Variability in porcine microRNA genes and its association with mRNA expression phenotypes

Emilio Mármol-Sánchez[1], Dailu Guan[1], Raquel Quintanilla[2], Raul Tonda[3], Marcel Amills[1,4]

[1]Centre for Research in Agricultural Genomics (CRAG), CSIC-IRTA-UAB-UB, Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain.

[2]Animal Breeding and Genetics Program, Institute for Research and Technology in Food and Agriculture (IRTA), Torre Marimon, 08140 Caldes de Montbui, Spain.

[3]CNAG-CRG, Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology (BIST), Barcelona, Spain.

[4]Departament de Ciència Animal i dels Aliments, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain.

Corresponding autor: Marcel Amills. [1]Centre for Research in Agricultural Genomics (CRAG), CSIC-IRTA-UAB-UB, Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain. Phone: +34 935636600. Email: marcel.Amills@uab.cat

Abstract

Background

Mature microRNAs (miRNAs) play an important role in repressing the expression of a wide range of protein coding transcripts by promoting their degradation or inhibiting their translation into functional proteins. The presence of segregating polymorphisms inside miRNA loci and their corresponding 3'UTR binding sites might disrupt canonical conserved miRNA-mRNA pairing, thus modifying gene expression patterns.

Results

We aimed to investigate the variability of miRNA genes and their putative binding sites by analyzing whole-genome sequences from 120 pigs and wild boars from Europe and Asia. In total, 285 SNPs residing within miRNA loci were detected. From these, 221 were located in precursor regions, whereas 52 and 12 mapped to mature and seed regions, respectively. Moreover, a total of 109,724 polymorphisms were identified in predicted 7mer-m8 miRNA binding sites within porcine 3'UTRs. A principal components analysis revealed a clear genetic divergence between Asian and European samples, which was particularly strong for 3'UTR sequences. We also observed that miRNA genes show reduced polymorphism compared with other non-miRNA regions. To assess the potential consequences of miRNA polymorphisms, we sequenced the genomes of 5 Duroc pigs and, by doing so, we identified 15 miRNA SNPs that were genotyped in the offspring (N = 345) of the five boars. Association analyses between miRNA SNPs and hepatic and muscle microarray data allowed us to identify 4 polymorphisms displaying significant associations. Particularly interesting was the rs319154814 polymorphism (G/A), located in the apical loop of the ssc-miR-326

precursor sequence. This polymorphism is predicted to cause a subtle hairpin rearrangement that improves the accessibility to processing enzymatic factors.

Conclusions

Porcine miRNA genes show a reduced variability, particularly in the seed region which plays a critical role in miRNA binding. Although it is generally assumed that SNPs mapping to the seed region are the ones with the strongest consequences on mRNA expression, we show that a SNP mapping to the apical region of ssc-miR-326 is associated with the mRNA expression of several of its predicted targets. This result suggests that porcine miRNA variability mapping within and outside the seed region could have important regulatory effects on gene expression.

Background

Mature microRNA transcripts (miRNAs) are short (~22 nt) non-coding RNAs which play an essential role in the regulation of gene expression [1]. During the biogenesis of miRNAs, one strand of the miRNA duplex binds to the guide-strand channel of an Argonaute protein forming an miRNA-induced silencing complex (miRISC) with the ability of repressing mRNA expression through binding to specific 3'UTR target sites [2]. In postembryonic cells, this repressor mechanism mainly acts by destabilizing the mRNA through decapping and poly(A)-tail shortening [3,4] and less often by hindering translation [2]. The binding of the miRNA to its 3'UTR target site depends critically on the sequence of the seed region, which encompasses nucleotides (nt) 2-8 from the 5'end of the miRNA and interacts with the target site through Watson-Crick pairing [1]. Polymorphisms located within the seed region have the potential to cause strong effects on target recognition, due to the emergence of novel target sites complementary to the mutated seed and the ablation of canonical wild-type miRNA-mRNA interactions [5,6]. Additionally, polymorphisms affecting miRNA binding sites in the 3'UTR of targeted mRNAs can also notably influence the miRNA-dependent expression of these genes, [5,7,8]. Nevertheless, imperfect seed matches can still be compensated by nucleotides 13-16 of the miRNA, thus providing additional anchor pairing to the seed region [2]. Other sites relevant for miRNA processing and function are the basal UG, flanking CNNC and apical UGU motifs [9], as well as a mismatched GHG motif [10], all of which contribute to facilitate miRNA processing [2].

Saunders et al. (2007) [5] investigated the variability of 474 human miRNA genes and found that SNP density within such loci is lower than that of surrounding regions. Moreover, they found that ~90% of human miRNA genes do not contain

polymorphisms, and the majority of SNPs mapping to miRNAs are outside the seed region, thus evidencing that the variability of this critical functional motif evolves under strong selective constraints [5]. Indeed, polymorphisms in the first 14 nucleotides of the mature miRNA, and particularly those within the seed region, might abrogate the binding of the miRNA to its 3'UTR targets, leading to an extensive rewiring of the miRNA-mediated regulatory network and, in some instances, to detrimental consequences [11]. Mammalian miRNA knockouts often display abnormal phenotypes, reduced viability and clinical disorders [2], although functional redundancy amongst miRNA family members might mitigate to some extent the severity of such manifestations [2]. Polymorphisms within miRNA loci laying outside the seed region can also affect the processing and stability of miRNAs during their maturation and loading into the functional silencing complex [12,13]. Moreover, Saunders et al. (2007) [5] showed that a broad array of predicted miRNA target sites in the 3'UTR of mRNAs are polymorphic, a finding that suggests that purifying selection on these regions is less intense than in miRNA genes [11].

Wild boars emerged as a species in Southeast Asia 5.3–3.5 Mya and dispersed westwards until reaching Europe 0.8 Mya, leading to the establishment of two highly divergent Asian and European gene pools with different levels of variability [14,15]. The wild ancestors of pigs were independently domesticated in the Near East and China 10,000 YBP [14]. Domestic pigs spread worldwide, becoming one of the most important sources of animal protein for humans and diversifying in an extensive array of breeds with distinct morphological and productive features [16]. Pig phenotypes might be explained, at least partially, by modifications in the microRNA regulation of gene expression [17]. Here, we wanted to investigate the patterns of variability of miRNA genes in European and Asian wild boars and domestic pigs. Moreover, we

aimed to elucidate the potential associations of miRNA polymorphisms with the expression profiles of protein-coding genes and complex phenotypes in domestic pigs.

Methods

Characterizing the polymorphisms of miRNA genes and their binding sites in a worldwide sample of pig and wild boar genome sequences

Retrieval of Porcine Whole-Genome Sequences

Whole-genome sequences from a total of 120 wild and domestic pigs (*Sus scrofa*) were retrieved from the NCBI Sequence Read Archive (SRA) database (https://www.ncbi.nlm.nih.gov/sra). Detailed information about each sequenced porcine sample is available at **Additional file 1: Table S1**. The 120 selected genome sequences were classified into Asian domestic pigs (ADM, N = 40), Asian wild boars (AWB, N = 20), European domestic pigs (EDM, N = 40) and European wild boars (EWB, N = 20). More specifically, the European and Asian domestic individuals encompassed a selection of different representative porcine breeds, i.e. Meishan, Tongchen, Jinhua, Rongchan, Wuzhishan, Tibetan, Sichuan, Hetao, Minzhu, Bamaixang and Laiwu for ADM population, and Pietrain, Mangalitza, Iberian, Duroc, American Yucatan (from America, but with a European origin), Yorkshire, Landrace, Hampshire and Large-white for EDM population, each of them including ~1-7 individuals, depending on the availability of good-quality whole-genome sequenced samples. Regarding European and Asian wild boars, samples were selected according to their location of origin, spanning a broad proportion of Europe and the Far East (**Additional file 1: Table S1**).

The European pool contained one sample from the Near East. Raw data in SRA format were downloaded from SRA public repositories and converted into fastq format by using the fastq-dump 2.8.2 tool available in the SRA-toolkit package (ncbi.github.io/sra-tools/).

Whole-genome sequence data processing and calling of single nucleotide polymorphisms

Fastq paired-end files generated from SRA data were quality-checked filtered and any remaining sequence adapters were trimmed by making use of the Trimmomatic software (v.0.36) with default parameters [18]. Trimmed paired-end sequences that successfully passed previous filters were aligned against the *Sus scrofa* reference genome (Sscrofa11.1) [19] with the BWA-MEM algorithm [20] and default settings. Sequence alignment map (SAM) formatted files were sorted and transformed into binary (BAM) formatted files and PCR duplicates were subsequently marked and removed to perform INDEL realignment with the Picard tool (https://broadinstitute.github.io/picard/). Base quality score recalibration (BQSR), variant calling and quality hard filtering of the variants called by default were implemented with the Genome Analysis Toolkit (GATK v.3.8) [21], according to GATK best practices recommendations. Individual gVCF formatted files, including both polymorphic and homozygous blocks, were generated in this way, and they were subsequently merged into separate multi-individual variant calling format (VCF) files containing single polymorphic and INDEL sites, respectively.

Cataloguing the repertoire of single nucleotide polymorphisms in miRNA genes and the 3'UTRs of mRNA genes

Single nucleotide polymorphisms (SNPs) mapping to annotated porcine miRNA loci (N = 370) were retrieved by making use of the curated Sscrofa11.1 annotation for miRNA regions available at miRCarta v1.1 database [22]. Additionally, annotated mature miRNA loci (N = 409) inside miRNA genes (N = 370) were retrieved and SNPs within miRNA genes residing in mature and seed regions (2nd to 8th positions at 5' end in the mature miRNA) were identified. Moreover, we expanded our analysis to putative miRNA target sites in the 3'UTR of protein-coding mRNAs. To this end, we downloaded the current annotation of the 3'UTR of porcine mRNA transcripts from Ensembl repositories (http://www.ensembl.org/info/data/ftp/index.html) and interrogated the corresponding set of sequences against seed regions of miRNAs available at miRBase database [23]. Seed sequences were reverse-complemented and searched along the 3'UTR sequences of mRNA genes by making use of the SeqKit toolkit [24]. All 7mer-m8 canonical seed pairing (2nd to 8th 5' nts perfect match) as well as 8mer seed pairing (adding an additional adenine to the 1st position of the matched region within the mRNA 3'UTR, which is used as a complementary anchor to the Argonaute miRISC complex), were identified for each miRNA seed and the corresponding putative miRNA-mRNA target pairs were defined. Subsequently, the genomic location of both 7mer-m8 and 8mer matching regions were determined, and SNP variants residing in such targeted sites were retrieved. Principal components analyses (PCA) based on autosomal whole-genome SNPs, as well as on autosomal SNPs inside miRNAs genes, 8mer and 7mer-m8 3'UTR sites and whole 3'UTRs were performed with the smartPCA software [25]. When implementing the whole-genome PCA, polymorphisms complying with the following parameters were retained: 1) minimum allele frequency (MAF) > 0.05, 2) Hardy-Weinberg equilibrium exact test *P*-

value > 0.001. In contrast, in the PCA based on miRNA SNPs no filtering was performed and all retrieved autosomal miRNA-related polymorphic sites were included.

Identifying signatures of selection in miRNA genes

Whole-genome SNP data sets generated for the four ADM, AWB, EDM and EWB populations were filtered on the basis of the following parameters: 1) MAF > 0.01, 2) Hardy-Weinberg equilibrium exact test *P*-value > 0.001, and 3) missing rate < 0.9. Only bi-allelic SNPs were considered. The Weir and Cockerham's F_{ST} statistic [26] was employed for evaluating the presence of signatures of selection between the four defined contrasts (i.e., ADM vs AWB, EDM vs EWB, ADM vs EDM and AWB vs EWB). In this regard, the F_{ST} estimate can be defined as a measure of population differentiation due to changes in genetic structure based on the variance of allele frequency. The wcFst calculator implemented in the Genotype Association Toolkit (GPAT++, https://github.com/vcflib/vcflib/) was used to estimate the magnitude of F_{ST} in each of the defined pairwise contrasts. Statistical significance was determined with a permutation test (1,000 permutations). An empirical *P*-value ≤ 0.05 was set as the cutoff of significance.

Frequency and distribution of SNPs in miRNA genes

To assess the patterns of SNP distribution across miRNA-related loci, polymorphic sites mapping to miRNA genes were classified according to their location, i.e., SNPs inside the seed region of mature miRNAs were flagged as "*seed*", whereas the rest of SNPs located in mature miRNAs but outside the seed were classified as

"*mature*". The remaining SNPs inside miRNA loci, which did not belong to any of the two previous categories, were assigned to the "*precursor*" class. Allele frequencies for reference and alternative alleles were separately calculated for the whole set of porcine samples (N = 120), as well as for each defined porcine population separately (**Additional file 2: Table S2**).

Additionally, SNPs inside mature miRNAs but outside the seed region were assigned to the following subtypes: 1) "*anchor*" (1st position at 5' end), and 2) "*supplemental pairing*" (13th to 18th position from 5' end). In order to calculate SNP density in precursor, mature and seed miRNA regions, we first calculated the total length of each of these regions. *Seed length* was obtained by considering 7 positions in each of the annotated mature miRNAs (N = 409) mapping to the 370 miRNA loci considered in this study. The total length of the 409 mature miRNAs (8,916 bp) was calculated, and total *seed length* (7 bp × 409 = 2,863 bp) was subtracted from this number to obtain the *mature length* (6,053 bp), which corresponds to the total length of all mature miRNA sequences excluding the seeds. *Precursor length* (22,229 bp) was calculated by subtracting *seed length* (2,863 bp) and *mature length* (6,053 bp) from the total length (31,145 bp) of all miRNA loci (N = 370). Then, SNP density in each of these regions was calculated as follows:

$$D = \frac{N_{snp} \times 100_{bp}}{N_r}$$

where N_{snp} is the total number of miRNA SNPs (N = 285), as well as the number of detected SNPs in each of the defined regions, (i.e. 221, 52 and 12 polymorphisms in

precursor, mature and seed regions, respectively); N_r is the total nucleotide length of all miRNA loci (31,145 nts), as well as the total number of nucleotides in miRNA loci belonging to either "*precursor*" (22,229 bp), "*mature*" (6,053 bp) or "*seed*" (2,863 bp) regions. These calculations would yield the number of SNPs per bp for each category. We decided to adjust this estimate to a window of 100 bp, and this is why N_{snp} is multiplied by 100 in the above formula. Furthermore, the average SNP density at the whole-genome level was compared with the SNP density within miRNA genes. To this end, we retrieved all detected SNPs in our whole-genome sequenced dataset (N = 120) and applied the SNP density formula considering the length of Sscrofa11.1 whole-genome assembly (~2.48 Gb) available at Ensembl repositories (http://www.ensembl.org/info/data/ftp/index.html).

Investigating the association of miRNA polymorphisms with gene expression and phenotype data recorded in Duroc pigs

Whole-Genome sequencing of five Duroc pigs

In 2003, five Duroc boars were selected and used as founders of a half-sib population of purebred Duroc pigs devoted to the production of high-quality cured ham and phenotypic and genotypic recording for subsequent analyses. In the current work, we aimed to characterize the variability of these five individuals by whole-genome sequencing, identify SNPs located in miRNA genes and investigate their association with mRNA levels and lipid phenotypes recorded in their offspring (Lipgen population, N = 345). Extraction of DNA was performed for the five Duroc founders and genome sequencing was carried at the Centro Nacional de Análisis Genómico (CNAG, Barcelona, Spain). Paired-end multiplex libraries were prepared according to the

instructions of the manufacturer with the KAPA PE Library Preparation kit (Kapa Biosystems, Wilmington, MA). Libraries were loaded to Illumina flow-cells for cluster generation prior to producing 100 bp paired-end reads on a HiSeq2000 instrument following the Illumina protocol. Base calling and quality control analyses were performed with the Illumina RTA sequence analysis pipeline according to the instructions of the manufacturer. Quality-checked filtered reads were mapped to the *Sus scrofa* genome version 11.1 and processed for SNP calling according to GATK best practices recommendations and the protocol implemented for Asian/European wild boars and domestic pigs.

Description of the Duroc Lipgen population and phenotype recording

A total of 350 Duroc barrows, sired by the five Duroc founder boars mentioned before were used as a resource population (Lipgen population [27,28]) to further investigate the segregation of SNPs affecting miRNA loci and to evaluate their association with the mRNA expression of miRNA predicted targets. The five sequenced boars were mated with 400 sows in three different farms and one offspring per litter was selected for phenotypic recording (only 350 individuals provided valid records). All selected piglets were weaned, castrated and subsequently fattened at IRTA pig experimental farm in Monells (Girona, Spain) under intensive standard conditions in four contemporary batches. Once they reached ~122 kg of live weight (~190 days of age), they were slaughtered in a commercial abattoir following recommended animal welfare guidelines. After slaughtering, tissue samples from *gluteus medius* (GM) and *longissimus dorsi* (LD) skeletal muscles and liver were obtained as previously described [29–31].

Total DNA was extracted from each sample following Vidal et al. 2005 [32]. A total of 345 DNA samples from the initial set of 350 pigs were successfully obtained and processed for further genotyping. Total RNA was extracted from a number of GM (N = 89 pigs) and liver (N = 87 pigs) tissue samples obtained from the Lipgen population following the acid/phenol RNA extraction method [33] implemented in the Ribopure isolation kit (Ambion, Austin, TX). Expression mRNA profiles were then characterized by means of hybridization to the GeneChip porcine arrays (Affymetrix Inc., Santa Clara, CA), as reported by Cánovas et al. [29]. Further details about tissue collection, sample selection, RNA isolation and microarray hybridization procedures can be found in [29]. Microarray data pre-processing, background correction, normalization and $\log_2$-transformation of expression estimates were performed with a robust multi-array average (RMA) method per probe [34]. The *mas5calls* function from affy R package [34] was then applied to infer probe intensity significance level in order to detect gene expression above background noise. This function applies a Wilcoxon signed rank-based gene expression presence/absence detection algorithm for labeling expressed probes in each sample. Control probes and those with expression levels below the detection threshold in more than 50% of samples were discarded from further analyses.

Besides gene expression phenotypes, we also obtained lipid-related phenotypes in the Duroc population. Backfat thickness measured between 3rd and 4th ribs and at last rib, ham fat thickness, as well as fatty acids (FA) composition phenotypes from GM and LD skeletal muscle samples (N = 345) were determined as described in [28,30]. Briefly, intramuscular fatty acid (IMF) percentages in GM and LD muscles were calculated with the Near Infrared Transmittance technique (NIT, Infratec 1625, Tecator Hoganas, Sweden), while muscle cholesterol measurements were inferred following Cayuela et al.

[35]. A gas chromatography of methyl esters protocol was used to determine muscle fatty acids composition of saturated (SFA), unsaturated (UFA), monounsaturated (MUFA) and polyunsaturated (PUFA) fatty acids. Weight (kg), backfat and ham fat thickness (mm) were measured on a regular basis prior and after slaughter. Mean and standard deviation values for the Lipgen population (N = 345) and further details about measured phenotypes are described in **Additional file 3: Table S3**.

Genotyping of a panel of single nucleotide polymorphisms mapping to microRNA genes in the Lipgen population

Whole-genome sequencing of the 5 Duroc founders yielded 54 polymorphisms mapping to miRNA loci (**Table 1**). From these, we selected 15 SNPs on the basis of their location at relevant annotated miRNA loci. Selected miRNA SNPs and their flanking regions (60 upstream and downstream nts) were evaluated with the Custom TaqMan Assay Design Tool website (https://www5.appliedbiosystems.com/tools/cadt/; Life Technologies) and genotyped in our purebred Duroc population (N = 345) at the Servei Veterinari de Genètica Molecular of the Universitat Autònoma of Barcelona (http://sct.uab.cat/svgm/en) by using a QuantStudio 12K Flex Real-Time PCR System (Thermo Fisher Scientific, Barcelona, Spain).

Association analyses between miRNA SNPs and mRNA expression and lipid phenotypes

Genotype data from 15 miRNA SNPs (**Table 1**) were processed with the PLINK software [36] in order to generate formatted files for subsequent analyses. The Genome-Wide Efficient Mixed-Model Association (GEMMA) software [37] was used to

implement association analyses between genotyped SNPs and lipid phenotypes and microarray expression data in GM and liver tissues. The following univariate mixed model was used:

$$y = W\alpha + x\delta + u + \varepsilon$$

Where y is the phenotypic vector of recorded phenotypes for each individual; α is a vector indicating the intercept plus the considered fixed effects, i.e. batch effect with 4 categories (all traits), farm of origin effect with 3 categories (all traits) and laboratory of processing with 2 categories (GM and liver microarray expression data). The α vector also includes the regression coefficients on IMF in LD and GM tissues (for LD and GM fatty acid composition traits, respectively), as well as on live and carcass weight (for backfat and ham fat thickness measures before and after slaughter); W corresponds to the incidence matrix relating phenotypes with their corresponding effects; x is the genotype vector for selected miRNA polymorphisms; δ is the allele substitution effect for each polymorphism; u is a vector indicating random individual effects with a n-dimensional multivariate normal distribution MVN_n $(0, \lambda\ \tau^{-1}\ K)$, where τ^{-1} corresponds to the variance of the residual errors, λ is the ratio between the two variance components and K is the known relatedness matrix derived from SNP information; and ε is the vector of residual errors.

With regard to the mRNA expression phenotypes, it is worth to mention that expressed probes mapping to mRNA genes in GM and hepatic tissues were identified

using BioMart databases [38]. Furthermore, probes/genes were filtered based on the following conditions: (1) they contain in their 3'UTR 7mer-m8 sites for any of the polymorphic miRNAs defined in **Table 1** and (2) such interaction has been experimentally validated in humans. To this end, predicted porcine targets on the basis of 7mer-m8 seed matches inferred with the seqkit software [24] were selected. If the analyzed polymorphism affected the seed region, putative targets were predicted according to the novel mutated seed. Moreover, miRNA-mRNA target pairs experimentally validated in humans were obtained from the Tarbase v.8 database [39] (i.e. directly repressed mRNA targets on the basis of CLASH, PAR-CLIP, HITS-CLIP and Luciferase assays). When no validated human targets were available for any of the miRNAs harboring the analyzed SNPs, only predicted 7mer-m8 targets were used. This was the case of ssc-miR-9792 and ssc-miR-9802 (**Table 1**).

The existence of associations between lipid-related traits and gene expression data with the analyzed SNPs were assessed on the basis of the estimated allele substitution effects (β). More specifically,. the alternative hypothesis $H_1 : \beta \neq 0$ was contrasted against the null hypothesis $H_0 : \beta = 0$ with a likelihood ratio test.

The statistical significance of the associations between miRNA SNPs and lipid and mRNA expression phenotypes were assessed by using a false discovery rate (FDR) approach [40] which corrects for multiple testing.

miRNA structural inference with RNAfold

The potential effects of SNPs on pri-miRNA structural conformation as well as on miRNA stability and processing availability were predicted with the RNAfold tool

from the ViennaRNA Package 2.0 [41] based on the Boltzmann-weighted centroid structure ensemble of the RNA sequence [42].

Results

microRNA-related polymorphisms follow differential segregation patterns in pigs and wild boars from Europe and Asia

Roughly, 58.54 million SNPs were identified with the GATK haplotype caller tool [21] in a data set comprising 120 whole-genome sequences from 40 EDM, 40 ADM, 20 EWB and 20 AWB individuals (**Additional file 1: Table S1**) retrieved from public repositories. The majority of these SNPs were biallelic (98.68%) but 770,806 of them showed 3 or more alleles. Alternative allele frequencies were consistently high (> 0.5) for 9.25% of variants, whereas low (between 0.05 and 0.01) and very low (< 0.01) alternative allele frequencies were detected in 27.85% and 19.62% of SNPs, respectively.

After filtering, 19,720,314 autosomal whole-genome SNPs were selected for assessing population structure based on PCA clustering techniques. The spatial representation of whole-genome data principal components showed a strong genetic differentiation among Asian and European populations (**Fig. 1A**). In contrast, domestic pigs and wild boars, and particularly those with a European origin, did not show such stark divergence. Asian pigs and wild boars displayed some level of genetic differentiation and they were more diverse than their European counterparts.

With regard to miRNA variability, the 370 porcine miRNA genes annotated in the manually curated miRCarta v1.1 database [22] were selected and SNPs within those regions were retrieved. A total of 285 SNPs residing in 139 (37.56% of the total count) miRNAs were identified (**Additional file 2: Table S2**), implying that most of miRNAs are monomorphic. The majority of these 139 miRNA loci (76.98%) presented 1-2 SNPs located inside their predicted genomic boundaries, while 18.70% contained between 3 up to 5 variants, and 4.32% of them displayed more than 7 polymorphisms (**Additional file 4: Fig. S1**). Only 43 miRNA SNPs (15.09%) were shared among EDM, ADM, EWB and AWB populations (**Additional file 2: Table S2**), showing alternative alleles in at least one of the analyzed individuals in each group. The number of SNPs segregating in each of the four defined groups were 129 (EDM), 201 (ADM), 76 (EWB) and 172 (AWB), respectively (**Additional file 2: Table S2**). With regard to precursor and mature regions, 41 and 2 SNPs where shared amongst the four populations under consideration, respectively (**Additional file 5: Fig. S2A** and **S2B**). None of the SNPs in the seed regions were shared by the four porcine populations **Additional file 5: Fig. S2C)**. Only three miRNA variants were found in the European data set but not in the Asian one. In strong contrast, 55 miRNA SNPs were detected in the Asian data set but not in the European one. Principal component analyses based on identified autosomal miRNA SNPs (N = 260) showed the existence of a poor differentiation between pigs and wild boars (**Fig. 1B**), while the genetic divergence between European and Asian individuals was still apparent using whole-genome autosomal SNPs (**Fig. 1A**).

When we analyzed population structure based on whole-genome autosomal 3'UTR SNPs (N = 709,343 SNPs, **Fig. 1C**), 3'UTR 7mer-m8 site SNPs (N = 107,196 SNPs, **Fig. 1D**) and 3'UTR 8mer site SNPs (N = 33,511 SNPs, **Fig. 1E**), genetic differentiation between Asian vs European populations was evident, in close

concordance with results shown in **Fig. 1A** and **1B**. However, we also detected a more pronounced differentiation between domestic pigs and wild boars, and this observation was particularly true for Asian pigs and wild boars (**Fig. 1C-D**).

The analysis of European and Asian populations shows reduced variability in porcine microRNAs

About 47.76%, 57.36%, 44.77% and 36.84% of miRNA SNPs showed alternative allele frequencies ≤ 0.1 in the ADM, EDM, AWB and EWB populations, respectively (**Fig. 2, Additional file 2: Table S2**). Variations located at mature miRNA and seed regions were enriched in rare or very rare variants when compared to the variability of miRNA precursor regions (**Additional file 2: Table S2**), with average alternative allele frequencies of ~0.1 for mature and seed miRNA polymorphisms. In contrast, the average alternative allele frequency observed for SNPs in precursor areas was ~0.15.

Moreover, the observed SNP density adjusted to 100 bp for miRNA precursor, mature and seed regions consistently followed the order *precursor* > *mature* > *seed* when we considered all whole-genome sequenced pigs (N = 120). Indeed, ~1 SNP per 100 bp was detected in precursor regions, whereas ~0.86 and ~0.42 SNPs per 100 bp were observed in the mature and seed regions, respectively (**Fig. 3A**). These results implied a slight difference in SNP density between precursor and mature regions, while for seed regions, which are critical determinants of miRNA-mRNA interaction, the observed SNP density was ~2.4 fold lower than in precursor regions. This differential distribution of the SNP density across miRNA regions (*precursor* > *mature miRNA* > *seed*) was also observed in each of the analyses performed in the ADM, EDM, AWB

and EWB groups (**Fig. 3A**). With regard to variants located within mature miRNAs (N = 64), both inside (N = 12) and outside seed regions (N = 52), their observed distribution within the whole body of the mature miRNA (~22 nts) showed a characteristic pattern (**Fig. 3B**): among all the detected SNPs, the 1st position of the mature miRNA 5' end, which binds to the MID domain of the Argonaute protein in the miRISC complex, exhibited a SNP density of ~0.49 SNPs per 100 bp. Such scarcity in polymorphic sites was also observed when considering the next 2nd to 8th positions in the mature miRNA sequence (seed region), where up to ~0.73 SNPs per 100 bp were observed in the 6th position of the mature miRNA, and an average of ~0.42 SNPs per 100 bp where found among all analyzed miRNA seeds. In contrast, the interval comprising positions 9th to 12th showed an average SNP density of ~0.98 SNPs per 100 bp. Regarding the next 13th to 18th positions of the mature miRNA, which roughly corresponds to a functional region providing additional anchor pairing to the seed region, we observed a decreased SNP density, more prominent at positions 16th to 17th (**Fig. 3B**). Finally, an increased SNP density was found at positions 19th to 22nd.

We have compared the SNP density of miRNA loci (seed, mature miRNA and precursor miRNA alltogether) with the global genomic SNP density. As previously mentioned, we identified ~58.54 million SNP sites across a whole length of 2,478 million bp in the Sscrofa11.1 assembly, i.e. the average estimated whole-genome SNP density was ~2.36 SNP per 100 bp. Conversely, the average SNP density in miRNA loci was ~0.92 SNPs per 100 bp. These results indicated that the overall probability of finding a SNP in any region of the pig genome was approximately 2.5-fold higher than inside miRNA loci.

We also carried out a selection scan based on the F_{ST} statistic in each one of the four defined contrasts between porcine populations (i.e., ADM vs AWB, EDM vs EWB,

ADM vs EDM and AWB vs EWB) with the aim of identifying positive selection signals coinciding with the location of miRNA genes. After filtering, a set of 6,408,611 SNPs were retrieved, from which 206 SNPs mapped to miRNA loci. Very few selection signals inside miRNA genes were detected with this approach (**Additional file 6: Table S4**). A SNP located at the precursor region of ssc-miR-4335 showed a significant F_{ST} in the ADM vs AWB contrast. Besides, the rs330981259 polymorphism, located in the precursor region of ssc-miR-9835 segregated in all populations except in EWB (**Additional file 2: Table S2**). This SNP showed a significant F_{ST} value when contrasting AWB vs EWB, as shown in **Additional file 6: Table S4**.

Statistics of the whole-genome sequencing of five Duroc boars

As previously explained, we sequenced the genomes of five Duroc pigs which founded a population of 350 offspring with the goal of identifying SNPs in miRNA genes and investigating their association with mRNA expression and lipid phenotypes. Mean coverage values of the five pig genomes ranged from 37.67× to 46.6×, with more than 98.6% of the genome bp covered by at least 10 reads in all five samples, and 96.71% of the bp covered by at least 15 reads. Details about coverage and genome mapping are shown in **Additional files 7** and **8: Table S5** and **Fig. S3**. After performing variant calling on mapped reads, a total of 13,839,422 SNPs passed the established quality filters, whereas 3,721,589 insertions and deletions (INDELs) were detected. From these, 54 SNPs and 5 INDELs were located inside miRNAs (N = 370) annotated in the Sscrofa.11.1 genome assembly according to the miRCarta database [22]. Moreover, a total of 1,643,861 INDELs (44.17%) and 6,034,548 SNPs (43.60%) resided inside annotated protein coding loci.

The rs319154814 polymorphism in the apical loop of ssc-miR-326 is associated with the mRNA expression of several of its putative gene targets

From the set of 15 SNPs listed in **Table 1**, only 4 SNPs showed significant associations with liver (N = 87) and/or GM muscle (N = 89) expression data. It is important to emphasize that we only considered probes corresponding to genes fulfilling two conditions: (1) Their 3'UTRs contain 7mer-m8 sites matching to positions 2-8 of at least one of the mature miRNAs under consideration, and (2) the corresponding miRNA-mRNA interaction has been experimentally validated in humans, as reported in the Tarbase v.8 database [37] (**Additional file 9: Table S6**). When we analyzed the association between the rs319154814 (G/A) polymorphism located in the apical loop of ssc-miR-326 and gene expression data (**Table 2**), several significant results were obtained after multiple testing correction (*q*-value < 0.1). More specifically, we detected seven significant associations between this SNP and the hepatic mRNA expression of experimentally confirmed 7mer-m8 targets of this miRNA (**Table 2**). For instance, the protein phosphatase 1 catalytic subunit γ (*PPP1CC*), the cellular FLICE-like inhibitory protein (*CFLAR*), the splicing factor 3A subunit 3 (*SF3A3*) and the Follistatin-like 1 (*FSTL1*) mRNAs showed the most significant associations (**Table 2**). No significant associations were found for GM tissue expression data. The expression levels of six significantly associated mRNAs targeted by ssc-miR-326 (**Table 2**) were reduced in pigs homozygous for the mutated allele (N = 32), as depicted in **Fig. 4**.

We further aimed to predict the potential consequences of the rs319154814 SNP on the secondary structure of the ssc-miR-326 transcript by means of the RNAfold centroid-based algorithm [41]. According to structural folding inference (**Fig. 5**), the presence of the rs319154814 polymorphism would have an impact on the steric forces contributing to the adequate base-pairing selection of the hairpin, thus generating an

unpaired bulge with a considerable size at the base of the pri-miRNA folded structure generated immediately after transcription (**Fig. 5**). More importantly, the accessibility of the two observed contiguous CNNC motifs (located at positions -16/-21 from the miRNA gene boundaries and inside the originated bulge at the base of the pri-miRNA) would be facilitated by the open unpaired bulge, hence improving its recognition by the Serine-rich splicing factor 3 (SRSF3) protein, a relevant miRNA processing factor that binds to the CNNC motifs of the pri-miRNA sequence and contributes to the correct positioning of Drosha endonuclease for miRNA maturation.

The rs322514450 polymorphism in the seed of ssc-miR-9792-5p is associated with the mRNA expression of several of its potential targets

Polymorphisms in the seed of mature miRNAs might alter the repertoire of their potential targets by abolishing existing miRNA-mRNA interactions and creating new ones. Only one SNP (rs322514450, G/A) located in the seed region of the ssc-miR-9792-5p segregated in our Duroc population. After genotyping, the observed allele frequency for the alternative mutated allele was of 19.89% (**Table 1**). In this case, we considered as ssc-miR-9792-5p targets those probes corresponding to mRNAs with predicted 7mer-m8 binding sites (we did not consider experimental validation in humans due to a lack of information in Tarbase v.8 database, **Additional file 9: Table S6**). Significant associations between this SNP and the expression levels of targeted mRNAs were detected (**Table 2**). The levels of the retinaldehyde binding protein 1 (*RBLP1*) mRNA in the liver and of the nudix hydrolase 6 (*NUDT6*) mRNA in the GM muscle were significantly associated with rs322514450 genotypes (**Table 2**). Moreover,

we observed a decreased expression of *RBLP1* and *NUTD6* mRNAs in pigs homozygous for the mutated allele (**Fig. 6**).

Polymorphisms outside the mature microRNA region are associated with mRNA expression

The rs333787816 (T/C) polymorphism, located in the precursor region, immediately downstream the mature ssc-miR-23a sequence, was significantly associated with several experimentally confirmed targeted genes in the GM muscle tissue (**Table 2, Additional file 9: Table S6**). From these, it is worth mentioning the Argonaute RISC component 1 (*AGO1*) and the Argonaute RISC catalytic component 2 (*AGO2*) mRNAs. Both *AGO1* and *AGO2* genes showed lower mRNA expression levels in homozygous CC pigs with respect to their TT and TC counterparts (**Table 2**). On the other hand, the rs335924546 (C/T) variant located at ssc-miR-1224 was significantly associated with the mRNA expression of the E3 ubiquitin ligase makorin ring finger protein 1 (MKRN1) in the liver. Pigs homozygous for the alternative allele of rs335924546 polymorphism showed a reduced expression of the *MKRN1* transcript (data not shown).

Porcine lipid phenotypes are associated with the genotypes of miRNA genes

We also evaluated the association between miRNA SNPs and several lipid-related phenotypes recorded in the Lipgen population (**Additional file 3: Table S3**). Only the rs319154814 variant inside ssc-miR-326 was significantly associated (*q*-value < 0.1) with lipid traits (**Table 3**, **Additional file 10: Table S7**). More specifically, we

found significant associations with the myristic acid (C14:0) content in both LD and GM muscles, as well as with the gadoleic acid (C20:1) content and the ratio between PUFA and MUFA in the LD muscle (**Table 3, Additional file 10: Table S7**). We also observed several associations between the rs319154814 SNP and fatty acid composition traits but they were significant only at the nominal level (*P*-value < 0.01), as shown in **Table 3**. Other apical loop SNPs like (1) rs81349391 at ssc-miR-339 and (2) rs344472188 at ssc-miR-130a were significantly associated at the nominal level (*P*-value < 0.01) with palmitic acid (C16:0) content and SFA and UFA proportion in the GM muscle, as well as with backfat thickness. Besides, a SNP located in the precursor 3' stem of ssc-miR-30d showed a nominally significant association with the content of arachidic acid (C20:0) (**Table 3**). Other relevant examples of significant associations at the nominal level were, for instance, those between the rs322514450 (G/A) polymorphism in the seed of the ssc-miR-9792-5p and several fatty acid composition phenotypes in the LD muscle, such as palmitic (C16:0), palmitoleic (C16:1), linoleic (C18:2), α-linolenic (C18:3) and arachidonic acids (C20:4) content. Besides, the rs334590580 (T/C) SNP located at the precursor stem region of ssc-miR-335 was associated with palmitic and arachidic acids content in GM tissue (**Table 3**).

Discussion

Divergent patterns of variation for microRNA and 3'UTR polymorphisms in Asian and European pigs and wild boars

The PCA revealed the existence of a detectable genetic differentiation between Asian and European populations (**Fig. 1**), with the later showing reduced levels of diversity when compared to the former. Groenen et al. (2012) [43] investigated the variability of pig genomes and found that Asian pigs and wild boars are more diverse than their European counterparts and that both gene pools split during the mid-Pleistocene 1.6–0.8 Myr ago. Calabrian glacial intervals probably favored a restricted gene flow between these two pools [43]. The high variability of Asian populations could be explained by the fact that *Sus scrofa* emerged as a species in Southeast Asia (5.3-3.5 Mya) and then dispersed westwards until reaching Europe around 0.8 Mya [15]. This initial founder effect combined with the occurrence of strong bottlenecks reduced the genetic diversity of European wild boars [43].

While genetic differentiation between wild boar and pig populations is clearly discernible in **Fig. 1A** (total SNP data set), this is less evident in the PCA based on miRNA SNPs (**Fig. 1B**), probably because the low number (285 SNPs) of markers employed in this analysis limits the resolution with which population differentiation can be detected. We have investigated whether there is any miRNA SNP under positive selection by using an F_{ST} test (**Additional file 6: Table S4**). According to the results of the selection scan, only 2 SNP yielded significant F_{ST} values, suggesting that the majority of miRNA genes have not been targeted by positive selection. This result implies that the genetic differentiation that we observe in **Fig. 1B** is mostly produced by non-neutral evolutionary forces.

We have also found that the degree of population differentiation between Asian domestic pigs and Asian wild boars increases when PCAs are built on the basis of SNPs located in the 3'UTR (709,343 SNPs), 3'UTR 7mer-m8 sites (107,196 SNPs) or 3'UTR 8 mer (33,511 SNPs) sites (**Figs. 1C-E**). The potential effects of 3'UTR SNPs are the

modulation of mRNA expression, secondary structure, stability, localization, translation, and binding to miRNAs and RNA-binding proteins [44], so in general they are not expected to have drastic consequences on gene function. In humans, only a small fraction (3.7%) of the polymorphisms associated with phenotypes reside in 3'UTRs [44]. Thus, it is reasonable to assume that the intensity of purifying selection is lower in 3'UTRs than in protein-coding regions, meaning that 3'UTRs evolve faster and accumulate a larger fraction of recent polymorphisms contributing to population differentiation. In this regard, Bachtiar et al. (2019) [45] reported that ~95% of population-differentiated polymorphisms reside in non-genic regions, compared to the proportion of all SNPs (58%) found in non-genic regions.

The majority of microRNA polymorphisms are shared by at least two porcine populations

The percentage of SNPs located within precursor miRNAs that were shared by all four populations was approximately 18.55%, while the percentages of SNPs found exclusively in ADM (13.57%), AWB (12.67%), EDM (10.40%) and EWB (5.43%) populations were lower (**Additional file 5: Fig. S2A**). Moreover, the examination of **Figs. S2B** and **S2C** indicates that in mature miRNAs and, to a lesser extent, in seed regions, the percentages of group-specific SNPs exceed that of SNPs shared by all four populations, being ADM pigs the population which displays a higher percentage of exclusive SNPs. By using a data set of 133 porcine whole-genome sequences, Bianco and coworkers [46] computed the percentages of total SNPs that were private to ADM (4.8%), AWB (23.4%), EDM (12.6%) and EWB (2.6%) populations, as well as the fraction of SNPs shared by all four of them (11.7%). In close similarity with our results,

the majority of SNPs (> 80%) were shared by two or more populations, and EWB pigs was the population that showed a lower percentage of private SNPs. It should be noticed, however, that the number of genomes used in our study and in the one of Bianco et al. (2015) [46] is moderate (~120-130 genomes), so these estimates might change if additional whole-genome sequences were incorporated to the investigated data sets. The sharing of SNPs between the four populations under consideration could be due to common ancestry as well as to the occurrence of an extensive gene flow between (1) domestic vs wild pigs, and (2) European vs Asian populations [47]. Reproductive isolation between wild and domestic pigs was disrupted to some extent during and after domestication in Asia and Europe [47]. Even in present times, significant levels of domestic pig introgression have been reported in northwest European [48], Sardinian [49] and Romanian [50] wild boars. On the other hand, there are abundant evidences of domestic pig exchanges between Europe and Asia [47,51,52]. The high frequency of Asian mitochondrial variants in European pigs has been interpreted in the light of the massive importation of Chinese sows into England in the 18th-19th centuries with the goal of increasing fatness and reproductive efficiency [51–53]. Conversely, the importation of improved European breeds into Asia explains the admixture of certain Chinese breeds with European blood [54]. In close agreement with our results, Bianco et al. (2015) identified Asian wild boars as the population with a higher percentage of private SNPs [46]. As previously said, the most likely explanation for this result is the Asian origin of *Sus scrofa* and the occurrence of a strong founder effect in Europe [15,43].

Low SNP density in microRNA genes and lack a uniform SNP distribution across sites

We have found that, in general, miRNA loci have a substantially lower SNP density than that of the global pig genome (2.5-fold reduction), a result that is concordant with data presented by Saunders et al. (2007) [5]. This reduction in SNP density was stronger in the seeds compared with mature and precursor regions (**Fig. 3A**). The low variability of miRNA genes, a feature that was particularly evident in the seed region, is probably due to the intense effects of purifying selection. Indeed, the importance of the miRNA seeds is revealed by the high conservation of their sequence across species [17,55,56], as this sequence ultimately determines the success of the miRNA-mRNA interactions [2]. In our study, a total of 221, 52 and 12 SNPs were found in precursor, mature and seed regions within miRNA loci, respectively (**Additional files 2** and **5: Table S2, Fig. S2**). Gong et al. (2012) [57] described the existence of 40% polymorphic miRNAs in the human genome but only 16% of them displayed more than one SNP. In a more recent study, He et al. (2018) [58] reported 1,879 SNPs in 1226 (43.6%) human miRNA seed regions, and 97.5% of these polymorphisms had frequencies below 5%, results in accordance with the overall frequency distribution of miRNA SNPs in the European and Asian populations analyzed in the current work (**Fig. 2**). He et al. (2018) also demonstrated that 1,587, 749, 340, 102, 31, and 4 miRNAs harbored zero, one, two, three, four, and five SNPs, respectively, in their seed regions, reflecting that mutations in this critical functional region are not well tolerated [58]. This distribution is similar to the one that we have observed in domestic pigs and wild boars, with 81, 31, 11, 9 and 5 miRNAs harboring one, two, three, four and five SNPs (**Additional file 4: Fig. S1**). Only 4 and 2 miRNAs showed a total of seven and ten polymorphisms within their sequences.

We have found a trend towards a decreased variability in porcine miRNA genes, as well as in their precursor, mature and seed regions. Moreover, we have detected a high heterogeneity in the SNP density across mature miRNA sites (**Fig. 3B**). Gong et al. (2012) [57] showed that SNPs tend to concentrate in the middle region of the mature miRNA gene rather than in its 5' and 3'ends, but they also detected a non-uniform distribution of variability across the mature miRNA sequence. Moreover, the same authors described an increased SNP density at positions 9 and 15 of the mature miRNA, a result that closely matches ours (**Fig. 3B**) However, we have also identified an elevated number of polymorphic sites at positions 11, 19 and 20, a finding that does not match human data presented by Gong et al. (2012) [57].

Several of the sites showing a reduced variability in porcine miRNAs exert critical functions (**Fig. 3B**). For instance, the 1st nucleotide of mature miRNAs plays an important role in the loading process of the mature miRNA within the Argonaute protein in the miRISC complex. Indeed, the 5' end (1st nucleotide) of the mature miRNAs is thought to interact with a structural pocket of the Argonaute MID domain, anchoring the miRNA in its position and thus being inaccessible to pairing with targeted mRNAs [59]. Nucleotides 2nd to 8th in the mature miRNA correspond to the seed, where we found a consistently reduced SNP density (**Fig. 3A** and **3B**) compared with other miRNA regions. This result was expected because this region has a crucial role in shaping the interaction between the mature miRNA and short sequences within 3'UTRs of targeted mRNAs. The presence of polymorphic sites inside the seed region has the potential to disrupt the proper miRNA-mRNA pairing and thus alter biologically relevant regulatory pathways, which tend to be evolutionary conserved [55]. Alternatively, polymorphisms in the seed might favor the emergence of novel miRNA-mRNA interactions, thus modifying gene regulatory networks.

In contrast with the first 5' position of the miRNA and the seed region, we found a high SNP density in positions 9th to 12th, which do not contribute substantially to miRNA target recognition [2] (**Fig. 3B**). Following positions 13th to 16th have been previously described to facilitate 3'-compensatory pairing between the mature miRNA and targeted 3'UTRs [60], although only at marginal levels [61]. Nevertheless, in the region comprising 13th to 18th nucleotides within the mature miRNA, only positions 16th-17th showed a reduced SNP density in our porcine dataset (**Fig. 3B**).

Polymorphisms in microRNA genes show associations with the mRNA expression of several of their predicted targets

By sequencing the genomes of five parental Duroc boars, we selected a total of 15 SNPs mapping to miRNA genes and segregating in their offspring (N = 345). We were interested in determining if any of these SNPs was associated with microarray mRNA expression data recorded in skeletal muscle and liver samples from 87-89 offspring individuals. The putative mRNA targets of the miRNAs were selected according to *in silico* prediction of 7mer-m8 binding sites using porcine 3'UTR sequences. We also identified those interactions that were experimentally confirmed in humans according to the Tarbase v.8 database [39].

No experimentally confirmed targets were available for the porcine ssc-miR-9272-5p, which contains the rs322514450 SNP in its seed. If we just consider mRNA targets on the basis of the 7mer-m8 criterion (without requiring experimental confirmation in humans), two significant associations in GM and liver tissues were detected for the *NUDT6* and *RBLP1* genes, respectively (**Table 2, Additional file 9: Table S6**). The *NUDT6* gene overlaps the fibroblast growth factor 2 (*FGF2*) gene in an

antisense manner, and is thought to negatively regulate its expression [62]. The *FGF2* gene is involved in myoblast proliferation and is a potent inhibitor of skeletal muscle cells differentiation through the activation of the phosphorylation cascade triggered by the protein kinase C (PCK) [63]. Both *NUDT6* and *RBLP1* transcripts showed a reduced expression in pigs homozygous for the mutated A-allele (**Fig. 6**).

In principle, mutations in the seed region should affect target recognition and, by this reason, seed variability known to evolve under strong evolutionary constraints. Abnormal phenotypes have been observed in knockout mice for specific miRNAs, sometimes with severe effects on viability [64]. However, we did not detect a drastic effect of this seed polymorphism on hepatic or muscle mRNA expression. In a previous experiment, Gong et al. (2012) [57] investigated the consequences of SNPs in the seeds of eight miRNAs by using a dual luciferase assay and found that there was a total or partial abrogation of target binding function for four miRNAs, and a gain of target binding for a fifth one, implying that seed polymorphisms not always induce a complete rewiring of the network of genes regulated by the miRNA. Indeed, 3'-compensatory sites within mature miRNAs can mitigate the consequences of imperfect pairing in the seed [2]. Moreover, the existence of miRNA clusters encompassing multiple, highly similar miRNAs is fairly frequent in mammals [64], introducing some degree of functional redundancy with compensatory effects on gene repression.

Regarding other polymorphisms located in the miRNA precursor region, the rs333787816 (T/C) polymorphism within the ssc-miR-23a showed several significant associations, with *AGO1* and *AGO2* transcripts amongst them, two essential components of the miRNA-mediated cell metabolism regulation [65,66]. Besides, the hepatic expression of *MKRN1* transcript, whose depletion promotes glucose usage and

reduces lipid accumulation via AMPK stabilization and activation [67] was also associated with the rs335924546 SNP at ssc-miR-1224 (**Table 2**).

Interestingly, the rs319154814 polymorphism in the apical loop region of ssc-miR-326 showed a significant association, after multiple testing correction, with the hepatic mRNA expression of several of its predicted and experimentally confirmed targets (**Table 2**). In contrast, no association with GM muscle mRNA expression was observed. Differences in the expression of miRNAs or their targets across tissues might explain such outcome. Indeed, the analysis of the distribution of miRNA expression across human tissues has shown that only a minority of miRNAs are expressed ubiquitously [68]. The hepatic mRNA targets showing the most significant association with rs319154814 genotype were *PPP1CC, CFLAR, FSTL1, SF3A3, NAA50, ELAVL1* and *EIF4G2* (**Table 2, Additional file 9: Table S6**). The protein encoded by the *PPP1CC* gene belongs to the protein phosphatase PP1 subfamily, which is a ubiquitous serine/threonine phosphatase involved in regulating multiple cellular processes through dephosphorylation signaling. Among them, it is worth to mention insulin signaling [69], post-translational localization of circadian clock components [70] and lipids [71,72] or glycogen metabolism regulation [73]. Last but not least, the cellular FLICE-like inhibitory protein (*CFLAR*) encodes the cFLIP protein and is involved in the inhibition of Fas-mediated apoptosis [74], while the Follistatin-like 1 (FSTL1) plays a role in the immune inflammatory signaling and fibrosis in the liver [75].

Although the apical region of the miRNA does not have a function as critical as the seed, polymorphisms located in this particular region can have relevant effects on the structural conformation of the pre-miRNA hairpin [13]. In other words, SNPs in the apical region can modify the efficiency with which the Drosha processing machinery, mediating the initial slicing of the hairpin, is recruited. We found particularly interesting

that pigs homozygous for the derived A-allele of the rs319154814 SNP showed a consistent downregulation of the hepatic mRNA expression of the *PPP1CC*, *CFLAR*, *FSTL1*, *SF3A3*, *NAA50* and *ELAVL1* genes (**Fig. 4**), thus suggesting that this variant may increase the repressive activity of ssc-miR-326. We might hypothesize that the rs319154814 A-allele in the apical region of ssc-miR-326 might enhance the processing and the expression of this miRNA. However, a functional test will be required to demonstrate the hypothesis outlined in **Fig. 5**, where we indicate the potential mechanism by which the rs319154814 SNP might affect the processing of the miRNA. The A allele produces a change in the organization of the hairpin that allows the creation of a bulge around the CNNC processing motif. Such specific region, located at the basis of the pri-miRNA hairpin, is recognized by the SRSF3 protein, which induces Drosha to correctly slice both stems of the pri-miRNA, thus generating the pre-miRNA hairpin [76,77] that will be subsequently exported to the cytoplasm and further processed by Dicer into the mature miRNA transcript [2].

The aforementioned downstream CNNC motif, jointly with the basal UG motif, was first reported by Auyeung and collaborators (2013). These authors described both sequence motifs located around the miRNA precursor hairpins and assessed their contribution to the miRNA maturation process [9]. Additional surveys further reported other motifs affecting the expression of miRNAs [10] and analyzed the structural specifications of RNA-protein interactions between miRNAs and protein subunits in the Microprocessor complex [77–79], as well as the influence of such processing motifs in miRNA prediction analyses [80]. Particularly relevant was the study by Fernandez et al. (2017) [13], where they described a mutation in the apical loop of hsa-miR-30c (G/A) that creates a steric disruption of the pri-miRNA folding structure of the hairpin, hence creating a bulge around the CNNC motif that facilitates the SRSF3 factor accessibility

to the RNA sequence. The SRSF3 protein incorporates an RNA-recognition motif (RRM), broadly conserved across many bilaterian animals, which recognizes a degenerate CNNC motif in a base-specific manner [81]. Interestingly, the rs319154814 (G/A) polymorphism detected in porcine ssc-miR-326 might have structural consequences similar to those described for the hsa-miR-30c apical loop variant [13]. This interpretation is further corroborated by structural analyses of the ssc-miR-326 hairpin with and without incorporating the rs319154814 variant (**Fig. 5**). Such similarity is compatible with a conserved mechanism for enhanced miRNA processing via restructuration of the hairpin through sequence modifications.

A polymorphism in the apical loop of microRNA 326 is associated with fatty acid composition traits

The only miRNA SNP showing significant associations (*q*-value < 0.1) with lipid-related phenotypes was, once again, the rs319154814 in the apical loop of ssc-miR-326. As shown in **Table 3**, this SNP was associated with myristic content in the GM and LD muscles of Duroc pigs. To the best of our knowledge, no direct effect of ssc-miR-326 on the metabolism of myristic fatty acid has been described so far, but there are reports suggesting that several of the targets of this miRNA might be involved in diverse carbohydrate and lipid metabolism pathways. For instance, increased expression of miR-326 has been detected in type 1 diabetic patients with ongoing islet autoimmunity [82] and there are evidences that this miRNA represses PKM2, an enzyme that catalyzes the final and limiting step in glycolysis [83]. Moreover, targets of miR-326 are enriched in pathways related with sphingolipid metabolism and arachidonic acid metabolism [71]. In this regard, the *PPP1CC* transcript, one of the

predicted targets of miR-326, encodes a subunit of protein phosphatase-1 which activates acetyl-CoA carboxylase α and 6-phosphofructo2-kinase/fructose-2,6-bisphosphatase, the main regulators of fatty acid synthesis and glycolysis, respectively [72]. Protein phosphatase-1 also activates lipogenic transcription factors such as sterol regulatory element-binding protein 1 (SREBF1), carbohydrate-responsive element-binding protein (MLXIPL) and, moreover, it dephosphorylates the DNA-dependent protein kinase encoded by the *PRKDC* gene, which is another main determinant of hepatic lipogenesis [72]. In summary, a potential effect of the rs319154814 SNP in the synthesis or degradation of myristic acid can be envisaged, but this hypothesis still needs to be confirmed at the functional level.

Conclusions

MicroRNA genes show divergent patterns of variation between Asian and European pigs and wild boars and, in general, they display low levels of polymorphism. As expected, this reduced miRNA variability was particularly prevalent in the seed region, a finding that is likely explained by the strong effects of purifying selection aiming to preserve the conservation of this critical site. In the light of this, it could be expected that SNPs in the seed region might have more drastic consequences on mRNA expression than those in other miRNA regions. However, in the analyzed Lipgen population, the SNP displaying the most significant associations with mRNA expression and lipid-related phenotypes was located in the apical loop of the ssc-miR-326, whereas a SNP in the seed region of ssc-miR-9792-5p showed very few relevant associations. Recent studies have demonstrated that polymorphisms in the apical region of

microRNAs can affect their processing and expression patterns, thus highlighting the importance of the maturation process in the fine-tuning of the miRNA regulatory function.

Declarations

Ethics approval

Animal care and management procedures were in accordance with national guidelines for Good Experimental Practices and they were approved by the Ethical Committee of the Institut de Recerca I Tecnologia Agroalimentàries (IRTA).

Consent of publication

Not applicable.

Availability of data materials

Microarray expression data used in the current study were deposited in the Gene Expression Omnibus (GEO) public repository and are accessible through GEO Series Accession Number GSE115484. Phenotypic and genotypic data sets generated and analyzed during the current study have been deposited in the Figshare public repository available at https://figshare.com/projects/SNPs_miRNA/78690. Whole-genome sequencing dataset from 5 Duroc boars is available at the Sequence Read Archive (SRA) database (BioProject: PRJNA624870).

Competing interests

The authors declare that they have no competing interests.

Funding

The research presented in the current publication was funded by Grants AGL2013-48742-C2-1-R and AGL2013-48742-C2-2-R awarded by the Spanish Ministry of Economy and Competitivity. We also acknowledge the support of the Spanish Ministry of Economy and Competitivity for the *Center of Excellence Severo Ochoa* 2016-2019 (SEV-2015-0533) grant awarded to the Centre for Research in Agricultural Genomics (CRAG). Emilio Mármol-Sánchez was funded by a FPU PhD grant from the Spanish Ministry of Education (FPU15/01733). Dailu Guan was funded by a PhD fellowship from the Scholarship Council of China (CSC). The authors thank the CERCA Programme of the Generalitat de Catalunya (Barcelona, Spain) for their support and those who provided publicly available data.

Authors' contributions

MA conceived this study. MA and RQ designed the experimental protocols. RQ coordinated phenotyping recording and contributed to generate microarray expression data. EMS did DNA extractions and selected SNPs to be genotyped. EMS performed all bioinformatic and statistical analyses of the data. DG contributed to population structure and signature of selection analyses. RT performed whole-genome sequencing analyses of the five Duroc boars. EMS and MA drafted the manuscript. All authors read and approved the content of the final manuscript.

References

1. Gebert LFR, MacRae IJ. Regulation of microRNA function in animals. Nat Rev Mol Cell Biol. 2019; 20:21–37.

2. Bartel DP. Metazoan MicroRNAs. Cell. 2018; 173:20–51.

3. Chen CYA, Shyu A Bin. Mechanisms of deadenylation-dependent decay. RNA. 2011; 167–83.

4. Eichhorn SW, Guo H, McGeary SE, Rodriguez-Mias RA, Shin C, Baek D, et al. MRNA destabilization is the dominant effect of mammalian microRNAs by the time substantial repression ensues. Mol Cell. 2014; 56:104–15.

5. Saunders MA, Liang H, Li WH. Human polymorphism at microRNAs and microRNA target sites. Proc Natl Acad Sci U S A. 2007; 104:3300–5.

6. Mencía A, Modamio-Høybjør S, Redshaw N, Morín M, Mayo-Merino F, Olavarrieta L, et al. Mutations in the seed region of human miR-96 are responsible for nonsyndromic progressive hearing loss. Nat Genet. 2009; 41:609–13.

7. Clop A, Marcq F, Takeda H, Pirottin D, Tordoir X, Bibé B, et al. A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. Nat Genet. 2006; 38:813–8.

8. Moszyńska A, Gebert M, Collawn JF, Bartoszewski R. SNPs in microRNA target sites and their potential role in human disease. Open Biol. 2017; 7:170019.

9. Auyeung VC, Ulitsky I, McGeary SE, Bartel DP. Beyond secondary structure: primary-sequence determinants license pri-miRNA hairpins for processing. Cell. 2013; 152:844–58.

10. Roden C, Gaillard J, Kanoria S, Rennie W, Barish S, Cheng J, et al. Novel determinants of mammalian primary microRNA processing revealed by systematic evaluation of hairpin-containing transcripts and human genetic variation. Genome Res. 2017; 27:374–84.

11. Li J, Zhang Z. MiRNA regulatory variation in human evolution. Trends Genet. 2013; 116–24.

12. Sun G, Yan J, Noltner K, Feng J, Li H, Sarkis DA, et al. SNPs in human miRNA genes affect biogenesis and function. RNA. 2009; 15:1640–51.

13. Fernandez N, Cordiner RA, Young RS, Hug N, MacIas S, Cáceres JF. Genetic variation and RNA structure regulate microRNA biogenesis. Nat Commun. 2017; 8:1–12.

14. Larson G, Cucchi T, Dobney K. Genetic aspects of pig domestication. Genet pig. 2011; 14–37.

15. Frantz LAF, Schraiber JG, Madsen O, Megens HJ, Bosse M, Paudel Y, et al. Genome sequencing reveals fine scale diversification and reticulation history during speciation in Sus. Genome Biol. 2013; 14:R107.

16. Ramos-Onsins SE, Burgos-Paz W, Manunza A, Amills M. Mining the pig genome to investigate the domestication process. Heredity. 2014; 471–84.

17. Penso-Dolfin L, Moxon S, Haerty W, Di Palma F. The evolutionary dynamics of microRNAs in domestic mammals. Sci Rep. 2018; 8:1–13.

18. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics. 2014; 30:2114–20.

19. Warr A, Affara N, Aken B, Beiki H, Bickhart DM, Billis K, et al. An improved pig reference genome sequence to enable pig genetics and genomics research. bioRxiv. 2019; 668921.

20. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. 2013. http://arxiv.org/abs/1303.3997.

21. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010; 20:1297–303.

22. Backes C, Fehlmann T, Kern F, Kehl T, Lenhof H-P, Meese E, et al. miRCarta: a central repository for collecting miRNA candidates. Nucleic Acids Res. 2018; 46:160–7.

23. Kozomara A, Birgaoanu M, Griffiths-Jones S. miRBase: from microRNA sequences to function. Nucleic Acids Res. 2019; 47:155–62.

24. Shen W, Le S, Li Y, Hu F. SeqKit: A cross-platform and ultrafast toolkit for FASTA/Q file manipulation. PLoS One. 2016; 11:e0163962.

25. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006; 38:904–9.

26. Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. Evolution. 1984; 38:1358–70

27. Gallardo D, Pena RN, Amills M, Varona L, Ramírez O, Reixach J, et al. Mapping of quantitative trait loci for cholesterol, LDL, HDL, and triglyceride serum concentrations

in pigs. Physiol Genomics. 2008; 35:199–209.

28. Gallardo D, Quintanilla R, Varona L, Díaz I, Ramírez O, Pena RN, et al. Polymorphism of the pig *acetyl-coenzyme A carboxylase α* gene is associated with fatty acid composition in a Duroc commercial line. Anim Genet. 2009; 40:410–7.

29. Cánovas A, Quintanilla R, Amills M, Pena RN. Muscle transcriptomic profiles in pigs with divergent phenotypes for fatness traits. BMC Genomics. 2010; 11:372.

30. Quintanilla R, Pena RN, Gallardo D, Cánovas A, Ramírez O, Díaz I, et al. Porcine intramuscular fat content and composition are regulated by quantitative trait loci with muscle-specific effects. J Anim Sci. 2011; 89:2963–71.

31. González-Prendes R, Mármol-Sánchez E, Quintanilla R, Castelló A, Zidi A, Ramayo-Caldas Y, et al. About the existence of common determinants of gene expression in the porcine liver and skeletal muscle. BMC Genomics. 2019; 20.

32. Vidal O, Noguera JL, Amills M, Varona L, Gil M, Jiménez N, et al. Identification of carcass and meat quality quantitative trait loci in a Landrace pig population selected for growth and leanness1. J Anim Sci. 2005; 83:293–300.

33. Chomczynski P, Sacchi N. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal Biochem. 1987; 162:156–9.

34. Gautier L, Cope L, Bolstad BM, Irizarry RA. affy--analysis of Affymetrix GeneChip data at the probe level. Bioinformatics. 2004; 20:307–15.

35. Cayuela JM, Garrido MD, Bañón SJ, Ros JM. Simultaneous HPLC analysis of α-tocopherol and cholesterol in fresh pig meat. J Agric Food Chem. 2003; 51:1120–4.

36. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007; 81:559–75.

37. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. Nat Genet. 2012; 44:821–4.

38. Smedley D, Haider S, Durinck S, Pandini L, Provero P, Allen J, et al. The BioMart community portal: an innovative alternative to large, centralized data repositories. Nucleic Acids Res. 2015; 43:W589–98.

39. Karagkouni D, Paraskevopoulou MD, Chatzopoulos S, Vlachos IS, Tastsoglou S, Kanellos I, et al. DIANA-TarBase v8: A decade-long collection of experimentally supported miRNA-gene interactions. Nucleic Acids Res. 2018; 46:D239–45.

40. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Stat Soc Ser B. 1995; 289–300.

41. Lorenz R, Bernhart SH, Höner zu Siederdissen C, Tafer H, Flamm C, Stadler PF, et al. ViennaRNA Package 2.0. Algorithms Mol Biol. BioMed Central. 2011; 6:26.

42. Ding Y, Chan CY, Lawrence CE. RNA secondary structure prediction by centroids in a Boltzmann weighted ensemble. RNA. 2005; 11:1157–66.

43. Groenen MAM, Archibald AL, Uenishi H, Tuggle CK, Takeuchi Y, Rothschild MF, et al. Analyses of pig genomes provide insight into porcine demography and evolution. Nature. 2012; 491:393–8.

44. Steri M, Idda ML, Whalen MB, Orrù V. Genetic variants in mRNA untranslated regions. Wiley Interdiscip Rev RNA. 2020; 9:e1474.

45. Bachtiar M, Jin Y, Wang J, Tan TW, Chong SS, Ban KHK, et al. Architecture of population-differentiated polymorphisms in the human genome. PLoS One. 2019; 14:e0224089.

46. Bianco E, Nevado B, Ramos-Onsins SE, Pérez-Enciso M. A deep catalog of autosomal single nucleotide variation in the pig. PLoS One. 2015; 10:e0118867.

47. Frantz LAF, Schraiber JG, Madsen O, Megens HJ, Cagan A, Bosse M, et al. Evidence of long-term gene flow and selection during domestication from analyses of Eurasian wild and domestic pig genomes. Nat Genet. 2015; 47:1141–8.

48. Goedbloed DJ, Megens HJ, Van Hooft P, Herrero-Medrano JM, Lutz W, Alexandri P, et al. Genome-wide single nucleotide polymorphism analysis reveals recent genetic introgression from domestic pigs into Northwest European wild boar populations. Mol Ecol. 2013; 856–66.

49. Iacolina L, Scandura M, Goedbloed DJ, Alexandri P, Crooijmans RPMA, Larson G, et al. Genomic diversity and differentiation of a managed island wild boar population. Heredity. 2016; 116:60–7.

50. Manunza A, Amills M, Noce A, Cabrera B, Zidi A, Eghbalsaied S, et al. Romanian wild boars and Mangalitza pigs have a European ancestry and harbour genetic signatures compatible with past population bottlenecks. Sci Rep. 2016; 6:29913.

51. Giuffra E, Kijas JMH, Amarger V, Carlborg Ö, Jeon JT, Andersson L. The origin of the domestic pig: Independent domestication and subsequent introgression. Genetics. 2000; 154:1785–91.

52. White S. From globalized pig breeds to capitalist pigs: a study in animal cultures and evolutionary history. Environ Hist. 2011; 16:94–120.

53. Fang M, Andersson L. Mitochondrial diversity in European and Chinese pigs is consistent with population expansions that occurred prior to domestication. Proc R Soc B Biol Sci. 2006; 273:1803–10.

54. Chen M, Su G, Fu J, Zhang Q, Wang A, Sandø Lund M, et al. Population admixture in Chinese and European *Sus scrofa*. Sci Rep. 2017; 7:1–9.

55. Friedman RC, Farh KKH, Burge CB, Bartel DP. Most mammalian mRNAs are conserved targets of microRNAs. Genome Res. 2009; 19:92–105.

56. Simkin A, Geissler R, McIntyre ABR, Grimson A. Evolutionary dynamics of microRNA target sites across vertebrate evolution. PLoS Genet. 2020; 16:e1008285.

57. Gong J, Tong Y, Zhang HM, Wang K, Hu T, Shan G, et al. Genome-wide identification of SNPs in microRNA genes and the SNP effects on microRNA target binding and biogenesis. Hum Mutat. 2012; 33:254–63.

58. He S, Ou H, Zhao C, Zhang J. Clustering pattern and functional effect of SNPs in human miRNA seed regions. Int J Genomics. 2018; 2018:2456076.

59. Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. Cell. 2005; 15–20.

60. Bartel DP. MicroRNAs: target recognition and regulatory functions. Cell. 2009; 136:215–33.

61. Salomon WE, Jolly SM, Moore MJ, Zamore PD, Serebrov V. Single-molecule imaging reveals that Argonaute reshapes the binding properties of its nucleic acid guides. Cell. 2015; 162:84–95.

62. Baguma-Nibasheka M, Macfarlane LA, Murphy PR. Regulation of fibroblast growth factor-2 expression and cell cycle progression by an endogenous antisense RNA. Genes. 2012; 3:505–20.

63. Lu Y, Chen S, Yang N. Expression and methylation of FGF2, TGF-β and their downstream mediators during different developmental stages of leg muscles in chicken. PLoS One. 2013; 8:e79495.

64. Park CY, Choi YS, McManus MT. Analysis of microRNA knockouts in mice. Hum Mol Genet. 2010; 19:R169–75.

65. Ruda VM, Chandwani R, Sehgal A, Bogorad RL, Akinc A, Charisse K, et al. The roles of individual mammalian Argonautes in RNA interference in vivo. PLoS One. 2014; 9:e101749.

66. Huang V, Zheng J, Qi Z, Wang J, Place RF, Yu J, et al. Ago1 interacts with RNA polymerase II and binds to the promoters of actively transcribed genes in human cancer cells. PLoS Genet. 2013; 9:e1003821.

67. Lee MS, Han HJ, Han SY, Kim IY, Chae S, Lee CS, et al. Loss of the E3 ubiquitin ligase MKRN1 represses diet-induced metabolic syndrome through AMPK activation. Nat Commun. 2018; 9:1–14.

68. Ludwig N, Leidinger P, Becker K, Backes C, Fehlmann T, Pallasch C, et al. Distribution of miRNA expression across human tissues. Nucleic Acids Res. 2016; 44:3865–77.

69. Brady MJ, Saltiel AR. The role of protein phosphatase-1 in insulin action. Recent Prog. Horm. Res. 2001; 56:157–73.

70. Schmutz I, Wendt S, Schnell A, Kramer A, Mansuy IM, Albrecht U. Protein Phosphatase 1 (PP1) is a post-translational regulator of the mammalian circadian clock. PLoS One. 2011; 6:e21325.

71. Liu X, Song B, Li S, Wang N, Yang H. Identification and functional analysis of the risk microRNAs associated with cerebral low-grade glioma prognosis. Mol Med Rep. 2017; 16:1173–9.

72. Wang Y, Viscarra J, Kim SJ, Sul HS. Transcriptional regulation of hepatic lipogenesis. Nat. Rev. Mol. Cell Biol. 2015; 16:678–89.

73. Toole BJ, Cohen PTW. The skeletal muscle-specific glycogen-targeted protein phosphatase 1 plays a major role in the regulation of glycogen metabolism by adrenaline in vivo. Cell Signal. 2007; 19:1044–55.

74. Ram DR, Ilyukha V, Volkova T, Buzdin A, Tai A, Smirnova I, et al. Balance between short and long isoforms of cFLIP regulates Fas-mediated apoptosis in vivo. Proc Natl Acad Sci U S A. 2016; 113:1606–11.

75. Mattiotti A, Prakash S, Barnett P, van den Hoff MJB. Follistatin-like 1 in development and human diseases. Cell. Mol. Life Sci. 2018; 75:2339–54.

76. Han J, Lee Y, Yeom KH, Nam JW, Heo I, Rhee JK, et al. Molecular basis for the recognition of primary microRNAs by the Drosha-DGCR8 complex. Cell. 2006; 125:887–901.

77. Kim K, Duc Nguyen T, Li S, Anh Nguyen T. SRSF3 recruits DROSHA to the basal junction of primary microRNAs. RNA. 2018; 24:892–8.

78. Kwon SC, Baek SC, Choi YG, Yang J, Lee Y suk, Woo JS, et al. Molecular basis

for the single-nucleotide precision of primary microRNA processing. Mol Cell. 2019; 73:505-518.e5.

79. Jin W, Wang J, Liu C-P, Wang H-W, Xu R-M. Structural basis for pri-miRNA recognition by Drosha. Mol Cell. 2020; S1097-2765(20)30144-1.

80. Mármol-Sánchez E, Cirera S, Quintanilla R, Pla A, Amills M. Discovery and annotation of novel microRNAs in the porcine genome by using a semi-supervised transductive learning approach. Genomics. 2020; 112:2107–18.

81. Hargous Y, Hautbergue GM, Tintaru AM, Skrisovska L, Golovanov AP, Stevenin J, et al. Molecular basis of RNA recognition and TAP binding by the SR proteins SRp20 and 9G8. EMBO J. 2006; 25:5126–37.

82. Sebastiani G, Grieco FA, Spagnuolo I, Galleri L, Cataldo D, Dotta F. Increased expression of microRNA miR-326 in type 1 diabetic patients with ongoing islet autoimmunity. Diabetes Metab Res Rev. 2011; 27:862–6.

83. Kefas B, Comeau L, Erdle N, Montgomery E, Amos S, Purow B. Pyruvate kinase M2 is a target of the tumor-suppressive microRNA-326 and regulates the survival of glioma cells. Neuro Oncol. 2010; 12:1102–12.

Table 1: List of miRNA polymorphisms genotyped in a population of Duroc pigs (N = 345).

microRNA	SSC[a]	Start	End	Strand	SNP	Type	Alt. Allele[b]	Frequency[c]
ssc-miR-339	1	164025972	164026057	-	rs81349391	apical loop	G	0.4138
ssc-miR-130a	2	13296695	13296773	-	rs344472188	apical loop	G	0.1364
ssc-miR-23a	2	65308117	65308186	+	rs333787816	precursor stem	C	0.5398
ssc-miR-30d	4	6948669	6948747	+	rs340704946	precursor stem	G	0.4762
ssc-miR-371	6	56427208	56427285	-	rs320008166	precursor stem	C	0.2955
ssc-miR-429	6	63491921	63492001	+	rs323906663	precursor stem	A	0.2955
ssc-miR-9802	7	6393454	6393532	-	rs337567928	mature region	T	0.0852
ssc-miR-9792	8	110922737	110922830	+	rs322514450	seed region	A	0.1989
ssc-miR-326	9	9581944	9582034	-	rs319154814	apical loop	A	0.5852
ssc-miR-34c	9	39280278	39280357	+	rs321151601	mature region	A	0.0805
ssc-miR-378-2	12	36947443	36947510	+	rs341950320	precursor stem	A	0.1176
ssc-miR-15b	13	100083172	100083269	+	rs334680106	precursor stem	T	0.2706
ssc-miR-1224	13	122141042	122141149	+	rs327603919	precursor stem	T	0.1724
ssc-miR-486	17	10758818	10758899	-	rs335924546	precursor stem	T	0.3391
ssc-miR-335	18	18341568	18341659	-	rs334590580	precursor stem	C	0.1875

[a]SSC: porcine chromosome; [b]Alt. Allele: alternative allele of the genotyped SNP; [c]Frequency: alternative allele frequency in the genotyped population of 345 Duroc pigs.

Table 2: Significant associations (q-value < 0.1) between 15 genotyped miRNA SNPs and the mRNA expression of their targets in the *gluteus medius* (GM) skeletal muscle (N = 89) and liver (N = 87) tissue samples from Duroc pigs.

SNP	Type	Tissue	Probe	ID	Gene	δ[a]	se[b]	*P*-value	*q*-value[c]
rs333787816 (2:65308181)	ssc-miR-23a precursor stem (T/C)	GM	Ssc.1790.1.S1_at	ENSSSCG00000000019	*NUP50*	-0.1052	0.0288	3.849E-04	9.160E-02
			Ssc.12493.1.A1_at	ENSSSCG00000024027	*PAFAH1B2*	-0.2021	0.0565	3.921E-04	9.160E-02
			Ssc.8682.2.A1_at	ENSSSCG00000014240	*CSNK1G3*	-0.1532	0.0434	4.515E-04	9.160E-02
			Ssc.4948.1.S1_at	ENSSSCG00000034725	*UBE2R2*	0.1726	0.0493	5.906E-04	9.160E-02
			Ssc.21303.1.S1_at	ENSSSCG00000003630	*AGO1*	0.1016	0.0275	6.775E-04	9.160E-02
			Ssc.24035.2.A1_at	ENSSSCG00000005935	*AGO2*	0.2896	0.0864	8.339E-04	9.395E-02
rs322514450 (8:110922752)	ssc-miR-9792 seed region (G/A)	GM	Ssc.23813.1.S1_at	ENSSSCG00000009085	*NUDT6*	-0.3577	0.0836	2.805E-05	5.073E-02
		LIVER	Ssc.11164.1.A1_at	ENSSSCG00000001836	*RLBP1*	-0.1619	0.0390	5.113E-05	9.479E-02
rs319154814 (9:9581989)	ssc-miR-326 apical loop (G/A)	LIVER	Ssc.9544.2.S1_a_at	ENSSSCG00000016101	*CFLAR*	0.3246	0.0997	1.264E-03	7.061E-02
			Ssc.11661.2.S1_at	ENSSSCG00000009828	*PPP1CC*	0.4508	0.1436	1.784E-03	7.061E-02
			Ssc.11044.1.A1_at	ENSSSCG00000003643	*SF3A3*	0.1808	0.0561	2.598E-03	7.061E-02
			Ssc.23242.1.A1_at	ENSSSCG00000039426	*FSTL1*	0.2407	0.0780	2.606E-03	7.061E-02
			Ssc.12222.1.S1_at	ENSSSCG00000013233	*CELF1*	-0.1714	0.0574	3.210E-03	7.061E-02
			Ssc.10946.2.A1_at	ENSSSCG00000011917	*NAA50*	0.2853	0.0993	4.070E-03	7.462E-02
			Ssc.4516.2.S1_at	ENSSSCG00000013592	*ELAVL1*	0.1666	0.0598	6.226E-03	9.783E-02
rs335924546 (13:122141078)	ssc-miR-1224 precursor stem (C/T)	LIVER	Ssc.11539.1.A1_at	ENSSSCG00000016498	*MKRN1*	0.2105	0.0594	4.661E-04	2.983E-02

[a]δ: estimated allele substitution effect; [b]se: standard error of the substitution effect; [c]q-value: q-value calculated with the false discovery rate (FDR) approach by considering experimentally confirmed 7mer-m8 targets of miRNAs harboring the genotyped SNP.

Table 3: Significant associations at the nominal level (*P*-value < 0.01) and/or after multiple testing (*q*-value < 0.1; in bold) between 15 genotyped SNPs and lipid metabolism-related phenotypes recorded in a population of Duroc pigs (N = 345).

SNP	Type	Trait	δ	se	*P*-value	*q*-value
rs81349391 (1:164026014)	ssc-miR-339 apical loop (A/G)	GM (C16:0)	-0.3234	0.1396	2.005E-02	3.772E-01
		GM SFA	-0.4394	0.2030	2.969E-02	3.772E-01
		GM UFA	0.4392	0.2030	2.978E-02	3.772E-01
rs344472188 (2:13296736)	ssc-miR-130a apical loop (T/C)	BFTLR	-1.2084	0.5914	4.087E-02	9.683E-01
rs340704946 (4:6948743)	ssc-miR-30d precursor stem (A/G)	GM (C20:0)	-0.0219	0.0098	2.541E-02	5.779E-01
rs320008166 (6:56427227)	ssc-miR-371 precursor stem (T/C)	GM IMF	0.5368	0.2461	3.104E-02	9.060E-01
rs323906663 (6:63491948)	ssc-miR-429 precursor stem (G/A)	GM IMF	0.5682	0.2482	2.348E-02	8.923E-01
rs337567928 (7:6393469)	ssc-miR-9802-3p seed region (G/T)	GM (C14:0)	0.0802	0.0379	3.692E-02	5.854E-01
rs322514450 (8:110922752)	ssc-miR-9792-5p seed region (G/A)	LD (C18:3)	-0.0261	0.0087	6.434E-03	1.298E-01
		LD Cholesterol	2.2471	0.9022	1.301E-02	1.298E-01
		LD UFA	-0.5428	0.2169	1.582E-02	1.298E-01
		LD (C14:0)	0.0624	0.0253	1.728E-02	1.298E-01
		LD SFA	0.5292	0.2177	1.907E-02	1.298E-01
		LD (C20:4)	-0.3518	0.1569	2.435E-02	1.298E-01
		LD (C16:0)	0.3151	0.1372	2.623E-02	1.298E-01
		LD PUFA	-1.3593	0.6184	2.732E-02	1.298E-01
		LD (C18:2)	-0.8890	0.4258	3.598E-02	1.519E-01
		LD (C16:1)	0.1208	0.0590	4.019E-02	1.527E-01
rs319154814 (9:9581989)	**ssc-miR-326 apical loop (G/A)**	**LD (C14:0)**	**0.0885**	**0.0213**	**8.722E-05**	**3.314E-03**
		GM (C14:0)	**0.0700**	**0.0184**	**3.942E-04**	**7.489E-03**
		LD PUFA/MUFA	**-0.0506**	**0.0188**	**7.109E-03**	**7.653E-02**
		LD (C20:1)	**0.0349**	**0.0125**	**8.055E-03**	**7.653E-02**
		LD (C20:4)	-0.2918	0.1268	2.092E-02	1.438E-01
		LD MUFA	0.9512	0.3971	2.485E-02	1.438E-01

		LD (C18:1)	0.8349	0.3624	2.814E-02	1.438E-01
		LD PUFA	-1.0670	0.5004	3.222E-02	1.438E-01
		LD (C18:2)	-0.7266	0.3443	3.405E-02	1.438E-01
		GM (C18:0)	-0.2245	0.1115	4.423E-02	1.681E-01
rs321151601 (9:39280312)	ssc-miR-34c mature region (C/A)	LD (C20:0)	-0.0262	0.0119	2.655E-02	8.171E-01
rs335924546 (13:122141078)	ssc-miR-1224 precursor stem (C/T)	LD (C18:3)	0.0270	0.0116	2.026E-02	4.480E-01
		LD (C18:0)	-0.3578	0.1663	3.091E-02	4.480E-01
rs335924546 (17:10758828)	ssc-miR-486 precursor stem (C/T)	GM (C20:4)	0.3396	0.1465	2.100E-02	4.555E-01
		LD (C14:0)	-0.0586	0.0256	2.937E-02	4.555E-01
		LD IMF	0.3300	0.1553	4.748E-02	4.555E-01
rs334590580 (18:18341582)	ssc-miR-335 precursor stem (T/C)	GM (C16:0)	-0.4305	0.1558	5.862E-03	1.201E-01
		GM (C20:0)	0.0331	0.0121	6.321E-03	1.201E-01
		GM UFA	0.5231	0.2278	2.147E-02	2.045E-01
		GM SFA	-0.5229	0.2278	2.153E-02	2.045E-01

[a]δ: estimated allele substitution effect; [b]se: standard error of the substitution effect; [c]*q*-value: q-value calculated with the false discovery rate (FDR) approach. Trait acronyms are defined in Additional file 3: Table S3.

Additional Files legends

Additional file 1: Table S1: List of whole-genome sequences from European domestic pigs (EDM, N = 40), Asian domestic pigs (ADM, N = 40), European wild boars (EWB, N = 20) and Asian wild boars (AWB, N = 20).

Additional file 2: Table S2: Single nucleotide polymorphisms located in microRNA genes and their frequencies in European (E) and Asian (A) domestic pigs (DM) and wild boars (WB).

Additional file 3: Table S3: Means and standard deviations (SD) of fatness and intramuscular fat content and composition traits recorded in the *gluteus medius* (GM) and *longissimus dorsi* (LD) muscles of 345 half-sib Duroc pigs.

Additional file 4: Fig. S1: Number of SNPs present in each of the annotated polymorphic miRNA genes.

Additional file 5: Fig. S2: Venn Diagrams depicting the degree of sharing of SNPs located at miRNA loci in (**A**) *precursor*, (**B**) *mature* and (**C**) *seed* regions, for all defined porcine populations (i.e. ADM, AWB, EDM and EWB).

Additional file 6: Table S4: Significant signatures of selection in miRNA genes inferred with an F_{ST} statistic test and a data set of 6,408,611 SNPs (206 inside miRNA loci). The contrasts with significant signatures were: Asian domestic pigs (ADM) vs Asian wild boars (AWB) and European wild boars (EDM) vs Asian wild boars (AWB).

Additional file 7: Table S5: Whole-genome sequencing statistics for five Duroc boars.

Additional file 8: Fig. S3: Boxplot distribution depicting whole-genome sequencing statistics for five Duroc boars.

Additional file 9: Table S6: Results of the association analysis between miRNA SNPs and the mRNA expression of 7mer-m8 targets experimentally validated in humans and expressed in the porcine *gluteus medius* (GM) skeletal muscle and liver tissues.

Additional file 10: Table S7: Results of the association analysis between miRNA SNPs and fatness and intramuscular fat content and composition traits recorded in the *gluteus medius* (GM) and *longissimus dorsi* (LD) skeletal muscles of 345 Duroc pigs.

Fig. 1: Principal component analysis plots based on SNPs mapping to: (**A**) the whole-genome, (**B**) miRNA genes, (**C**) Full 3'UTRs, (**D**) 3'UTR 7mer-m8 sites and (**E**) 3'UTR 8mer sites, respectively.

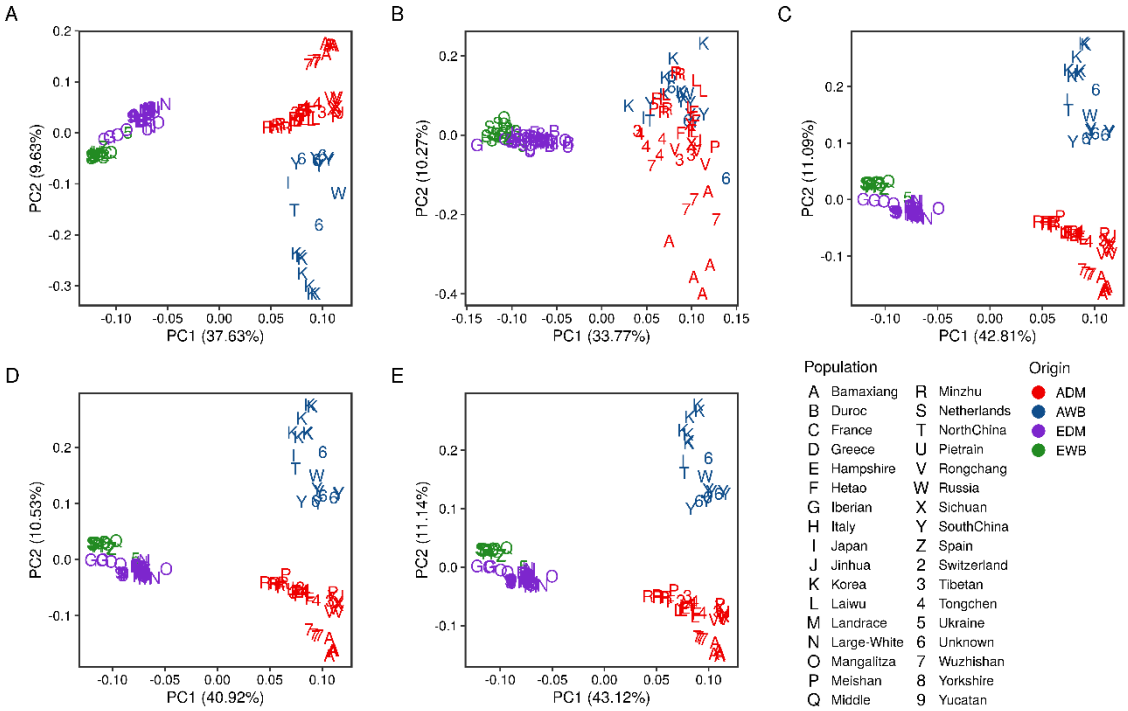

Fig. 2: Alternative allele frequency distribution of polymorphisms located at miRNA loci in (**A**) Asian domestic pigs (ADM), (**B**) European domestic (EDM) pigs, (**C**) Asian wild boars (AWB) and (**D**) European wild boars (EWB).

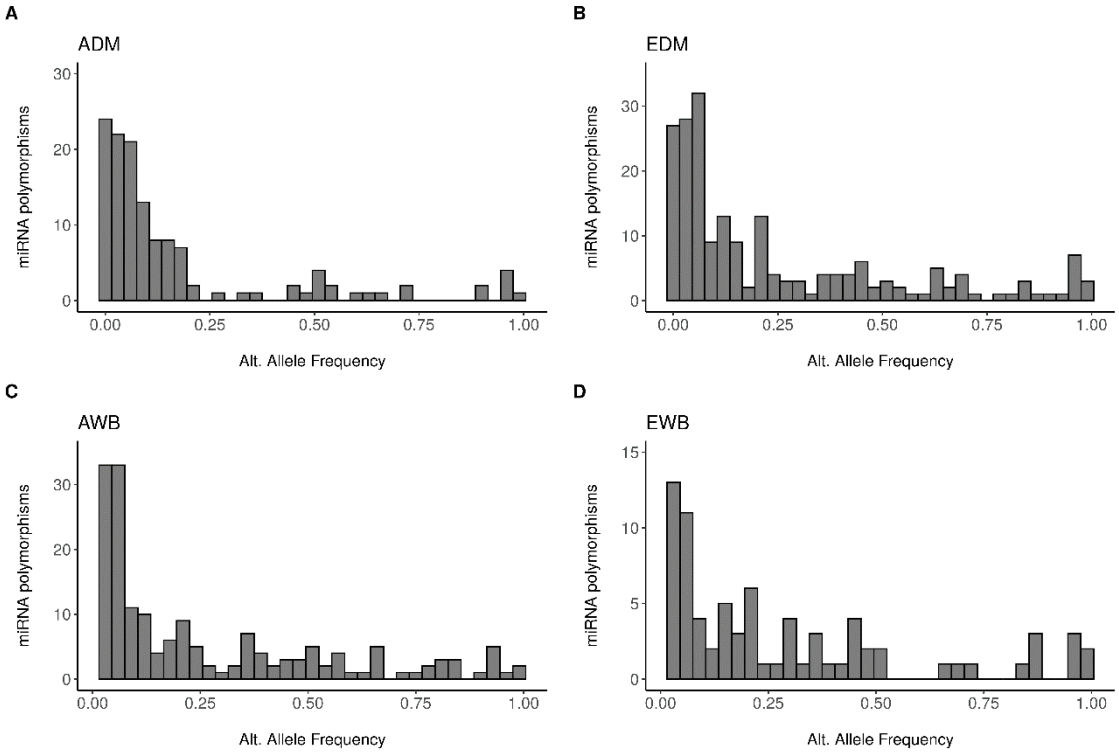

Fig. 3: (**A**) SNP density per 100 bp for each analyzed miRNA region considering the full set of 120 whole-genome sequenced porcine samples as well as each of the ADM, EDM, AWB and EWB populations. (**B**) SNP density across mature miRNA regions. 1: anchor (1st 5' end position), 2: seed (2nd to 8th position) and 3: supplemental pairing (13th to 18th position).

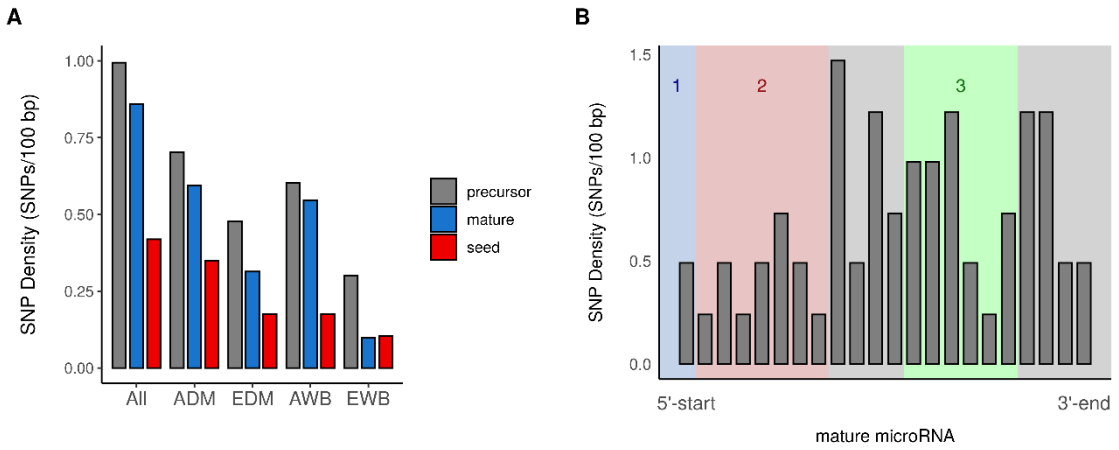

Fig. 4: Hepatic mRNA expression levels of the *CFLAR*, *ELAVL1*, *FSTL1*, *NAA50*, *PPP1CC* and *SF3A3* genes according to the genotype of the rs319154814 apical loop polymorphism in the ssc-miR-326 gene. The number of individuals representing each genotype were: GG (N = 17), GA (N = 37) and AA (N = 32).

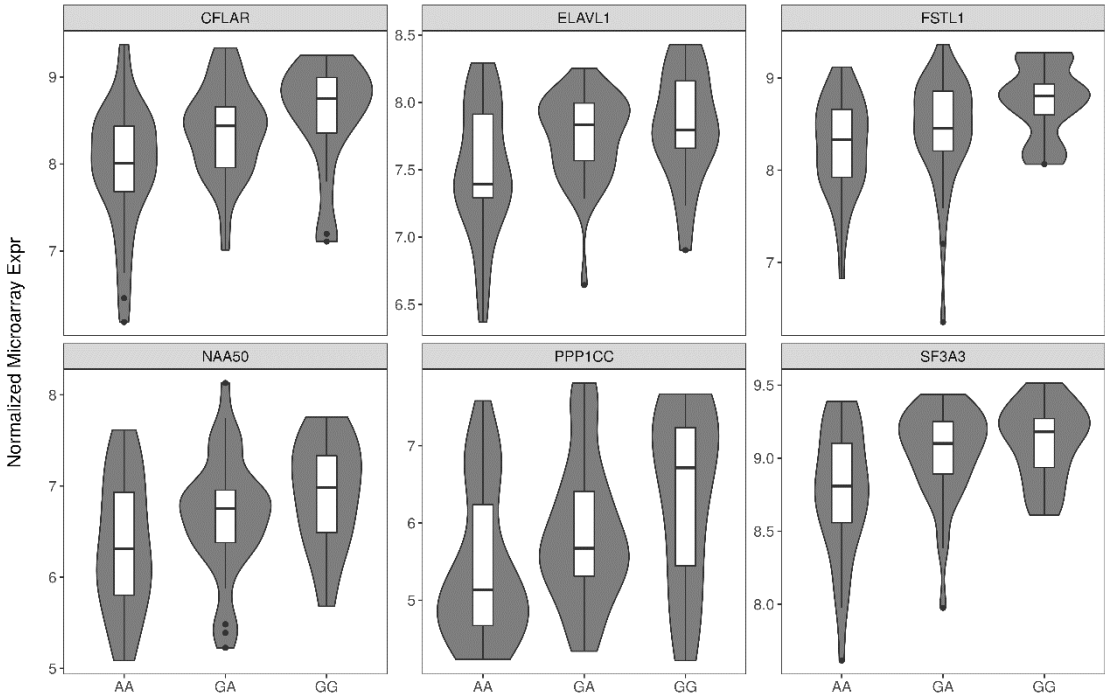

Fig. 5: Secondary structural folding of the initial ssc-miR-326 pri-miRNA predicted with the RNAfold centroid-based algorithm. The rs319154814 (G/A) polymorphism in the apical loop of the ssc-miR-326 is thought to produce a reorganization to the base-pairing selection of the hairpin, thus generating an unpaired bulge at the base of the pri-miRNA hairpin. Such alteration of the hairpin structure would facilitate the accessibility of the SRSF3 protein to two consecutive CNNC motifs located at the basal bulge of the polymorphic miRNA. The SRSF3 factor then promotes the recruitment of Drosha slicing factor allowing an increased maturation rate of ssc-miR-326. This interaction would imply a higher repression in the expression of its predicted mRNA targets (see Fig. 4). Precursor miRNA (pre-miRNA) sliced after Drosha processing is highlighted in blue, whereas mature ssc-miR-326 is depicted in red.

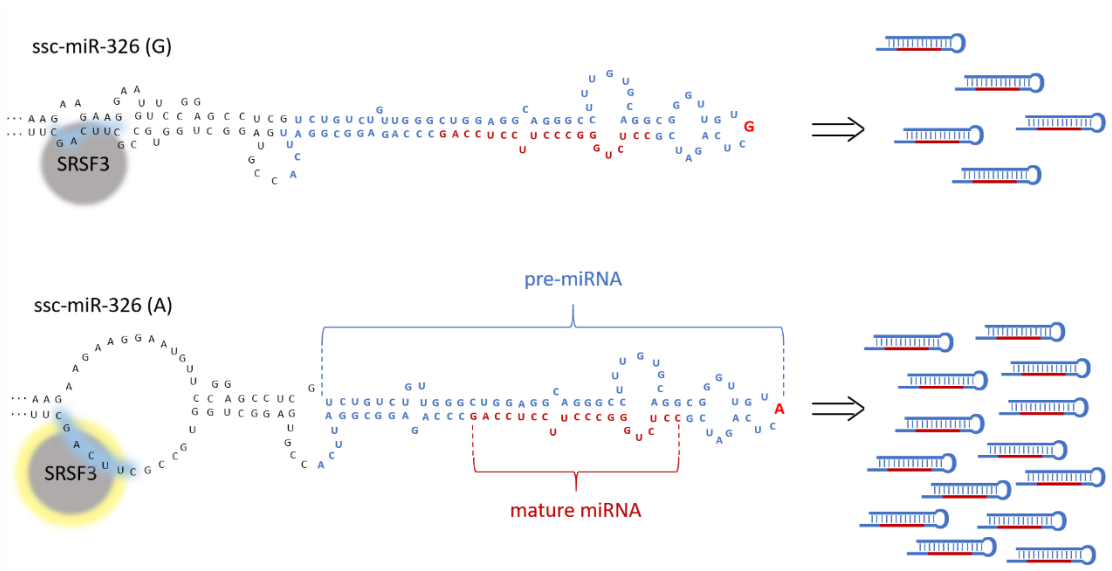
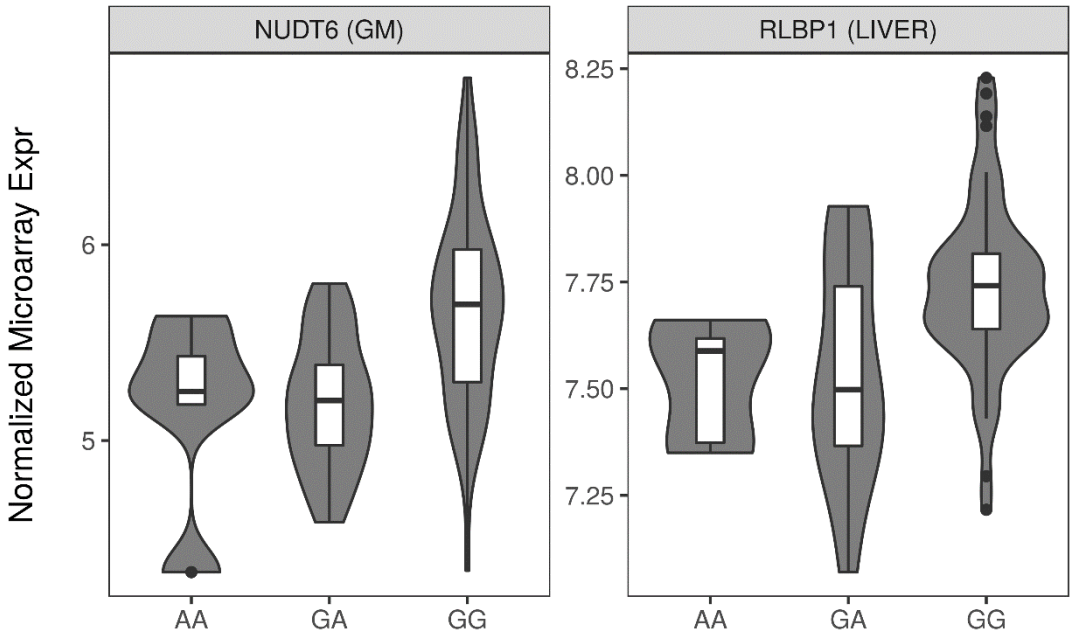
Fig. 6: *Gluteus medius* (GM) skeletal muscle and hepatic mRNA expression levels of the *NUDT6* and *RBLP1* genes according to the genotype of the rs322514450 polymorphism located in the seed of the ssc-miR-9792-5p gene. The number of pigs representing each genotype were: GG (N = 56 in GM, 58 in liver), GA (N = 25) and AA (N = 5).