness and growth traits [33]. NR4A1 can affect insulin resistance and downregulated intramuscular lipid content [34]. RB1 has a direct interaction relationship related to adipogenesis growth [35]. RPL11 has been revealed to play a role in fat storage [36]. SRC and STAT3 can respond to adipogenesis through the TNF- α pathway [37]. We inferred that PELP1 may influence IMF by interacting with other proteins.

In previous research, the interacting regions of PELP1 and ESR, AR, GR, RB, and STAT3 were amino acids 1–400, or LXXLL motifs, amino acids 1–600, or amino acids 1–330 [38]. Hence, we then studied whether the ASs affected the 3D structure of the PELP1 protein and affected the interaction between PELP1 and its interactive proteins. The structures of the two proteins coded by PELP 1 ASs were predicted using AlphaFold2. AlphaFold2 has been used to predict the structures of many difficult protein targets at or near experimental resolution [22]. Our results may have high reliability. The results indicated that, in the variation location of amino acids 83–105, a helix was unfolded in F1RFT3 (coded by ENSSSCT00000019597). In A0A5G2R420 (coded by ENSSSCT00000075280), the structure located in the variation region had very high confidence (predicted LDDT (pLDDT)>90). Moreover, this helix was between two LXXLL motifs (located on 69–74 AA and 111–116 AA). We inferred that the structure changes potentially caused by CNV150 may affect the interaction of PELP1 and its interactive proteins. However, the function and molecular regulatory mechanisms between CNV150 and IMF content require further experimental research, such as gene knockdown/editing, co-immunoprecipitation, and so on.

Conclusions

In this study, a CNV-based GWAS was performed between CNVs and IMF, and a total of 19 CNVRs were found to be significantly associated with IMF. Some CNVRs, such as the three CNVRs overlapped with known QTLs, may be useful candidate markers for IMF selection. CNV150, which was located on the 3'UTR of PELP1, may influence IMF content by regulating the alternative splicing of the PELP1 gene and the structure of PELP1 protein. These findings suggest a novel mechanistic approach for meat quality improvement in animals and potential treatment of insulin resistance in human beings.

Materials And Methods

Animals and phenotype determination

The pigs used in the experiment were all from the Large white pig $\times$ Min F2 resource population, raised in the Changping pig farm of the Institute of Animal Science, Chinese Academy of Agricultural Sciences. All 592 F2 individuals were raised to market age (240 ± 7 days) and slaughtered for commercial purposes. Longissimus dorsi muscles (LDM) were collected for IMF content measurement and DNA sequencing. IMF content was measured using an ether extraction method (Soxtec Avanti 2055 Fat Extraction System; Foss Tecator, Hillerød, Denmark).

Tissue DNA extraction and sequencing

DNA of LDM tissue was extracted using the phenol imitation method. The NANODROP 1000 was used to detect the concentration and quality of DNA (Thermo Scientific, USA). An Illumina Hi-seq2500 was used for paired-end sequencing, and the sequencing depth was 5–7 $\times$.

CNV detection, genotyping, and genomic association analysis

CNVcaller software was used to detect and genotype all individual CNVs with a sliding window of 500 bp [17]. In the step of CNV determination, we used a strict standard of silhouette_score >0.6. When genotyping, each CNV region was divided into no more than three genotypes, and each sample was labeled 1, 2, and 3 to represent the genotype cluster to which the individual belonged.

CNV-based GWAS was performed using the GLM, mixed linear model (MLM), and FarmCPU model approaches with rMVP software [39]. In the GLM method, we used the fixed effects of sex and batches. In the method of MLM, we used PCA as a fixed effect, and in the FarmCPU method, we also used PCA as a fixed effect. The selected threshold for genomic level significance was $-\log_{10}(0.05/N)$, where N was the total number of CNVRs after quality control.

RNA data acquisition

Six RNA-seq datasets, as in our previous research, were used for RNA expression analysis [40]. These data have been submitted to the Genome Sequence Archive with the accession number CRA001645.

Validation of the CNVs and PELP1 RNA sequencing results using qPCR

DNA of five individuals in each CNVR genotype group were used for qPCR amplification for CNVR validation. The primers were designed using the Primer 6 software. The glucagon gene (*GCG*) was used as the single-copy control. RNA extract from five individuals' LDM in each CNVR genotype group was used for expression validation. The glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) gene was used as the housekeeping gene. Copy number and relative expression were calculated using the $2^{-\Delta\Delta CT}$ method [41], where ΔCT is the differential value of target region cycle threshold (CT) and the control region CT. Moreover, $2^{-\Delta\Delta CT}$ stood for the comparison of the ΔCT value of samples with CNV to those without CNV. The PCR cycle was as follows: 3 min at 50°C, 10 min at 95°C, 40 cycles of 15 sec at 95°C, and 1 min at 60°C. The primer sequences are shown in Supplementary Table S3. One-way ANOVA was used to determine the statistical differences between any two groups, followed by Tukey's test for multiple comparisons. $P < 0.05$ was considered to indicate a significant difference.

Annotation of CNVRs

For the combined analysis of CNVRs and QTLs, QTLs were retrieved from the PigQTLdb [6]. Bedtools (v2.27.1) [42] software was used and the 'intersection' command was used: intersectBed -a -b -wa -wb > QTLs.txt. The PPI network contained the top 25 interactions of PELP1 retrieved from the STRING website: (<https://version11.string-db.org/cgi/network.pl?taskId=NzfSlslTYWvr>) [43].

Protein alignment and structure prediction

Consensus sequences and alignment of A0A5G2R420 and F1RFT3 were built using ClustalX in MEGA X [44]; the structures of these two sequences were predicted using alphafold2 [22] and visualized using iCn3D [45].

List Of Abbreviations

$\Delta\Delta CT$: delta delta cycle threshold

ASs: alternative splices;

AR: androgen receptor;

CNVs: copy number variations

CNVRs: copy number variation regions

ESR1: estrogen receptor

FCH02: FCH And Mu Domain Containing Endocytic Adaptor 2

GAPDH: Glyceraldehyde-3-phosphate dehydrogenase

GCG: glucagon

GSA: genome sequence archive

GR, *NR3C1*: glucocorticoid receptor

GWAS: genome-wide association study

IMF: Intramuscular fat

NR4A1: nuclear receptor subfamily 4 group A member 1

PELP1: Proline, Glutamate, and Leucine Rich Protein 1

qPCR: Quantitative real-time polymerase chain reaction

QTLs: quantitative trait loci

QQ: quantile–quantile

RB1: retinoblastoma-associated protein

RD: read depth

RPL 11: 60S ribosomal protein L11

SH3BP4: SH3 Domain Binding Protein 4;

SNPs: single nucleotide polymorphisms

SPATA22: Spermatogenesis Associated 22;

SRC: proto-oncogene tyrosine-protein kinase Src

STAT3: signal transducer and activator of transcription 3

ULK4: Unc-51 Like Kinase 4.

Declarations

Ethics approval and consent to participate

All animals used in this study were handled and kept according to the standard guidelines for experimental animals established by the Ministry of Science and Technology (Beijing, China). All animal experiments in our research were

carried out with the ethical approval (No. IAS2020-109) of the Animal Ethics Committee of the Institute of Animal Science, Chinese Academy of Agricultural Sciences.

Consent for publication

Not applicable.

Availability of data and materials

Some of the sequencing data used in the current study have been submitted to the Genome Sequence Archive, with the accession number CRA001645. The other data presented in this study are available from the corresponding author upon reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was jointly supported by the National Natural Science Foundation of China (Grant No. 31872337, 31501919) and Agricultural Science and Technology Innovation Project (ASTIP-IAS02).

Authors' contributions

Conceptualization, L.W. Ligang Wang and L.W. Lixian Wang.; methodology, L.W. Ligang Wang.; software, T.Z.; validation, T.Z., N.L. and L.Z.; formal analysis, X.H. and H.Y.; writing—original draft preparation, L.W. Ligang Wang.; writing—review and editing, L.W. Ligang Wang and L.W. Lixian Wang.

Acknowledgements

We thank the researchers at our laboratories for their dedication and hard work. We thank MDPI for their professional English language editing services.

References

1. Bergman BC, Perreault L, Hunerdosse DM, Koehler MC, Samek AM, Eckel RH. Intramuscular lipid metabolism in the insulin resistance of smoking. *Diabetes*. 2009;58(10):2220–7.
2. Madeira M, Costa P, Alfaia C, Lopes P, Bessa R, Lemos JC, Prates J. The increased intramuscular fat promoted by dietary lysine restriction in lean but not in fatty pig genotypes improves pork sensory attributes. *J Anim Sci*. 2013;91(7):3177–87.
3. Valsta LM, Tapanainen H, Mannisto S. Meat fats in nutrition. *Meat Sci*. 2005;70(3):525–30.
4. Gao G, Gao N, Li S, Kuang W, Zhu L, Jiang W, Yu W, Guo J, Li Z, Yang C, et al. Genome-Wide Association Study of Meat Quality Traits in a Three-Way Crossbred Commercial Pig Population. *Front Genet*. 2021;12:614087.
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.
6. Hu Z-L, Park CA, Reecy JM. Building a livestock genetic and genomic information knowledgebase through integrative developments of Animal QTLdb and CorrDB. *Nucleic acids research*. 2019;47(D1):D701–10.

7. Xie L, Qin J, Rao L, Tang X, Cui D, Chen L, Xu W, Xiao S, Zhang Z, Huang L. Accurate prediction and genome-wide association analysis of digital intramuscular fat content in longissimus muscle of pigs. *Anim Genet.* 2021;52(5):633–44.
8. Zhuang Z, Ding R, Qiu Y, Wu J, Zhou S, Quan J, Zheng E, Li Z, Wu Z, Yang J. A large-scale genome-wide association analysis reveals QTL and candidate genes for intramuscular fat content in Duroc pigs. *Anim Genet.* 2021;52(4):518–22.
9. Rubin CJ, Megens HJ, Martinez Barrio A, Maqbool K, Sayyab S, Schwochow D, Wang C, Carlborg O, Jern P, Jorgensen CB, et al. Strong signatures of selection in the domestic pig genome. *Proc Natl Acad Sci U S A.* 2012;109(48):19529–36.
10. Fowler KE, Pong-Wong R, Bauer J, Clemente EJ, Reitter CP, Affara NA, Waite S, Walling GA, Griffin DK. Genome wide analysis reveals single nucleotide polymorphisms associated with fatness and putative novel copy number variants in three pig breeds. *BMC Genom.* 2013;14:784.
11. Chen C, Liu C, Xiong X, Fang S, Yang H, Zhang Z, Ren J, Guo Y, Huang L. Copy number variation in the MSRB3 gene enlarges porcine ear size through a mechanism involving miR-584-5p. *Genetics Selection Evolution.* 2018;50(1):1–18.
12. Krüger L, Stillfried M, Prinz C, Schröder V, Neubert LK, Denner J. Copy number and prevalence of porcine endogenous retroviruses (PERVs) in German wild boars. *Viruses.* 2020;12(4):419.
13. Revilla M, Puig-Oliveras A, Castello A, Crespo-Piazuelo D, Paludo E, Fernandez AI, Ballester M, Folch JM. A global analysis of CNVs in swine using whole genome sequence data and association analysis with fatty acid composition and growth traits. *PLoS One.* 2017;12(5):e0177014.
14. Stafuzza NB, Silva RMO, Fragomeni BO, Masuda Y, Huang Y, Gray K, Lourenco DAL. A genome-wide single nucleotide polymorphism and copy number variation analysis for number of piglets born alive. *BMC Genom.* 2019;20(1):321.
15. Wang L, Xu L, Liu X, Zhang T, Li N, Hay el H, Zhang Y, Yan H, Zhao K, Liu GE, et al. Copy number variation-based genome wide association study reveals additional variants contributing to meat quality in Swine. *Sci Rep.* 2015;5:12535.
16. Huang T, Li J, Jia B, Sang H. CNV-MEANN: A Neural Network and Mind Evolutionary Algorithm-Based Detection of Copy Number Variations From Next-Generation Sequencing Data. *Front Genet.* 2021;12:700874.
17. Wang X, Zheng Z, Cai Y, Chen T, Li C, Fu W, Jiang Y. CNVcaller: highly efficient and widely applicable software for detecting copy number variations in large populations. *Gigascience.* 2017;6(12):1–12.
18. Zheng X, Zhao P, Yang K, Ning C, Wang H, Zhou L, Liu J. CNV analysis of Meishan pig by next-generation sequencing and effects of AHR gene CNV on pig reproductive traits. *J Anim Sci Biotechnol.* 2020;11:42.
19. Wu Q, Zhou Y, Wang Y, Zhang Y, Shen Y, Su Q, Gao G, Xu H, Zhou X, Liu B. Whole-genome sequencing reveals breed-differential CNVs between Tongcheng and Large White pigs. *Anim Genet.* 2020;51(6):940–4.
20. De Luca A, Avena P, Sirianni R, Chimento A, Fallo F, Pilon C, Casaburi I, Pezzi V. **Role of Scaffold Protein Proline-, Glutamic Acid-, and Leucine-Rich Protein 1 (PELP1) in the Modulation of Adrenocortical Cancer Cell Growth.** *Cells* 2017, 6(4).
21. Zhao Y, Gao N, Li X, El-Ashram S, Wang Z, Zhu L, Jiang W, Peng X, Zhang C, Chen Y, et al. Identifying candidate genes associated with sperm morphology abnormalities using weighted single-step GWAS in a Duroc boar population. *Theriogenology.* 2020;141:9–15.
22. Tunyasuvunakool K, Adler J, Wu Z, Green T, Zielinski M, Židek A, Bridgland A, Cowie A, Meyer C, Laydon A, et al. Highly accurate protein structure prediction for the human proteome. *Nature.* 2021;596(7873):590–6.
23. Sudmant PH, Huddleston J, Catacchio CR, Malig M, Hillier LW, Baker C, Mohajeri K, Kondova I, Bontrop RE, Persengiev S, et al. Evolution and diversity of copy number variation in the great ape lineage. *Genome Res.* 2013;23(9):1373–82.

24. Wu Y, Li Y, Murtaza G, Zhou J, Jiao Y, Gong C, Hu C, Han Q, Zhang H, Zhang Y, et al. Whole-exome sequencing of consanguineous families with infertile men and women identifies homologous mutations in SPATA22 and MEIOB. *Hum Reprod.* 2021;36(10):2793–804.
25. Kim YM, Stone M, Hwang TH, Kim YG, Dunlevy JR, Griffin TJ, Kim DH. SH3BP4 is a negative regulator of amino acid-Rag GTPase-mTORC1 signaling. *Mol Cell.* 2012;46(6):833–46.
26. Sheu WH, Kuo JZ, Lee IT, Hung YJ, Lee WJ, Tsai HY, Wang JS, Goodarzi MO, Klein R, Klein BE, et al. Genome-wide association study in a Chinese population with diabetic retinopathy. *Hum Mol Genet.* 2013;22(15):3165–73.
27. Latomanski EA, Newton P, Khoo CA, Newton HJ. The Effector Cig57 Hijacks FCHO-Mediated Vesicular Trafficking to Facilitate Intracellular Replication of *Coxiella burnetii*. *PLoS Pathog.* 2016;12(12):e1006101.
28. de Lemos MVA, Peripolli E, Berton MP, Feitosa FLB, Olivieri BF, Stafuzza NB, Tonussi RL, Kluska S, Chiaia HLJ, Mueller L, et al. Association study between copy number variation and beef fatty acid profile of Nellore cattle. *J Appl Genet.* 2018;59(2):203–23.
29. Ying F, Gu H, Xiong Y, Zuo B. Analysis of Differentially Expressed Genes in Gastrocnemius Muscle between DGAT1 Transgenic Mice and Wild-Type Mice. *Biomed Res Int.* 2017;2017:5404682.
30. Soleimanzad H, Montaner M, Ternier G, Lemitre M, Silvestre JS, Kassis N, Giacobini P, Magnan C, Pain F, Gurden H. Obesity in Midlife Hampers Resting and Sensory-Evoked Cerebral Blood Flow in Mice. *Obesity (Silver Spring).* 2021;29(1):150–8.
31. Poessel M, Freiherr J, Wiencke K, Villringer A, Horstmann A. **Insulin Resistance Is Associated with Reduced Food Odor Sensitivity across a Wide Range of Body Weights.** *Nutrients* 2020, 12(8).
32. Jenks MZ, Fairfield HE, Johnson EC, Morrison RF, Muday GK. Sex Steroid Hormones Regulate Leptin Transcript Accumulation and Protein Secretion in 3T3-L1 Cells. *Sci Rep.* 2017;7(1):8232.
33. Puig-Oliveras A, Revilla M, Castelló A, Fernández AI, Folch JM, Ballester M. Expression-based GWAS identifies variants, gene interactions and key regulators affecting intramuscular fatty acid content and composition in porcine meat. *Sci Rep.* 2016;6:31803.
34. Müssig K, Machicao F, Machann J, Schick F, Claussen CD, Stefan N, Fritsche A, Häring HU, Staiger H. No association between variation in the NR4A1 gene locus and metabolic traits in white subjects at increased risk for type 2 diabetes. *BMC Med Genet.* 2010;11:84.
35. Lim D, Lee SH, Kim NK, Cho YM, Chai HH, Seong HH, Kim H. Gene Co-expression Analysis to Characterize Genes Related to Marbling Trait in Hanwoo (Korean) Cattle. *Asian-Australas J Anim Sci.* 2013;26(1):19–29.
36. Liu S, Kim TH, Franklin DA, Zhang Y. Protection against High-Fat-Diet-Induced Obesity in MDM2(C305F) Mice Due to Reduced p53 Activity and Enhanced Energy Expenditure. *Cell Rep.* 2017;18(4):1005–18.
37. Tada A, Islam MA, Kober AH, Fukuyama K, Takagi M, Igata M, Albarracin L, Ikeda-Ohtsubo W, Miyazawa K, Yoda K, et al: **Transcriptome Modifications in the Porcine Intramuscular Adipocytes during Differentiation and Exogenous Stimulation with TNF- α and Serotonin.** *Int J Mol Sci* 2020, 21(2).
38. Girard BJ, Daniel AR, Lange CA, Ostrander JH. PELP1: a review of PELP1 interactions, signaling, and biology. *Mol Cell Endocrinol.* 2014;382(1):642–51.
39. Yin L, Zhang H, Tang Z, Xu J, Yin D, Zhang Z, Yuan X, Zhu M, Zhao S, Li X, et al: **rMVP: A Memory-efficient, Visualization-enhanced, and Parallel-accelerated tool for Genome-Wide Association Study.** *Genomics Proteomics Bioinformatics* 2021.
40. Wang L, Zhou ZY, Zhang T, Zhang L, Hou X, Yan H, Wang L. IRLnc: a novel functional noncoding RNA contributes to intramuscular fat deposition. *BMC Genom.* 2021;22(1):95.
41. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2^{(-Delta Delta C(T))} Method. *Methods.* 2001;25(4):402–8.

42. Quinlan AR. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Curr Protoc Bioinformatics*. 2014;47:11.12.11–34.
43. Szklarczyk D, Gable AL, Lyon D, Junge A, Wyder S, Huerta-Cepas J, Simonovic M, Doncheva NT, Morris JH, Bork P, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*. 2019;47(D1):D607-d613.
44. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol Biol Evol*. 2018;35(6):1547–9.
45. Wang J, Youkharibache P, Zhang D, Lanczycki CJ, Geer RC, Madej T, Phan L, Ward M, Lu S, Marchler GH, et al. iCn3D, a web-based 3D viewer for sharing 1D/2D/3D representations of biomolecular structures. *Bioinformatics*. 2020;36(1):131–5.

Figures

Figure 1

The distribution of pig CNVRs. The outer circle presents the lengths of the chromosomes; the inner circle presents the CNVR distributions. The length of the histogram bars indicates the ratio of different types of CNVRs to the total number of CNVRs.

Figure 2

Circular Manhattan plot and quantile–quantile (QQ) plot of associated CNVs for IMF. (A): Circular Manhattan plot of associated CNVs for IMF using GLM, MLM, and FarmCPU methods, the outer circle is the number of CNVRs in the 1 Mb size region. Significant CNVRs in the same location are marked in line. (B): QQ plot of CNVs associated with IMF using GLM, MLM, and Farm-CPU methods.

Figure 3

Relative expression of validated CNVRs and alternative splices of PELP1 (A): Relative expression of normal and duplicate (Dup) status of CNV150 (B): Relative expression of normal and duplicate (Dup) status of CNV223. (C): Relative expression of normal and duplicate (Dup) status of CNV148. (D): Relative expression of normal and deletion (Del) status of CNV11. (E): Relative expression of normal and duplicate (Dup) status of PELP1 alternative splice 1 (ENSSSCT00000075280). (F): Relative expression of normal and duplicate (Dup) status of PELP1 alternative splice 2 (ENSSSCT00000019507). * represents significant difference (P<0.05).

Figure 4

CNVRs on chromosome 12 in Large white and Min pigs. (A): CNVRs on pig chromosome 12, where the colored lines represent the numbers of CNVRs in the 1 Mb region; (B): Relative copy numbers in the 1 kb sliding windows across the CNV150 peak region for Large white (Green) and Min pigs (Orange).

Figure 5

PPI network of PELP1. This figure displays data obtained from three databases and illustrated using STRING.

Figure 6

Comparison of the sequence and structure of PELP1 alternative-splice proteins. (A): Sequence alignment of the two proteins. A0A5G2R420: protein translated by ENSSSCT00000075280. F1RFT3: protein translated by ENSSSCT00000019507; (B): 3D structure of A0A5G2R420, where the structure marked by a yellow line indicates a helix with very high confidence (predicted LDDT (pLDDT)>90). (C): 3D structure of F1RFT3, where the structure marked by a yellow line indicates an unfolded helix with low confidence (70>pLDDT>50).

Supplementary Files

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

- [Additionalfile1.docx](#)