Identification of a missense variant in the porcine AGPAT gene family that affects intramuscular fat content through whole-genome sequencing

Eduard Molinero
Universitat de Lleida - Agrotecnio-CERCA Center

Ramona N. Pena
Universitat de Lleida - Agrotecnio-CERCA Center

Joan Estany
Universitat de Lleida - Agrotecnio-CERCA Center

Roger Ros-Freixedes (roger.ros@udl.cat)
Universitat de Lleida - Agrotecnio-CERCA Center

Research Article

Keywords: AGPAT5, association study, backfat thickness, fatty acids, intramuscular fat, pig, whole-genome sequencing

Posted Date: February 2nd, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1321611/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Identification of a missense variant in the porcine AGPAT gene family that affects intramuscular fat content through whole-genome sequencing

Eduard Molinero, Ramona N. Pena, Joan Estany, Roger Ros-Freixedes*

Department of Animal Science, University of Lleida-Agrotecnio-CERCA Center, Lleida, Catalonia, Spain.

*Corresponding author: RRF - roger.ros@udl.cat

Abstract

The 1-acylglycerol-3-phosphate O-acyltransferases (AGPAT) are enzymes that catalyze the conversion of lysophosphatidic acid to phosphatidic acid, which is a precursor of triacylglycerol, the main fat reservoir in mammals. In pigs, there are 5 annotated genes in the AGPAT gene family. We used whole-genome sequencing to identify genetic variants in the AGPAT gene family in pigs and then we validated their association with fat content and fatty acid composition. Among the 6,639 variants identified across the AGPAT gene family through whole-genome sequencing of 205 Duroc pigs, we preselected a missense single nucleotide polymorphism in exon 6 of AGPAT5 (rs196952262, A>G) based on minor allele frequency, association analyses, and linkage disequilibrium. The effect of this variant was validated in 1,105 pigs from the same Duroc line genotyped using a high-resolution melt
protocol. The A allele showed a positive additive effect for intramuscular fat (+1.12% ± 0.21, p < 0.001, for gluteus medius and +0.89% ± 0.33, p < 0.01, for longissimus). We also observed significant effects on fatty acid composition that were, at least in part, independent of the increased intramuscular fat content. Accounting for intramuscular fat content, the A allele resulted in more monounsaturated fatty acids (+0.34% ± 0.15, p < 0.05, for longissimus) and a greater monounsaturated/polyunsaturated fatty acids ratio (+0.11 ± 0.04, p < 0.01, for gluteus medius and +0.13 ± 0.05, p < 0.05, for longissimus).

Although further studies would be needed before the causality of the variant on intramuscular fat content and fatty acid composition can be confirmed, this variant can be used as a marker in assisted selection for modulating pig fat deposition and fatty acid content.

Introduction

Intramuscular fat content (IMF) is related to organoleptic attributes and consumer acceptance of pork (Fernandez et al., 1999; Huff-Lonergan et al., 2002; Schwab et al., 2006). In turn, intramuscular fatty acid composition has implications on the nutritional value of pork. Despite that saturated (SFA) and monounsaturated fatty acids (MUFA) provide more favorable organoleptic and technological attributes than polyunsaturated fatty acids (PUFA) (Cameron et al., 2000; Wood et al., 2008; Bekhit et al., 2013; Wood and Enser, 2017), the intake of saturated fatty acids has been associated to increased risk of cardiovascular disease (Christophersen and Haug, 2011; Calder, 2015; Kapoor et al., 2021). Intramuscular fat content and fatty acid composition are especially relevant traits for consumer preferences for high-quality products (Estany et al., 2017) such as premium fresh pork or dry-cured products, and, as a consequence, they have been included in the selection objectives of pig lines aimed at this production. However, selection for IMF has been hindered by its unfavorable positive genetic correlation to backfat thickness (Newcom et al., 2005; Suzuki et al., 2005; Schwab et al., 2010; Ros-
Freixedes et al., 2013), which is typically included in selection objectives of pig breeding programs as an indicator of carcass composition on production efficiency. Similarly, the technical handicaps for routinely measuring fatty acid composition in live selection candidates and the strong correlation structure between fatty acid contents (Ros-Freixedes and Estany, 2014; Zhang et al., 2019) have limited selection for fatty acid composition. To overcome these difficulties, a lot of efforts have been put into finding genetic variants that affect IMF and fatty acid composition (e.g., Puig-Oliveras et al., 2016; Ros-Freixedes et al., 2016; Won et al., 2018; Ding et al., 2019; Pena et al., 2019; Zhang et al., 2021) that can be used for marker-assisted selection.

In mammals, the main fat reservoir is in the form of triacylglycerol. Within skeletal muscle, triacylglycerols are mainly stored in the adipocytes but they can also be stored as droplets in the myocyte cytoplasm (Machann et al., 2004; Gardan et al., 2006). Triacylglycerols are primarily synthesized by sequential esterification of a glycerol backbone mediated first by glycerol 3-phosphate acyltransferase (GPAT), then by 1-acylglycerol-3-phosphate O-acyltransferase (AGPAT), and later by diacylglycerol acyltransferase (DGAT) (Coleman and Lee, 2004) (Figure 1). Each of these enzymes has multiple isoforms that are encoded by a family of genes. In pig, there are 4 annotated GPAT genes, 5 annotated AGPAT genes, and 2 annotated DGAT genes.

A recent study showed that a single nucleotide polymorphism (SNP) in the DGAT2 gene is associated with intramuscular palmitoleic acid content in pigs (Solé et al., 2021). Because the AGPAT genes are involved in the same metabolic pathway of triacylglycerol synthesis, those genes are plausible candidates for underlying the genetic variation of fat content and composition traits. In particular, the AGPAT enzymes catalyze the conversion of lysophosphatidic acid to phosphatidic acid. The phosphatidic group is hydrolyzed by a phosphatidic phosphatase (LPIN and PLPP) to produce diacylglycerol that later the DGAT enzyme takes as a substrate. Phosphatidic acid is not only a precursor of triacylglycerol, but also a precursor of various glycerophospholipids and of signaling
molecules involved in multiple regulatory processes, such as phosphatidylinositol, which is involved in insulin signaling (Coleman and Lee, 2004). Mutations in the AGPAT genes have been associated to body fat mass in humans (AGPAT2; Agarwal et al., 2002), mice (AGPAT1 and AGPAT2; Cortés et al., 2009; Vogel et al., 2011; Agarwal et al., 2017), and buffalo (AGPAT1 and AGPAT6; Xiaoya Ma et al., 2022), and to milk fat content in goat (AGPAT6; He et al., 2011). However, only scarce information can be found on AGPAT genes in pigs. One study in Berkshire pigs described a missense SNP (rs19695226) in the AGPAT5 gene that contributes to meat color, cooking loss and carcass temperature (Park et al., 2017).

Whole-genome sequencing is a powerful tool for variant detection. Whole-genome sequence data include rare and population-specific variants, including causative mutations (Daetwyler et al., 2014; Nicod et al., 2016; Schaid et al., 2018). The objective of this study was to use whole-genome sequencing to identify variants in the AGPAT genes and then to validate the association of preselected variants with carcass and meat quality traits, mainly backfat thickness and IMF content and composition, in pigs.

**Materials and Methods**

**Ethics statement**

Pigs used in this study were raised and slaughtered in commercial units complying with the regulations and good practice guidelines on the protection of animals kept for farming purposes, during transport and slaughter (Royal Decree 37/2014, Spain). Tissue samples used in this study were collected at the slaughterhouse. The Ethical Committee on Animal Experimentation of the University of Lleida approved all experimental procedures.

**Animals and phenotypes**
A total of 1,105 Duroc pigs from 182 sires and 585 dams of the same line were used in this experiment. Animals were raised in 16 batches between 2002 and 2019 following a common protocol for data recording and tissue sampling (Ros-Freixedes et al., 2013; Ros-Freixedes et al., 2016). Pigs from each batch were raised from 75 days until slaughter age (210 days, SD 9.4) under identical conditions. During this time, animals had ad libitum access to commercial feed (Esporc, Riudarenes, Girona, Spain). At 180 days of age (SD 8.8), pigs were weighted and backfat and loin thickness were measured between the third and fourth last ribs by a portable ultrasonic scanner (Piglog 105, Frontmatec, Kolding, Denmark). All pigs were slaughtered in the same abattoir, where carcass weight, carcass backfat and carcass loin thickness at 6 cm off the midline between the third and fourth last ribs were recorded using an on-line ultrasound automatic scanner (AutoFOM, SFK-Technology, Denmark).

After chilling for about 24 h at 2 °C, samples of the muscles *gluteus medius* (n=1,105) and *longissimus* (n=492) were collected, vacuum-packaged, and stored at −20 °C until required. Samples of subcutaneous fat (n=354) were also collected and stored in the same way. The IMF content in *gluteus medius* and *longissimus*, as well as the fatty acid composition of both muscles and of subcutaneous fat, were determined in duplicate by quantitative gas chromatography (Bosch et al., 2009). Fatty acids were expressed as percentages relative to total fatty acid content. The proportion of SFA (C14:0, C16:0, C18:0 and C20:0), MUFA (C16:1n-7, C18:1n-7, C18:1n-9 and C20:1n-9), and PUFA (C18:2n-6, C18:3n-3, C20:2n-6 and C20:4n-6) were calculated.

**Whole-genome sequencing**

Genomic DNA from a subset of 205 pigs was isolated from *gluteus medius* samples using a standard protocol. The DNA samples were submitted to Centre Nacional d’Anàlisi Genòmica (CNAG-CRG, Barcelona, Spain) for sequencing. The short-insert paired-end libraries for the whole-genome sequencing were prepared with PCR free protocol using KAPA HyperPrep kit (Roche, Basel,
Switzerland) with some modifications. In short, in 0.4-1.0 µg of genomic DNA was sheared on a Covaris™ LE220-Plus focused-ultrasonicator (Covaris, Brighton, UK) in order to reach the fragment size of ~400 bp. The fragmented DNA was size-selected for the fragment size of 220-550 bp with AMPure XP beads (Agencourt, Beckman Coulter, Nyon, Switzerland). The size-selected genomic DNA fragments were end-repaired, adenylated and ligated to adaptors with unique dual indexes and unique molecular identifiers that were compatible with the Illumina platform (Integrated DNA Technologies, Leuven, Belgium). The libraries were quality controlled on an Agilent 2100 Bioanalyzer with the DNA 7500 assay (Agilent, Madrid, Spain) for size and quantified by Kapa Library Quantification Kit for Illumina platforms (Roche).

The libraries were sequenced on NovaSeq6000 (Illumina, San Diego, CA, USA) in paired-end mode with a read length of 2x151+17+8 bp following the manufacturer’s protocol for dual indexing. Image analysis, base calling and quality scoring of the run were processed using the manufacturer’s software Real Time Analysis (RTA 3.4.4, Illumina) and followed by generation of FASTQ sequence files. A minimum of 20 Gb of sequencing data was generated per sample.

**Sequencing data processing and identification of variants**

DNA sequence reads were pre-processed using Trimmomatic (Bolger et al., 2014) to remove adapter sequences from the reads. We mapped the reads to the reference genome Sscrofa11.1 (GenBank accession: GCA_000003025.6; Warr et al., 2020) using the BWA-MEM algorithm (Li, 2013). Duplicates were marked with Picard (http://broadinstitute.github.io/picard). SNPs and short insertions and deletions (indels) were identified with the variant caller GATK HaplotypeCaller (GATK 3.8.0) (DePristo et al., 2011; Poplin et al., 2017) using default settings. The average realized sequencing coverage was 7.9x (SD 2.4x). Variant discovery with GATK HaplotypeCaller was performed separately for each individual and then a joint variant set for all the individuals in each population was
obtained by extracting the variant positions from all the individuals. We retained all biallelic variants for further analyses with VCFtools (Danecek et al., 2011).

Variants in the porcine AGPAT genes were retrieved including all variants in the transcription units and 500 bp upstream of the proximal promoter of the gene (for AGPAT1: SSC7, 24,204,902 to 24,213,693 bp; for AGPAT2: AEMK02000682.1, 936,421 to 1,106,715 bp; for AGPAT3: SSC13, 206,750,603 to 206,906,353 bp; for AGPAT4: SSC1, 6,772,006 to 6,883,555 bp; and for AGPAT5: SSC15, 37,800,616 to 37,853,472 bp). Note that the AGPAT2 gene is located in an unplaced scaffold in the current genome assembly. Although the GPAT4 was previously named AGPAT6, we did not include it in our study because it is no longer considered a member of the AGPAT family and its ortholog showed no acylglycerol acyltransferase activity in humans (Chen et al., 2008).

Preselection of candidate variants for fat traits

To preselect those variants with greater evidence of association with the studied traits among all called variants, we performed an association study on the 205 sequenced pigs. To do so, we used a single-marker regression approach based on a univariate linear mixed model that accounted for the genomic relationship matrix. We used the GEMMA 0.96 software (Zhou and Stephens, 2012) to fit the following model:

\[ y = Wb + x_j \alpha_j + Zu + e \]

where \( y \) is the vector of phenotypic values (backfat thickness, IMF, or fatty acid composition); \( W \) is the incidence matrix for fixed effects; \( b \) is the vector of the fixed effects; \( x_j \) is the vector of genotypes of the \( j^{th} \) SNP; \( \alpha_j \) is the allele substitution effect of the \( j^{th} \) SNP; \( Z \) is the incidence matrix for the random individual polygenic effects; \( u \) is the vector of the random individual polygenic effects; and \( e \) is the vector of residual terms. The fixed effects included the intercept, the fattening batch (14 levels), the
genotypes of two polymorphisms with major impact on fat content and fatty acid composition (rs709596309 for LEPR and rs80912566 for SCD; 3 genotypes each) (Ros-Freixedes et al., 2016), and the slaughter age as a covariate. Even though DGAT2 is involved in the same metabolic pathway, we did not include it in the model since no interaction was observed. The random individual polygenic effects and residual terms were assumed to follow normal distributions $\mathbf{u} \sim N(0, \mathbf{K}\sigma_u^2)$ and $\mathbf{e} \sim N(0, \mathbf{I}\sigma_e^2)$, respectively, where $\mathbf{K}$ is the genomic relationship matrix, $\sigma_u^2$ is the additive genetic variance, $\mathbf{I}$ is an identity matrix, and $\sigma_e^2$ is the residual variance. We focused the analyses on the variants with minor allele frequency equal or greater than 0.20 so that all genotypes were sufficiently represented in the 205 pigs. Those variants with p-value equal or lower than 0.001 for any of the tested traits were prioritized as candidate variants. Among those, we prioritized candidate variants in promoters, 5’ and 3’ untranslated regions (UTR) and exons based on their predicted functional annotation using VEP (McLaren et al., 2016). Candidate variants were also examined for linkage disequilibrium to avoid redundancy using PLINK 1.9 (Chang et al., 2015). Based on these criteria, a single variant was preselected for further validation.

Genotyping of the candidate variant

The SNP rs196952262 (A>G) in the AGPAT5 gene (SSC15 at 37,843,344 bp) was genotyped in 1,105 pigs using 5’-GTCCCTTCGAAAGCCACTGT-3’ as forward primer and 5’-CACCAAGAATAAAGGCCAACCCA-3’ as reverse primer. Amplifications were performed by real-time PCR (QuantStudio3, Applied Biosystems, Thermo Scientific, Waltham, MA, USA) with High-Resolution Melt analysis (Luminaris Colour HRM Master Mix, Thermo Scientific) using 20 ng of genomic DNA and 0.4 µM of each primer in 5 µL final volume reaction. Thermocycling conditions were 95 °C for 10 min, and 40 cycles of 95 °C for 15 sec, 60 °C for 1 min, followed by a high-resolution melting curve starting with a denaturation at 95 °C for 15 sec, annealing at 60 °C for 1 min and a slow ramp at 0.015 °C/sec up to 95 °C. High Resolution Melt software v3.1 (Applied Biosystems, Thermo
Scientific) was used for the analysis of melting data and sample genotyping. All pigs were also genotyped for SNPs in genes SCD (rs80912566, C>T, on SSC14) and LEPR (rs709596309, C>T, on SSC6) following the protocols described in (Estany et al., 2014; Solé et al., 2022) and (Ros-Freixedes et al., 2016), respectively. In a subset of 807 pigs, the genotype for DGAT2 (ss7315407085, G>A, on SSC9) was also available (Solé et al., 2021).

Validation of the AGPAT5 variant

The effect of the AGPAT5 rs196952262 SNP genotype on production traits (body weight, carcass weight, backfat thickness, loin thickness, and IMF) and fatty acid composition was estimated using a model with the same fixed effects as above (with 16 batches) plus the genotype for AGPAT5 (GG, AG and AA). The effect of the genotype was tested using the F-statistic. Multiple pairwise comparisons among AGPAT5 genotypes were tested with the Tukey HSD test. The additive and dominant effects of the AGPAT5 SNP were also estimated by replacing the genotype with two covariates coded as (-1, 0, 1) and (0, 1, 0) for the GG, AG and AA genotypes, respectively. The IMF content was added as a covariate to test whether the effect of the AGPAT5 SNP on fatty acid composition was due to changes in fat content. Finally, we also tested the interaction of the AGPAT5 SNP with LEPR and SCD, as well as with DGAT2, by adding this latter SNP to previous models. All the analyses were performed using the statistical package JMP Pro 15 (SAS Institute Inc., Cary, NC, USA).

Results

Preselection of candidate variants for fat traits

Our analysis of the genetic variation of the AGPAT gene family revealed that AGPAT5 contained the strongest candidate variants for fat content and fatty acid composition traits. We found large differences in the number of variants located in the transcription unit of the genes of the AGPAT family (Table 1).
In total, 6,639 variants were detected across the 205 pigs, 2,220 of which had minor allele frequency equal or greater than 0.20 (referred to here as ‘common’ variants). Of these, 171 were prioritized as candidate variants because they showed associations with p-value ≤ 0.001 for at least one of the tested traits. Most of these candidate variants (166 out of 210 common variants) were located on *AGPAT5*, only 1 (out of 120 common variants) was located on *AGPAT2*, and 4 (out of 442 common variants) on *AGPAT4*. The five candidate variants in *AGPAT2* and *AGPAT4* were intronic.

Of the 210 common variants detected in *AGPAT5*, only 11 were in the promoter, 5’ and 3’ untranslated and exonic regions of the gene (Table 2). Three of these variants were located in exons and were predicted to be missense mutations. One of these (SNP3) was predicted to be missense for both annotated transcripts for the gene, while the other two were missense for only one of the two transcripts. SNP3 was also the missense variant that showed the highest association with the tested traits (p ≤ 0.001). Regarding the other 8 variants, one was located in the promoter region, one in the 5’-UTR and 6 (3 SNPs and 3 indels) in the 3’-UTR. Linkage disequilibrium of the candidate variants showed that all markers were in moderate linkage disequilibrium (r² > 0.7) with SNP3 (Table 2). Taking into account these results, together with the fact that SNP3 (rs196952262) was previously reported in Berkshire pigs as associated to meat quality traits (Park *et al.*, 2017), we preselected this variant as a tag SNP for further validation.

**Validation of the AGPAT5 variant**

The effect of the *AGPAT5* SNP rs196952262 on fat content was validated using data from 1,105 pigs (Table 3). The *AGPAT5* SNP was associated with production traits at 180 d of age, with the A allele showing a positive additive effect on backfat thickness (+0.60 mm ± 0.27, p < 0.05) and a negative effect on loin thickness (−0.85 mm ± 0.35, p < 0.05). These effects were less clearly detected when measured on carcass around 4 weeks later and only loin thickness showed a similar trend (−0.96 mm
The A allele also was positively associated with IMF, both in *gluteus medius* (+1.12% ± 0.52, p < 0.10) and in *longissimus* (+0.89% ± 0.31, p < 0.001) and in *longissimus* (+0.89% ± 0.31, p < 0.001).

Differences in intramuscular fatty acid composition were also detected across AGPAT5 genotypes. The A allele had a consistent effect in both muscles (Tables 4 and 5). The A allele had a positive additive effect on MUFA (+0.35% ± 0.13, p < 0.01, in *gluteus medius* and +0.50% ± 0.16, p < 0.01, in *longissimus*), mainly due to its effect on C18:1n-9. Also, the A allele showed a negative effect on PUFA (–0.45% ± 0.13, p < 0.001, in *gluteus medius*, and –0.54% ± 0.18, p < 0.01, in *longissimus*), with decreased values for C18:2n-6, C18:3n-3 and C20:4n-6. As a result, the A allele showed a positive effect on the MUFA/PUFA ratio (0.22 ± 0.04, p < 0.001, in *gluteus medius*, and 0.34 ± 0.09, p < 0.001, in *longissimus*). Since the effects of the A allele on MUFA and PUFA counteract each other, the AGPAT5 genotype did not influence the SFA/(MUFA+PUFA) ratio. We observed no differences in fatty acid composition for subcutaneous fat (Supplementary Table S1).

With the exception of MUFA and C18:1n-9, the effects of the A allele on fatty acid composition were mostly due to changes in fat content, since they were not detected when adjusted for IMF (Supplementary Tables S2 and S3). The effects of the A allele on MUFA, which was driven mainly by its effect on C18:1n-9, was in part independent to IMF, although this was more evident in *longissimus* (+0.34% ± 0.15, p < 0.05, for MUFA and +0.27% ± 0.15, p < 0.10, for C18:1n-9) than in *gluteus medius* (+0.21% ± 0.12, p < 0.10, for MUFA and +0.16% ± 0.10, p = 0.11, for C18:1n-9). As a result, the A allele displayed a positive additive effect on the MUFA/PUFA ratio in both *gluteus medius* (+0.11 ± 0.04, p < 0.01) and *longissimus* (+0.13 ± 0.05, p < 0.05) even after adjusting for IMF.

We detected a significant interaction between AGPAT5 and LEPR for IMF (Supplementary Table S4). Although the effect of AGPAT5 on IMF showed the same trend within each LEPR genotype, the additive effect of the A allele of AGPAT5 was larger in the LEPR-TT pigs (+1.47% ± 0.72, p < 0.05),
which are fatter, than in the *LEPR*-CC or CT pigs (+0.99% ± 0.34, p < 0.01). We also confirmed the
effect of *AGPAT5* on IMF in the *DGAT2*-GG pigs (Supplementary Table S5), although statistical
evidence for its effect within the other *DGAT2* genotypes was limited to assess the interaction between
these two genes. We detected some significant interactions for fatty acid composition between
*AGPAT5* and genes *LEPR* (Supplementary Table S4) and *SCD* (Supplementary Table S6), but they did
not show clear patterns that were biologically meaningful. There were no interactions between
*AGPAT5* and *DGAT2* for fatty acid composition (Supplementary Table S5).

**Discussion**

In this study, we identified variants in the family of *AGPAT* genes using whole-genome sequencing
and preselected a candidate variant for fat content and composition traits. Using a large number of
individuals, we validated the effect of the rs196952262 SNP as a tag SNP for the haplotype in *AGPAT5*.
In the following we will discuss: (1) the suitability of whole-genome sequencing for preselecting
candidate variants in the *AGPAT* gene family associated to traits of interest, and (2) the effect of the
*AGPAT5* SNP on fat content and composition.

**Preselection of candidate variants in the *AGPAT* gene family**

Whole-genome sequencing is a powerful tool for detecting large numbers of variants that could be
associated with complex traits. However, the identification of causal variants remains challenging
because the high density of variants that are in high linkage disequilibrium hinders the disentangling
of the causal variant. We found the strongest evidence of association with fat content and composition
traits for *AGPAT5*. In contrast, we did not find any candidate variant in *AGPAT1* or *AGPAT3*, and the
few candidate variants in *AGPAT2* and *AGPAT4* were discarded because they were located in introns.
A typical criterion for prioritizing candidate variants is to limit the search to coding and promoter
regions, since the prediction of the potential effects of variants located in non-coding regions from
DNA sequence is not straightforward (Johnsson and Jungnickel, 2021). However, in some instances non-coding variants, which may have regulatory functions, have been proposed as candidate variants (Van Laere et al., 2003; Ryan et al., 2012; Solé et al., 2021), while variants in coding regions have often been found to have a small impact on complex traits (Koufariotis et al., 2018; Xiang et al., 2019). Despite that mutations in AGPAT2 that affected triacylglycerol synthesis and storage have been described in humans (Agarwal et al., 2002), we did not detect any orthologous polymorphism in pigs with significant association to the studied traits.

The results for AGPAT5 are consistent with the fact that this gene is the most expressed of the AGPAT gene family in skeletal muscle in pigs. Using data from a previous RNA-Seq experiment (Solé et al., 2022), we found that AGPAT5 and AGPAT2 are the most expressed genes in muscle semimembranosus in pig (Supplementary Figure S1). Those results are also in line with previous gene expression data reported in pig skeletal muscle (Freeman et al., 2012). In contrast, expression analysis in humans showed that AGPAT1 and AGPAT2 were highly expressed in adipocytes and skeletal muscle (Agarwal et al., 2002; Prasad, Garg and Agarwal, 2011), while in mice AGPAT1 and AGPAT2 and AGPAT4 showed the highest overall expression (Vergnes et al., 2006). Although our expression data were insufficient to detect any potential expression differences between AGPAT5 genotypes (Supplementary Figure S2), the fact that in pigs AGPAT5 is the most expressed gene of the AGPAT family, in contrast to human and mice, supports the contribution of this isoform in the conversion of lysophosphatidic into phosphatidic acid in pig skeletal muscle, and hence, in triacylglycerol synthesis. If these results were confirmed, they could help understand differences in the metabolic pathways of fat deposition of pigs compared to other mammals.

We preselected the rs196952262 SNP as a tag variant for the AGPAT5 gene based on several criteria. Coincidentally, the same SNP was previously reported to be associated with meat quality in Berkshire pigs (Park et al., 2017) for traits such as meat color, cooking loss and carcass temperature, albeit not
for backfat thickness. In particular, these authors found that the A allele was associated to a lighter pork color (greater CIE L* parameter). Although these authors did not analyze IMF, the lighter color associated to the A allele is compatible with greater levels of IMF (Schwab et al., 2006; Suárez-Mesa et al., 2021), which adds yet another layer of evidence from an independent population in support of this variant.

**Effect of the AGPAT5 SNP on fat content and composition**

We validated the association of rs196952262 SNP from *AGPAT5* on fat content and composition. This association had not been identified in previous analyses in Duroc. We observed that the A allele had a positive additive effect on backfat thickness at 180 d and on IMF in both muscles. Alterations on the gene activity led to lower triacylglycerol synthesis, and therefore to a reduced fat accumulation (Cortés et al., 2009). Although we expected to observe the same additive effect on carcass backfat thickness, we were not able to detect this effect on backfat thickness, likely because of differences in age and measuring methods for live and carcass backfat thickness, but collectively results indicate an effect on overall carcass fatness. The rs196952262 SNP explained 1.2% and 1.4% of the genetic variance for backfat and loin thickness at 180 days, respectively, but much less at 210 days (0.03% and 0.6%, respectively). The rs196952262 SNP had a higher impact on IMF than on backfat thickness. The SNP explained 2.3% and 2.5% of the genetic variance for IMF in *gluteus medius* and *longissimus*, respectively. Moreover, we observed differences in fatty acid composition that were specific of intramuscular fat and did not affect subcutaneous fat, which might further indicate a lower impact on backfat thickness. Since the A allele also had a negative effect on loin thickness and thus, attention should be paid to any unfavorable correlated responses in carcass lean content.

The A allele segregated at high frequency in Iberian pigs (0.81, n=18), which is consistent with the body fatness that characterizes this breed, and at intermediate frequency in much leaner breeds such as
Pietrain (0.47, n=28) and a Large White × Landrace crossbred (0.50, n=49). Despite this, it segregated at lower frequencies in other heavily marbled breeds such as Duroc, both the population studied here (0.24) and in two other Duroc populations (0.15, n=26, and 0.25, n=20), and Berkshire (0.21; Park et al., 2017). The lower frequency in Duroc pigs indicated that selection assisted by this marker could increase the population average IMF in *gluteus medius* by up to +1.7% if the A allele was fixed.

The rs196952262 SNP was also associated with intramuscular fatty acid composition. The effects that we observed for *AGPAT5* on fatty acid composition were consistent with the faster accumulation of SFA and MUFA relative to PUFA as fat reserves grow (De Smet et al., 2004; Ros-Freixedes et al., 2016; Zhang et al., 2019). However, we found some evidence of an effect of *AGPAT5* on fatty acid composition that was, at least in part, independent to its effect on fat content. In particular, *AGPAT5* showed an IMF-independent effect on MUFA, mainly driven by C18:1n-9. The rs196952262 SNP explained 0.5% of the genetic variance for MUFA in *gluteus medius* and 1.0% in *longissimus*. In humans, functional studies of the AGPAT proteins showed that both AGPAT3 and AGPAT5 had a higher affinity for C18:1n-9 as a substrate, which could explain our results (Prasad et al., 2011). As a consequence of this, we found significant differences for the MUFA/PUFA ratio between genotypes and the SNP could be used as a marker for selecting pork with a higher monounsaturated fatty acid profile.

The effect of *AGPAT5* on IMF may become more noticeable in fatter pigs. Thus, *AGPAT5* may show significant interactions with genes that affect overall fatness, such as *LEPR*. In turn, this can produce changes on fatty acid composition, to the extent that composition indirectly reflects changes in fat content. However, there were no clear interactions between *AGPAT5* and *SCD* despite that both genes affect MUFA. Even though the *AGPAT5* gene intervenes in the same metabolic pathway as *DGAT2*, in a previous study we found no effect of *DGAT2* on IMF (Solé et al., 2021) and therefore the interaction of these two genes for IMF seems unlikely. In contrast, *DGAT2* has a specific effect on
intramuscular C16:1n-7 (Solé et al., 2021) but we did not detect any effect of \textit{AGPAT5} on this fatty acid.

In conclusion, our analysis of the genetic variation in the sequence of the genes of the \textit{AGPAT} family revealed that \textit{AGPAT5} contained the strongest candidate variant for fat content and fatty acid composition traits. The A allele of the rs196952262 variant increases IMF and MUFA but it negatively affects loin thickness. The effect of \textit{AGPAT5} on IMF is more noticeable in fatter pigs and, therefore, \textit{AGPAT5} could interact with genes that affect overall fatness, such as \textit{LEPR}. Although further studies would be needed before the causality of the variant can be confirmed, this variant can be used as a selection marker for modulating pig fat deposition and fatty acid composition of pork.

Data Availability Statement

The original contributions presented in the study are included in the article and supplementary files. Further inquiries can be directed to the corresponding author.

Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author Contributions

JE and RRF designed the study; EM and RNP processed the samples; EM and RRF performed the statistical analyses; EM and RRF wrote the first draft; RNP and JE contributed to the interpretation of the results and provided comments on the manuscript. All authors read and approved the final manuscript.
Funding

This research was supported by the Spanish Ministry of Science, Innovation & Universities and the EU Regional Development Funds (grant RTI2018-101346-B-I00). EM is recipient of a UdL-Santander Predoc scholarship.

Acknowledgments

We acknowledge the personnel at Selección Batallé for their cooperation for the recording of on-farm data and sample collection. We gratefully acknowledge Pilar Sopeña from the Animal Breeding group, University of Lleida, for laboratory assistance.

References


Figure 1. Triacylglycerol synthesis pathway. GPAT: glycerol-3-phosphate acyltransferase; AGPAT: 1-acyl-glycerol-3-phosphate O-acyltransferase; LPIN: lipin; PLPP: phospholipid phosphatase; DGAT: diacylglycerol acyltransferase; and P:\ phosphate.
### Table 1. Number of total called, common and candidate variants in the gene transcription unit plus 500-bp upstream of the transcription start sites of the pig *AGPAT* gene family.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Total called variants</th>
<th>Common variants</th>
<th>Candidate variants</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGPAT1</td>
<td>55</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>AGPAT2</td>
<td>1469</td>
<td>120</td>
<td>1</td>
</tr>
<tr>
<td>AGPAT3</td>
<td>2045</td>
<td>1448</td>
<td>0</td>
</tr>
<tr>
<td>AGPAT4</td>
<td>2722</td>
<td>442</td>
<td>4</td>
</tr>
<tr>
<td>AGPAT5</td>
<td>348</td>
<td>210</td>
<td>166</td>
</tr>
<tr>
<td>Total</td>
<td>6639</td>
<td>2220</td>
<td>171</td>
</tr>
</tbody>
</table>

Common variants are variants with a minor allele frequency ≥ 0.2. Candidate variants are variants that showed an association with at least one trait (p ≤ 0.001).
<table>
<thead>
<tr>
<th>Variant</th>
<th>Position (bp)</th>
<th>Location</th>
<th>Major/minor allele</th>
<th>MAF⁠a</th>
<th>Association study⁠b</th>
<th>Predicted consequence⁠c</th>
<th>Linkage disequilibrium⁠d</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNP1</td>
<td>37,800,993</td>
<td>Promoter</td>
<td>C/T</td>
<td>0.266</td>
<td>2.75</td>
<td>-</td>
<td>0.85</td>
</tr>
<tr>
<td>SNP2</td>
<td>37,801,250</td>
<td>5'-UTR</td>
<td>A/C</td>
<td>0.263</td>
<td>4.23</td>
<td>-</td>
<td>0.79</td>
</tr>
<tr>
<td>SNP3</td>
<td>37,843,344</td>
<td>Exon 6</td>
<td>G/A</td>
<td>0.259</td>
<td>3.38</td>
<td>missense (2/2)</td>
<td>1.00</td>
</tr>
<tr>
<td>SNP4</td>
<td>37,850,594</td>
<td>Exon 8</td>
<td>G/C</td>
<td>0.263</td>
<td>2.88</td>
<td>missense (1/2)</td>
<td>0.77</td>
</tr>
<tr>
<td>SNP5</td>
<td>37,850,596</td>
<td>Exon 8</td>
<td>G/T</td>
<td>0.263</td>
<td>2.88</td>
<td>missense (1/2)</td>
<td>0.77</td>
</tr>
<tr>
<td>INDEL6</td>
<td>37,851,301</td>
<td>3'-UTR</td>
<td>G/GAAAC</td>
<td>0.278</td>
<td>3.99</td>
<td>-</td>
<td>0.76</td>
</tr>
<tr>
<td>SNP7</td>
<td>37,852,012</td>
<td>3'-UTR</td>
<td>C/A</td>
<td>0.281</td>
<td>3.61</td>
<td>-</td>
<td>0.70</td>
</tr>
<tr>
<td>SNP8</td>
<td>37,852,233</td>
<td>3'-UTR</td>
<td>C/T</td>
<td>0.231</td>
<td>3.26</td>
<td>-</td>
<td>0.70</td>
</tr>
<tr>
<td>SNP9</td>
<td>37,852,886</td>
<td>3'-UTR</td>
<td>C/T</td>
<td>0.278</td>
<td>4.30</td>
<td>-</td>
<td>0.76</td>
</tr>
<tr>
<td>INDEL10</td>
<td>37,853,314</td>
<td>3'-UTR</td>
<td>G/GT</td>
<td>0.276</td>
<td>3.74</td>
<td>-</td>
<td>0.80</td>
</tr>
<tr>
<td>INDEL11</td>
<td>37,853,443</td>
<td>3'-UTR</td>
<td>G/GA</td>
<td>0.272</td>
<td>3.13</td>
<td>-</td>
<td>0.78</td>
</tr>
</tbody>
</table>

⁠aMAF: minor allele frequency.

⁠bP-value, expressed as –log(p-value), of the most significant association of the variant with any of the studied traits.

⁠cPredicted consequence on the angiogene transcripts. In parentheses, the number of transcripts affected over the total number of transcripts.

⁠dLinkage disequilibrium between the variant and SNP3, measured as the squared coefficient of correlation between alleles (r²).
Table 3. Least-square means (± standard errors) for live and carcass weight and composition by *AGPAT5* (rs196952262, A>G) genotype.

<table>
<thead>
<tr>
<th>Trait</th>
<th>AGPAT5 genotype</th>
<th>Additive effect ¹</th>
<th>Dominant effect</th>
<th>p-value</th>
<th>d</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AA (n=71)</td>
<td>AG (n=390)</td>
<td>GG (n=644)</td>
<td>a</td>
<td>p-value</td>
<td>d</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Live measurements (180 d)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Body weight, kg</td>
<td>110.5±1.5</td>
<td>110.4±0.6</td>
<td>109.5±0.5</td>
<td>0.51±0.78</td>
<td>0.51</td>
<td>0.38±0.95</td>
</tr>
<tr>
<td>Backfat thickness, mm</td>
<td>20.5±0.5</td>
<td>19.7±0.2</td>
<td>19.3±0.2</td>
<td>0.60±0.27</td>
<td>0.02</td>
<td>-0.20±0.32</td>
</tr>
<tr>
<td>Loin thickness, mm</td>
<td>43.2±0.7</td>
<td>45.0±0.3</td>
<td>44.9±0.2</td>
<td>-0.85±0.35</td>
<td>0.02</td>
<td>0.98±0.43</td>
</tr>
<tr>
<td>Carcass measurements (210 d)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Carcass weight, kg</td>
<td>96.9±1.2</td>
<td>98.0±0.52</td>
<td>96.8±0.4</td>
<td>0.01±0.63</td>
<td>0.98</td>
<td>1.15±0.78</td>
</tr>
<tr>
<td>Backfat thickness, mm</td>
<td>24.0±0.4</td>
<td>24.3±0.18</td>
<td>23.8±0.1</td>
<td>0.09±0.22</td>
<td>0.66</td>
<td>0.36±0.27</td>
</tr>
<tr>
<td>Loin thickness, mm</td>
<td>42.2±1.0</td>
<td>42.8±0.43</td>
<td>44.1±0.3</td>
<td>-0.96±0.52</td>
<td>0.07</td>
<td>-0.32±0.64</td>
</tr>
<tr>
<td>Lean content, %</td>
<td>40.4±0.7</td>
<td>40.1±0.3</td>
<td>40.9±0.2</td>
<td>-0.29±0.36</td>
<td>0.43</td>
<td>0.57±0.45</td>
</tr>
<tr>
<td>Intramuscular fat, % dry matter</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M. gluteus medius</td>
<td>20.42±0.59</td>
<td>18.85±0.26</td>
<td>18.18±0.20</td>
<td>1.12±0.31</td>
<td>&lt;0.001</td>
<td>-0.45±0.39</td>
</tr>
<tr>
<td>M. longissimus ²</td>
<td>15.65±0.64</td>
<td>14.45±0.31</td>
<td>13.86±0.26</td>
<td>0.89±0.33</td>
<td>0.007</td>
<td>-0.31±0.41</td>
</tr>
</tbody>
</table>

¹ Additive allele substitution of G by A.
² Sample size was 38, 187 and 267 for AA, AG, and GG, respectively.
Bold font indicates statistical significance (p ≤ 0.05).

*ab* Within trait, means with different superscripts indicate statistically significant differences (p ≤ 0.05).
Table 4. Least-square means (± standard errors) for fatty acid composition in muscle gluteus medius by AGPAT5 (rs196952262, A>G) genotype.

<table>
<thead>
<tr>
<th>Trait</th>
<th>AGPAT5 genotype</th>
<th>Additive effect&lt;sup&gt;1&lt;/sup&gt;</th>
<th>Dominant effect</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AA (n=71)</td>
<td>AG (n=390)</td>
<td>GG (n=644)</td>
</tr>
<tr>
<td>Fatty acid, %</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SFA</td>
<td>37.92±0.24</td>
<td>37.83±0.11</td>
<td>37.73±0.08</td>
</tr>
<tr>
<td>C16:0</td>
<td>24.39±0.15</td>
<td>24.43±0.07</td>
<td>24.31±0.05</td>
</tr>
<tr>
<td>C18:0</td>
<td>11.73±0.11</td>
<td>11.56±0.05</td>
<td>11.59±0.04</td>
</tr>
<tr>
<td>MUFA</td>
<td>50.37±0.25&lt;sup&gt;a&lt;/sup&gt;</td>
<td>49.75±0.11&lt;sup&gt;b&lt;/sup&gt;</td>
<td>49.66±0.09&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>C16:1n-7</td>
<td>3.70±0.07</td>
<td>3.72±0.03</td>
<td>3.66±0.02</td>
</tr>
<tr>
<td>C18:1n-7</td>
<td>4.26±0.04</td>
<td>4.30±0.02</td>
<td>4.29±0.02</td>
</tr>
<tr>
<td>C18:1n-9</td>
<td>42.02±0.19&lt;sup&gt;a&lt;/sup&gt;</td>
<td>41.66±0.10&lt;sup&gt;ab&lt;/sup&gt;</td>
<td>41.50±0.09&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>PUFA</td>
<td>11.71±0.24&lt;sup&gt;a&lt;/sup&gt;</td>
<td>12.41±0.11&lt;sup&gt;b&lt;/sup&gt;</td>
<td>12.61±0.08&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>C18:2n-6</td>
<td>9.43±0.19&lt;sup&gt;a&lt;/sup&gt;</td>
<td>9.89±0.08&lt;sup&gt;ab&lt;/sup&gt;</td>
<td>10.03±0.06&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>C18:3n-3</td>
<td>0.56±0.01</td>
<td>0.57±0.01</td>
<td>0.58±0.01</td>
</tr>
<tr>
<td>C20:4n-6</td>
<td>1.20±0.07&lt;sup&gt;a&lt;/sup&gt;</td>
<td>1.37±0.03&lt;sup&gt;b&lt;/sup&gt;</td>
<td>1.46±0.02&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
<tr>
<td>Fatty acid ratio</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MUFA/SFA</td>
<td>1.35±0.02</td>
<td>1.33±0.01</td>
<td>1.34±0.01</td>
</tr>
<tr>
<td>MUFA/PUFA</td>
<td>4.57±0.09&lt;sup&gt;a&lt;/sup&gt;</td>
<td>4.23±0.04&lt;sup&gt;b&lt;/sup&gt;</td>
<td>4.14±0.03&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>SFA/(MUFA+PUFA) (x10)</td>
<td>6.16±0.06</td>
<td>6.14±0.03</td>
<td>6.11±0.02</td>
</tr>
</tbody>
</table>

SFA: saturated fatty acids, MUFA: monounsaturated fatty acids, PUFA: polyunsaturated fatty acids, C16:0: palmitic acid, C18:0: stearic acid, C16:1n-7: palmitoleic acid, C18:1n-7: vaccenic acid, C18:1n-9: oleic acid, C18:2n-6: linoleic acid, C18:3n-3: linolenic acid, and C20:4n-6: arachidonic acid.

<sup>1</sup> Additive allele substitution of G by A.

Bold font indicates statistical significance (p ≤ 0.05).

<sup>a</sup>-<sup>c</sup> Within trait, means with different superscripts indicate statistically significant differences (p ≤ 0.05).
Table 5. Least-square means (± standard errors) for fatty acid composition in muscle *longissimus* by *AGPAT5* (rs196952262, A>G) genotype

<table>
<thead>
<tr>
<th>Trait</th>
<th>AGPAT5 genotype</th>
<th>Additive effect</th>
<th>Dominant effect</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>AA (n=38)</td>
<td>AG (n=187)</td>
<td>GG (n=267)</td>
</tr>
<tr>
<td>Fatty acid, %</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SFA</td>
<td>39.56±0.34</td>
<td>39.51±0.16</td>
<td>39.48±0.14</td>
</tr>
<tr>
<td>C16:0</td>
<td>25.40±0.20</td>
<td>25.29±0.10</td>
<td>25.29±0.18</td>
</tr>
<tr>
<td>C18:0</td>
<td>12.46±0.17</td>
<td>12.45±0.08</td>
<td>12.41±0.07</td>
</tr>
<tr>
<td>MUFA</td>
<td>51.22±0.32</td>
<td>50.50±0.16</td>
<td>50.22±0.13</td>
</tr>
<tr>
<td>C16:1n-7</td>
<td>4.00±0.09</td>
<td>3.92±0.04</td>
<td>3.97±0.03</td>
</tr>
<tr>
<td>C18:1n-7</td>
<td>4.34±0.06</td>
<td>4.33±0.03</td>
<td>4.34±0.03</td>
</tr>
<tr>
<td>C18:1n-9</td>
<td>42.07±0.29</td>
<td>41.73±0.15</td>
<td>41.30±0.14</td>
</tr>
<tr>
<td>PUFA</td>
<td>9.21±0.34</td>
<td>9.99±0.17</td>
<td>10.30±0.14</td>
</tr>
<tr>
<td>C18:2n-6</td>
<td>7.18±0.24</td>
<td>7.73±0.12</td>
<td>7.93±0.10</td>
</tr>
<tr>
<td>C18:3n-3</td>
<td>0.34±0.01</td>
<td>0.37±0.01</td>
<td>0.37±0.01</td>
</tr>
<tr>
<td>C20:4n-6</td>
<td>1.31±0.09</td>
<td>1.46±0.05</td>
<td>1.57±0.04</td>
</tr>
<tr>
<td>Fatty acid ratio</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MUFA/SFA</td>
<td>1.31±0.02</td>
<td>1.29±0.01</td>
<td>1.29±0.01</td>
</tr>
<tr>
<td>MUFA/PUFA</td>
<td>5.78±0.18</td>
<td>5.26±0.09</td>
<td>5.11±0.07</td>
</tr>
<tr>
<td>SFA/(MUFA+PUFA) (x10)</td>
<td>6.59±0.09</td>
<td>6.57±0.04</td>
<td>6.57±0.04</td>
</tr>
</tbody>
</table>

SFA: saturated fatty acids, MUFA: monounsaturated fatty acids, PUFA: polyunsaturated fatty acids, C16:0: palmitic acid, C18:0: stearic acid, C16:1n-7: palmitoleic acid, C18:1n-7: vaccenic acid, C18:1n-9: oleic acid, C18:2n-6: linoleic acid, C18:3n-3: linolenic acid, and C20:4n-6: arachidonic acid.

1 Additive allele substitution of G by A.

Bold font indicates statistical significance (p ≤ 0.05).

a,b Within trait, means with different superscripts indicate statistically significant differences (p ≤ 0.05).
Figure 1

Triacylglycerol synthesis pathway. GPAT: glycerol-3-phosphate acyltransferase; AGPAT: 1-acyl-glycerol-3-phosphate O-acyltransferase; LPIN: lipin; PLPP: phospholipid phosphatase; DGAT: diacylglycerol acyltransferase; and Pi: phosphate.
Supplementary Files

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

- AGPAT5frontiersEMG20220202preprintSUPPpdf