1 **Variability in porcine microRNA genes and its association with mRNA** 2 **expression and lipid phenotypes**

3

4 Emilio Mármol-Sanchez¹, María Gracia Luigi-Sierra¹, Anna Castelló^{1,4}, Dailu Guan¹,
5 Raquel Quintanilla², Raul Tonda³, Marcel Amills^{1,4}

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

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

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

12 ⁴Departament de Ciència Animal i dels Aliments, Universitat Autònoma de Barcelona,
13 08193 Bellaterra, Barcelona, Spain.

14

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

18

19

20

21

22

23

24 **Abstract**

25 **Background**

26 Mature microRNAs (miRNAs) play an important role in repressing the expression of a
27 wide range of mRNAs. The variability of miRNA genes and their corresponding 3'UTR
28 binding sites might disrupt canonical conserved miRNA-mRNA pairing, thus modifying
29 gene expression patterns. The presence of polymorphic sites in miRNA genes and their
30 association with gene expression phenotypes and complex traits has been poorly
31 characterized in pigs so far.

32 **Results**

33 By analyzing whole-genome sequences from 120 pigs and wild boars from Europe and
34 Asia, we have identified 285 single nucleotide polymorphisms (SNPs) mapping to
35 miRNA loci, as well as 109,724 SNPs located in predicted 7mer-m8 miRNA binding
36 sites within porcine 3'UTRs. Porcine miRNA genes show a reduced SNP density
37 compared with their flanking non-miRNA regions. By sequencing the genomes of 5
38 Duroc boars, we have identified 12 miRNA SNPs that have been subsequently
39 genotyped in their offspring (N = 345, Lipgen population). Association analyses
40 between miRNA SNPs and 38 lipid-related traits as well as hepatic and muscle
41 microarray expression phenotypes recorded in the Lipgen population were carried out.
42 The most relevant association detected was the one between the genotype of the
43 rs319154814 (G/A) SNP located in the apical loop of the ssc-miR-326 hairpin precursor
44 and *PPP1CC* mRNA levels in the liver (q -value = 0.058). This result was subsequently
45 confirmed by qPCR (P -value = 0.027). The rs319154814 (G/A) genotype was also
46 associated with several fatty acid composition traits.

47 **Conclusions**

48 Porcine miRNA genes show a reduced variability consistent with strong purifying
49 selection, particularly in the seed region, which plays a critical role in miRNA binding.
50 Although it is generally assumed that SNPs mapping to the seed region are the ones
51 with the strongest consequences on mRNA expression, we show that a SNP mapping to
52 the apical region of ssc-miR-326 is significantly associated with the hepatic mRNA
53 levels of the *PPP1CC* gene, one of its predicted targets. Although experimental
54 confirmation of such interaction has been obtained in humans but not in pigs, this result
55 highlights the need of further investigating the functional effects of miRNA
56 polymorphisms located outside the seed region on gene expression in pigs.

57

58

59 **Background**

60 Mature microRNA transcripts (miRNAs) are short (~22 nt) non-coding RNAs
61 which play an essential role in the regulation of gene expression [1]. During the
62 biogenesis of miRNAs, one strand of the miRNA duplex binds to the guide-strand
63 channel of an Argonaute protein forming a miRNA-induced silencing complex
64 (miRISC) with the ability to repress mRNA expression through binding to specific
65 3'UTR target sites [2]. In postembryonic cells, this repressor mechanism mainly acts by
66 destabilizing the mRNA through decapping and poly(A)-tail shortening [3,4] and less
67 often by hindering translation [2]. The binding of the miRNA to its 3'UTR target site
68 depends critically on the sequence of the seed region, which encompasses nucleotides
69 (nt) 2-8 from the 5' end of the mature miRNA and interacts with the target site through
70 Watson-Crick pairing [1]. Polymorphisms located within the seed region might affect

71 the proper recognition of mRNA targets [5,6]. Nevertheless, imperfect seed matches can
72 be compensated by nucleotides 13th-to-16th of the mature miRNA, thus providing
73 additional anchor pairing to the seed region [2]. Other sites relevant for miRNA
74 processing and function are the basal UG, flanking CNNC and apical UGU motifs [7],
75 as well as a mismatched GHG motif [8], all of which contribute to facilitate miRNA
76 processing [2]. Variability in the 3'UTR miRNA binding sites can also affect gene
77 expression at the post-transcriptional level [5,9,10].

78 The evolutionary conservation of miRNAs across species is associated with their
79 levels of expression [11,12] as well as with the functional importance of the regulatory
80 networks they modulate [13,14]. Besides, the conservation of the repertoire of mRNAs
81 targeted by a given miRNA depends mostly on the age of the associated miRNA gene,
82 with novel miRNAs acquiring targets more rapidly than ancient miRNAs [15].
83 Conserved sites are particularly enriched in the 5'end of 3'UTR [12], although selective
84 pressure on miRNA binding sites is not uniform across miRNA genes [16].

85 Saunders et al. (2007) [5] investigated the variability of 474 human miRNA
86 genes and found that the single nucleotide polymorphism (SNP) density is lower in
87 miRNA loci when compared to their flanking regions. Moreover, Saunders et al. (2007)
88 [5] found that ~90% of human miRNA genes do not contain polymorphisms, and the
89 majority of SNPs mapping to miRNAs are located outside the seed region, thus
90 evidencing that the variability of this critical functional motif evolves under strong
91 selective constraints [5]. Indeed, polymorphisms in the first 14 nucleotides of the
92 mature miRNA, and particularly those within the seed region, might abrogate the
93 binding of the miRNA to its 3'UTR targets, leading to an extensive rewiring of the
94 miRNA-mediated regulatory network [17]. Mammalian miRNA knockouts often
95 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 [18,19]. 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 [17]. Overall, these findings suggest functional conserved mechanisms of miRNA-mediated gene regulation influenced by polymorphic sites at both miRNA loci and their 3'UTR binding sites [20].

The wild ancestors of pigs were independently domesticated in the Near East and China 10,000 YBP [21]. Domestic pigs then 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 [22]. Pig phenotypes might be explained, at least partially, by modifications in the microRNA-mediated regulation of gene expression [23]. Indeed, several studies have reported associations between SNPs in miRNA binding sites and porcine phenotypic variation [24–29], while fewer studies have investigated the association between SNPs in miRNA genes and complex traits in pigs [30–33]. For instance, Chai et al. (2018) reported a significant causal relationship between a SNP in the seed of ssc-miR-378-3p and extensive modifications of its target mRNA repertoire [33]. Moreover, the authors associated the presence of the observed miRNA seed variant to a structural reconfiguration of the hairpin, leading to an enhanced expression of the mature miRNA [33].

The polymorphism of miRNA genes has not been systematically characterized in pigs despite its potential impact on gene regulation and phenotypic variation. In the

current work, we aimed to elucidate the patterns of variability of miRNA genes in European and Asian wild boars and domestic pigs as well as to infer whether such patterns are influenced by purifying selection. Moreover, we wanted to infer whether SNPs in miRNA genes are associated with liver and muscle gene expression traits and lipid phenotypes recorded in a Duroc pig population.

Methods

Characterizing the polymorphisms of miRNA genes and their 3'UTR binding sites in a worldwide sample of pigs and wild boars

Retrieval of Porcine Whole-Genome Sequences

Whole-genome sequences (WGS) from 120 wild boars and domestic pigs (*Sus scrofa*) were retrieved from the NCBI Sequence Read Archive (SRA) database (<https://www.ncbi.nlm.nih.gov/sra>). Most of these WGS have been reported in previous publications [34–37] and detailed information can be found in **Additional file 1: Table S1**. The 120 WGS corresponded to 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). The ADM group included WGS from Meishan, Tongchen, Jinhua, Rongchan, Wuzhishan, Tibetan, Sichuan, Hetao, Minzhu, Bamaixang and Laiwu pigs, while the EDM group was composed of WGS from Pietrain, Mangalitza, Iberian, Duroc, American Yucatan (from America, but with a European origin), Yorkshire, Landrace, Hampshire and Large-white pigs. Both AWB and EWB samples were selected according to their geographic distribution (**Additional file 1: Table S1**).

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

148

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

151 Fastq paired-end files generated from SRA data were filtered according to their
152 quality, and sequence adapters were trimmed by making use of the Trimmomatic
153 software (v.0.36) with default parameters [38]. Trimmed paired-end sequences were
154 aligned against the *Sus scrofa* reference genome (Sscrofa11.1) [39] with the BWA-
155 MEM algorithm [40] and default settings. Sequence alignment map (SAM) formatted
156 files were sorted and transformed into binary (BAM) formatted files. PCR duplicates
157 were subsequently identified with the *MarkDuplicates* package from the Picard tool
158 (<https://broadinstitute.github.io/picard/>) and removed to perform INDEL realignment
159 with the *IndelRealigner* package from the Genome Analysis Toolkit (GATK v.3.8) [41].
160 Base quality score recalibration (BQSR) was implemented with the *BaseRecalibrator*
161 package from GATK v.3.8. Variant calling was implemented with the *HaplotypeCaller*
162 function according to GATK best practices [41]. Individual gVCF formatted files,
163 including both polymorphic and homozygous blocks, were generated with the
164 *GenotypeGVCFs* package from the GATK v.3.8 tool [41], and they were merged into
165 separate multi-individual variant calling format (VCF) files containing single
166 polymorphic and INDEL sites for each defined population (i.e. ADM, EDM, AWB and
167 EWB). Finally, variants were distributed into SNP and INDEL files with the

168 *SelectVariants* package and quality filtered using the *FilterVariants* function from
 169 GATK v.3.8 [41]. For SNPs, we used the following parameters: QD < 2.0, FS > 60.0,
 170 MQ < 40.0, MQRankSum < -12.5, ReadPosRankSum < -8.0, SOR > 3.0. INDELs were
 171 not taken into consideration in the present study because INDEL detection based on
 172 commonly used pipelines for variant calling have shown low reliability and
 173 reproducibility across tools and analyzed datasets [42,43].

174

175 *Identification of single nucleotide polymorphisms in miRNA genes and in their 3'UTR* 176 *binding sites*

177 Single nucleotide polymorphisms mapping to annotated porcine miRNA loci (N
 178 = 370) were retrieved by making use of the curated Sscrofa11.1 annotation for miRNA
 179 regions available in the miRCarta v1.1 [44] and miRBase [45] databases. Additionally,
 180 annotated mature miRNA loci (N = 409) within miRNA genes (N = 370) were
 181 retrieved, and SNPs located in the mature and seed regions (2nd to 8th positions at 5' end
 182 in the mature miRNA) were identified. A comprehensive description of miRNA gene
 183 regions and miRNA-mRNA interactions can be found in **Fig. 1A** and **1B**, respectively.

184 We also retrieved the annotated 3'UTR porcine mRNA transcripts from
 185 Ensembl repositories (v.100, <http://www.ensembl.org/info/data/ftp/index.html>) and the
 186 corresponding set of sequences was interrogated against the seed regions of the 409
 187 porcine mature miRNAs considered in our study. Seed sequences were reverse-
 188 complemented and searched along the 3'UTR sequences of mRNA genes by making
 189 use of the *locate* function of the SeqKit toolkit [46]. All 7mer-m8 canonical seed pairing
 190 sites, as well as 8mer seed pairing sites (**Fig. 1**), were identified for each miRNA seed,
 191 and, based on such information, the corresponding putative miRNA-mRNA target pairs

were established. Additional non-canonical miRNA-mRNA interactions were not considered (**Fig. 1B**) as they are regarded as less biologically significant than 7mer-m8 or 8mer interactions [2,47,48]. Subsequently, the genomic location of both 7mer-m8 and 8mer matching regions were determined, and SNPs residing in such predicted miRNA binding sites were gathered.

Principal components analyses (PCA) were performed with the smartPCA software [49]. These PCA analyses were based on either (1) autosomal whole-genome SNPs, (2) autosomal SNPs from miRNA genes, (3) autosomal SNPs from 3'UTRs, or (4) autosomal SNPs from 8mer and 7mer-m8 3'UTR sites. When implementing the PCA based on whole-genome data, SNPs complying with the following parameters were retained: 1) minimum allele frequency > 0.05, and 2) Hardy-Weinberg equilibrium exact test *P*-value > 0.001. In contrast, no filtering was performed in the PCA based on miRNA SNPs because there are a few hundreds.

Frequency and distribution of SNPs in miRNA genes

To investigate the distribution of SNPs along miRNA genes, SNPs were classified according to their location, i.e., SNPs within the seed region of mature miRNAs were flagged as “*seed*”, whereas SNPs located in mature miRNAs but outside the seed were classified as “*mature*”. Additionally, SNPs classified as “*mature*” were assigned to the following subtypes: 1) “*anchor*” (1st position at 5' end), and 2) “*supplemental pairing*” (13th to 18th position from 5' end). The remaining SNPs were assigned to the “*precursor*” class. Allele frequencies were estimated in the whole set of porcine samples (N = 120), as well as in each EDM, ADM, EWB and AWB group independently (**Additional file 2: Table S2**).

216 In order to calculate the SNP density in precursor, mature and seed miRNA
217 regions, we first calculated the total nucleotide length of each of these regions in the set
218 of porcine miRNA loci. *Seed length* was calculated by considering 7 nucleotide
219 positions in each of the annotated mature miRNAs ($N = 409$) mapping to the 370
220 miRNA loci considered in our study ($7 \text{ bp} \times 409 = 2,863 \text{ bp}$). The total length of the
221 409 mature miRNAs was determined ($\sim 22 \text{ bp} \times 409 = 8,916 \text{ bp}$), and the total *seed*
222 *length* (2,863 bp) was subtracted from 8,916 bp to obtain the *mature length* (6,053 bp),
223 which corresponds to the total length of all mature miRNA sequences excluding the
224 seeds. The *precursor length* (22,229 bp) was calculated by subtracting the *seed length*
225 (2,863 bp) and *mature length* (6,053 bp) from the total length (31,145 bp) of all miRNA
226 loci ($N = 370$). Then, the SNP density (D) in each of these regions was calculated as
227 follows:

$$D = \frac{N_r \times 100}{L_r}$$

229
230 where N_r is the total number of miRNA SNPs ($N = 285$), or the number of SNPs
231 detected in each of the defined regions, (i.e., 221, 52 and 12 SNPs in precursor, mature
232 and seed regions, respectively); L_r is the total nucleotide length of all miRNA loci
233 (31,145 nts), or the total number of nucleotides in miRNA loci belonging to the
234 “*precursor*” (22,229 bp), “*mature*” (6,053 bp) or “*seed*” (2,863 bp) regions. These
235 calculations would yield the number of SNPs per bp for each category. We decided to
236 adjust this estimate to a window of 100 bp, which results in the multiplication of N_r by

100 in the above formula. We also compared the SNP density between miRNA loci ($N = 370$) and their flanking genomic regions (± 1 kb) by considering 10 upstream and downstream nucleotide bins comprising 100 bp each. Differences in SNP density across miRNA loci, as well as compared with surrounding regions, were estimated with a non-parametric approach using a Mann Whitney U test [50].

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

Whole-Genome sequencing of five Duroc pigs

In the year 2,003, 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. By whole-genome sequencing of these 5 individuals, we aimed to identify miRNA SNPs and investigate their association with gene expression traits and lipid phenotypes recorded in the offspring of the 5 boars (Lipgen population, $N = 345$). Genomic DNA was extracted [51] and sequenced at the Centre Nacional d'Anàlisi Genòmica (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 [41] and the same protocol defined in a previous section

(please see the section describing Whole-genome sequence data processing and calling of single nucleotide polymorphisms). Since only five boars were sequenced, SNPs within miRNA loci were retrieved without quality filtering criteria, in an effort to maximize the number of identified miRNA SNPs.

Recording of phenotypes related with gene expression and lipid traits in the Lipgen population

A total of 350 Duroc barrows, sired by the five Duroc founder boars mentioned before were used as a resource population (Lipgen population [52,53]) to evaluate associations between miRNA SNPs and phenotypes related with gene expression and lipid traits. 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 in four contemporary batches at the IRTA pig experimental farm in Monells (Girona, Spain) under intensive standard conditions. Once they reached ~122 kg of live weight (~190 days of age), they were slaughtered in a commercial abattoir following the guidelines of current Spanish legislation (<https://www.boe.es/buscar/doc.php?id=BOE-A-1995-3942>). After slaughtering, tissue samples from *gluteus medius* (GM) and *longissimus dorsi* (LD) skeletal muscles and liver were obtained according to previously described methods [54–56].

Total DNA was extracted from each sample following Vidal et al. 2005 [51]. A subset of 345 DNA samples from the initial set of 350 pigs were successfully obtained and processed for further genotyping. Total RNA was extracted from GM (N = 89 pigs) and liver (N = 87 pigs) tissue samples by using the Ribopure isolation kit (Ambion,

285 Austin, TX). Expression mRNA profiles were characterized by hybridizing total RNA
286 to GeneChip porcine arrays (Affymetrix Inc., Santa Clara, CA), which encompass a
287 total of 23,998 probes [54,57]. Further details about tissue collection, sample selection,
288 RNA isolation and microarray hybridization procedures can be found in Cánovas et al.
289 (2010) [54]. Microarray data pre-processing, background correction, normalization and
290 $\log_2$ -transformation of expression estimates were performed with a robust multi-array
291 average method per probe [58]. The *mas5calls* function from the affy R package [58]
292 was then applied to identify probes displaying intensity signals above the background
293 noise. This function applies a Wilcoxon signed rank-based gene expression
294 presence/absence detection algorithm for labeling expressed probes in each sample.
295 Control probes and those with expression levels below the detection threshold in more
296 than 50% of samples were discarded from further analyses. By doing so, we only
297 retained probes corresponding to mRNAs that were robustly expressed in GM muscle
298 (12,131 probes) and liver (12,787 probes) tissues. Probes from GeneChip Porcine
299 Genomic array identifiers (Affymetrix, Inc., Santa Clara, CA) of GM skeletal muscle
300 and liver tissues were assigned to mRNA genes annotated in the Sscrofa11.1 assembly
301 [39] using the BioMart tool [59].

302 With regard to lipid-related phenotypes, 38 traits were measured in the Duroc
303 Lipgen population, i.e. backfat thickness measured between 3rd and 4th ribs, backfat
304 thickness at the last rib, ham fat thickness, and intramuscular fat content and
305 composition of GM and LD skeletal muscle samples (N = 345) [53,55] (**Additional file**
306 **3: Table S3**). Briefly, intramuscular fatty acid (IMF) content in the GM and LD
307 muscles was measured with the Near Infrared Transmittance technique (NIT, Infratec
308 1625, Tecator Hoganas, Sweden), while muscle cholesterol measurements were inferred
309 following Cayuela et al. [60]. Gas chromatography of methyl esters was used to

determine muscle fatty acids (FA) composition, including the percentages of saturated (SFA), unsaturated (UFA), monounsaturated (MUFA) and polyunsaturated (PUFA) fatty acids. Live and carcass weights (used as covariates in the statistical analyses), as well as backfat and ham fat thickness, were measured on a regular basis prior (live weight and backfat thickness) and after slaughter (carcass weight and ham fat thickness). Mean and standard deviations of lipid traits recorded the Lipgen population (N = 345) are described in **Additional file 3: Table S3**.

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

Twelve SNPs mapping to miRNA loci and segregating in the five sequenced parental Duroc boars were genotyped in the 345 Duroc pigs from the Lipgen population (**Table 1**). Briefly, selected miRNA SNPs and their flanking regions (60 upstream and downstream bp) were used for assay design with the Custom TaqMan Assay Design Tool website (<https://www5.appliedbiosystems.com/tools/cadt/>; Life Technologies). Genotyping tasks were carried out at the Servei Veterinari de Genètica Molecular of the Universitat Autònoma of Barcelona (<http://sct.uab.cat/svbm/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 corresponding to the 12 miRNA SNPs mentioned before (**Table 1**) were processed with the PLINK software [61] in order to generate formatted files for subsequent analyses. The Genome-Wide Efficient Mixed-Model Association

(GEMMA) software [62] was used to implement association analyses between genotyped SNPs and phenotypes related with microarray gene expression data and lipid traits. 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 following covariates: IMF of LD (for LD fatty acid composition traits), IMF of GM (for GM fatty acid composition traits), live weight (for backfat thickness) and carcass weight (for ham fat thickness); W corresponds to the incidence matrix relating phenotypes with their corresponding effects; x is the genotype vector for the 12 miRNA polymorphisms; δ is the allele substitution effect for each polymorphism; u is a vector of random individual genetic effects with an 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.

Association analyses were performed between each miRNA polymorphism (**Table 1**) and the 38 lipid-related traits listed in **Additional file 3: Table S3**. Moreover,

we explored the associations between miRNA SNPs and expression levels of mRNAs that fulfilled the following conditions: 1) the 3'UTR of the mRNA contains at least one 7mer-m8 site complementary to the seed of the miRNA harboring the SNP to be tested (as predicted with the *locate* tool from SeqKit software [46]), 2) the existence of experimentally validated mRNA-miRNA interactions between the given mRNA and the tested miRNA have been confirmed in humans according to information provided in the DIANA-Tarbase v8 database [63]. The first condition was established because we were interested in focusing our analysis on miRNA-mRNA pairs for which a direct physical interaction (through complementarity between the seed of the miRNA and a target site in the 3'UTR of the mRNA) is feasible. The second condition was imposed because there is extensive evidence that in silico prediction of miRNA binding sites can have high false positive rates [64]. We considered as valid miRNA-mRNA interactions those experimentally validated in humans by using wet lab methods such as cross-linking, ligation and sequencing of hybrids (CLASH), photoactivatable ribonucleoside-enhanced and high-throughput sequencing of RNA isolated by crosslinking and immunoprecipitation (PAR-CLIP and HITS-CLIP) and luciferase assays [63].

Importantly, the seeds of the porcine miRNAs harboring the 12 genotyped SNPs were completely conserved with regard to the seeds of the human orthologous miRNA genes (**Additional file 4: Fig. S1**). This feature further supports the extrapolation of miRNA-mRNA interactions experimentally validated in humans [63] to pigs. For the sake of completeness, we also carried out a complementary association analysis between miRNA SNPs and the mRNA levels of the whole sets of genes expressed in the GM and liver tissues (i.e., without applying the two conditions mentioned before) but the results of such analyses should be interpreted with caution due to reasons stated above.

379 In association analyses with lipid traits or gene expression phenotypes, the
380 alternative hypothesis $H_1 : \delta \neq 0$ was contrasted against the null hypothesis $H_0 : \delta = 0$
381 with a likelihood ratio test. The statistical significance of the associations between
382 miRNA SNPs and lipid and mRNA expression phenotypes was assessed with a false
383 discovery rate (FDR) approach [65].

384

385 *Pathway enrichment analysis*

386 The lists of probes/genes significantly associated at the nominal level (P -value < 0.05)
387 with miRNA SNPs genotypes ($N = 12$) after considering the criteria for miRNA target
388 inference (see previous Methods section) were used as inputs for pathway enrichment
389 analyses. The ClueGO v2.5.0 plug-in application [66] within the Cytoscape 3.8.2
390 software [67] was used for determining enriched Reactome and KEGG pathways. A
391 one-sided hypergeometric test of significance was applied for determining enriched
392 terms and multiple testing correction was implemented with a false discovery rate
393 approach [65].

394

395 *Confirming associations between ssc-miR-326 rs319154814 genotypes and gene* 396 *expression data by RT-qPCR*

397 The hepatic mRNA levels of three of the mRNA transcripts amongst the most
398 significantly associated with ssc-miR-326 rs319154814 (G/A) genotype were analyzed
399 by reverse transcription-quantitative polymerase chain reaction (RT-qPCR). The β -actin
400 (*ACTB*) and TATA-Box binding protein (*TBP*) genes were used as endogenous
401 controls, as previously reported [68]. In brief, 100 mg of liver tissue from 10 selected

402 samples (5 from each GG and AA genotypes) were submerged in 1 mL TRIzol reagent
 403 (Invitrogen Corp., Carlsbad, CA, United States) and subsequently homogenized using
 404 the Lysing Matrix D reagent (MP Biomedicals, Santa Ana, CA) in a Precellys 24 tissue
 405 homogenizer (Bertin Instruments, Rockville, MD). Total RNA isolation was performed
 406 following the protocol described by Rio et al. (2010) [69]. Reverse transcription was
 407 achieved with the High-Capacity cDNA Reverse Transcription Kit (Applied
 408 Biosystems, Foster City, CA) following the instructions of the manufacturer. We used 1
 409 µg of total RNA as a template in a final volume of 20 µl and the synthesized cDNA was
 410 diluted to 1:20 in Milli-Q water. One pair of primers spanning exon-exon junctions
 411 were designed for each gene (**Additional file 5: Table S4**) with the Primer Express
 412 software (Life Technologies Corporation). Real-time qPCR reactions contained 10 µl of
 413 SYBR Select Master Mix, 300 nM of each primer, and 5 µl of 1:20 cDNA dilution, in a
 414 final volume of 20 µl. Reactions were run in an ABI PRISM 7900HT instrument
 415 (Applied Biosystems, Foster City, CA). The thermal cycle was set as follows: 2 min at
 416 50 °C, one denaturing step at 95 °C during 10 min, followed by 40 cycles of 15 sec at 95
 417 °C and 1 min at 60 °C. Moreover, a melting curve analysis was performed (i.e., 95 °C
 418 for 15 sec, 60 °C for 15 sec and a gradual increase in temperature, with a ramp rate of
 419 1% up to 95 °C, followed by a final step of 95 °C for 15 sec), in order to assess the
 420 specificity of the reactions. A standard curve with serial dilutions from a pool including
 421 all the cDNA samples was used to assess whether amplification efficiencies of the three
 422 genes under analysis were comprised in the 90 to 110% range. All reactions were run in
 423 triplicate.

424 The $2^{-\Delta\Delta C_t}$ method [70] was used for the relative quantification of gene expression (Rq),
 425 using the group of GG samples as a calibrator. Subsequently, Rq values were log₂
 426 transformed and the significance of expression differences between GG and AA

427 genotypes was assessed using Welch's t-test for unpaired groups of samples [71]
428 implemented in the *t.test* R function.

429

430 *Measuring the expression of the ssc-miR-326 gene in pigs with different rs319154814*
431 *genotypes.*

432 Liver samples from 9 AA and 9 GG pigs (rs319154814 genotype) were used to measure
433 the expression of the ssc-miR-326 gene. Total RNA was isolated with TRIzol as
434 previously described and reverse transcribed with the TaqMan Advanced MicroRNA
435 Reverse Transcription Kit (Applied Biosystems, Foster City, CA), following the
436 instructions of the manufacturer. Total RNA (10 ng) diluted in a volume of 2 µl, was
437 used as a template to generate and pre-amplify cDNA by carrying out four consecutive
438 reactions (poly-A tailing, adaptor ligation, reverse transcription and pre-amplification).
439 The resulting pre-amplified cDNA was diluted to 1:50 in Milli-Q water. To measure
440 ssc-miR-326 expression levels, both let-7a and miR-26a-5p were chosen as endogenous
441 controls following Timoneda et al. (2012) [72]. To conduct the experiments for the
442 three miRNAs under investigation (one target and two controls), available inventoried
443 TaqMan advanced miRNA assays (Applied Biosystems, Foster City, CA) for each
444 miRNA were purchased. The reactions were performed in a final volume of 15 µl that
445 contained: 3.75 µl of 1:50 cDNA, 3 µl of nuclease-free water, 7.5 µl of TaqMan Fast
446 Advanced Master Mix (2×), and 0.75 µl TaqMan Advanced miRNA Assay (20×). Each
447 reaction was carried out in triplicate in a 384-well plate. Reactions were run in a
448 QuantStudio 12K Flex Real-Time PCR System (Applied Biosystems, Foster City, CA).
449 The thermal cycle program was the same used for qPCR experiments on mRNA
450 transcripts. No standard dissociation curve was performed. The results were analyzed
451 applying the $2^{-\Delta\Delta C_t}$ method [70] for the relative quantification of miRNA expression

(Rq), using the group of GG samples as a calibrator. Subsequently, Rq values were log₂ transformed and the significance of expression differences between GG and AA genotypes was assessed using a Welch's t-test for unpaired groups [71].

Results

Differential segregation of SNPs mapping to microRNA genes in pigs and wild boars from Europe and Asia

A total of 58,537,491 million SNPs were identified with the GATK haplotype caller tool [41] in a data set comprising 120 WGS from 40 EDM, 40 ADM, 20 EWB and 20 AWB individuals (**Additional file 1: Table S1**). The distribution of these SNPs within different annotated regions (i.e. protein coding genes, exons, introns, 3'UTRs, miRNAs, lncRNAs, snRNAs, snoRNAs and pseudogenes) can be found in **Additional file 6: Table S5**. The majority of these SNPs were biallelic (96.82%), 777,008 (1.33%) were tri-allelic, 1,864,388 (3.18%) had a deletion allele and 57,878 (0.098%) showed four or more alleles (multi-allelic). From the set of multi-allelic SNPs, 1,765 SNPs mapped to exons, and 531 of them were clustered in a set of 74 genes enriched in olfactory receptors (**Additional file 7: Table S6**). This latter result is consistent with the potential presence of copy number variation (CNV) in such locations [73]. 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.

474 After filtering, 19,720,314 autosomal whole-genome SNPs were selected for
475 assessing population structure based on PCA clustering techniques. The PCA showed a
476 strong genetic differentiation amongst Asian and European populations (**Fig. 2A**). In
477 contrast, domestic pigs and wild boars, and particularly those with European origin, did
478 not show such stark divergence. Asian pigs and wild boars displayed some level of
479 genetic differentiation and were more diverse than their European counterparts.

480 With regard to miRNA variability, the 370 porcine miRNA genes annotated in
481 the manually curated miRCarta v1.1 database [44] were selected, and SNPs within these
482 genes were retrieved. A total of 285 SNPs residing in 139 miRNAs (37.56% of the total
483 count) were identified (**Additional file 2: Table S2**), implying that most of miRNAs are
484 monomorphic. The majority of these 139 miRNA loci (76.98%) contained 1-2 SNPs
485 located inside their predicted genomic boundaries, while 18.70% contained between 3
486 up to 5 variants, and 4.32% of them displayed more than 7 SNPs (**Additional file 8:**
487 **Fig. S2**). Only 43 miRNA SNPs (15.09%) were shared amongst EDM, ADM, EWB and
488 AWB populations (**Additional file 2: Table S2**), showing alternative alleles in at least
489 one of the analyzed individuals in each group. The number of SNPs segregating in each
490 of the four defined groups were 129 (EDM), 201 (ADM), 76 (EWB) and 172 (AWB),
491 respectively (**Additional file 2: Table S2**). With regard to precursor and mature
492 regions, 41 and 2 SNPs were shared amongst the four populations under consideration,
493 respectively (**Additional file 9: Fig. S3A and S3B**). None of the SNPs in the seed
494 regions were shared by the four porcine populations (**Additional file 9: Fig. S3C**). Only
495 three miRNA SNPs were found in the European data set but not in the Asian one. In
496 strong contrast, 55 miRNA SNPs were detected in the Asian data set but not in the
497 European one. Principal component analyses based on identified autosomal miRNA
498 SNPs (N = 260, **Additional file 2: Table S2**) showed the existence of a poor

499 differentiation between pigs and wild boars (**Fig. 2B**), while the genetic divergence
500 between European and Asian individuals was still apparent.

501 When we analyzed the population structure based on whole-genome autosomal
502 3'UTR SNPs (N = 709,343 SNPs, **Fig. 2C**), 3'UTR 7mer-m8 site SNPs (N = 107,196
503 SNPs, **Fig. 2D**) and 3'UTR 8mer site SNPs (N = 33,511 SNPs, **Fig. 2E**), the genetic
504 differentiation between Asian vs European populations was evident, in close
505 concordance with results shown in **Fig. 2A** and **2B**. However, we also detected a more
506 pronounced differentiation between domestic pigs and wild boars, and this observation
507 was particularly applicable to Asian pigs and wild boars (**Fig. 2C-D**).

508

509 **The analysis of European and Asian populations shows reduced variability in** 510 **porcine microRNAs**

511 About 47.76%, 57.36%, 44.77% and 36.84% of miRNA SNPs showed
512 alternative allele frequencies ≤ 0.1 in the ADM, EDM, AWB and EWB populations,
513 respectively (**Fig. 3, Additional file 2: Table S2**). Variations located at mature miRNA
514 and seed regions were enriched in rare or very rare variants when compared to the
515 variability of miRNA precursor regions (**Additional file 2: Table S2**), with average
516 alternative allele frequencies of ~ 0.1 for mature and seed miRNA polymorphisms. In
517 contrast, the average alternative allele frequency observed for SNPs in precursor areas
518 was ~ 0.15 .

519 Moreover, the observed SNP density adjusted to 100 bp for miRNA precursor,
520 mature and seed regions consistently followed the order *precursor* > *mature* > *seed*
521 when we considered the full set of 120 WGS. Indeed, ~ 1 SNP per 100 bp was detected
522 in precursor regions, whereas ~ 0.86 and ~ 0.42 SNPs per 100 bp were observed in the

523 mature and seed regions, respectively (**Fig. 4A**). After computing a non-parametric
524 comparison of SNP density between the mature and seed regions with a Mann Whitney
525 U test [50], a significant difference (P -value = 0.016) was detected. These results
526 implied a 1.16-fold reduction in SNP density between precursor and mature regions,
527 while for seed regions, which are critical determinants of miRNA-mRNA interaction,
528 the observed SNP density was ~2.4 fold lower than in precursor regions. This
529 differential distribution of the SNP density across miRNA regions (*precursor* > *mature*
530 > *seed*) was also observed in each of the analyses performed in the ADM, EDM, AWB
531 and EWB groups (**Fig. 4A**). With regard to variants located within mature miRNAs (N
532 = 64), both inside (N = 12) and outside seed regions (N = 52), their distribution along
533 the sequence of the mature miRNA (~22 nts) showed a characteristic pattern (**Fig. 4B**):
534 amongst all the detected SNPs, the 1st position of the mature miRNA 5' end, which
535 binds to the MID domain of the Argonaute protein in the miRISC complex, exhibited a
536 SNP density of ~0.49 SNPs per 100 bp. Such scarcity in polymorphic sites was also
537 observed when considering the next 2nd to 8th positions in the mature miRNA sequence
538 (seed region), with an average of ~0.42 SNPs per 100 bp across the whole seed and up
539 to ~0.73 SNPs per 100 bp in the 6th position of the mature miRNA. In contrast, the
540 interval comprising positions 9th to 12th (with no functional implications in terms of
541 mRNA targeting) showed an increased average SNP density of ~0.98 SNPs per 100 bp.
542 Positions 13th to 18th of the mature miRNA, which roughly correspond to a functional
543 region providing additional anchor pairing to the seed region, showed a decreased SNP
544 density, particularly at positions 16th to 17th (**Fig. 4B**). Additionally, an increased SNP
545 density was found at positions 19th to 22nd. Furthermore, the SNP density in miRNA
546 genes (N = 370, **Fig. 4C**) was ~2.6 fold lower (P -value < 0.01, see Methods) than in
547 their flanking regions ($\pm$ 1 kb). A list of the miRNA SNPs (N = 64) located at mature

548 miRNAs and their genomic coordinates within the mature miRNA sequence is available
549 at **Additional file 10: Table S7**.

550

551 **Statistics of the whole-genome sequencing of five Duroc boars**

552 As previously explained, we sequenced the genomes of five Duroc pigs which
553 founded a population of 350 offspring with the goal of identifying SNPs in miRNA
554 genes and investigating their association with mRNA expression and lipid phenotypes.
555 Mean coverage values of the five pig genomes ranged from 37.67× to 46.6×, with more
556 than 98.6% of the genome covered by at least 10 reads in all five samples, and 96.71%
557 of the genome covered by at least 15 reads. More details about coverage and genome
558 mapping parameters are shown in **Additional files 11 and 12: Table S8 and Fig. S4**.
559 After performing variant calling on mapped reads, a total of 13,839,422 SNPs passed
560 the established quality filters, whereas 3,721,589 insertions and deletions (INDELs)
561 were detected. Moreover, a total of 1,643,861 INDELs (44.17%) and 6,034,548 SNPs
562 (43.60%) were located in annotated protein coding loci. A total of 55 SNPs and 5
563 INDELs were located in the 370 miRNA loci annotated in the Sscrofa.11.1 genome
564 assembly according to the miRCarta database [44]. From these, 53 variants (96.36%)
565 were shared with the set of 285 miRNA SNPs found in the 120 WGS reported before
566 (**Additional file 13: Table S9**). After performing quality check of the variants mapping
567 to miRNA genes, 49 SNPs and 5 INDELs passed all established filtering criteria.

568 Amongst the 49 identified miRNA polymorphisms, we selected 12 SNPs on the
569 basis of their location at relevant annotated miRNA loci (**Table 1, Additional file 13:**
570 **Table S9**) for genotyping and performing association analyses. Microarray expression

571 data comprised a total of 23,998 probes, of which 12,131 (50.55%) and 12,787
572 (53.28%) corresponded to genes expressed in GM and liver tissues, respectively.

573 To provide a comprehensive view about the association between miRNA
574 genotypes and liver and muscle mRNA levels, we have carried out an association
575 analysis between miRNA genotypes and the whole set of expressed probes/genes in
576 these two tissues (**Table 2**). It can be seen that there are hundreds of associations at the
577 nominal level, while the number of significant associations is strongly reduced after
578 multiple testing correction. This analysis does not take into account neither the
579 existence of complementarity between the seed of the miRNA and the porcine 3'UTR
580 of target mRNAs, nor functional validation data obtained in humans (please see the
581 Methods section). A second analysis taking into account these two criteria revealed that
582 from the set of 12 genotyped SNPs listed in **Table 1**, only 2 SNPs showed significant
583 associations (q -value < 0.1) with the levels of mRNAs predicted to be miRNA targets
584 (**Table 3, Additional file 14: Table S10**). Further details of target sites at the 3'UTR of
585 mRNA transcript expression profiles significantly associated with miRNA SNP
586 genotypes (**Table 3**) are available at **Additional file 15: Table S11**). All genes
587 displaying significant associations (q -value < 0.1) in this second analysis ranked
588 amongst the top 1% of significance (**Additional file 16: Table S12**) in the condition-
589 free analysis (Analysis 1), except for transcripts associated with ssc-miR-326 in the
590 liver, which ranked amongst the top 5% (**Table 1, Additional file 16: Table S12**). We
591 observed that for several of the significant associations between miRNA and mRNA
592 pairs reported in **Table 3** and **Additional file 14: Table S10**, the interaction was
593 classified as 7-mer-m8 by our targeting approach with the *locate* tool of SeqKit toolkit
594 [46] when, in reality, it was of type 8mer with one mismatch (data not shown).

595

596 **Two SNPs in the apical loop of ssc-miR-326 and in the precursor region of ssc-**
597 **miR-23a are associated with the mRNA expression of several of their putative**
598 **targets.**

599 When we analyzed the association between the rs319154814 (G/A)
600 polymorphism located in the apical loop of ssc-miR-326 and gene expression data
601 (**Table 3**), significant results were obtained after multiple testing correction (q -value <
602 0.1). More specifically, we detected eight significant associations between this SNP and
603 the hepatic mRNA expression of experimentally confirmed targets of this miRNA
604 (**Table 3**). For instance, the hepatic mRNA levels of the cellular FLICE-like inhibitory
605 protein (*CFLAR*), the protein phosphatase 1 catalytic subunit γ (*PPP1CC*), the syndecan
606 1 (*SDC1*), the splicing factor 3A subunit 3 (*SF3A3*) or the Follistatin-like 1 (*FSTLI*)
607 were amongst the most significantly associated with ssc-miR-326 genotypes (**Table 3**).
608 The expression levels of seven of these mRNAs associated with ssc-miR-326
609 rs319154814 (G/A) genotypes (**Table 3**) were reduced in pigs homozygous for the
610 mutated allele (AA, N = 32), as depicted in **Fig. 5**.

611 The mRNA levels of the *PPP1CC*, *CFLAR* and *SF3A3* mRNA transcripts were
612 measured by RT-qPCR in 10 liver samples (5 belonging to each AA or GG
613 rs319154814 genotypes). After comparing the mean expression for each of the three
614 analyzed mRNA transcripts in both AA and GG genotype groups, only the *PPP1CC*
615 transcripts showed a significantly reduced expression in AA pigs compared with their
616 GG counterparts (P -value = 0.027, **Fig. 6A**). The other two transcripts (*CFLAR* and
617 *SF3A3*) also showed a reduced expression in pigs with AA genotype (**Fig. 6A**), but
618 results were not significant. We further investigated whether the ssc-miR-326
619 rs319154814 (G/A) genotypes were associated with the levels of this very same miRNA
620 in the liver by using a specific Taqman Advanced miRNA probe assay. In this qPCR

analysis, pigs with AA genotypes for the rs319154814 polymorphism showed an increased ~1.9-fold expression of ssc-miR-326 transcripts compared with their GG counterparts, although such difference did not reach statistical significance (P -value = 0.178, **Fig. 6B**).

The rs333787816 (T/C) polymorphism, located in the precursor region, immediately downstream the mature ssc-miR-23a sequence, was also significantly associated with 36 experimentally confirmed targeted genes in the GM muscle tissue (q -value < 0.1, **Table 3, Additional file 14: Table S10**). 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 3**).

633

Pathway enrichment analyses of significantly associated mRNA targets provide insights of putative metabolic functions of miRNAs.

After performing pathway enrichment analyses on the sets of putative targeted mRNA transcripts significantly associated with miRNA SNP genotypes (P -value < 0.05), several biological categories were identified and reported at **Additional file 17: Table S13**. To mention a few of the most relevant findings, the ssc-miR-23a genotype was associated with several regulatory pathways (post-transcriptional silencing by small RNAs, competing endogenous RNAs regulate PTEN translation, regulation of PTEN mRNA translation, regulation of RUNX1 Expression and Activity etc.) involving genes, such as *AGO1*, *AGO2* or *TNRC6C* genes, which play relevant roles in miRNA-mediated gene regulation [74]. Amongst the sets of mRNAs nominally associated with ssc-miR-

645 30d genotype in GM, we identified the regulation of ornithine decarboxylase activity or
646 cellular response to hypoxia, as well as mature mRNA transport after splicing
647 **(Additional file 17: Table S13).**

648 In the liver tissue, the ssc-miR-326 genotype was significantly associated with the
649 expression of several circadian clock regulatory transcripts e.g., histone deacetylase 3
650 (*HDAC3*), period 1 (*PER1*) and the already mentioned *PPP1CC* gene. Moreover,
651 negative regulation of DDX58/IFIH1 signaling of type I interferon response and p53
652 signaling were pathways enriched in the sets of mRNAs nominally associated with ssc-
653 miR-130a genotypes **(Additional file 17: Table S13).**

654

655 **Porcine lipid phenotypes are associated with the genotypes of miRNA genes**

656 We also evaluated the association between miRNA SNPs and several lipid-
657 related phenotypes recorded in the Lipgen population **(Additional file 3: Table S3).**
658 Only the rs319154814 (G/A) variant in the ssc-miR-326 gene was significantly
659 associated (q -value < 0.1) with lipid traits **(Table 4, Additional file 18: Table S14).**
660 More specifically, we found significant associations with myristic acid (C14:0) content
661 in both LD and GM muscles, as well as with the gadoleic acid (C20:1) content and the
662 ratio between PUFA and MUFA in the LD muscle **(Table 4, Additional file 18: Table S14).**
663 We also observed several associations between the rs319154814 (G/A) SNP and
664 fatty acid composition traits, but they were significant only at the nominal level (P -
665 value < 0.01 , **Table 4**). Other apical loop SNPs like rs81349391 (A/G) at ssc-miR-339
666 and rs344472188 (T/C) at ssc-miR-130a were significantly associated at the nominal
667 level (P -value < 0.01) with palmitic acid (C16:0) content and SFA and UFA proportion
668 in the GM muscle, as well as with backfat thickness. As shown in **Table 4**, a SNP

669 located in the precursor 3' stem of ssc-miR-30d showed a nominally significant
670 association with the content of arachidic acid (C20:0). Other significant associations at
671 the nominal level were, for instance, those between the rs334590580 (T/C) SNP located
672 at the precursor stem region of ssc-miR-335 and palmitic and arachidic acids content in
673 GM tissue (**Table 4**).

674

675 **Discussion**

676 **Divergent patterns of variation for microRNA and 3'UTR polymorphisms in Asian** 677 **and European pigs and wild boars**

678 The PCA revealed the existence of a detectable genetic differentiation between
679 Asian and European populations, with the latter showing reduced levels of diversity
680 when compared to the former. Groenen et al. (2012) [75] investigated the variability of
681 pig genomes and found that Asian pigs and wild boars are more diverse than their
682 European counterparts and that both gene pools split during the mid-Pleistocene 1.6–0.8
683 Myr ago. Calabrian glacial intervals probably favored a restricted gene flow between
684 these two gene pools [75]. The high variability of Asian populations could be explained
685 by the fact that *Sus scrofa* emerged as a species in Southeast Asia (5.3-3.5 Mya) and
686 then dispersed westwards until reaching Europe around 0.8 Mya [76]. This initial
687 founder effect combined with the occurrence of strong bottlenecks reduced the genetic
688 diversity of European wild boars [75].

689 While genetic differentiation between wild boar and pig populations was clearly
690 discernible considering the whole-genome SNP data set (**Fig. 2A**), this was less evident
691 in the PCA based on miRNA SNPs (**Fig. 2B**), 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 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 8mer (33,511 SNPs) sites (**Figs. 2C-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 [77], so in general, they are not expected to have drastic consequences on gene function [77]. Purifying selection is less intense in 3'UTRs than in protein-coding regions, implying that 3'UTRs evolve faster and accumulate a larger fraction of recent polymorphisms contributing to population differentiation [78].

703

704 **Low SNP density in microRNA genes and lack of uniform SNP distribution across** 705 **sites**

We have found that, in general, miRNA loci have a substantially lowered SNP density in their seeds when compared with mature and precursor regions (**Fig. 4A**). When we analyzed the SNP density in miRNA loci and their flanking regions (± 1 kb), a significantly reduced amount of SNPs per 100 bp was found in miRNAs compared with upstream and downstream flanking sequences (**Fig. 4C**). These results were in agreement with data reported by Omariba et al. (2020) [79] and Saunders et al. (2007) [5].

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

716 across species [16,23,80], as this sequence ultimately determines the success of
 717 miRNA-mRNA interactions [2]. In our study, a total of 221, 52 and 12 SNPs were
 718 found in precursor, mature and seed regions within miRNA loci, respectively
 719 (**Additional files 2 and 9: Table S2, Fig. S3**). Gong et al. (2012) [81] described the
 720 existence of 40% polymorphic miRNAs in the human genome but only 16% of them
 721 displayed more than one SNP. In a more recent study, He et al. (2018) [82] reported
 722 1,879 SNPs in 1226 (43.6%) human miRNA seed regions, and 97.5% of these
 723 polymorphisms had frequencies below 5%. These results agree well with the overall
 724 frequency distribution of miRNA SNPs in the European and Asian populations analyzed
 725 in the current work (**Fig. 3**). He et al. (2018) also demonstrated that 1,587, 749, 340,
 726 102, 31, and 4 miRNAs harbored zero, one, two, three, four, and five SNPs,
 727 respectively, in their seed regions, reflecting that mutations in this critical functional
 728 region are not well tolerated [82]. This distribution is similar to the one that we have
 729 observed in domestic pigs and wild boars, with 81, 31, 11, 9 and 5 miRNAs harboring
 730 one, two, three, four and five SNPs (**Additional file 8: Fig. S2**). Only 4 and 2 miRNAs
 731 showed a total of seven and ten polymorphisms within their sequences (**Additional file**
 732 **2: Table S2**).

733 We have also detected a non-uniform SNP density along the sequence of mature
 734 miRNAs (**Fig. 4B**). Gong et al. (2012) [81] showed that SNPs tend to concentrate in the
 735 middle region of the mature miRNA gene rather than in its 5' and 3' ends. Moreover, the
 736 same authors described an increased SNP density at positions 9 and 15 of the mature
 737 miRNA, a result that agrees well with ours (**Fig. 4B**). However, we have also identified
 738 an elevated number of polymorphic sites at positions 11, 19 and 20, a finding that does
 739 not match human data presented by Gong et al. (2012) [81].

740 Several of the sites showing a reduced variability in porcine miRNAs exert
741 critical functions (**Fig. 4B, Additional file 10: Table S7**). For instance, the 1st
742 nucleotide of mature miRNAs plays an important role in the loading process of the
743 mature miRNA on the Argonaute protein to form the miRISC complex [83].
744 Nucleotides 2nd to 8th in the mature miRNA correspond to the seed, where we found a
745 consistently reduced SNP density (**Fig. 4A and 4B**) compared with other miRNA
746 regions. This result was expected because this region has a crucial role in modulating
747 the interaction between the mature miRNA and its 3'UTR binding sites. Polymorphisms
748 in the seed region have the potential to disrupt the proper miRNA-mRNA pairing and
749 thus alter biologically relevant regulatory pathways, which tend to be evolutionarily
750 conserved [80]. Hence, these variants located in the miRNA seeds (**Additional file 10:**
751 **Table S7**) might favor the emergence of novel miRNA-mRNA interactions as well as
752 the abolishment of conserved ones, thus modifying gene regulatory networks.

753 In contrast with the first eight 5' positions of the miRNA, we found a higher
754 SNP density in positions 9th to 12th, which do not contribute substantially to miRNA
755 target recognition [2] (**Fig. 4B**). Positions 13th to 16th facilitate 3'-compensatory pairing
756 between the mature miRNA and targeted 3'UTRs [84], although only at marginal levels
757 [85]. Nevertheless, in our porcine data set only positions 16th-17th showed a reduced
758 SNP density (**Fig. 4B**).

759

760 **Polymorphisms in microRNA genes show associations with the mRNA expression** 761 **of several of their predicted targets**

762 We have investigated the association of 12 SNPs mapping to miRNA genes and
763 hepatic and muscle mRNA expression in Duroc pigs. Several hundreds of significant

764 associations at the nominal level (P -value < 0.05) were detected in both GM muscle and
 765 liver tissues when considering the whole set of microarray expression profiles of mRNA
 766 transcripts (**Table 2**). One pitfall of this analysis is that it does not take into account
 767 whether there is molecular evidence supporting the existence of direct miRNA-mRNA
 768 interactions displaying significant associations. On the other hand, indirect interactions
 769 between miRNAs and mRNAs might also exist i.e., a SNP regulating the expression of
 770 a miRNA might repress the translation of a given mRNA through a direct interaction
 771 (an event that would not necessarily imply any transcriptional decay, hence being
 772 undetectable in our experimental system) and, in turn, this might indirectly affect the
 773 expression of other multiple mRNAs regulated by the translationally repressed mRNA.
 774 Despite this consideration, we decided to restrict our association analyses to miRNA-
 775 mRNA pairs with molecular evidence of potential interactions because we believe that
 776 with this stringent approach, we are able to minimize the chances of detecting spurious
 777 false positive results. The genomic position of miRNA-binding sequences in the 3'UTR
 778 of targeted mRNAs showing significant associations (q -value < 0.1 , **Table 3**) with
 779 miRNA genotypes is available in **Additional file 15: Table S11**).

780 After correction for multiple testing (q -value < 0.1 , **Table 3**), only SNPs
 781 mapping to the ssc-miR-23a and ssc-miR-326 genes displayed significant associations
 782 with the expression levels of some of their putative mRNA targets. More in detail, the
 783 rs333787816 (T/C) polymorphism in the ssc-miR-23a gene showed significant
 784 associations with several putative target mRNAs, amongst which are *AGO1* and *AGO2*
 785 transcripts, two essential components of the miRNA-mediated silencing machinery
 786 [86,87]. The rs319154814 (G/A) polymorphism in the apical loop region of ssc-miR-
 787 326 also showed a significant association, after multiple testing correction, with the
 788 hepatic mRNA expression of several of its predicted and experimentally confirmed

789 targets (**Table 3**). In contrast, no association with GM muscle mRNA expression was
 790 observed. Tissue-specific differences in the expression of the miRNA or its mRNA
 791 targets might explain such outcome. Indeed, the analysis of the distribution of miRNA
 792 expression across human tissues has shown that only a minority of miRNAs are
 793 expressed ubiquitously [88]. The hepatic mRNA targets showing the most significant
 794 association with rs319154814 (G/A) ssc-miR-326 genotype were *CFLAR*, *PPP1CC*,
 795 *SDC1*, *SF3A3* and *FSTL1*. The protein encoded by the *PPP1CC* gene belongs to the
 796 protein phosphatase PP1 subfamily, which is a ubiquitous serine/threonine phosphatase
 797 involved in regulating multiple cellular processes through dephosphorylation signaling.
 798 Amongst them, it is worth mentioning insulin signaling [89], post-translational
 799 localization of circadian clock components [90] and lipids [91,92] or glycogen
 800 metabolism regulation [93]. Moreover, the cellular FLICE-like inhibitory protein
 801 (*CFLAR*) gene encodes the cFLIP protein and is involved in the inhibition of Fas-
 802 mediated apoptosis [94], while the Follistatin-like 1 (*FSTL1*) gene plays a role in the
 803 immune inflammatory signaling and fibrosis in the liver [95].

804 Pigs homozygous for the derived A-allele of the rs319154814 SNP showed a
 805 consistent downregulation of the hepatic mRNA expression of the *CFLAR*, *PPP1CC*,
 806 *SDC1*, *SF3A3*, *FSTL1*, *NAA50* and *ELAVL1* genes (**Fig. 5**), a result that was confirmed
 807 by qPCR for *PPP1CC* mRNA transcripts (*P*-value = 0.027, **Fig. 6A**). Although in the
 808 qPCR experiment the *CFLAR* and *SF3A3* genes also showed decreased mRNA levels in
 809 pigs with homozygous AA genotypes for the rs319154814 SNP, such associations did
 810 not reach statistical significance (**Fig. 6A**). The significant downregulation of at least
 811 one top mRNA putative target (i.e., *PPP1CC*) might suggest that the rs319154814
 812 (G/A) variant could increase the repressive activity of ssc-miR-326. Indeed, when we
 813 assessed, with a dedicated qPCR assay, the expression level of ssc-miR-326 transcripts

814 in the liver tissue, we found that ssc-miR-326 log₂ relative expression is increased by
815 ~1.9 fold in pigs homozygous for the A-allele, although such result was not statistically
816 significant (*P*-value = 0.178, **Fig. 6B**).

817 As previously discussed, the rs319154814 (G/A) polymorphism is located in the
818 apical loop region of ssc-miR-326. Although the apical region of the miRNA does not
819 have a function as critical as the seed, polymorphisms located in this specific region can
820 have relevant effects on the structural conformation of the pre-miRNA hairpin. Indeed,
821 Fernandez et al. (2017) [19] described a mutation in the apical loop of hsa-miR-30c
822 (G/A) that induces a steric disruption of the pri-miRNA folding structure of the hairpin,
823 hence creating a bulge around the flanking downstream CNNC motif that facilitates the
824 SRSF3 factor accessibility to the RNA sequence [96]. In other words, SNPs in the
825 apical region can modify the efficiency with which the Drosha processing machinery, is
826 recruited. The rs319154814 (G/A) polymorphism detected in our study might have
827 structural consequences similar to those described for the hsa-miR-30c apical loop
828 variant [19].

829 One clear limitation of our experimental design is that we did not test the
830 existence of miRNA-mRNA interactions by using functional assays. Such experiments
831 would be necessary to demonstrate our hypothesis that the rs319154814 (G/A)
832 polymorphism in the apical loop of ssc-miR-326 has a role in promoting the processing
833 and maturation efficiency of the hairpin precursor, thus modifying the expression
834 profiles of ssc-miR-326 and its putative mRNA targets.

835

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

838 In our study, the only miRNA polymorphism showing significant associations
839 (q -value < 0.1) with lipid-related phenotypes was the rs319154814 (G/A) SNP in the
840 apical loop of ssc-miR-326. This SNP was associated with myristic fatty acid content in
841 the GM and LD muscles of Duroc pigs, as well as with PUFA/MUFA ratio or gadoleic
842 fatty acid content in LD. To the best of our knowledge, no direct effect of ssc-miR-326
843 on the metabolism of myristic fatty acid has been described so far, but there are reports
844 suggesting that several of the targets of this miRNA might be involved in carbohydrate
845 and lipid metabolism pathways [97,98]. Moreover, the *PPP1CC* transcript, one of the
846 predicted targets of miR-326, encodes a subunit of protein phosphatase-1, which
847 activates acetyl-CoA carboxylase α (ACACA) and 6-phosphofructo2-kinase/fructose-
848 2,6-bisphosphatase (PFKFB), the main regulators of fatty acid synthesis and glycolysis,
849 respectively [92]. The protein phosphatase-1 also activates lipogenic transcription
850 factors such as sterol regulatory element-binding protein 1 (SREBF1), carbohydrate-
851 responsive element-binding protein (MLXIPL) and, moreover, it dephosphorylates the
852 DNA-dependent protein kinase encoded by the *PRKDC* gene, which is another main
853 determinant of hepatic lipogenesis [92]. In summary, a potential effect of the
854 rs319154814 (G/A) SNP in the synthesis or degradation of myristic acid can be
855 envisaged, but this hypothesis still needs to be confirmed at the functional level.
856 Interestingly, Cardoso et al. (2021) [99] explored the relationship between miRNA
857 SNPs located in the seed and mature regions of bovine miRNA loci and fatty acid
858 composition traits. By combining lipid-related phenotype data with mRNA and protein
859 abundances in a multi-omics approach, they described three miRNA SNPs significantly
860 associated with several unsaturated and polyunsaturated fatty acid traits, as well as with
861 the expression levels of some of their predicted targets [99].

862

863 **Conclusions**

864 MicroRNA genes show divergent patterns of variation between Asian and
865 European pigs and wild boars and, in general, they display low levels of polymorphism.
866 As expected, this reduced miRNA variability was particularly prevalent in the seed
867 region, a finding that is likely explained by the strong effects of purifying selection
868 aiming to preserve the conservation of this critical site. We have detected one SNP in
869 the apical loop of ssc-miR-326 and another one in the precursor region of ssc-miR-23a
870 that are associated with the mRNA expression of several of their putative targets. If
871 confirmed with functional assays, these results would reinforce the need of exploring
872 the role of miRNA variation within and outside the seed in the fine-tuning of mRNA
873 expression in pigs.

874

875 **Declarations**

876 **Ethics approval**

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

880

881 **Consent of publication**

882 Not applicable.

883

884 **Availability of data materials**

885 Microarray expression data used in the current study were deposited in the Gene
886 Expression Omnibus (GEO) public repository and are accessible through GEO Series
887 Accession Number GSE115484. Phenotypic and genotypic data sets generated and
888 analyzed during the current study have been deposited in the Figshare public repository
889 available at https://figshare.com/projects/SNPs_miRNA/78690. Whole-genome
890 sequencing dataset from the 5 Duroc boars is available at the Sequence Read Archive
891 (SRA) database (BioProject: PRJNA626370). Variant Calling Format (VCF) files from
892 the 120 European and Asian pigs belonging to domestic breeds and wild boars, as well
893 as from the 5 sequenced Duroc boars are available at the following link:
894 https://figshare.com/projects/VCF_PIGs/93140.

895

896 **Competing interests**

897 The authors declare that they have no competing interests.

898

899 **Funding**

900 The research presented in the current publication was funded by Grants AGL2013-
901 48742-C2-1-R and AGL2013-48742-C2-2-R awarded by the Spanish Ministry of
902 Economy and Competitiveness. We also acknowledge the support of the Spanish Ministry
903 of Economy and Competitiveness for the *Center of Excellence Severo Ochoa* 2016-2019
904 (SEV-2015-0533) grant awarded to the Centre for Research in Agricultural Genomics
905 (CRAG). Emilio Mármol-Sánchez was funded by a FPU PhD grant from the Spanish
906 Ministry of Education (FPU15/01733). Dailu Guan was funded by a PhD fellowship

907 from the Scholarship Council of China (CSC). The authors thank the CERCA
908 Programme of the Generalitat de Catalunya (Barcelona, Spain) for their support and
909 those who provided publicly available data.

910

911 **Authors' contributions**

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

919

920 **References**

- 921 1. Gebert LFR, MacRae IJ. Regulation of microRNA function in animals. *Nat Rev Mol*
922 *Cell Biol.* 2019;20:21–37.
- 923 2. Bartel DP. Metazoan MicroRNAs. *Cell.* 2018;173:20–51.
- 924 3. Chen CYA, Shyu A Bin. Mechanisms of deadenylation-dependent decay. *RNA.*
925 2011;2:167–83.
- 926 4. Eichhorn SW, Guo H, McGeary SE, Rodriguez-Mias RA, Shin C, Baek D, et al.
927 mRNA destabilization is the dominant effect of mammalian microRNAs by the time
928 substantial repression ensues. *Mol Cell.* 2014;56:104–15.

- 929 5. Saunders MA, Liang H, Li WH. Human polymorphism at microRNAs and
930 microRNA target sites. *Proc Natl Acad Sci U S A*. 2007;104:3300–5.
- 931 6. Mencía A, Modamio-Høybjør S, Redshaw N, Morín M, Mayo-Merino F, Olavarrieta
932 L, et al. Mutations in the seed region of human miR-96 are responsible for
933 nonsyndromic progressive hearing loss. *Nat Genet*. 2009;41:609–13.
- 934 7. Auyeung VC, Ulitsky I, McGeary SE, Bartel DP. Beyond secondary structure:
935 primary-sequence determinants license pri-miRNA hairpins for processing. *Cell*.
936 2013;152:844–58.
- 937 8. Roden C, Gaillard J, Kanoria S, Rennie W, Barish S, Cheng J, et al. Novel
938 determinants of mammalian primary microRNA processing revealed by systematic
939 evaluation of hairpin-containing transcripts and human genetic variation. *Genome Res*.
940 2017;27:374–84.
- 941 9. Clop A, Marcq F, Takeda H, Pirottin D, Tordoir X, Bibé B, et al. A mutation creating
942 a potential illegitimate microRNA target site in the myostatin gene affects muscularity
943 in sheep. *Nat Genet*. 2006;38:813–8.
- 944 10. Moszyńska A, Gebert M, Collawn JF, Bartoszewski R. SNPs in microRNA target
945 sites and their potential role in human disease. *Open Biol*. 2017;7:170019.
- 946 11. Liang H, Li WH. Lowly expressed human MicroRNA genes evolve rapidly. *Mol*
947 *Biol Evol*. 2009;26:1195–8.
- 948 12. Xiong P, Schneider RF, Hulsey CD, Meyer A, Franchini P. Conservation and
949 novelty in the microRNA genomic landscape of hyperdiverse cichlid fishes. *Sci Rep*.
950 2019;9:1–12.

- 951 13. Peterson KJ, Dietrich MR, McPeck MA. MicroRNAs and metazoan
952 macroevolution: Insights into canalization, complexity, and the Cambrian explosion.
953 *BioEssays*. 2009;31:736–47.
- 954 14. Lee CT, Risom T, Strauss WM. Evolutionary conservation of microRNA regulatory
955 circuits: An examination of microRNA gene complexity and conserved microRNA-
956 target interactions through metazoan phylogeny. *DNA Cell Biol*. 2007;26:209–18.
- 957 15. Junjie LUO, Wang Y, Yuan J, Zhao Z, Jian LU. MicroRNA duplication accelerates
958 the recruitment of new targets during vertebrate evolution. *RNA*. 2018;24:787–802.
- 959 16. Simkin A, Geissler R, McIntyre ABR, Grimson A. Evolutionary dynamics of
960 microRNA target sites across vertebrate evolution. *PLOS Genet*. 2020;16:e1008285.
- 961 17. Li J, Zhang Z. MiRNA regulatory variation in human evolution. *Trends Genet*.
962 2013;29:116–24.
- 963 18. Sun G, Yan J, Noltner K, Feng J, Li H, Sarkis DA, et al. SNPs in human miRNA
964 genes affect biogenesis and function. *RNA*. 2009;15:1640–51.
- 965 19. Fernandez N, Cordiner RA, Young RS, Hug N, MacIas S, Cáceres JF. Genetic
966 variation and RNA structure regulate microRNA biogenesis. *Nat Commun*. 2017;8:1–
967 12.
- 968 20. Jin Y, Lee CGL. Single nucleotide polymorphisms associated with microRNA
969 regulation. *Biomolecules*. 2013;3:287–302.
- 970 21. Larson G, Cucchi T, Dobney K. Genetic aspects of pig domestication. In:
971 Rothschild MF, Ruvinsky A, editors. *Genet of the pig*. 2011. p.14–37.
- 972 22. Ramos-Onsins SE, Burgos-Paz W, Manunza A, Amills M. Mining the pig genome

- 973 to investigate the domestication process. *Heredity*. 2014;471–84.
- 974 23. Penso-Dolfin L, Moxon S, Haerty W, Di Palma F. The evolutionary dynamics of
975 microRNAs in domestic mammals. *Sci Rep*. 2018;8:1–13.
- 976 24. Wei W, Wang G, Cheng Y, Yang R, Song J, Huang S, et al. A miR-511-binding site
977 SNP in the 3'UTR of IGF-1 gene is associated with proliferation and apoptosis of PK-
978 15 cells. *Vitr Cell Dev Biol – Anim*. 2019;55:323–30.
- 979 25. Liu R, Deng D, Liu X, Xiao Y, Huang J, Wang F, et al. A miR-18a binding-site
980 polymorphism in *CDC42* 3'UTR affects *CDC42* mRNA expression in placentas and is
981 associated with litter size in pigs. *Mamm Genome*. 2019;30:34–41.
- 982 26. Ma C, Song H, Yu L, Guan K, Hu P, Li Y, et al. MiR-762 promotes porcine
983 immature Sertoli cell growth via the ring finger protein 4 (RNF4) gene. *Sci Rep*.
984 2016;6:32783.
- 985 27. Stachowiak M, Szydlowski M, Flisikowski K, Flisikowska T, Bartz M, Schnieke A,
986 et al. Polymorphism in 3' untranslated region of the pig PPARA gene influences its
987 transcript level and is associated with adipose tissue accumulation. *J Anim Sci*.
988 2014;92:2363–71.
- 989 28. Bartz M, Szydlowski M, Kociucka B, Salamon S, Jeleń HH, Switonski M.
990 Transcript abundance of the pig stearoyl-CoA desaturase gene has no effect on fatty
991 acid composition in muscle and fat tissues, but its polymorphism within the putative
992 microRNA target site is associated with daily body weight gain and feed conversion
993 ratio. *J Anim Sci*. 2013;91:10–9.
- 994 29. Shao GC, Luo LF, Jiang SW, Deng CY, Xiong YZ, Li FE. A C/T mutation in
995 microRNA target sites in BMP5 gene is potentially associated with fatness in pigs. *Meat*

996 Sci. 2011;87:299–303.

997 30. Bartz M, Koscianska E, Szczerbal I, Nowacka-Woszek J, Kociucka B, Salamon S,
998 et al. Polymorphism of the porcine miR-30d is associated with adipose tissue
999 accumulation, its fatty acid profile and the ME1 gene expression. *Livest Sci.*
1000 2015;182:54–7.

1001 31. Lee JS, Kim JM, Lim KS, Hong JS, Hong KC, Lee YS. Effects of polymorphisms in
1002 the porcine microRNA MIR206 / MIR133B cluster on muscle fiber and meat quality
1003 traits. *Anim Genet.* 2013;44:101–6.

1004 32. Kim JM, Lim KS, Hong JS, Kang JH, Lee YS, Hong KC. A polymorphism in the
1005 porcine miR-208b is Asso. with microRNA biogenesis and expressions of SOX-6 and
1006 MYH7 with effects on muscle fibre characteristics and meat quality. *Anim Genet.*
1007 2015;46:73–7.

1008 33. Chai J, Chen L, Luo Z, Zhang T, Chen L, Lou P, et al. Spontaneous single
1009 nucleotide polymorphism in porcine microRNA-378 seed region leads to functional
1010 alteration. *Biosci Biotechnol Biochem.* 2018;82:1081–9.

1011 34. Choi JW, Chung WH, Lee KT, Cho ES, Lee SW, Choi BH, et al. Whole-genome
1012 resequencing analyses of five pig breeds, including Korean wild and native, and three
1013 European origin breeds. *DNA Res.* 2015;22:259–67.

1014 35. Kim H, Song KD, Kim HJ, Park W, Kim J, Lee T, et al. Exploring the Genetic
1015 Signature of Body Size in Yucatan Miniature Pig. Johnson N, editor. *PLoS One.*
1016 2015;10:e0121732.

1017 36. Frantz LAF, Schraiber JG, Madsen O, Megens HJ, Cagan A, Bosse M, et al.
1018 Evidence of long-term gene flow and selection during domestication from analyses of

1019 Eurasian wild and domestic pig genomes. *Nat Genet.* 2015;47:1141–8.

1020 37. Bosse M, Megens HJ, Madsen O, Crooijmans RPMA, Ryder OA, Austerlitz F, et al.
1021 Using genome-wide measures of coancestry to maintain diversity and fitness in
1022 endangered and domestic pig populations. *Genome Res.* 2015;25:970–81.

1023 38. Bolger AM, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina
1024 sequence data. *Bioinformatics.* 2014;30:2114–20.

1025 39. Warr A, Affara N, Aken B, Beiki H, Bickhart DM, Billis K, et al. An improved pig
1026 reference genome sequence to enable pig genetics and genomics research. *GigaScience.*
1027 2020;9:giaa051.

1028 40. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-
1029 MEM. 2013. <http://arxiv.org/abs/1303.3997>.

1030 41. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al.
1031 The genome analysis toolkit: A MapReduce framework for analyzing next-generation
1032 DNA sequencing data. *Genome Res.* 2010;20:1297–303.

1033 42. O’Rawe J, Jiang T, Sun G, Wu Y, Wang W, Hu J, et al. Low concordance of
1034 multiple variant-calling pipelines: Practical implications for exome and genome
1035 sequencing. *Genome Med.* 2013;5:28.

1036 43. Hasan MS, Wu X, Zhang L. Performance evaluation of indel calling tools using real
1037 short-read data. *Hum Genomics.* 2015;9:20.

1038 44. Backes C, Fehlmann T, Kern F, Kehl T, Lenhof H-P, Meese E, et al. miRCarta: a
1039 central repository for collecting miRNA candidates. *Nucleic Acids Res.* 2018;46:160–7.

1040 45. Kozomara A, Birgaoanu M, Griffiths-Jones S. miRBase: from microRNA sequences

1041 to function. Nucleic Acids Res. 2019;47:155–62.

1042 46. Shen W, Le S, Li Y, Hu F. SeqKit: A cross-platform and ultrafast toolkit for
1043 FASTA/Q file manipulation. PLoS One. 2016;11:e0163962.

1044 47. Grimson A, Farh KK-H, Johnston WK, Garrett-Engele P, Lim LP, Bartel DP.
1045 MicroRNA targeting specificity in mammals: determinants beyond seed pairing. Mol
1046 Cell. 2007;27:91–105.

1047 48. Friedman RC, Farh KKH, Burge CB, Bartel DP. Most mammalian mRNAs are
1048 conserved targets of microRNAs. Genome Res. 2009;19:92–105.

1049 49. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal
1050 components analysis corrects for stratification in genome-wide association studies. Nat
1051 Genet. 2006;38:904–9.

1052 50. Mann HB, Whitney DR. On a test of whether one of two random variables is
1053 stochastically larger than the other. Ann Math Stat. 1947;18:50–60.

1054 51. Vidal O, Noguera JL, Amills M, Varona L, Gil M, Jiménez N, et al. Identification of
1055 carcass and meat quality quantitative trait loci in a Landrace pig population selected for
1056 growth and leanness. J Anim Sci. 2005;83:293–300.

1057 52. Gallardo D, Pena RN, Amills M, Varona L, Ramírez O, Reixach J, et al. Mapping of
1058 quantitative trait loci for cholesterol, LDL, HDL, and triglyceride serum concentrations
1059 in pigs. Physiol Genomics. 2008;35:199–209.

1060 53. Gallardo D, Quintanilla R, Varona L, Díaz I, Ramírez O, Pena RN, et al.
1061 Polymorphism of the pig *acetyl-coenzyme A carboxylase α* gene is associated with fatty
1062 acid composition in a Duroc commercial line. Anim Genet. 2009;40:410–7.

1063 54. Cánovas A, Quintanilla R, Amills M, Pena RN. Muscle transcriptomic profiles in
1064 pigs with divergent phenotypes for fatness traits. BMC Genomics. 2010;11:372.

1065 55. Quintanilla R, Pena RN, Gallardo D, Cánovas A, Ramírez O, Díaz I, et al. Porcine
1066 intramuscular fat content and composition are regulated by quantitative trait loci with
1067 muscle-specific effects. J Anim Sci. 2011;89:2963–71.

1068 56. González-Prendes R, Mármol-Sánchez E, Quintanilla R, Castelló A, Zidi A,
1069 Ramayo-Caldas Y, et al. About the existence of common determinants of gene
1070 expression in the porcine liver and skeletal muscle. BMC Genomics. 2019;20:518.

1071 57. Ramayo-Caldas Y, Mármol-Sánchez E, Ballester M, Sánchez JP, González-Prendes
1072 R, Amills M, et al. Integrating genome-wide co-association and gene expression to
1073 identify putative regulators and predictors of feed efficiency in pigs. Genet Sel Evol.
1074 2019;51:48.

1075 58. Gautier L, Cope L, Bolstad BM, Irizarry RA. affy--analysis of Affymetrix
1076 GeneChip data at the probe level. Bioinformatics. 2004;20:307–15.

1077 59. Smedley D, Haider S, Durinck S, Pandini L, Provero P, Allen J, et al. The BioMart
1078 community portal: an innovative alternative to large, centralized data repositories.
1079 Nucleic Acids Res. 2015;43:W589–98.

1080 60. Cayuela JM, Garrido MD, Bañón SJ, Ros JM. Simultaneous HPLC analysis of α -
1081 tocopherol and cholesterol in fresh pig meat. J Agric Food Chem. 2003;51:1120–4.

1082 61. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al.
1083 PLINK: a tool set for whole-genome association and population-based linkage analyses.
1084 Am J Hum Genet. 2007;81:559–75.

1085 62. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association
1086 studies. *Nat Genet.* 2012;44:821–4.

1087 63. Karagkouni D, Paraskevopoulou MD, Chatzopoulos S, Vlachos IS, Tastsoglou S,
1088 Kanellos I, et al. DIANA-TarBase v8: A decade-long collection of experimentally
1089 supported miRNA-gene interactions. *Nucleic Acids Res.* 2018;46:D239–45.

1090 64. Leclercq M, Diallo AB, Blanchette M. Prediction of human miRNA target genes
1091 using computationally reconstructed ancestral mammalian sequences. *Nucleic Acids*
1092 *Res.* 2017;45:556–66.

1093 65. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and
1094 Powerful Approach to Multiple Testing. *J R Stat Soc Ser B.* 1995;57:289–300.

1095 66. Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, et al.
1096 ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and
1097 pathway annotation networks. *Bioinformatics.* 2009;25:1091–3.

1098 67. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape:
1099 A software environment for integrated models of biomolecular interaction networks.
1100 *Genome Res.* 2003;13:2498–504.

1101 68. Ballester M, Ramayo-Caldas Y, Revilla M, Corominas J, Castelló A, Estellé J, et al.
1102 Integration of liver gene co-expression networks and eGWAs analyses highlighted
1103 candidate regulators implicated in lipid metabolism in pigs. *Sci Rep.* 2017;7:46539.

1104 69. Rio DC, Ares M, Hannon GJ, Nilsen TW. Purification of RNA by SDS
1105 solubilization and phenol extraction. *Cold Spring Harb Protoc.*
1106 2010;2010:prot5438.

1107 70. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time
1108 quantitative PCR and the $2^{-\Delta\Delta CT}$ method. *Methods*. 2001;25:402–8.

1109 71. Welch BL. The generalization of ‘student’s’ problem when several different
1110 population variances are involved. *Biometrika*. 1947;34:28-35.

1111 72. Timoneda O, Balcells I, Córdoba S, Castelló A, Sánchez A. Determination of
1112 Reference microRNAs for relative quantification in porcine tissues. *PLoS One*.
1113 2012;7:e44413.

1114 73. Paudel Y, Madsen O, Megens H-J, Frantz LAF, Bosse M, Crooijmans RPMA, et al.
1115 Copy number variation in the speciation of pigs: a possible prominent role for olfactory
1116 receptors. *BMC Genomics*. 2015;16:330.

1117 74. Dueck A, Ziegler C, Eichner A, Berezikov E, Meister G. MicroRNAs associated
1118 with the different human Argonaute proteins. *Nucleic Acids Res*. 2012;40:9850–62.

1119 75. Groenen MAM, Archibald AL, Uenishi H, Tuggle CK, Takeuchi Y, Rothschild MF,
1120 et al. Analyses of pig genomes provide insight into porcine demography and evolution.
1121 *Nature*. 2012;491:393–8.

1122 76. Frantz LAF, Schraiber JG, Madsen O, Megens HJ, Bosse M, Paudel Y, et al.
1123 Genome sequencing reveals fine scale diversification and reticulation history during
1124 speciation in *Sus*. *Genome Biol*. 2013;14:R107.

1125 77. Steri M, Idda ML, Whalen MB, Orrù V. Genetic variants in mRNA untranslated
1126 regions. *Wiley Interdiscip Rev RNA*. 2020;9:e1474.

1127 78. Bachtiar M, Jin Y, Wang J, Tan TW, Chong SS, Ban KHK, et al. Architecture of
1128 population-differentiated polymorphisms in the human genome. *PLoS One*.

1129 2019;14:e0224089.

1130 79. Omariba G, Xu F, Wang M, Li K, Zhou Y, Xiao J. Genome-Wide analysis of
1131 microRNA-related single nucleotide polymorphisms (SNPs) in mouse genome. Sci Rep.
1132 2020;10:5789.

1133 80. Friedman RC, Farh KKH, Burge CB, Bartel DP. Most mammalian mRNAs are
1134 conserved targets of microRNAs. Genome Res. 2009;19:92–105.

1135 81. Gong J, Tong Y, Zhang HM, Wang K, Hu T, Shan G, et al. Genome-wide
1136 identification of SNPs in microRNA genes and the SNP effects on microRNA target
1137 binding and biogenesis. Hum Mutat. 2012;33:254–63.

1138 82. He S, Ou H, Zhao C, Zhang J. Clustering pattern and functional effect of SNPs in
1139 human miRNA seed regions. Int J Genomics. 2018;2018:2456076.

1140 83. Lewis BP, Burge CB, Bartel DP. Conserved seed pairing, often flanked by
1141 adenosines, indicates that thousands of human genes are microRNA targets. Cell.
1142 2005;120:15–20.

1143 84. Bartel DP. MicroRNAs: target recognition and regulatory functions. Cell.
1144 2009;136:215–33.

1145 85. Salomon WE, Jolly SM, Moore MJ, Zamore PD, Serebrov V. Single-molecule
1146 imaging reveals that Argonaute reshapes the binding properties of its nucleic acid
1147 guides. Cell. 2015;162:84–95.

1148 86. Ruda VM, Chandwani R, Sehgal A, Bogorad RL, Akinc A, Charisse K, et al. The
1149 roles of individual mammalian Argonautes in RNA interference in vivo. PLoS One.
1150 2014;9:e101749.

- 1151 87. Huang V, Zheng J, Qi Z, Wang J, Place RF, Yu J, et al. Ago1 interacts with RNA
1152 polymerase II and binds to the promoters of actively transcribed genes in human cancer
1153 cells. PLoS Genet. 2013;9:e1003821.
- 1154 88. Ludwig N, Leidinger P, Becker K, Backes C, Fehlmann T, Pallasch C, et al.
1155 Distribution of miRNA expression across human tissues. Nucleic Acids Res.
1156 2016;44:3865–77.
- 1157 89. Brady MJ, Saltiel AR. The role of protein phosphatase-1 in insulin action. Recent
1158 Prog. Horm. Res. 2001;56:157–73.
- 1159 90. Schmutz I, Wendt S, Schnell A, Kramer A, Mansuy IM, Albrecht U. Protein
1160 Phosphatase 1 (PP1) is a post-translational regulator of the mammalian circadian clock.
1161 PLoS One. 2011;6:e21325.
- 1162 91. Liu X, Song B, Li S, Wang N, Yang H. Identification and functional analysis of the
1163 risk microRNAs associated with cerebral low-grade glioma prognosis. Mol Med Rep.
1164 2017;16:1173–9.
- 1165 92. Wang Y, Viscarra J, Kim SJ, Sul HS. Transcriptional regulation of hepatic
1166 lipogenesis. Nat. Rev. Mol. Cell Biol. 2015;16:678–89.
- 1167 93. Toole BJ, Cohen PTW. The skeletal muscle-specific glycogen-targeted protein
1168 phosphatase 1 plays a major role in the regulation of glycogen metabolism by
1169 adrenaline in vivo. Cell Signal. 2007;19:1044–55.
- 1170 94. Ram DR, Ilyukha V, Volkova T, Buzdin A, Tai A, Smirnova I, et al. Balance
1171 between short and long isoforms of cFLIP regulates Fas-mediated apoptosis in vivo.
1172 Proc Natl Acad Sci U S A. 2016;113:1606–11.

1173 95. Mattiotti A, Prakash S, Barnett P, van den Hoff MJB. Follistatin-like 1 in
1174 development and human diseases. *Cell. Mol. Life Sci.* 2018;75:2339–54.

1175 96. Hargous Y, Hautbergue GM, Tintaru AM, Skrisovska L, Golovanov AP, Stevenin J,
1176 et al. Molecular basis of RNA recognition and TAP binding by the SR proteins SRp20
1177 and 9G8. *EMBO J.* 2006;25:5126–37.

1178 97. Sebastiani G, Grieco FA, Spagnuolo I, Galleri L, Cataldo D, Dotta F. Increased
1179 expression of microRNA miR-326 in type 1 diabetic patients with ongoing islet
1180 autoimmunity. *Diabetes Metab Res Rev.* 2011;27:862–6.

1181 98. Kefas B, Comeau L, Erdle N, Montgomery E, Amos S, Purow B. Pyruvate kinase
1182 M2 is a target of the tumor-suppressive microRNA-326 and regulates the survival of
1183 glioma cells. *Neuro Oncol.* 2010;12:1102–12.

1184 99. Cardoso TF, Coutinho LL, Bruscadin JJ, Diniz WJ da S, Petrini J, Andrade BGN, et
1185 al. Multi-omics approach reveals mir-SNPs affecting muscle fatty acids profile in
1186 Nelore cattle. *Genes.* 2021;12:67.

1187

1188

1189

1190 **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-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-486	17	10758818	10758899	-	rs335924546	precursor stem	T	0.3391
ssc-miR-335	18	18341568	18341659	-	rs334590580	precursor stem	C	0.1875

1191

1192 ^aSSC: porcine chromosome; ^bAlt. Allele: alternative allele of the genotyped SNP; ^cFrequency: alternative allele frequency in the genotyped
1193 population of 345 Duroc pigs.

1194

1195

Table 2: Number of significant associations between the 12 genotyped miRNA SNPs and the whole set of mRNA expressed in the *gluteus medius* (GM) skeletal muscle (N = 89) and liver (N = 87) tissue samples from Duroc pigs.

SNP	Type	Significant Associations			
		GM (<i>P</i> -value < 0.05)	LIVER (<i>P</i> -value < 0.05)	GM (<i>q</i> -value < 0.1)	LIVER (<i>q</i> -value < 0.1)
rs81349391 (1:164026014)	ssc-miR-339 apical loop (A/G)	456	356	0	0
rs344472188 (2:13296736)	ssc-miR-130a apical loop (A/G)	320	1252	0	1
rs333787816 (2:65308181)	ssc-miR-23a precursor stem (T/C)	2075	751	6	1
rs340704946 (4:6948743)	ssc-miR-30d precursor stem (A/G)	563	398	0	0
rs320008166 (6:56427227)	ssc-miR-371 precursor stem (T/C)	270	365	2	8
rs323906663 (6:63491948)	ssc-miR-429 precursor stem (G/A)	328	406	2	8
rs319154814 (9:9581989)	ssc-miR-326 apical loop (G/A)	372	2400	0	0
rs321151601 (9:39280312)	ssc-miR-34c mature region (C/A)	947	392	0	0
rs341950320 (12:36947510)	ssc-miR-378-2 precursor stem (G/A)	792	322	0	0
rs334680106 (13:100083229)	ssc-miR-15b precursor stem (C/T)	277	205	0	0
rs335924546 (17:10758828)	ssc-miR-486 precursor stem (C/T)	1232	171	0	0
rs334590580 (18:18341582)	ssc-miR-335 precursor stem (T/C)	665	219	1	0

Table 3: Top significant associations after correction for multiple testing (q -value < 0.1) between miRNA SNPs and the mRNA expression of their potential 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	<i>NUP50</i>	-0.1052	0.0288	3.849E-04	6.157E-02
			Ssc.12493.1.A1_at	ENSSSCG000000024027	<i>PAFAH1B2</i>	-0.2021	0.0565	3.921E-04	6.157E-02
			Ssc.13786.1.A1_at	ENSSSCG000000003857	<i>ZYG11B</i>	0.2614	0.0730	4.341E-04	6.157E-02
			Ssc.8682.2.A1_at	ENSSSCG000000014240	<i>CSNK1G3</i>	-0.1532	0.0434	4.515E-04	6.157E-02
			Ssc.23976.1.S1_at	ENSSSCG000000012252	<i>DDX3X</i>	-0.4318	0.1217	5.095E-04	6.157E-02
			Ssc.4948.1.S1_at	ENSSSCG000000034725	<i>UBE2R2</i>	0.1726	0.0493	5.906E-04	6.157E-02
			Ssc.21303.1.S1_at	ENSSSCG000000003630	<i>AGO1</i>	0.1016	0.0275	6.775E-04	6.157E-02
			Ssc.24035.2.A1_at	ENSSSCG000000005935	<i>AGO2</i>	0.2896	0.0864	8.339E-04	6.157E-02
			Ssc.3706.1.S2_at	ENSSSCG000000040317	<i>SOD2</i>	-0.2987	0.0896	8.880E-04	6.157E-02
rs319154814 (9:9581989)	ssc-miR-326 apical loop (G/A)	LIVER	Ssc.3706.1.S1_at	ENSSSCG000000040317	<i>SOD2</i>	-0.4224	0.1261	8.884E-04	6.157E-02
			Ssc.9544.2.S1_a_at	ENSSSCG000000016101	<i>CFLAR</i>	0.3246	0.0997	1.264E-03	5.785E-02
			Ssc.11661.2.S1_at	ENSSSCG000000009828	<i>PPP1CC</i>	0.4508	0.1436	1.784E-03	5.785E-02
			Ssc.31172.3.S1_at	ENSSSCG000000008601	<i>SDC1</i>	0.2507	0.0808	2.486E-03	5.785E-02
			Ssc.11044.1.A1_at	ENSSSCG000000003643	<i>SF3A3</i>	0.1808	0.0561	2.598E-03	5.785E-02
			Ssc.23242.1.A1_at	ENSSSCG0000000039426	<i>FSTL1</i>	0.2407	0.078	2.606E-03	5.785E-02
			Ssc.12222.1.S1_at	ENSSSCG000000013233	<i>CELF1</i>	-0.1714	0.0574	3.210E-03	5.939E-02
			Ssc.10946.2.A1_at	ENSSSCG000000011917	<i>NAA50</i>	0.2853	0.0993	4.070E-03	6.454E-02
			Ssc.4516.2.S1_at	ENSSSCG000000013592	<i>ELAVL1</i>	0.1666	0.0598	6.226E-03	8.639E-02

1204

1205 ^a δ : estimated allele substitution effect; ^bse: standard error of the substitution effect; ^cq-value: q-value calculated with the false discovery rate
1206 (FDR) approach [63].

1207 **Table 4:** Significant associations at the nominal level (P -value < 0.01) and after multiple testing correction (q -value < 0.1; in bold) between 12
1208 genotyped miRNA SNPs and lipid 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
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 (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

1209		GM SFA	-0.5229	0.2278	2.153E-02	2.045E-01
1210	^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					
1211	(FDR) approach [63]. GM: <i>gluteus medius</i> skeletal muscle; LD: <i>longissimus dorsi</i> skeletal muscle; BFTLR: backfat thickness at the last rib;					
1212	IMF: intramuscular fat content; SFA: saturated fatty acids content; UFA: unsaturated fatty acids content; PUFA: polyunsaturated fatty acids					
1213	content; MUFA: monounsaturated fatty acids content; C14:0: myristic content; C16:0: palmitic content; C18:0: stearic content; C18:1: oleic					
1214	content; C18:2: linoleic content; C20:0 arachidic content; C20:1: gadoleic content; C20:4: arachidonic content.					
1215						

1216 **Figure legends**

1217 **Fig. 1:** Schematic representation of (A) precursor miRNA sequence hairpin
1218 conformation, with mature miRNAs located in the stems (highlighted in blue) and their
1219 corresponding seed regions (in red). (B) Canonical miRNA-mRNA interactions where
1220 each binding site is conformed by 6-7 Watson-Crick pairings (vertical lines) to the seed
1221 region of the mature miRNA (2nd to 8th 5' nucleotides). As reported elsewhere [2,47,48],
1222 7mer-m8 and 8mer binding sites are considered to be the most functionally active
1223 canonical miRNA-mRNA interactions.

1224 **Fig. 2:** Principal component analysis plots based on SNPs mapping to: (A) the whole-
1225 genome (N = 19,720,314 SNPs), (B) miRNA genes (N = 285 SNPs), (C) Full 3'UTRs
1226 (N = 709,343 SNPs), (D) 3'UTR 7mer-m8 sites (N = 107,196 SNPs) and (E) 3'UTR
1227 8mer sites (N = 33,511 SNPs), respectively.

1228 **Fig. 3:** Alternative allele frequency distribution of polymorphisms located at miRNA
1229 loci in (A) Asian domestic pigs (ADM), (B) European domestic (EDM) pigs, (C) Asian
1230 wild boars (AWB) and (D) European wild boars (EWB).

1231 **Fig. 4:** (A) SNP density per 100 bp for each analyzed miRNA region considering the
1232 full set of 120 porcine whole-genome sequences (WGS) as well as each WGS from the
1233 ADM, EDM, AWB and EWB groups. (B) SNP density across mature miRNA regions.
1234 1: anchor (1st 5' end position), 2: seed (2nd to 8th position) and 3: supplemental pairing
1235 (13th to 18th position). (C) SNP density in miRNA loci and flanking regions considering
1236 a window of $\pm$ 1kb divided in 10 upstream and downstream 100 bp bins.

1237 **Fig. 5:** Hepatic mRNA expression levels of the *CFLAR*, *PPP1CC*, *SDC1*, *SF3A3*,
1238 *FSTL1*, *NAA50* and *ELAVL1* genes according to the genotype of the rs319154814 (G/A)

1239 apical loop SNP in the ssc-miR-326 gene. The number of individuals representing each
1240 genotype were: GG (N = 17), GA (N = 37) and AA (N = 32).

1241 **Fig. 6:** (A) Barplots depicting qPCR log₂ transformed relative quantities (Rq) for
1242 *CFLAR*, *PPP1CC* and *SF3A3* mRNA transcripts measured in the liver tissue. The
1243 number of individuals representing each genotype were: GG (N = 5) and AA (N = 5).
1244 (B) Barplots depicting qPCR log₂ transformed relative quantities (Rq) for ssc-miR-326
1245 transcripts measured in the liver tissue. The number of individuals representing each
1246 genotype were: GG (N = 9) and AA (N = 9).

1247

1248

1249 Additional Files legends

1250 **Additional file 1: Table S1:** List of whole-genome sequences from European domestic
1251 pigs (EDM, N = 40), Asian domestic pigs (ADM, N = 40), European wild boars (EWB,
1252 N = 20) and Asian wild boars (AWB, N = 20).

1253 **Additional file 2: Table S2:** Single nucleotide polymorphisms located in microRNA
1254 genes and their frequencies in European (E) and Asian (A) domestic pigs (DM) and
1255 wild boars (WB).

1256 **Additional file 3: Table S3:** Means and standard deviations (SD) of fatness and
1257 intramuscular fat content and composition traits recorded in the *gluteus medius* (GM)
1258 and *longissimus dorsi* (LD) muscles of 345 Duroc pigs.

1259 **Additional file 4: Fig. S1:** Sequence homology between 12 porcine miRNAs harboring
1260 SNPs that have been genotyped in the Lipgen population (N = 345) and their
1261 corresponding human miRNA orthologous sequences.

1262 **Additional file 5: Table S4:** Set of primers used for the qPCR quantification of three
1263 mRNAs putatively targeted by ssc-miR-326.

1264 **Additional file 6: Table S5:** Distribution of SNPs mapping to specific genomic
1265 annotated regions (i.e., protein coding genes, exons, introns, 3'UTRs, miRNAs,
1266 lncRNAs, snRNAs, snoRNAs and pseudogenes) and segregating in the set of whole-
1267 genome sequences from 120 Asian and European domestic pigs and wild boars.

1268 **Additional file 7: Table S6:** List of porcine genes showing 3 or more multi-allelic
1269 SNPs in the set of 120 whole-genome sequences from Asian and European domestic
1270 pigs and wild boars.

1271 **Additional file 8: Fig. S2:** Number of SNPs present in each of the annotated porcine
1272 miRNA loci that display at least 1 SNP.

1273 **Additional file 9: Fig. S3:** Venn Diagrams illustrating the sharing of SNPs located at
1274 the (A) *precursor*, (B) *mature* and (C) *seed* regions of miRNA loci amongst Asian
1275 domestic pigs (ADM), Asian wild boars (AWB), European domestic pigs (EDM) and
1276 European wild boars (EWB).

1277 **Additional file 10: Table S7:** Description of SNPs mapping to porcine mature miRNA
1278 loci (N = 409) and segregating in the set of 120 whole-genome sequences from Asian
1279 and European domestic pigs and wild boars.

1280 **Additional file 11: Table S8:** Whole-genome sequencing statistics for the five Duroc
1281 boars that sired the Lipgen population (N = 345).

1282 **Additional file 12: Fig. S4:** Boxplot distribution depicting whole-genome sequencing
1283 statistics for the five Duroc boars that sired the Lipgen population (N = 345).

1284 **Additional file 13: Table S9:** List of SNPs mapping to 370 porcine miRNA loci and
1285 segregating in the 5 Duroc boars that sired the Lipgen population (N = 345).

1286 **Additional file 14: Table S10:** Association analyses between SNPs in miRNA genes
1287 and the mRNA levels of their predicted targets in the *gluteus medius* (GM) skeletal
1288 muscle and liver tissues of Lipgen pigs.

1289 **Additional file 15: Table S11:** Genomic coordinates and sequence of the seed
1290 (miRNA) and its predicted 3'UTR binding site (mRNA) for putative miRNA-mRNA
1291 pairs showing significant associations after multiple testing correction (q -value < 0.1).

1292 **Additional file 16: Table S12:** Association analyses between miRNA SNPs and the
1293 mRNA expression profiles of the whole set of expressed probes/genes in the porcine
1294 *gluteus medius* (GM) skeletal muscle and liver tissues from Lipgen pigs.

1295 **Additional file 17: Table S13:** Pathway enrichment analyses of the lists of
1296 probes/genes significantly associated at the nominal level (P -value < 0.05) with miRNA
1297 SNP genotypes.

1298 **Additional file 18: Table S14:** Association analyses between miRNA SNPs and fatness
1299 and intramuscular fat content and composition traits recorded in the *gluteus medius*
1300 (GM) and *longissimus dorsi* (LD) skeletal muscles of 345 Duroc pigs.

1301

1302

1303

Fig. 2:

1328 **Fig. 3:**

1329

1330

1331

1332

1333

1334

1335

1336

1337

1338

1339

1340 **Fig. 4:**

1341

1342

1343 **Fig. 5:**

1354 **Fig. 6:**

1355

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

Emilio Mármol-Sanchez¹, María Gracia Luigi-Sierra¹, Anna Castelló^{1,4}, Dailu Guan¹, Raquel Quintanilla², Raul Tonda³, Marcel Amills^{1,4}

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

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

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

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

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

Additional Figures

Additional file 4: Fig. S1: Sequence homology between 12 porcine miRNAs harboring SNPs that have been genotyped in the Lipgen population (N = 345) and their corresponding human miRNA orthologous sequences.

hsa-miR-339-5p	UCCUGUCCUCCAGGAGCUCACG	hsa-miR-326	CCUCUGGCCCCUCCUCCAG
ssc-miR-339	UCCUGUCCUCCAGGAGCUCA	ssc-miR-326	CCUCUGGCCCCUCCUCCAG
hsa-miR-130a-3p	CAGUGCAAUGUUAAGGGCAU	hsa-miR-34c-5p	AGGCAGUGUAGUAGCUGAUUGC
ssc-miR-130a	CAGUGCAAUGUUAAGGGCAU	ssc-miR-34c	AGGCAGUGUAGUAGCUGAUUGC
hsa-miR-23a-3p	AUCACAUUGCCAGGGAUUUCC	hsa-miR-378a-3p	ACUGGACUUGGAGUCAGAAGGC
ssc-miR-23a	AUCACAUUGCCAGGGAUUUCC	ssc-miR-378-2	ACUGGACUUGGAGUCAGAAGGC
hsa-miR-30d-5p	UGUAAACAUCCCGACUGGAAG	hsa-miR-15b-5p	UAGCAGCAUCAUGGUUACA
ssc-miR-30d	UGUAAACAUCCCGACUGGAAG	ssc-miR-15b	UAGCAGCAUCAUGGUUACA
hsa-miR-371a-5p	ACUCAACUGUGGGGACACU	hsa-miR-486-5p	UCCUGUACUGAGCUGCCCCGAG
ssc-miR-371-5p	ACUCAACUGUGGGGACUUU	ssc-miR-486	UCCUGUACUGAGCUGCCCCGAG
hsa-miR-429	UAAUACUGUCUGUAAAACCGU	hsa-miR-335-5p	UCAAGAGCAAUAACGAAAAUG
ssc-miR-429	UAAUACUGUCUGUAAUGCCGU	ssc-miR-335	UCAAGAGCAAUAACGAAAAUG

Additional Figures

Additional file 8: Fig. S2: Number of SNPs present in each of the annotated porcine miRNA loci that display at least 1 SNP.

Additional Figures

Additional file 9: Fig. S3: Venn Diagrams illustrating the sharing of SNPs located at the (A) *precursor*, (B) *mature* and (C) *seed* regions of miRNA loci amongst Asian domestic pigs (ADM), Asian wild boars (AWB), European domestic pigs (EDM) and European wild boars (EWB).

A

B

C

Additional Figures

Additional file 12: Fig. S4: Boxplot distribution depicting whole-genome sequencing statistics for the five Duroc boars that sired the Lipgen population (N = 345).