Genomics 115 (2023) 110589

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Assessing the relationship between the in silico predicted consequences of 97 missense mutations mapping to 68 genes related to lipid metabolism and their association with porcine fatness traits

Rayner González-Prendes [a,*], Martijn F.L. Derks [a], Martien A.M. Groenen [a], Raquel Quintanilla [c], Marcel Amills [b,d]

[a] Animal Breeding and Genomics, Wageningen University & Research, Droevendaalsesteeg 1, 6708 PB Wageningen, the Netherlands
[b] Centre for Research in Agricultural Genomics (CRAG), CSIC-IRTA-UAB-UB, Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain
[c] Animal Breeding and Genetics Program, Institute of Agrifood Research and Technology (IRTA), Torre Marimon, 08140 Caldes de Montbui, Spain
[d] Departament de Ciència Animal i dels Aliments, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain

ARTICLE INFO

Original content: Assessing the relationship between the predicted deleteriousness of 97 missense mutations in 68 lipid genes and phenotypic variation of complex fatness traits in pigs (Original data)
Genotype and Phenotype data for association analyses of miRNA SNPs with gluteus medius and longissimus dorsi skeletal muscle phenotype data (Reference data)

Keywords:
Missense
Polymorphism,pig
Lipid
Functional
Prediction

ABSTRACT

In general, the relationship between the predicted functional consequences of missense mutations mapping to genes known to be involved in human diseases and the severity of disease manifestations is weak. In this study, we tested in pigs whether missense single nucleotide polymorphisms (SNPs), predicted to have consequences on the function of genes related to lipid metabolism are associated with lipid phenotypes. Association analysis demonstrated that nine out of 72 nominally associated SNPs were classified as "highly" or "very highly consistent" in silico-predicted functional mutations and did not show association with lipid traits expected to be affected by inactivation of the corresponding gene. Although the lack of endophenotypes and the limited sample size of certain genotypic classes might have limited to some extent the reach of the current study, our data indicate that present-day bioinformatic tools have a modest ability to predict the impact of missense mutations on complex phenotypes.

1. Introduction

Missense variants are a source of phenotypic variation by altering protein structure and activity. Because of this, many in silico methods have been implemented to predict the functional consequences of specific amino acid substitutions [1]. Computational methods devised to predict such consequences mostly rely on approaches based on sequence conservation, structural analysis, sequence and structure information, and meta-prediction i.e. predictors integrating data from multiple sources [1].

The relationship between the predicted effects of missense polymorphisms and their observable clinical and functional consequences has been explored in several studies mostly focused on monogenic human diseases. For instance, Tchernitchko et al. [2] reported that in silico predictions, with PolyPhen and SIFT, of the functional consequences of non-synonymous variants in the hemoglobin and glucose-6-phosphate dehydrogenase genes correlated weakly with the phenotypic manifestation of anemia and hemolytic disorders, respectively. In another study, Dorfman et al. [3] predicted the deleteriousness of missense polymorphisms located in the cystic fibrosis transmembrane conductance regulator (*CFTR*) gene with three in silico tools (PANTHER, SIFT and PolyPhen). Unfortunately, none of the three tools was able to accurately differentiate mutations producing cystic fibrosis from mutations found in individuals with related disorders or no disease [3], evidencing that clinically relevant prediction by coding variant classifiers was unreliable [4]. In mouse, only 4 out of 30 missense mutations

* Corresponding author at: Animal Breeding and Genomics, Wageningen University & Research, Droevendaalsesteeg 1, 6708 PB Wageningen, the Netherlands.
E-mail address: rayner.gonzalezprendes@wur.nl (R. González-Prendes).

https://doi.org/10.1016/j.ygeno.2023.110589
Received 20 August 2022; Received in revised form 29 January 2023; Accepted 21 February 2023
Available online 24 February 2023

identified in immune genes and anticipated to be deleterious had a phenotypic impact on target lymphocyte populations, thus revealing a strong discrepancy between the predicted and observed effects of these missense mutations [5]. More recently, the use of endophenotypes (e.g. enzyme catalytic activity or protein stability) resulted in an overall improvement of the predictive power of bioinformatic coding variant classifiers [4,6].

In farm animals, the relationship between the predicted functional consequences of missense mutations and their observable impact on the phenotypic variation of complex traits has not been systematically investigated. In the current work, we have predicted the functional consequences of 97 porcine missense single nucleotide polymorphisms (SNPs) mapping to 68 genes involved in lipid metabolism by using four in silico tools. Next, we have investigated in a population of 316 Duroc pigs the association of these missense mutations with 48 lipid traits, paying special attention to mutations that are robustly predicted to have functional consequences by at least three tools. Our expectation was that the phenotypic consequences of mutations consistently predicted to cause alterations in gene function should resemble those observed in relevant genetic models.

2. Materials and methods

2.1. Generation of a Duroc resource population and phenotype recording

In 2003, we generated a population of 350 Duroc pigs (Lipgen population) by mating five boars with 400 sows and keeping only one piglet from each litter [7]. After a number of setbacks encountered during the generation of the experimental population (litters without weaned males, mortality and illness in fattening period, etc.), the analyzed population consisted of 350 males born on 3 farms, belonging to 5 half-sib families and distributed in 4 fattening batches [7]. Piglets were castrated after weaning and moved to the IRTA pig experimental farm in Monells (Girona, Spain) and kept under intensive conditions [7]. During the first period of fattening (up to 90 kg of live weight, around 150 days of age) barrows were fed a standard diet (2450 kcal/kg) with 18% protein, 3.8% fiber, 7.0% fat, 1.0% lysine, and 0.3% methionine ad libitum [7]. In the last period of fattening (30–40 days before slaughter), pigs were fed ad libitum a standard diet (2375 kcal/kg) with 15.9% protein, 4.5% fiber, 5.2% fat, 0.7% lysine, and 0.2% methionine [7]. When pigs reached a live weight of ~122 kg (approximately 190 days of age), they were slaughtered in a commercial slaughterhouse following current Spanish legislation (https://www.boe.es/buscar/doc.php?id=BOE-A-1995-3942). Tissue samples from the *gluteus medius* (GM) and *longissimus dorsi* (LD) skeletal muscles were obtained after pigs were slaughtered [8]. All experimental procedures were approved by the Ethical Committee of IRTA.

In the Lipgen population, we recorded 48 lipid traits including: eight serum lipid concentrations measured in 45 and 190 days-old pigs by following the protocols reported by Gallardo et al. [7] and Manunza et al. [9], eight carcass traits described by Eusebi et al. [10], and 32 intramuscular fat content and composition traits measured in the *gluteus medius* and *longissimus dorsis* muscles as described by Gonzalez-Prendes et al. [11]. A list of the 48 traits used in the current experiment is shown in Tables S1 and S2.

2.2. Genome sequencing of five Duroc boars

We have sequenced the genomes of the five Duroc boars that founded the Lipgen population. Our goal was to characterize the missense variation segregating in the lipid metabolism genes of pigs from the Lipgen population. Total genomic DNA was purified from blood samples [12] and sequenced at the Centre Nacional d'Anàlisi Genòmica (CNAG, Barcelona, Spain). The synthesis of paired-end multiplex libraries was carried out with the KAPA PE Library Preparation kit (Kapa Biosystems, Wilmington, MA) in accordance with the instructions of the manufacturer. Libraries were loaded to Illumina flow-cells from a HiSeq2000 instrument for cluster generation to yield 150-bp paired-end reads following the Illumina protocol. Base calling and quality control analyses were carried out with the Illumina RTA sequence analysis pipeline following the guidelines of the manufacturer. Quality-checked filtered reads were mapped to the *Sus scrofa* genome version 11.1 [13] using the Burrows-Wheeler Aligner (BWA) [14].

2.3. Retrieval of missense SNPs mapping to genes involved in lipid metabolism and in silico prediction of their effects

Aligned reads in BAM files were sorted and indexed using SAMtools v0.1.19 [15] and variant calling was performed in a multi-sample calling with the Genome Analysis Toolkit or GTKA [16]. Single nucleotide polymorphisms were annotated with the SnpEff tool [17] (http://pingola.github.io/SnpEff/). For downstream analyses, we only retained missense SNPs mapping to lipid metabolism genes that displayed a heterozygous genotype in at least one of the five boars and with a quality score ≥ 30. Genes involved in lipid metabolism were identified through a systematic search in the LIPID MAPS Proteome Database [18] in 2020 and these are listed in Table S3. Protein sequences from selected genes related to lipid metabolism were extracted from the Ensembl database (Sscrofa 11.1), version 101 [19].

The predicted functional effect of each missense SNPs on the corresponding protein was assessed with 4 different bioinformatic tools: Sorting Intolerant From Tolerant (SIFT) [20], MutPred2 [21], Protein Variation Effect Analyzer (PROVEAN) [22], and Align GVGD [23,24]. In the case of the SIFT software, we just retrieved the automated predictions made by this tool as displayed in the Ensembl release 106 (Sscrofa 11.1), while in the case of the remaining tools all predictions were made manually. The SIFT software infers the probability that a mutation has functional effects by aligning the affected protein with other related proteins and assessing the evolutionary conservation of the position where the missense substitution occurred as well as the type of amino acid substitution [20,25]. According to the SIFT manual, amino acid substitutions with normalized probabilities <0.05 are predicted to be deleterious. Similarly, the PROVEAN algorithm uses an alignment-based score measuring the change in sequence similarity of a query sequence to a set of related protein sequences, before and after the introduction of an amino acid variation to the query sequence, in order to predict the damaging effect of such amino acid substitution [26]. When a score is equal to or below -2.5, the protein variant is predicted to have a deleterious effect. The Align GVGD tool [23] employs two parameters, the Grantham variation (GV) and Grantham deviation (GD), which are based on the concept of the Grantham difference [27], to classify missense mutations. Deleterious mutations are discerned from their neutral counterparts by taking into account both biophysical and evolutionary conservation criteria. Align GVGD prediction classes form a spectrum (C0, C15, C25, C35, C45, C55, C65) with C65 most likely to interfere with function and C0 least likely (here, we have considered that a mutation is deleterious when it belongs to class C65). Finally, the MutPred2 software relies on a machine learning-based method trained on a set of 53,180 pathogenic and 206,946 putatively neutral variants which takes into account specific aspects of protein structure and function to infer variant impact on a probabilistic framework [21]. Scores calculated with MutPred2 range between 0 and 1, and a score threshold equal or above 0.50 suggests deleteriousness.

In our study, missense SNPs predicted to have functional consequences by the four tools were considered as "very highly consistent" (*vhc*) functional mutations, while those predicted to be functional by three tools were defined as "highly consistent" (*hc*) functional mutations. We considered that mutations that are predicted to be functional by just two or less tools do not yield conclusive in silico evidence of having functional consequences. The degree of prediction consistency between each pair of tools was assessed by calculating the total number of concordant predictions divided by the sum of the concordant and

discordant predictions.

2.4. Genotyping of 97 missense SNPs mapping to genes related to lipid metabolism

We decided to perform the genotyping and association analysis of 97 missense mutations mapping to porcine genes involved in lipid metabolism. The suitability of missense mutations to be genotyped with a TaqMan Open Array multiplex assay was assessed by submitting SNPs, as well as their flanking sequences (60 nucleotides upstream and downstream), to the Custom TaqMan Assay Design Tool website (https://www.thermofisher.com/order/custom-genomic-products/tools/cadt/). From the 350 Duroc pigs that formed part of the Lipgen population in 2003, only 316 had available DNA samples to be genotyped with TaqMan assays. The genotyping of 97 missense SNPs in 316 Duroc pigs was carried out at the Servei Veterinari de Genètica Molecular of the Universitat Autònoma de Barcelona (http://sct.uab.cat/svgm/en) with the aid of a 12 K Flex QuantStudio equipment and by following the instructions of the manufacturer. Genotypes were visualized with the TAQMAN GENOTYPER software v.1.3 (Applied Biosystems, Foster City, CA).

2.5. SNP association analyses with lipid traits

Association analyses between missense SNPs and lipid phenotypes were carried out with the Genome-wide Efficient Mixed-Model Association (GEMMA) software [28]. The statistical model implemented in GEMMA to estimate the effects of missense SNPs on lipid phenotypes was as follows:

$$y = \mathbf{W}\alpha + \mathbf{x}\,\delta + \mathbf{u} + \boldsymbol{\varepsilon}$$

where $\boldsymbol{y}$ is the vector of lipid phenotypes for all individuals; $\mathbf{W}$ is a matrix of fixed effects (batch of fattening, with 4 categories), a column of 1 s, and a covariate that depends on the trait: (1) IMF content in GM, for fatty acid (FA) related traits measured in the GM muscle, (2) IMF content in LD, for FA related traits measured in the LD muscle, (3) backfat thickness, for IMF content measured in GM and LD, (4) live weight at slaughter, for CHOL, HDL, LDL, BFT34R, BFTLR, LE%, CW, HW, BFT34RPS, FHT and LD34R, and age at slaughter for TRIG; $\boldsymbol{\alpha}$ is a vector of the corresponding coefficients that include the intercept, the batch effects and the regression coefficient on the covariate; $\mathbf{x}$ is a vector of marker genotypes in each individual; $\boldsymbol{\delta}$ is the effect size of the marker (allele substitution effect); $\mathbf{u}$ is a vector of random individual genetic effects with a n-dimensional multivariate normal distribution $\mathbf{u} \sim N(\mathbf{0}, \lambda\,\tau^{-1}\,\mathbf{K})$, being τ^{-1} the variance of the residual error, λ is the ratio between the two variance components and $\mathbf{K}$ a known relatedness matrix derived from SNP genotypes; and $\boldsymbol{\varepsilon}$ is the vector of errors. Multiple testing was corrected with the false discovery rate approach reported by Benjamini and Hochberg [29].

3. Results

3.1. Detection of missense polymorphisms mapping to porcine genes related to lipid metabolism and in silico prediction of their functional effects

Whole-genome sequencing of the five boars resulted in a 37.67 to 46.6 × coverage, with >98% of the genome covered by at least 10 reads in all five samples. Moreover, about 78% of the reads were uniquely mapped. By using the SnpEff tool [17], 10,002,757 SNPs and 2,867,142 indels were annotated. From 66,572 SNPs mapping to 768 genes involved in lipid metabolism, 279 were classified as missense polymorphisms. Based on biological and technical criteria, we selected a subset of 97 missense SNPs (Table S4) located in 68 lipid metabolism genes to be genotyped in 316 Duroc pigs.

Visual inspection of the in silico predictions of the functional consequences of the 97 missense SNPs revealed the existence of five *vhc* functional mutations mapping to the *CAV2* (p.S108C), *DHCR24* (p.P302S), *LIPC* (p.P212L), *LRP4* (p.P1080L), and *PARK2* (p.R161G) genes (Table 1). Moreover, we identified seven *hc* functional mutations located in the *ACSM3* (p.I190T), *FABP3* (p.G25S and T104M), *JMJD1C* (p.S2018F), *MC4R* (p.D298N), *MLXIPL* (p.G683R) and *ZNF648* (p.S343W) genes (Table 1). Thirty-one and 11 missense SNPs were predicted to be functional by one or two in silico tools respectively, while 43 SNPs were predicted to be neutral by all four tools. With the PROVEAN, MutPred2 and SIFT software we predicted 15, 17 and 16 missense mutations to have functional effects respectively, while Align GVGD predicted 48 mutations to have such effects. Indeed, this software shows the lowest concordance with the remaining ones (Table 2, Fig. 1). Noteworthy, PROVEAN and SIFT display the highest level of consistency in their predictions (Table 2, Fig. 1).

3.2. Performance of a gene-centric association analysis between missense SNPs in genes related to lipid metabolism and phenotypic variation of lipid traits recorded in Duroc pigs

We genotyped, in the Lipgen population, the full set of 97 missense mutations and not only those that are strong candidates to have functional effects (Table S4). We did this to infer whether associations with lipid traits are more numerous or more significant in the group of mutations consistently predicted to be functional vs the group represented by those that do not yield conclusive in silico evidence of having functional effects.

The gene-centric association analysis between 97 missense SNPs and 48 lipid traits revealed that 65 missense SNPs displayed nominally significant associations (raw *P*-value <0.05) with intramuscular fat composition phenotypes (Table S5). With regard to the eight carcass traits and the six serum lipid traits, 23 and 12 SNPs, showed significant associations at the nominal level, respectively (Table S6 and Table S7). When using a *q*-value <0.05 as a threshold of significance, only the association between polymorphism p.G25S in the *FABP3* gene and GM C17:0 content remained significant (Table 3). When a more lenient threshold of statistical significance (*q*-value <0.10) was employed, 10 associations remained significant after correction for multiple testing (Table 3). From the set of missense polymorphisms associated with lipid traits (*q*-value <0.10), one mutation was predicted to be neutral by the 4 tools (p.I32L in *GALNT2*), four were predicted to be functional by just one tool (p.T485M in *AACS*, p.G77D in *PARK2*, p.S392C in *STAT5A* and p.G90R in *ZNF648*), and two were *hc* functional mutations i.e. p.G25S and T104M in the *FABP3* gene. The number of associations found between missense SNPs and lipid traits and their statistical significance did not show any evident difference between groups of mutations that differ in their predicted severity (Figs. 2 and 3).

Regarding the *vhc* functional mutations, only three showed associations with FA composition traits that were significant at the nominal level i.e. S108C in the *CAV2* gene, p.P302S in the *DHCR24* gene and p.P1080L in the *LRP4* gene (Table 4). No significant associations between the genotypes of *vhc* functional mutations and serum lipid concentrations or carcass traits were detected. When considering the set of *hc* functional mutations, the two missense polymorphisms in the *FABP3* gene (p.G25S and p.T104M) were the ones which showed the largest number of nominally significant associations with a broad array of intramuscular fat composition and carcass traits (Table 5, Fig. S1). A few additional nominally significant associations were also detected in the gene-centric association analysis i.e. p.I190T (*ACSM3* gene) and LD C22:6 content, p.S2018F (*JMJD1C* gene) and GM C18:3 content and p.LD34R; p.S343W (*ZNF648* gene) and LD C20:1 and GM C20:3 contents, and p.G683R (*MLXIPL* gene) and GM C17:0 and GM CHOL contents (Table 5).

For several genes containing either *vhc* or *hc* mutations, the phenotypes of knockout mice have been reported in the literature (see

Table 1
Missense mutations in porcine genes related to lipid metabolism classified as functional by three (highly consistent functional mutations) or four (very highly consistent functional mutations - shown in bold) in silico prediction tools.

Gene	SNP ID[1]	SSC[2]	Position (Mb)	Substitution	SIFT[3]	MutPred2[4]	Provean[5]	Align GVGD[6]
ACSM3	rs328465555	3	25,262,463	p.I190T	0	Neutral	−3.388	89.28 (Class C65)
CAV2	**rs332492536**	**18**	**29,706,153**	**p.S108C**	**0**	**0.734**	**−4.361**	**111.67 (Class C65)**
DHCR24	**rs328825271**	**6**	**157,502,221**	**p.P302S**	**0**	**0.757**	**−6.212**	**73.35 (Class C65)**
FABP3	rs326530011	17	16,028,772	p.G25S	0.01	0.906	−5.203	Neutral (Class C55)
	rs341327293	17	16,029,010	p.T104M	0.01	Neutral	−3.919	81.04 (Class C65)
JMJD1C	rs335786128	14	66,665,158	p.S2018F	0	Neutral	−2.894	154.81 (Class C65)
LIPC	**rs712720116**	**1**	**113,456,385**	**p.P212L**	**0.01**	**0.579**	**−9.320**	**97.78 (Class C65)**
LRP4	**rs328789611**	**2**	**15,651,335**	**p.P1080L**	**0.01**	**0.517**	**−4.246**	**97.78 (Class C65)**
MC4R	rs81219178	1	160,773,437	p.D298N	0.01	0.829	−4.330	Neutral (Class C15)
MLXIPL	rs788027215	3	10,900,920	p.G683R	0	Neutral	−2.951	125.13 (Class C65)
PARK2	**rs341517389**	**1**	**5,922,601**	**p.R161G**	**0.01**	**0.580**	**−4.474**	**125.13 (Class C65)**
ZNF648	rs341953239	9	123,508,330	p.S343W	0	Neutral	−3.846	176.58 (Class C65)

[1] SNPs were annotated as missense variants in the Ensembl database (Sscrofa 11.1) version 106 and Biomart portal (http://www.ensembl.org/biomart/).
[2] SSC: porcine chromosome.
[3] Positions with normalized probabilities <0.05 are predicted to be deleterious.
[4] The MutPred2 score ranges between 0 and 1, with a score threshold of 0.50 suggesting pathogenicity. However, a threshold of 0.68 yields a false positive rate of 10% and that of 0.80 yields a false positive rate of 5%.
[5] When the PROVEAN score is equal to or below −2.5, the protein variant is predicted to have a deleterious effect.
[6] The prediction classes form a spectrum (C0, C15, C25, C35, C45, C55, C65) with C65 most likely to interfere with function and C0 least likely. Here, we have considered that a mutation is deleterious when it belongs to class C65.

Table 2
Concordance[1] of in silico predictions about the functional consequences of missense mutations in porcine genes related to lipid metabolism.

Tool	SIFT	Provean	MutPred2	Align GVGD
SIFT	–	0.87	0.77	0.28
Provean		–	0.84	0.30
MutPred2			–	0.32
Align GVGD				–

[1] Ratio between the number of concordant predictions divided by the sum of concordant plus discordant predictions. Low-confidence variants from Aling GVGD were treated as neutral.

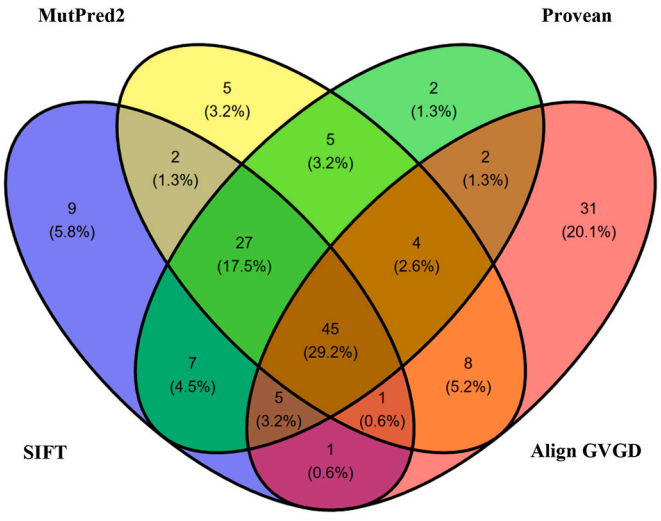

Fig. 1. Venn diagram showing the concordance of the in silico predictions made with SIFT, PROVEAN, MutPred2 and Align GVGD about the functional consequences of 97 missense SNPs mapping to 68 porcine genes related to lipid metabolism. Predictions from two tools are concordant when they concur in the classification of a given mutation as functional or non-functional. For this figure, non-deleterious SNPs were considered neutral.

Discussion for more details). Individuals with homozygous genotypes for mutations inactivating the *DHCR24, PARK2*, *JMJD1C*, *MLXIPL* and *LIPC* genes are expected to show alterations in serum lipid concentrations, and those homozygous for mutations suppressing *LIPC* or *MC4R* function are predicted to display changes in phenotypes related with obesity. In Figs. S2 to S7, we display the genotypic means for serum lipid and/or obesity traits recorded in pigs harbouring *vhc* or *hc* mutations in any of these six genes. These traits were selected according to the expected phenotypic consequences of gene abrogation in knockout mice. Contrast of genotypic means based on GEMMA, which implements a correction for multiple testing, did not show any significant result. These findings are consistent with the notion that *vhc* or *hc* mutations do not produce the alterations predicted by experiments characterizing the phenotypes of knockout mice (see Discusion for more details about such predictions).

4. Discussion

4.1. Limited concordance in the prediction of the consequences of amino acid substitutions in porcine genes related to lipid metabolism

In this study, we predicted the potential consequences of 97 missense mutations on protein function by using four in silico tools (SIFT PROVEAN, MutPred2 and Align GVGD) and subsequently assessing the concordance of these predictions. The two in silico prediction tools which displayed the highest level of concordance were SIFT and PROVEAN. The SIFT software employs Dirichlet mixtures retrieved from protein multiple sequence alignments to generate position specific scoring matrices and score missense substitutions [20,25]. SIFT yields a normalized probability for each amino acid replacement to occur by considering the evolutionary conservation of protein families at the sequence level [25]. PROVEAN is also a sequence and evolutionary conservation-based method that uses a protein sequence and amino acid variations as inputs [22,26]. In our view, the most probable reason for the high consistency between SIFT and PROVEAN predictions is that both tools use similar methodological principles (i.e. sequence and evolutionary conservation) to make inferences about the pathogenicity of missense mutations.

In our study, predictions made with Align GVGD showed the lowest concordance with those generated by the other three methods. The Align GVGD tool uses a multiple sequence alignment as a starting point and calculates a Grantham Variation score (GV), which estimates biochemical variation at each alignment position, and a Grantham Difference score (GD) assessing the difference in side chain atomic composition, polarity, and volume between two amino acids [30]. So Align GVGD

Table 3
Gene-centric association analysis[1] between missense polymorphisms in porcine genes related to lipid metabolism and lipid traits recorded in a population of 316 Duroc pigs (only associations with a *q*-value <0.10 are displayed).

Fatty acid composition traits												
Gene Symbol	Traits	SNP ID	SSC	Position (Mb)	Substitution	*P*-value	*q*-value	*B*	δ (SE)	A_1	MAF	In silico prediction
AACS	GM C17:0	rs343900702	14	28,129,365	p.T485M	0.001	0.065	0.133	0.052 (0.015)	G	0.295	AlignGVGD
FABP3	GM C17:0	rs326530011	17	16,028,772	p.G25S	0.000	0.040	0.040	0.093 (0.025)	C	0.083	SIFT, MutPred2, Provean
GALNT2	GM C14:0	rs341291169	14	60,046,668	p.I32L	0.001	0.078	0.078	−0.096 (0.028)	A	0.164	–
PARK2	LD C18:3	rs331967252	1	5,858,928	p.G77D	0.001	0.055	0.055	0.030 (0.008)	G	0.304	Align GVGD
ZNF648	GM C20:5	rs324288383	9	123,509,090	p.G90R	0.002	0.094	0.188	−0.043 (0.012)	A	0.434	Align GVGD

Serum lipids concentrations												
Gene Symbol	Traits	SNP ID	SSC	Pos	Substitution	*P*-value	*q*-value	Bonf	δ (SE)	A_1	MAF	In silico prediction
STAT5A	HDL45	rs321203224	12	20,480,895	p.S392C	0.001	0.080	0.080	0.035 (0.010)	G	0.379	Align GVGD

Carcass traits												
Gene Symbol	Traits	SNP ID	SSC	Pos	Substitution	*P*-value	*q*-value	Bonf	δ (SE)	A_1	MAF	In silico prediction
FABP3	BFT34R	rs341327293	**17**	16,029,010	p.T104M	0.002	0.057	0.172	2.579 (0.736)	G	0.224	SIFT, Provean, Align GVGD
PARK2	BFT34R	rs331967252	**1**	5,858,928	p.G77D	0.001	0.057	0.147	−2.110 (0.560)	A	0.305	Align GVGD
PARK2	BFTLR	rs331967252	**1**	5,858,928	p.G77D	0.002	0.094	0.188	−4.094 (1.139)	A	0.305	Align GVGD
STAT5A	BFTLR	rs321203224	**12**	20,480,895	p.S392C	0.001	0.078	0.078	4.289 (1.126)	G	0.378	Align GVGD

[1] **SSC**: porcine chromosome, **GM**: *gluteus medius* muscle, **LD**: *longissimus dors*i muscle, **C14:0**: Myristic, **C17:0:** Margaric, **C18:3**: α-Linolenic, **C20:5**: Eicosapentaenoic, **HDL45**: High density lipoproteins at 45 days (mg/dL), **BFT34R**: backfat thickness between the third and fourth ribs, **BFTLR**: backfat thickness measured in the last rib, ***P*-value**: nominal *P*-value, ***q*-value**: *q-value* calculated with a false discovery rate approach, ***B***: Bonferroni corrected *P*-values, δ: allelic effect and its standard error (SE), **A_1**: minority allele, **MAF**: frequency of the minority allele, In silico prediction: tools that predict that a given missense mutation has functional effects. All genotyped SNPs were annotated as missense variants in the Ensembl database (Sscrofa 11.1) version 106 and Biomart portal (http://www.ensembl.org/biomart/).

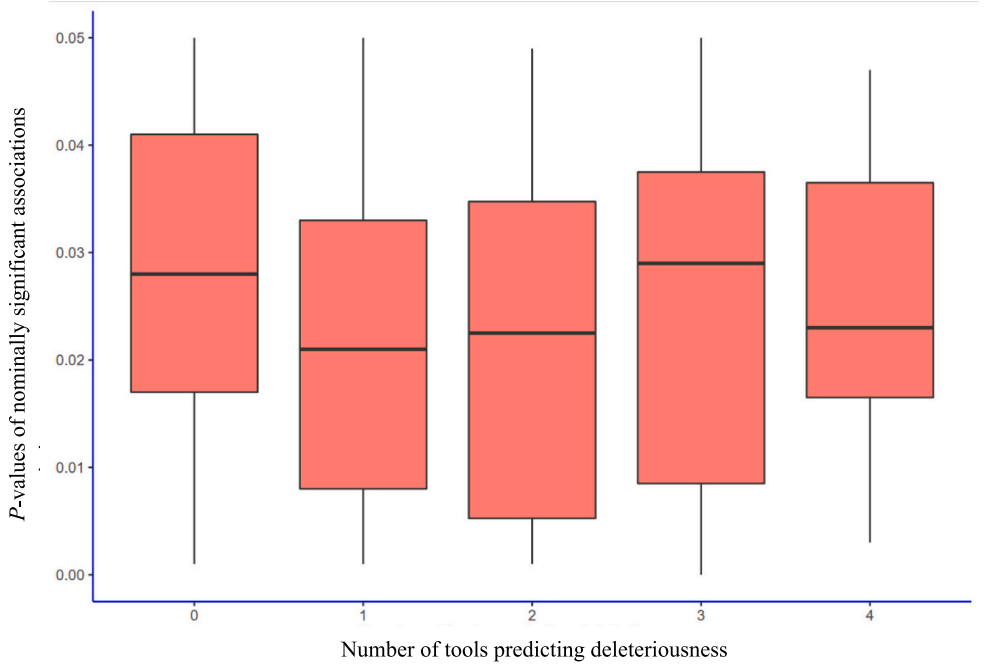

Fig. 2. Boxplot indicating the average *P*-values of nominally significant associations for missense SNP classified a having functional effects by 0, 1, 2, 3 or 4 in silico tools. Three groups of lipid phenotypes are jointly considered: serum lipid concentrations, carcass traits and intramuscular fat content and composition traits recorded in 316 Duroc pigs.

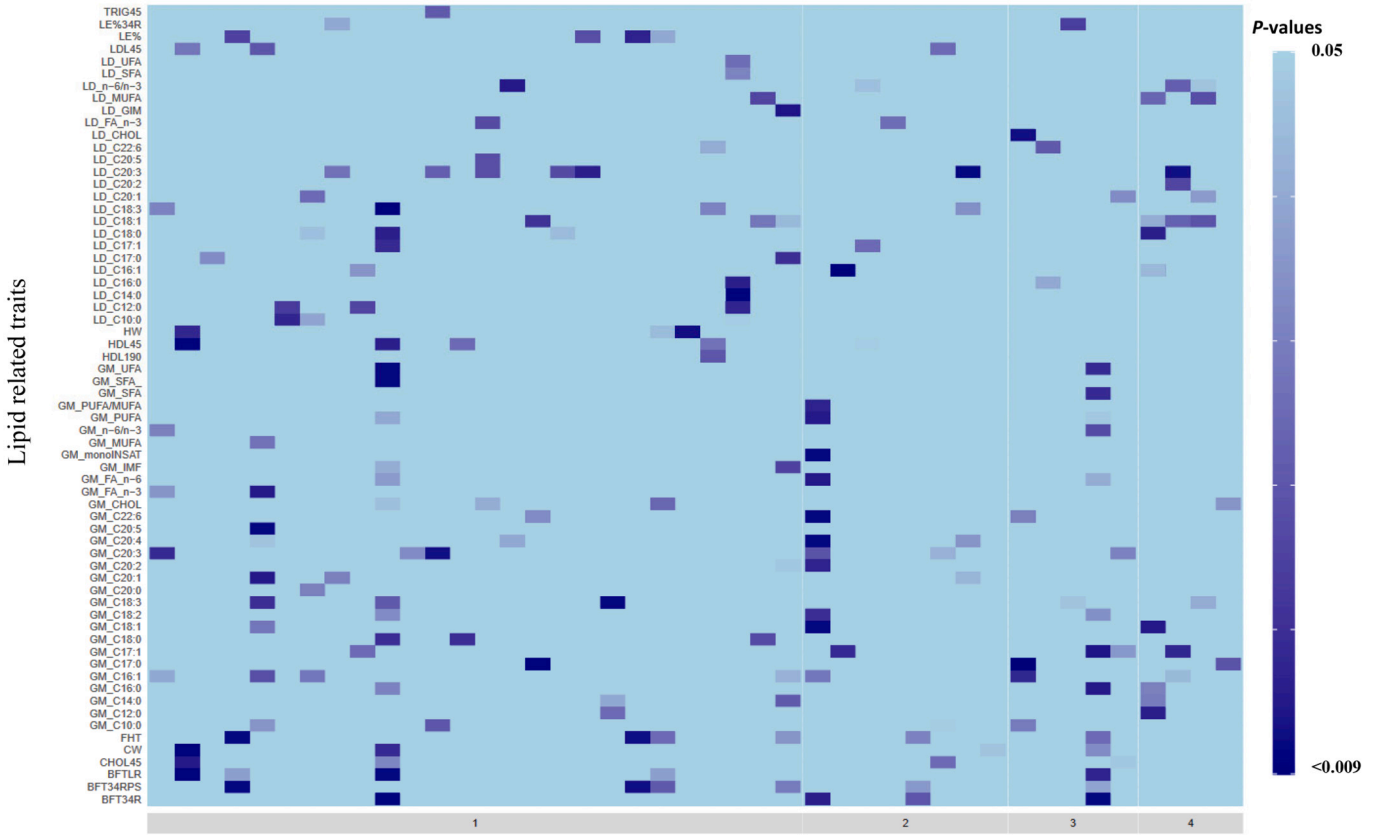

Fig. 3. Heatmap indicating the significance of nominal associations between missense SNP classified as having functional effects by 1, 2, 3 or 4 in silico tools and serum lipid concentration, carcass and intramuscular fat content and composition traits recorded in 316 Duroc pigs. For the sake of clarity, the 53 missense SNPs not predicted to be deleterious by any of the 4 in silico tools have not been taken into account.

differs from SIFT and PROVEAN in the sense that it does not rely exclusively on evolutionary conservation but also on the biophysical properties of amino acids to predict the consequences of missense mutations on gene function. This important methodological difference might explain the discrepancies observed in our study when comparing the predictions of AlignGVGD vs PROVEAN and SIFT. The second less concordant tool was MutPred2, something that could be anticipated because its methodological basis is completely different than that of the three other methods. Indeed, the MutPred2 software is based on a machine-learning approach taking into account sequence-based predictors (for over 50 structural and functional protein properties) which are trained on a large common dataset obtained from different sources and databases [21]. By integrating both genetic and molecular data, MutPred2 infers the damaging potential of amino acid substitutions by yielding a pathogenicity score [21].

Limited concordance of predictions made with different in silico tools has been reported in previous studies [31,32]. Grimm et al. (2015) compared the performance of ten in silico tools predicting the consequences of amino acid substitutions on five data sets, and reported that programs trained for the purpose of predicting pathogenicity generally performed better than those based on the calculation of conservation scores. Moreover, Grimm et al. [33] highlighted a problem of circularity in such predictions which makes it difficult to compare different analytical approaches. Importantly, the accuracy of predictions made with tools based on protein multiple sequence alignments are strongly dependent on the quality and informativeness of such alignments in terms of size, evolutionary depth, and careful construction and curation of the alignment [30]. Moreover, discrepancies in the predictions of in silico tools could be due, at least in part, to the fact that they build and use different multiple sequence alignments, although predictions can be different even when the same multiple sequence alignment is used by different tools [34]. Notably, tools do not always perform best when using their own alignment [34].

Another factor influencing the accuracy of in silico predictions is the gene being studied. For instance, Leong et al. [31] assessed the performance of five in silico tools to predict the consequences of non-synonymous mutations in the *KCNQ1*, *KCNH2* and *SCN5A* genes and found that predictions for the *SCN5A* locus were much more inaccurate than those for the *KCNQ1* or *KCNH2* loci. In summary, multiple technical and methodological issues might explain the limited concordance of predictions made with SIFT, PROVEAN, MutPred2 and Align GVGD about the functional consequences of missense substitutions mapping to porcine lipid metabolism genes.

4.2. The associations of very highly consistent functional mutations with lipid traits do not recapitulate phenotypes observed in knockout mice

We have detected 5 missense mutations in five porcine genes related to lipid metabolism that were consistently classified by the four in silico tools as deleterious and that, in consequence, are strongly predicted to alter gene function. These mutations were p.S108C (*CAV2*), p.P302S (*DHCR24*), p.P212L (*LIPC*), p.P1080L (*LRP4*) and p.R161G (*PARK2*). In principle, we would expect that if these mutations are truly functional, their associations with lipid traits in pigs should reflect, at least to some extent, alterations in lipid-related phenotypes observed in knockout mice or related genetic models. Our results show that, in general, this is not the case. For instance, the p.P302S mutation in the *DHCR24* gene is predicted to be highly pathogenic by the four tools and it is present in two of three porcine *DHCR24* transcripts defined in the Ensembl database (http://www.ensembl.org). The DHCR24 enzyme catalyzes the conversion of desmosterol to cholesterol, and the complete inactivation of the corresponding gene causes desmosterolosis in mice which is often

Table 4
Nominally significant associations (*P*-value <0.05) between porcine lipid traits and missense polymorphisms classified as functional by four in silico prediction tools (no association was significant after correction for multiple testing).

Gene	SNP ID[1]	SSC	Position (Mb)	Substitution	Nominal associations with lipid phenotypes[2]
CAV2	rs332492536	18	29,706,153	p.S108C	LD MUFA, LD C18:1, LD C20:1, LD n-6/n-3, GM C18:3
DHCR24	rs328825271	6	157,502,221	p.P302S	LD C18:1, LD C20:2, LD C20:3, LD n-6/n-3, GM C16:1, GM C17:1
LIPC	rs712720116	1	113,456,385	p.P212L	–
LRP4	rs328789611	2	15,651,335	p.P1080L	LD C16:1, LD C18:0, LD C16:1, LD C18:1, LD MUFA, GM C12:0, GM C14:0, GM C16:0, GM C18:1
PARK2	rs341517389	1	5,922,601	p.R161G	–

[1] All genotyped SNPs were annotated as missense variants in the Ensembl database (Sscrofa 11.1) version 106 and Biomart portal (http://www.ensembl.org/biomart/).

[2] **SSC**: porcine chromosome, **GM**: *gluteus medius* muscle, **LD**: *longissimus dorsi* muscle, **C14:0**: Myristic, **C16:1**: Palmitelaidic, **C17:0**: Margaric, **C17:1**: Heptadecenoic, **C18:0**: Stearic, **C18:1**: Octadecenoic, **C18:3**: α-Linolenic, **C20:1**: Gondoic, **C20:2**: Eicosadienoic, **C20:3**: Eicosatrienoic, **C20:5**: Eicosapentaenoic, **MUFA**: Monounsaturated FA, **n-6/n-3**: Omega-6 to −3 ratio.

Table 5
Nominally significant associations (*P*-value <0.05) between porcine lipid traits and missense polymorphisms classified as functional by three in silico prediction tools (associations with *q*-values <0.1 are shown in bold).

Gene	SNP ID[1]	SSC[2]	Position	Substitution	Nominal associations with lipid phenotypes
ACSM3	rs328465555	3	25,262,463	p.I190T	LD C22:6
FABP3	rs326530011	17	16,028,772	p.G25S	LD CHOL, GM C10:0, GM C16:1, **GM C17:0**, GM C22:6
	rs341327293	17	16,029,010	p.T104M	GM C16:0, GM C17:1, GM SFA, GM UFA, GM n-6/n-3, GM C18:2
					GM n-6, GM PUFA, **BFT34R**, **BFTLR**, **FHT**, **CW**, **BFT34RPS**
JMJD1C	rs335786128	14	66,665,158	p.S2018F	GM C18:3, **LD34R**
MC4R	rs81219178	1	160,773,437	p.D298N	–
MLXIPL	rs788027215	3	10,900,920	p.G683R	GM C17:0, GM CHOL
ZNF648	rs341953239	9	123,508,330	p.S343W	LD C20:1, GM C20:3

[1] All genotyped SNPs were annotated as missense variants in the Ensembl database (Sscrofa 11.1) version 106 and Biomart portal (http://www.ensembl.org/biomart/).

[2] **SSC**: porcine chromosome, **GM**: *gluteus medius* muscle, **LD**: *longissimus dorsi* muscle, **C10:0**: Capric, **C16:0**: Palmitic, **C16:1**: Palmitelaidic, **C17:0**: Margaric, **C17:1**: Heptadecenoic, **C18:2**: Linoleic, **C18:3**: α-Linolenic, **C20:1**: Gondoic, **C20:3**: Eicosatrienoic, **C22:6**: Docosahexaenoic, **SFA:** Saturated fatty acid, **UFA:** Unsaturated fatty acid, **PUFA:** Polyunsaturated fatty acid, GM n-6: Omega-6, **n-6/n-3**: Omega-6 to −3 ratio, CHOL: Cholesterol, **BFT34R**: backfat thickness between the third and fourth ribs, **BFTLR**: backfat thickness measured in the last rib, **HW**: ham weight, **LE%**: lean percentage, **CW**: carcass weight, **BFT34RPS**: BFT between the third and fourth ribs prior slaughter, **FHT**: fresh fat ham thickness and **LD34R**: loin depth between the third and fourth ribs.

lethal soon after birth [35]. In the Lipgen population, we did not observe any association between the p.P302S mutation and serum cholesterol concentrations at either 45 days or 190 days suggesting that this mutation does not impair DHCR24 function (Fig. S2). Data shown in Fig. S2 show that this lack of association is not due to the absence of pigs with homozygous genotypes for the mutation predicted to be functional. Another relevant case is represented by the *LIPC* gene which encodes an enzyme hydrolyzing triacylglycerols in plasma lipoproteins to release free FA to be used by cells as a source of energy [36]. We have detected a p.P212L missense mutation that is consistently predicted to be pathogenic by SIFT, PROVEAN, MutPred2 and Align GVGD, and it is present in one of three porcine *LIPC* transcripts defined in the Ensembl database (http://www.ensembl.org/). The inactivation of the *LIPC* gene in mouse is associated with reduced daily feed intake and increased leanness [36], and the hepatic knockout of this gene causes a mild dyslipidemia [37]. In the association analysis performed in the current work, we did not detect any association between p.P212L genotype and fatness or serum lipid traits recorded in Lipgen pigs (Fig. S4).

Following the same line of reasoning, the p.R161G mutation, present in three of the six pig *PARK2* gene transcripts, was not associated with any fatness phenotype (Fig. S3), despite the fact that this gene is a fundamental regulator of fat uptake in mice and its inactivation leads to impaired intestinal absorption of lipids and reduced plasma triglycerides after intragastric fat challenge [38,39]. In this case, however, the lack of pigs homozygous for the mutation predicted to be functional could explain the lack of an association between this polymorphism and fatness traits (as long as complete dominance is assumed). The potential phenotypic consequences of the p.P1080L (present in one of the two *LRP4* transcripts) and p.S108C (present in the only *CAV2* transcript) genotypes, that were associated with multiple fatty acid traits, was difficult to predict because the abrogation of the function of these two genes causes embryonic lethality [40] and severe pulmonary disfunction [41], respectively. Moreover, there is evidence that the functional suppression of the caveolin 1 gene causes elevated circulating FA and reduced hepatic FA content [42]. Interestingly, the MAF of the p.S108C (CAV2, MAF = 0.147) and the p.P1080L (LRP4, MAF = 0.309) mutations are quite high, while as previously commented highly deleterious mutations tend to have very low frequencies because they are purged by purifying selection. Such observation is inconsistent with a highly deleterious effect on biological viability of the two missense mutations mentioned before.

4.3. Most of highly consistent functional mutations do not display associations concordant with their expected phenotypic consequences on lipid traits

By performing association studies between each variant and phenotype, we have found that two *hc* functional mutations (p.G25S and p.T104M) in the *FABP3* gene showed the highest number of nominally significant associations with muscle FA composition and carcass traits (Table 5, Fig. S1). Besides, the p.G25S mutation was also significantly associated with GM C17:0 (*q*-value = 0.04) after correction for multiple testing. In general, associations with minority FA should be interpreted with caution because it is difficult to measure their quantities with high precision. The FABP3 molecule is a small cytoplasmic protein highly

expressed in the skeletal muscle and heart which acts as a lipid "chaperone" modulating the solubility, mobility, and utilization of FA [43,44]. In previous studies, the polymorphism of the porcine *FABP3* gene was associated with intramuscular fat content [45–48] and composition traits [49] as well as with backfat thickness [50], an outcome that is consistent with our results and emphasizes the interest of characterizing the functional impact of this mutation with an in vitro assay.

In contrast, the remaining *hc* mutations did not show associations with lipid phenotypes that become altered in knockout mice or related genetic models, whenever such evidence was available. For instance, the *JMJD1C* gene induces lipogenesis in vivo and modulates hepatic and plasma triglyceride levels [51], but the p.S2018F genotype does not show any association with serum lipid traits in our study (Fig. S5). Similarly, MLXIPL is a transcription factor modulating lipogenesis and its variation is associated with plasma triglyceride levels in humans [52], but the p.G683R genotype does not show such association in our resource population (Fig. S7). More interesting is the case of the p.D298N mutation in the *MC4R* gene, because MC4R deficiency leads to hyperphagia, obesity and increased growth in mice [53] and, more importantly, the p.D298N mutation has been associated with fatness, growth, and feed intake traits in pigs [54]. In our study, we did not detect any association between this polymorphism and obesity traits (Fig. S6). Interestingly, Fan et al. [55] performed functional tests to assess the ligand binding and signaling properties of the p.D298N allelic variant in transiently transfected HEK293T cells and did not find any evidence of functional differences between the two D and N alleles. In contrast, Zhang et al. (2020) reported that the N allele has a decreased basal constitutive activity and less surface expression than the D allele [56].

4.4. Predicted functional effect severity of missense substitutions does not correlate with the outcome of association analyses

We have explored the associations of the 97 genotyped SNPs with 48 lipid traits to determine whether there is any relationship between the predicted severity of mutations and the number and/or significance of such associations. Obviously, not all lipid metabolism genes have the same impact on the determinism of fatness traits, but this would be very difficult to weigh from a biological standpoint. In fact, in this study, we did not observe co-localizations between missense SNPs showing significant associations with lipid traits and QTLs for lipid phenotypes detected in the same Lipgen population [9,57,58]. This may be due to the limitations of genome-wide association analyses in small sample size populations, which can fail to detect loci with small effects or rare variants with strong effects [57]. Furthermore, changes at the protein level may not always result in noticeable consequences on polygenic traits. Even the complete loss of protein function may only result in small changes of the phenotype. For example, Matsukawa et al. [59] reported that mouse knockout for the *NPR3* gene, which plays a key role in determining blood pressure, produces a change of not >8 mm of Hg in homozygous mutants but not in the heterozygous ones.

Our general assessment is that there is not a strong relationship between the predicted severity of missense mutations and their association with lipid phenotypes. Not only most of *vhc* and *hc* mutations did not show associations with traits that in the corresponding knockout mice models were altered, but, moreover, we do not observe an increased number of associations, or more significant associations, for *vhc* or *hc* mutations when compared to the set of mutations without enough evidence of having functional effects. This result could be partly due to the fact that, for several missense mutations analyzed in the current work, there is a lack of pigs homozygous for the variant predicted to be non-functional (under a complete dominance inheritance model heterozygous individuals should show a fully normal phenotype because the functional allele compensates the non-functional one) or their numbers are quite small (making difficult to obtain a significant result). Another important limitation of our study is the lack of endophenotypes defining protein activity for genes under investigation. Such source of information would have provided more valuable clues about the consequences of mutations predicted to be functional than raw phenotypes which have a highly complex genetic basis.

Despite these caveats, results obtained in humans and mouse are consistent with our observation that mutations mapping to porcine lipid metabolism genes and predicted to have functional consequences by 3 or 4 in silico tools do not seem to determine the phenotypic variation of fatness traits recorded in Duroc pigs. For instance, Ernst et al. [32] compared the performance of four prediction tools using a representative number of BRCA1/2 missense variants and concluded that poor specificity and high proportion of false positives made it very difficult to anticipate the pathogenicity of such missense mutations. Dorfman et al. [3] and Michels et al. [60] reached similar conclusions when contrasting in silico predictions about the functional consequences of missense mutations in the *CFTR* gene and the severity of cystic fibrosis. In another study, Miosge et al. [5] detected 33 de novo missense mutations in 23 genes with fundamental immunological functions by sequencing the genomes of mice treated with N-ethyl-N-nitrosourea. Miosge et al. [5] predicted the functional consequences of these mutations with PolyPhen2, SIFT, MutationAssessor, Panther, CADD and Condel, and they also determined their potential effects by analyzing individuals homozygous for the predicted in vivo knockout phenotype. Concordance between predictions and observations was low e.g. only 20% of missense substitutions classified as deleterious by PolyPhen2 showed an observable phenotype in homozygous individuals [5]. This outcome could be due to genetic compensation, but when they compared in silico predictions with "in vitro" functional data, concordance improved only to a limited extent. These results highlight the intrinsic limitations of anticipating the phenotypic consequences of missense substitutions based on in silico predictions made with dedicated bionformatic tools. In stark contrast with simple Mendelian phenotypes, complex traits depend on multiple genetic and environmental factors as well as on intricate interaction networks. This circumstance exacerbates even more the limitations of current bioinformatic tools to predict the potential impact of missense mutations on traits of economic interest in pigs and other farm species.

Support

Research presented in this publication was funded by grants AGL2013-48742-C2-1-R and AGL2013-48742-C2-2-R awarded by the Ministerio de Economía y Competitividad (Spain).

CRediT authorship contribution statement

Rayner González-Prendes: Writing - Original draft, Formal analysis, data curation, visualization, investigation. **Martijn F.L. Derks**: writing - reviewing and editing. **Martien A.M. Groenen:** writing-reviewing and editing. **Raquel Quintanilla**: Resources, funding acquisition writing- reviewing and editing. **Marcel Amills:** Writing - Original draft, conceptualization, data curation, visualization, investigation, funding acquisition, resources.

Declaration of Competing Interest

The authors declare no conflict of interest.

Data availability

figshare

Assessing the relationship between the predicted deleteriousness of 97 missense mutations in 68 lipid genes and phenotypic variation of complex fatness traits in pigs (Original data) (https://figshare.com)

Genotype and Phenotype data for association analyses of miRNA

SNPs with gluteus medius and longissimus dorsi skeletal muscle phenotype data (Reference data) (https://figshare.com)

Acknowledgements

The authors wish to thank Selección Batallé S.A. for providing the animal material. We also acknowledge the support of the Spanish Ministerio de Ciencia e Innovación for the Center of Excellence Severo Ochoa 2020-2023 (CEX2019-000902-S) grant awarded to the Centre for Research in Agricultural Genomics (CRAG, Bellaterra, Spain). Thanks also to the CERCA Programme of the Generalitat the Catalunya for their constant support.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ygeno.2023.110589.

References

[1] H. Tang, P.D. Thomas, PANTHER-PSEP: predicting disease-causing genetic variants using position-specific evolutionary preservation, Bioinformatics. 32 (2016) 2230–2232, https://doi.org/10.1093/bioinformatics/btw222.
[2] D. Tchernitchko, M. Goossens, H. Wajcman, In silico prediction of the deleterious effect of a mutation: proceed with caution in clinical genetics, Clin. Chem. 50 (2004) 1974–1978, https://doi.org/10.1373/clinchem.2004.036053.
[3] R. Dorfman, T. Nalpathamkalam, C. Taylor, T. Gonska, K. Keenan, X.W. Yuan, et al., Do common in silico tools predict the clinical consequences of amino-acid substitutions in the *CFTR* gene? Clin. Genet. 77 (2010) 464–473, https://doi.org/10.1111/j.1399-0004.2009.01351.x.
[4] D.L. Masica, R. Karchin, Towards increasing the clinical relevance of in silico methods to predict pathogenic missense variants, PLoS Comput. Biol. (2016) 12, https://doi.org/10.1371/journal.pcbi.1004725.
[5] L.A. Miosge, M.A. Field, Y. Sontani, V. Cho, S. Johnson, A. Palkova, et al., Comparison of predicted and actual consequences of missense mutations, Proc. Natl. Acad. Sci. U. S. A. 112 (2015) E5189–E5198, https://doi.org/10.1073/pnas.1511585112.
[6] S. Wettstein, J. Underhaug, B. Perez, B.D. Marsden, W.W. Yue, A. Martinez, et al., Linking genotypes database with locus-specific database and genotype-phenotype correlation in phenylketonuria, Eur. J. Hum. Genet. 23 (2015) 302–309, https://doi.org/10.1038/ejhg.2014.114.
[7] D. Gallardo, R.N. Pena, M. Amills, L. Varona, O. Ramírez, J. Reixach, et al., Mapping of quantitative trait loci for cholesterol, LDL, HDL, and triglyceride serum concentrations in pigs, Physiol. Genomics 35 (2008) 199–209, https://doi.org/10.1152/physiolgenomics.90249.2008.
[8] R. Quintanilla, R.N. Pena, D. Gallardo, A. Cánovas, O. Ramírez, I. Díaz, et al., Porcine intramuscular fat content and composition are regulated by quantitative trait loci with muscle-specific effects, J. Anim. Sci. 89 (2011) 2963–2971, https://doi.org/10.2527/JAS.2011-3974.
[9] A. Manunza, J. Casellas, R. Quintanilla, R. González-Prendes, R.N. Pena, J. Tibau, et al., A genome-wide association analysis for porcine serum lipid traits reveals the existence of age-specific genetic determinants, BMC Genomics (2014) 15, https://doi.org/10.1186/1471-2164-15-758.
[10] P.G. Eusebi, R. González-Prendes, R. Quintanilla, J. Tibau, T.F. Cardoso, A. Clop, et al., A genome-wide association analysis for carcass traits in a commercial Duroc pig population, Anim. Genet. 48 (3 Feb 2017) 466–469, https://doi.org/10.1111/age.12545.
[11] R. González-Prendes, R. Quintanilla, E. Mármol-Sánchez, R.N. Pena, M. Ballester, T.F. Cardoso, 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, https://doi.org/10.1186/s12864-019-5557-9.
[12] O. Vidal, J.L. Noguera, M. Amills, L. Varona, M. Gil, N. Jiménez, et al., Identification of carcass and meat quality quantitative trait loci in a landrace pig population selected for growth and leanness, J. Anim. Sci. 83 (2005) 293–300, https://doi.org/10.2527/2005.832293x.
[13] A. Warr, N. Affara, B. Aken, H. Beiki, D.M. Bickhart, K. Billis, et al., An improved pig reference genome sequence to enable pig genetics and genomics research, Gigascience. 9 (2020) 1–14, https://doi.org/10.1093/GIGASCIENCE/GIAA051.
[14] H. Li, R. Durbin, Fast and accurate long-read alignment with burrows-Wheeler transform, Bioinformatics. 26 (2010) 589–595, https://doi.org/10.1093/bioinformatics/btp698.
[15] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, et al., The sequence alignment/map format and SAMtools, Bioinformatics. 25 (2009) 2078–2079, https://doi.org/10.1093/bioinformatics/btp352.
[16] A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, et al., The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data, Genome Res. 20 (2010) 1297–1303, https://doi.org/10.1101/gr.107524.110.
[17] P. Cingolani, A. Platts, L.L. Wang, M. Coon, T. Nguyen, L. Wang, et al., A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3, Fly (Austin) 6 (2012) 80–92, https://doi.org/10.4161/fly.19695.
[18] D. Cotter, A. Maer, C. Guda, B. Saunders, S. Subramaniam, LMPD: LIPID MAPS proteome database, Nucleic Acids Res. 34 (2006) D507–D510, https://doi.org/10.1093/nar/gkj122.
[19] A.D. Yates, P. Achuthan, W. Akanni, J. Allen, J. Allen, J. Alvarez-Jarreta, et al., Ensembl 2020, Nucleic Acids Res. 48 (2020) D682–D688, https://doi.org/10.1093/nar/gkz966.
[20] P. Kumar, S. Henikoff, P.C. Ng, Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm, Nat. Protoc. 4 (2009) 1073–1082, https://doi.org/10.1038/nprot.2009.86.
[21] V. Pejaver, J. Urresti, J. Lugo-Martinez, K.A. Pagel, G.N. Lin, H.J. Nam, et al., Inferring the molecular and phenotypic impact of amino acid variants with MutPred2, Nat. Commun. 11 (2020) 1–13, https://doi.org/10.1038/s41467-020-19669-x.
[22] Y. Choi, A.P. Chan, PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels, Bioinformatics. 31 (2015) 2745–2747, https://doi.org/10.1093/bioinformatics/btv195.
[23] S.V. Tavtigian, A.M. Deffenbaugh, L. Yin, T. Judkins, T. Scholl, P.B. Samollow, et al., Comprehensive statistical study of 452 BRCA1 missense substitutions with classification of eight recurrent substitutions as neutral, J. Med. Genet. 43 (2006) 295–305, https://doi.org/10.1136/jmg.2005.033878.
[24] E. Mathe, M. Olivier, S. Kato, C. Ishioka, P. Hainaut, S.V. Tavtigian, Computational approaches for predicting the biological effect of p53 missense mutations: a comparison of three sequence analysis based methods, Nucleic Acids Res. 34 (2006) 1317–1325, https://doi.org/10.1093/nar/gkj518.
[25] P.C. Ng, S. Henikoff, SIFT: predicting amino acid changes that affect protein function, Nucleic Acids Res. 31 (2003) 3812–3814, https://doi.org/10.1093/nar/gkg509.
[26] Y. Choi, G.E. Sims, S. Murphy, J.R. Miller, A.P. Chan, Predicting the functional effect of amino acid substitutions and Indels, PLoS One (2012) 7, https://doi.org/10.1371/journal.pone.0046688.
[27] R. Grantham, Amino acid difference formula to help explain protein evolution, Science 185 (1974) 862–864, https://doi.org/10.1126/science.185.4154.862.
[28] X. Zhou, M. Stephens, Genome-wide efficient mixed-model analysis for association studies, Nat. Genet. 44 (2012) 821–824, https://doi.org/10.1038/ng.2310.
[29] Y. Benjamini, Y. Hochberg, Controlling the false discovery rate: a practical and powerful approach to multiple testing, J. R. Stat. Soc. Ser. B 57 (1995) 289–300, https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.
[30] S.V. Tavtigian, M.S. Greenblatt, F. Lesueur, G.B. Byrnes, In silico analysis of missense substitutions using sequence-alignment based methods, Hum. Mutat. (2008) 1327–1336, https://doi.org/10.1002/humu.20892.
[31] I.U.S. Leong, A. Stuckey, D. Lai, J.R. Skinner, D.R. Love, Assessment of the predictive accuracy of five in silico prediction tools, alone or in combination, and two metaservers to classify long QT syndrome gene mutations, BMC Med. Genet. 16 (2015), https://doi.org/10.1186/s12881-015-0176-z.
[32] C. Ernst, E. Hahnen, C. Engel, M. Nothnagel, J. Weber, R.K. Schmutzler, et al., Performance of in silico prediction tools for the classification of rare BRCA1/2 missense variants in clinical diagnostics, BMC Med. Genet. (2018) 11, https://doi.org/10.1186/s12920-018-0353-y.
[33] D.G. Grimm, C.A. Azencott, F. Aicheler, U. Gieraths, D.G. Macarthur, K. E. Samocha, et al., The evaluation of tools used to predict the impact of missense variants is hindered by two types of circularity, Hum. Mutat. 36 (2015) 513–523, https://doi.org/10.1002/humu.22768.
[34] S. Hicks, D.A. Wheeler, S.E. Plon, M. Kimmel, Prediction of missense mutation functionality depends on both the algorithm and sequence alignment employed, Hum. Mutat. 32 (2011) 661–668, https://doi.org/10.1002/humu.21490.
[35] B. Kanuri, V. Fong, S.R. Ponny, R. Weerasekera, K. Pulakanti, K.S. Patel, et al., Generation and validation of a conditional knockout mouse model for desmosterolosis, J. Lipid Res. 62 (2021), 100028, https://doi.org/10.1016/J.JLR.2021.100028.
[36] H.K. Chiu, K. Qian, K. Ogimoto, G.J. Morton, B.E. Wisse, N. Agrawal, et al., Mice lacking hepatic lipase are lean and protected against diet-induced obesity and hepatic steatosis, Endocrinology. 151 (2010) 993–1001, https://doi.org/10.1210/en.2009-1100.
[37] G.E. Homanics, H.V. De Silva, J. Osada, S.H. Zhang, H. Wong, J. Borensztajn, et al., Mild dyslipidemia in mice following targeted inactivation of the hepatic lipase gene, J. Biol. Chem. 270 (1995) 2974–2980, https://doi.org/10.1074/jbc.270.7.2974.
[38] D.K. Costa, B.R. Huckestein, L.R. Edmunds, M.C. Petersen, A. Nasiri, G.M. Butrico, et al., Reduced intestinal lipid absorption and body weight-independent improvements in insulin sensitivity in high-fat diet-fed Park2 knockout mice, Am. J. Physiol. Endocrinol. Metab. 311 (2016) E105–E116, https://doi.org/10.1152/ajpendo.00042.2016.
[39] K.Y. Kim, M.V. Stevens, M.H. Akter, S.E. Rusk, R.J. Huang, A. Cohen, et al., Parkin is a lipid-responsive regulator of fat uptake in mice and mutant human cells, J. Clin. Invest. 121 (2011) 3701–3712, https://doi.org/10.1172/JCI44736.
[40] C.M. Karner, M.F. Dietrich, E.B. Johnson, N. Kappesser, C. Tennert, F. Percin, et al., Lrp4 regulates initiation of ureteric budding and is crucial for kidney formation - a mouse model for cenani-Lenz syndrome, PLoS One 5 (2010), e10418, https://doi.org/10.1371/journal.pone.0010418.
[41] B. Razani, X.B. Wang, J.A. Engelman, M. Battista, G. Lagaud, X.L. Zhang, et al., Caveolin-2-deficient mice show evidence of severe pulmonary dysfunction without

disruption of caveolae, Mol. Cell. Biol. 22 (2002) 2329–2344, https://doi.org/10.1128/mcb.22.7.2329-2344.2002.
[42] J.P. Otis, M.C. Shen, V. Quinlivan, J.I. Anderson, S.A. Farber, Intestinal epithelial cell caveolin 1 regulates fatty acid and lipoprotein cholesterol plasma levels, DMM Dis. Model Mech. 10 (2017) 283–295, https://doi.org/10.1242/dmm.027300.
[43] M. Furuhashi, G.S. Hotamisligil, Fatty acid-binding proteins: role in metabolic diseases and potential as drug targets, Nat. Rev. Drug Discov. 7 (2008) 489–503, https://doi.org/10.1038/nrd2589.
[44] S.M. Lee, S.H. Lee, Y. Jung, Y. Lee, J.H. Yoon, J.Y. Choi, et al., FABP3-mediated membrane lipid saturation alters fluidity and induces ER stress in skeletal muscle with aging, Nat. Commun. 11 (2020) 1–15, https://doi.org/10.1038/s41467-020-19501-6.
[45] F. Gerbens, A.J.M. Van Erp, F.L. Harders, F.J. Verburg, T.H.E. Meuwissen, J. H. Veerkamp, et al., Effect of genetic variants of the heart fatty acid-binding protein gene on intramuscular fat and performance traits in pigs, J. Anim. Sci. 77 (1999) 846–852, https://doi.org/10.2527/1999.774846x.
[46] D. Nechtelberger, V. Pires, J. Sölkner, I. Stur, G. Brem, M. Mueller, et al., Intramuscular fat content and genetic variants at fatty acid-binding protein loci in Austrian pigs, J. Anim. Sci. 79 (2001) 2798–2804, https://doi.org/10.2527/2001.79112798x.
[47] S.H. Lee, Y.M. Choi, J.H. Choe, J.M. Kim, K.C. Hong, H.C. Park, et al., Association between polymorphisms of the heart fatty acid binding protein gene and intramuscular fat content, fatty acid composition, and meat quality in Berkshire breed, Meat Sci. 86 (2010) 794–800, https://doi.org/10.1016/j.meatsci.2010.06.024.
[48] B. Wang, P. Li, W. Zhou, C. Gao, H. Liu, H. Li, et al., Association of twelve candidate gene polymorphisms with the intramuscular fat content and average backfat thickness of chinese suhuai pigs, Animals. 9 (2019) 858, https://doi.org/10.3390/ani9110858.
[49] J. Hong, D. Kim, K. Cho, S. Sa, S. Choi, Y. Kim, et al., Effects of genetic variants for the swine FABP3, HMGA1, MC4R, IGF2, and FABP4 genes on fatty acid composition, Meat Sci. 110 (2015) 46–51, https://doi.org/10.1016/j.meatsci.2015.06.011.
[50] K.H. Cho, M.J. Kim, G.J. Jeon, H.Y. Chung, Association of genetic variants for FABP3 gene with back fat thickness and intramuscular fat content in pig, Mol. Biol. Rep. 38 (2011) 2161–2166, https://doi.org/10.1007/s11033-010-0344-3.
[51] J.A. Viscarra, Y. Wang, H.P. Nguyen, Y.G. Choi, H.S. Sul, Histone demethylase JMJD1C is phosphorylated by mTOR to activate de novo lipogenesis, Nat. Commun. 11 (2020) 1–16, https://doi.org/10.1038/s41467-020-14617-1.
[52] K. Nakayama, Y. Yanagisawa, A. Ogawa, Y. Ishizuka, L. Munkhtulga, P. Charupoonphol, et al., High prevalence of an anti-hypertriglyceridemic variant of the MLXIPL gene in Central Asia, J. Hum. Genet. 56 (2011) 828–833, https://doi.org/10.1038/jhg.2011.109.
[53] I.S. Farooqi, J.M. Keogh, G.S.H. Yeo, E.J. Lank, T. Cheetham, S. O'Rahilly, Clinical spectrum of obesity and mutations in the melanocortin 4 receptor gene, N. Engl. J. Med. 348 (2003) 1085–1095, https://doi.org/10.1056/nejmoa022050.
[54] K.S. Kim, N. Larsen, T. Short, G. Plastow, M.F. Rothschild, A missense variant of the porcine melanocortin-4 receptor (MC4R) gene is associated with fatness, growth, and feed intake traits, Mamm. Genome 11 (2000) 131–135, https://doi.org/10.1007/s003350010025.
[55] Z.-C. Fan, J.L. Sartin, Y.-X. Tao, Pharmacological analyses of two naturally occurring porcine melanocortin-4 receptor mutations in domestic pigs, Domest. Anim. Endocrinol. 34 (2008) 383–390, https://doi.org/10.1016/j.domaniend.2007.05.003.
[56] J. Zhang, J. Li, C. Wu, Z. Hu, L. An, Y. Wan, et al., The Asp298Asn polymorphism of melanocortin-4 receptor (MC4R) in pigs: evidence for its potential effects on MC4R constitutive activity and cell surface expression, Anim. Genet. 51 (2020) 694–706, https://doi.org/10.1111/AGE.12986.
[57] R. González-Prendes, E. Mármol-Sánchez, R. Quintanilla, A. Castelló, A. Zidi, Y. Ramayo-Caldas, et al., About the existence of common determinants of gene expression in the porcine liver and skeletal muscle, BMC Genomics (2019) 20, https://doi.org/10.1186/s12864-019-5889-5.
[58] P.G. Eusebi, R. González-Prendes, R. Quintanilla, J. Tibau, T.F. Cardoso, A. Clop, et al., A genome-wide association analysis for carcass traits in a commercial Duroc pig population, Anim. Genet. 48 (2017) 466–469, https://doi.org/10.1111/AGE.12545.
[59] M. Matsukawa, W.J. Grzesik, N. Takahashi, K.N. Pandey, S. Pang, M. Yamauchi, et al., The natriuretic peptide clearance receptor locally modulates the physiological effects of the natriuretic peptide system, Proc. Natl. Acad. Sci. U. S. A. 96 (1999) 7403–7408, https://doi.org/10.1073/PNAS.96.13.7403/ASSET/DA02F099-3AC1-457F-B6C7-599D4C78EA21/ASSETS/GRAPHIC/PQ1391661004.JPEG.
[60] M. Michels, U. Matte, L.R. Fraga, A.C.B. Mancuso, R. Ligabue-Braun, E.F. R. Berneira, et al., Determining the pathogenicity of cftr missense variants: multiple comparisons of in silico predictors and variant annotation databases, Genet. Mol. Biol. 42 (2019) 560–570, https://doi.org/10.1590/1678-4685-gmb-2018-0148.