UAB

Universitat Autònoma de Barcelona

Modulation of porcine production and molecular phenotypes by nutrition and genetics

Emilio Mármol Sánchez

Universitat Autònoma de Barcelona

Departament de Ciència Animal i dels Aliments

Facultat de Veterinària

CENTRE FOR RESEARCH IN AGRICULTURAL GENOMICS

Centre de Recerca en Agrigenòmica (CRAG)

Departament de Genètica Animal

Modulation of porcine production and molecular phenotypes by nutrition and genetics

Doctoral thesis to obtain the Ph.D. degree in Animal Production of the Universitat Autònoma de Barcelona (UAB), June 2020.

Emilio Mármol Sánchez

Supervisor:

Dr. Marcel Amills Eras

El Dr. Marcel Amills Eras, professor agregat del Departament de Ciència Animal i dels Aliments de la Universitat Autònoma de Barcelona,

fa constar:

que el treball de recerca i la redacció de la memòria de la tesi doctoral titulada:

“Modulation of porcine production and molecular phenotypes by nutrition and genetics”

han estat realitzats sota la seva direcció per

Emilio Mármol Sánchez

i certifica:

que aquest treball s’ha dut a terme al Departament de Ciència Animal i del Aliments de la Facultat de Veterinària de la Universitat Autònoma de Barcelona i al Departament de Genètica Animal del Centre de Recerca en Agrigenòmica (CRAG), i es considera que la memòria resultant és apta per optar al grau de Doctor en Producció Animal per la Universitat Autònoma de Barcelona.

I perquè en quedi constància, signen aquest document el juny del 2020.

Dr. Marcel Amills Eras

Emilio Mármol Sánchez

The present Ph.D. thesis was funded by grants AGL2010-22208-C02-02, AGL2013–48742-C2–1-R and AGL2013–48742-C2–2-R awarded by the Spanish Ministry of Economy and Competitivity. We acknowledge the support of the Spanish Ministry of Economy and Competitivity for the Center of Excellence Severo Ochoa 2016–2019 (SEV-2015-0533) grant awarded to the Centre for Research in Agricultural Genomics (CRAG). Emilio Mármol Sánchez was funded with a Ph.D. fellowship (FPU15/01733) awarded by the Spanish Ministry of Education and Culture (MECD).

A mi madre,

“Sometimes we can spend years without living at all,
and suddenly, our whole life is concentrated
in a single moment”

Oscar Wilde.

Content

Summary

The genetic modulators of porcine fatness and meat quality traits, as well as their mechanisms of action, are still poorly understood. In the first study of the present Ph.D. thesis, we aimed to investigate the variability of candidate genes located within QTL regions associated with meat quality traits, with a special emphasis on intramuscular fat content and composition. In this way, we used QTL mapping information to prioritize candidate genes for further analyses. Polymorphic sites located at selected candidate genes were identified based on RNA-seq data generated in previous studies and whole-genome sequencing of five Duroc boars that sired a commercial population formed by 350 Duroc pigs (Lipgen population). Statistical analyses revealed several significant nominal associations between *TGFBRAP1*, *SELENOI, ACADSB, GPR26*, *ATP1A2*, *ATP8B2* and *CREB3L4* genotypes and meat quality traits. However, only the association between *ATP1A2* genotype and electric conductivity in the *longissimus dorsi* muscle remained significant after correction for multiple testing, as well as in a chromosome-wide analysis. Our results suggest that the *ATP1A2* gene might be involved in the regulation of the electric conductivity of the skeletal muscle, but additional structural and functional studies will be needed to assess this hypothesis.

In the second study, we employed whole-genome sequencing data from the five Duroc boars mentioned before to identify putative stop gained mutations which might be segregating in the Lipgen population. By doing so, seven apparently healthy pigs homozygous for a potentially lethal nonsense recessive mutation in the *ASS1* gene (rs81212146, c.944T>A) were detected. In order to elucidate the possible underlying causes of such finding, we sequenced the region surrounding the mutation at the genomic and transcriptomic levels. Our results indicated the presence of an additional polymorphism (rs81212145, c.943T>C) located immediately before the nonsense mutation that disrupts the stop codon. Both SNPs segregated in complete linkage disequilibrium in a sample of 120 pigs with available whole-genome sequences. The rs81212146 and rs81212145 mutations form a dinucleotide polymorphism that causes a benign amino acid substitution (Leu315Gln) in the *ASS1* sequence. Such results illustrate the complexity of predicting loss-of-function effects due to compensatory mechanisms that limit the harmful consequences of deleterious mutations.

In the third study, we made use of previous RNA-seq differential expression data to investigate the association of several candidate genes with meat quality traits recorded in the Lipgen population. Polymorphisms in genes related to peripheral circadian clock regulation (*ARNTL2*, *CIART*, *CRY2*, *NPAS2*, *PER1* and *PER2*), glucose metabolism (*PCK1*) and energy homeostasis (*MIGA2*) were genotyped and association analyses were performed. Two polymorphisms located in the *CRY2* (rs320439526, c.-6C>T) and *MIGA2* (rs330779504, c.1455G>A) genes showed significant associations with stearic acid content in the *longissimus dorsi* skeletal muscle and with LDL serum concentration at ~190 days of age, respectively. Moreover, these polymorphic sites were also associated with the mRNA levels of the corresponding genes. Additional joint association analyses with chromosome-wide genotyping data showed that these polymorphisms (rs320439526 and rs330779504) are not the ones showing the most significant associations with stearic acid content and LDL serum concentration, respectively. Such results highlighted that these variants might not be the causal mutations explaining the phenotypic variation of these two phenotypes.

In the fourth study of this thesis, we extended our search for regulatory determinants of meat quality traits to polymorphisms residing in microRNA genes, which are known to play relevant roles in modulating the expression of protein-coding mRNAs. A total of 120 publicly available whole-genome sequences from European and Asian wild boars and domestic pigs were used for variant calling analyses, and polymorphisms within miRNA loci and targeted 3’-UTR binding sites were investigated. Distinctive segregation patterns were observed between pigs from Asian and European origins, while such differentiation was less evident when comparing wild boars with domestic pigs. Variability within miRNA loci was strongly reduced in the seed region compared with the rest of the miRNA sequence, and also with other regions in the genome. The most likely explanation for this result is the presence of strong purifying selection removing mutations that might alter the binding properties of the miRNA. Fifteen SNPs mapping to miRNA genes were also genotyped in the Lipgen population. Several significant associations between miRNA SNPs and mRNA levels in the *gluteus medius* skeletal muscle and liver tissues of their putative target mRNAs were observed. Of special relevance was the case of the rs319154814 (n.46G>A) polymorphism, located in the apical loop of ssc-miR-326. This SNP might contribute to a structural rearrangement of the miRNA hairpin pairing, thus modifying the efficiency of the miRNA maturation.

In the fifth study, we aimed to improve the yet poorly annotated porcine miRNAome by developing a bioinformatic pipeline for the discovery and annotation of miRNA genes from small RNA-seq data and homology-based search. This goal was achieved by selecting *bona fide* porcine miRNA genes, jointly with other non-miRNA loci closely resembling hairpin-like structures, typical of miRNA precursors. An additional set of unlabeled hairpin sequences were extracted from the porcine genome to help increase the biological information embedded in the prediction model. The small RNA fraction of 48 Duroc gilts was sequenced and used to detect novel and known expressed miRNAs. Subsequently, small RNA-seq transcripts and annotated human mature miRNAs were mapped to the porcine genome. Further reconstruction of candidate hairpin sequences was performed by applying a motif search correction approach. Moreover, a series of sequence and thermodynamic features were obtained from each sequence and a machine learning graph-based transductive algorithm was employed for predicting novel and annotated miRNA sequences. A total of 47 unreported putative porcine miRNAs were detected with this approach. Twenty of them corresponded to transcripts present in the porcine muscle small RNA-seq data set, while the remaining ones were inferred on the basis of an homology-based search using human miRNAs. The expression of three of the unreported miRNAs was assessed by using RT-qPCR analyses and their expression in an independent Göttingen minipig population was confirmed.

Finally, in the sixth study of the present thesis, we employed the muscle small RNA-seq data set generated in 36 Duroc guilts in order to determine miRNA differential expression patterns between pigs subjected to fasting conditions and after being fed during 5 and 7 hours, respectively. The expression profiles of mRNAs, miRNAs and lincRNAs were compared in terms of their abundance and intragroup variability. Protein coding transcripts were generally more expressed than miRNAs and lincRNAs, whereas miRNAs showed very stable expression patterns compared with mRNAs and lincRNAs. The reconstruction of gene regulatory networks for miRNA-mRNA interactions highlighted several co-expression modules containing genes related with lipid metabolism. Moreover, we described the potential influence of several differentially expressed miRNAs, such as ssc-miR-148a-3p, ssc-miR-151-3p, ssc-miR-30a-3p, ssc-miR-30e-3p, ssc-miR-421-5p and ssc-miR-493-5p, in regulating the expression of mRNA genes with key roles in glucose metabolism and energy homeostasis.

Resumen

Los reguladores genéticos del engrasamiento porcino y la calidad de la carne, así como sus mecanismos de acción, son aún poco conocidos. En el primer artículo de la presente tesis doctoral, nos propusimos investigar la variabilidad de genes candidatos situados en regiones QTL asociadas a diversos caracteres de calidad de la carne, con especial énfasis en el contenido y la composición de la grasa intramuscular. En este sentido, hicimos uso de información de cartografiado de QTL con el objetivo de priorizar genes candidatos para ser analizados. Identificamos regiones polimórficas en los genes candidatos seleccionados mediante datos de RNA-seq generados en estudios previos y a partir de la secuenciación del genoma de cinco verracos que dieron origen a una población comercial de 350 cerdos Duroc (población Lipgen). Los análisis estadísticos revelaron diversas asociaciones nominalmente significativas entre caracteres de calidad de la carne y los genotipos de los genes *TGFBRAP1, SELENOI, ACADSB, GPR26*, *ATP1A2*, *ATP8B2* y *CREB3L4.* Sin embargo, sólo la asociación entre el genotipo del gen *ATP1A2* y la conductividad eléctrica en el músculo *longissimus dorsi* permaneció significativa tras aplicar la corrección para tests múltiples, así como después de su análisis estadístico a nivel cromosómico. Nuestros resultados sugieren que el gen *ATP1A2* podría estar involucrado en la regulación de la conductividad eléctrica del músculo esquelético porcino. No obstante, estudios estructurales y funcionales serán necesarios para confirmar dicha hipótesis.

En el segundo artículo, hicimos uso de los datos de secuenciación del genoma de los cinco cerdos Duroc, con el objetivo de identificar posibles mutaciones con efecto *stop gain* que pudieran segregar en la población Lipgen. Dicho análisis reveló la existencia de siete cerdos aparentemente sanos y, sin embargo, homocigotos para una mutación recesiva potencialmente letal en el gen *ASS1* (rs81212146, c.944T>A). Con el objetivo de dilucidar las posibles causas de dicha observación, procedimos a secuenciar la región que contiene la mencionada mutación, tanto a nivel genómico como transcriptómico. Nuestros resultados indicaron la presencia de un polimorfismo adicional (rs81212145, c.943T>C) localizado en una posición inmediatamente anterior a la mutación *stop gain*, que propicia la eliminación del codón de parada prematura de la traducción. Se observó que ambos SNPs segregan en completo desequilibrio de ligamiento en una muestra de 120 cerdos con secuencias de genoma completo disponibles. Las mutaciones rs81212146 y rs81212145 constituyen un polimorfismo

dinucleotídico que causa una sustitución aminoacídica benigna (Leu315Gln) en la secuencia del gen *ASS1*. Dichos resultados ilustran la complejidad de predecir efectos funcionales debido a la existencia de mecanismos compensatorios que limitan las consecuencias potencialmente dañinas de las mutaciones deletéreas.

En el tercer artículo, hicimos uso de resultados previos de expresión diferencial a partir de datos de RNA-seq, con el objetivo de investigar la asociación de varios genes candidatos con caracteres de calidad de la carne medidos en la población Lipgen. Una selección de polimorfismos en genes relacionados con la regulación periférica del ciclo circadiano (*ARNTL2*, *CIART*, *CRY2*, *NPAS2*, *PER1* y *PER2*), el metabolismo de la glucosa (*PCK1*) y la homeostasis energética (*MIGA2*) fueron genotipados para realizar análisis de asociación. Dos polimorfismos localizados en los genes *CRY2* (rs320439526, c.-6C>T) y *MIGA2* *(*rs330779504, c.1455G>A) mostraron asociaciones significativas con el contenido de ácido esteárico en el músculo *longissimus dorsi* y con la concentración sérica de LDL medida a los ~190 días de edad, respectivamente. Además, se encontraron asociaciones significativas entre dichos polimorfismos y los niveles de expresión de los correspondientes mRNAs. Análisis estadísticos adicionales a nivel cromosómico relevaron que ambos polimorfismos (rs320439526 y rs330779504) no muestran las asociaciones más significativas con los caracteres de contenido de ácido esteárico y concentraciones séricas de LDL. Dichos resultados indican que estos polimorfismos probablemente no tengan efectos causales sobre la variación fenotípica de los caracteres bajo estudio.

En el cuarto artículo de la presente tesis, decidimos extender nuestra búsqueda de elementos reguladores de la calidad de la carne con la finalidad de abarcar polimorfismos en los genes que codifican microRNAs, los cuales poseen una función relevante en la modulación de la expresión de mRNAs codificantes de proteína. Un total de 120 secuencias genómicas, previamente descritas, de jabalíes y cerdos domésticos europeos y asiáticos, fueron utilizadas para identificar variantes genéticas. Mediante dicha información, se pudo investigar la presencia de polimorfismos en genes miRNA, así como en sus potenciales dianas de unión en las regiones 3’-UTRs de los mRNAs a los cuales regulan. Se observaron patrones diferenciados de segregación de variantes entre cerdos asiáticos y europeos, mientras que dichas diferencias fueron menos evidentes cuando se contrastaron cerdos domésticos y jabalíes. La variabilidad de los genes miRNA fue muy baja en la región *seed* al compararla con la de otras regiones del propio miRNA, así como respecto a otras regiones del genoma.

La explicación más plausible para este hallazgo es la presencia de una fuerte selección purificadora que eliminaría las mutaciones que pudieran alterar las secuencias a través de las cuales los miRNAs hibridan con sus mRNAs diana. Un total de quince SNPs localizados en genes miRNA fueron genotipados en la población Lipgen. Diversas asociaciones significativas fueron identificadas entre dichos SNPs y los niveles de expresión de sus mRNAs diana en el músculo *gluteus medius* y en el hígado. Especialmente relevante fue el caso de la variante rs319154814 (n.46G>A), localizada en el bucle apical del miRNA ssc-miR-326. Este SNP podría contribuir a la reestructuración del apareamiento de bases en la cadena en forma de horquilla (*hairpin*) del miRNA, modificando la eficiencia de la maduración del propio miRNA.

En el quinto artículo, nuestro objetivo fue mejorar la anotación del aún limitado miRNAoma porcino mediante el desarrollo de un procedimiento bioinformático para la identificación y la anotación de genes miRNA a partir de datos de small RNA-seq y búsqueda por homología. Para conseguir dicho objetivo, se seleccionaron los miRNAs porcinos con una anotación fiable, conjuntamente con otros loci no identificados como miRNAs, pero que aun así poseen una estructura secundaria en forma de horquilla, similar a la de los propios miRNAs. De forma adicional, un conjunto de secuencias con estructura de horquilla, pero sin anotación asignada, fueron extraídas del genoma porcino con el objeto de incrementar la información biológica incluida en el modelo predictivo. La fracción de RNAs pequeños de 48 cerdas Duroc fue secuenciada y utilizada para detectar miRNA expresados, tanto anotados como sin anotar. Seguidamente, los transcritos identificados a partir de datos de small RNA-seq, así como las secuencias de miRNA maduro anotadas en humano, fueron cartografiadas en el genoma porcino. Por otra parte, se reconstruyeron secuencias candidatas con estructura en forma de horquilla mediante una técnica de corrección posicional basada en la búsqueda de determinados motivos nucleotídicos. Además, se obtuvieron una serie de parámetros de secuencia y termodinámicos para cada secuencia candidata y se utilizó un algoritmo transductivo de machine learning basado en grafos para la predicción de miRNAs, tanto nuevos como ya conocidos. Un total de 47 miRNAs porcinos putativos y carentes de anotación fueron detectados mediante esta aproximación. Veinte de ellos se correspondieron con transcritos expresados en la fracción del RNA pequeño del músculo porcino, mientras que el resto fueron identificados mediante búsqueda por homología a partir de miRNAs anotados en humano. Mediante técnicas de RT-qPCR, se pudo confirmar, en una población independiente formada por cerdos de la raza Göttingen minipig, la expresión de tres de los nuevos miRNAs identificados con el algoritmo transductivo.

Por último, en el sexto artículo de la presente tesis, hicimos uso de los datos de secuenciación de RNAs pequeños generados en 36 cerdas Duroc, con el objetivo de identificar patrones de expresión diferencial de miRNAs en condiciones de ayuno y tras haber recibido alimento durante 5 y 7 horas. Se compararon los perfiles de expresión de mRNAs, miRNAs y lincRNAs respecto a su abundancia y variabilidad intragrupal. Los transcritos codificantes de proteínas presentaron una expresión mayor que la de los miRNAs y lincRNAs, mientras que los transcritos de miRNAs mostraron patrones de expresión muy estables comparados con mRNAs y lincRNAs. La reconstrucción de redes de regulación génica para interacciones miRNA-mRNA reveló diversos módulos de co-expresión formados por genes relacionados con el metabolismo de los lípidos. Además, se evidenció la posible influencia de diversos miRNAs diferencialmente expresados tales como ssc-miR-148a-3p, ssc-miR-151-3p, ssc-miR-30a-3p, ssc-miR-30e-3p, ssc-miR-421-5p o ssc-miR-493-5p, sobre la regulación de la expresión de genes mRNA con funciones importantes en el metabolismo de la glucosa y la homeostasis energética.

.

List of tables

CHAPTER I. INTRODUCTION

CHAPTER III. PUBLICATIONS

Paper I

Paper II

Paper III

Paper IV

Paper V

Paper VI

CHAPTER IV. DISCUSSION

List of Figures

CHAPTER I. INTRODUCTION

CHAPTER III. PUBLICATIONS

Paper I

Paper II

Paper III

Paper IV

Paper V

Paper VI

CHAPTER IV. DISCUSSION

List of publications

The present Ph.D. thesis is based on the work contained in the list of articles below:

I. **Mármol-Sánchez, E**., Quintanilla, R., Jordana, J., and Amills, M. (2019). An association analysis for 14 candidate genes mapping to meat quality quantitative trait loci in a Duroc pig population reveals that the *ATP1A2* genotype is highly associated with muscle electric conductivity. *Anim. Genet.* 51, 95-100.

II. **Mármol-Sánchez, E**., Luigi-Sierra, M. G., Quintanilla, R., and Amills, M. (2020). Detection of homozygous genotypes for a putatively lethal recessive mutation in the porcine argininosuccinate synthase 1 (*ASS1*) gene. *Anim. Genet.* 51, 106–110.

III. **Mármol-Sánchez, E**., Quintanilla, R., Cardoso, T. F., Jordana Vidal, J., and Amills, M. (2019). Polymorphisms of the cryptochrome 2 and mitoguardin 2 genes are associated with the variation of lipid-related traits in Duroc pigs. *Sci. Rep.* 9, 9025.

IV. **Mármol-Sánchez, E.**, Guan, D., Quintanilla, R., Tonda, R. and Amills, M. (2020) Variability in porcine microRNA genes and its association with mRNA expression phenotypes. *Genet. Sel. Evol*. Under review

V. **Mármol-Sánchez, E.,** Cirera, S., Quintanilla, R., Pla, A., and Amills, M. (2020). Discovery and annotation of novel microRNAs in the porcine genome by using a semi-supervised transductive learning approach. *Genomics* 112, 2107–2118.

VI. **Mármol-Sánchez, E**., Ramayo-Caldas, Y., Quintanilla, R., Cardoso, T. F., González-Prendes, R., Tibau, J., et al. (2020). Co-expression network analysis predicts a key role of microRNAs in the adaptation of the porcine skeletal muscle to nutrient supply. *J. Anim. Sci. Biotechnol.* 11, 10.

Other publications from the author

Not included in the present Ph.D. thesis:

- Luigi Luigi-Sierra, M. G., **Mármol-Sánchez, E**., and Amills, M. (2020). Comparing the diversity of the casein genes in the Asian mouflon and domestic sheep. *Anim. Genet.* 51, 470-475.

- Guan, D., Landi, V., Luigi-Sierra, M. G., Delgado, J. V., Such, X., Castelló, A., Cabrera, B., **Mármol-Sánchez, E**., Ferndández-Álvarez, J., Ruiz de la Torre, J. L., Martínez, A., Jordana, J. and Amills, M. (2020). Analyzing the genomic and transcriptomic architecture of milk traits in Murciano-Granadina goats. *J. Anim. Sci. Biotechnol.* 11, 35.

- Ramayo-Caldas, Y., **Mármol-Sánchez, E**., Ballester, M., Sánchez, J. P., González-Prendes, R., Amills, M. and Quintanilla, R. (2019). Integrating genome-wide co-association and gene expression to identify putative regulators and predictors of feed efficiency in pigs. *Genet. Sel. Evol.* 51, 48.

- González-Prendes, R., **Mármol-Sánchez, E**.*, Quintanilla, R., Castelló, A., Zidi, A., Ramayo-Caldas, Y., Cardoso, T. F., Manunza, A., Cánovas, A. and Amills, M. (2019). About the existence of common determinants of gene expression in the porcine liver and skeletal muscle. *BMC Genomics* 20, 518. *shared first co-authorship.

- Guan, D., **Mármol-Sánchez, E**., Cardoso, T. F., Such, X., Landi, V., Tawari, N. R. and Amills, M. (2019). Genomic analysis of the origins of extant casein variation in goats. *J. Dairy Sci.* 102, 5230–5241.

- González-Prendes, R., Quintanilla, R., **Mármol-Sánchez, E**., Pena, R. N., Ballester, M., Cardoso, T. F., Manunza, A., Casellas, J., Cánovas, A., Díaz, I., Noguera, J. L., Castelló, A., Mercadé, A. and Amills, M. (2019). Comparing the mRNA expression profile and the genetic determinism of intramuscular fat traits in the porcine *gluteus medius* and *longissimus dorsi* muscles. *BMC Genomics* 20, 170.

- Cardoso, T. F., Quintanilla, R., Castelló, A., **Mármol-Sánchez, E**., Ballester, M., Jordana, J. and Amills, M. (2018). Analysing the Expression of Eight Clock Genes in Five Tissues From Fasting and Fed Sows. *Front. Genet.* 9, 475.

- Cardoso, T. F., Quintanilla, R., Tibau, J., Gil, M., **Mármol-Sánchez, E**., González-Rodríguez, O. and Amills, M. (2017). Nutrient supply affects the mRNA expression profile of the porcine skeletal muscle. *BMC Genomics* 18, 603.

List of contributions to Congresses & Conferences

- **Mármol-Sánchez, E.**, Cirera, S., Quintanilla, R., Pla, A. and Amills, M. eMIRNA: A comprehensive pipeline for discovery and annotation of microRNAs in multiple species. 37th International Society for Animal Genetics Conference (ISAG) 2019, Lleida, Spain. Oral communication.

- **Mármol-Sánchez, E.**, Quintanilla, R., Cardoso, T. F., Tibau, J. and Amills, M. The variance of gene expression in the porcine skeletal muscle changes in response to food intake. 37th International Society for Animal Genetics Conference (ISAG) 2019, Lleida, Spain. Poster.

- Ramayo-Caldas, Y., **Mármol-Sánchez, E.**, Ballester, M., González-Prendes, R., Amills, M. and Quintanilla, R. Analysis of porcine muscle transcriptome reveals regulators and pathways associated with feed efficiency. 37th International Society for Animal Genetics Conference (ISAG) 2019, Lleida, Spain. Poster.

- Guan, D., Landi, V., Luigi-Sierra, M. G., Delgado, J. V., Castelló, A., Cabrera, B., **Mármol-Sánchez, E.**, Fernández-Álvarez, J., Martínez, A., Such, X., Jordana, J. and Amills, M. A genome-wide association analysis for dairy traits in Murciano-Granadina goats. 37th International Society for Animal Genetics Conference (ISAG) 2019, Lleida, Spain. Poster.

- Luigi-Sierra, M. G., Landi, V., Guan, D., Delgado, J. V., Castelló, A., Cabrera, B., **Mármol-Sánchez, E.**, Fernández-Álvarez, J., Martínez, A., Such, X., Jordana, J. and Amills, M. Identification of genomic regions associated with morphological traits un Murciano-Granadina goats. 37th International Society for Animal Genetics Conference (ISAG) 2019, Lleida, Spain. Poster.

- **Mármol-Sánchez, E.**, Quintanilla, R., Cardoso, T. F., Tibau, J. and Amills, M. Identificación de RNAs largos no-codificantes en respuesta a la ingesta de alimentos. XVIII Jornadas sobre Producción Animal AIDA-ITEA, 2019, Zaragoza. Oral Communication.

- Guan, D., **Mármol-Sánchez, E.**, Luigi-Sierra, M. G., Such, X., Jordana, J., Landi, V., Martínez, A., Delgado, J. V. and Amills, M. Transcriptional profile of mammary gland tissue during lactation in Murciano-Granadina goats. XVIII Jornadas sobre Producción Animal AIDA-ITEA, 2019, Zaragoza. Oral Communication.

- Luigi-Sierra, M. G., **Mármol-Sánchez, E.**, Cardoso, T. F. and Amills, M. Caracterización de la variabilidad de los genes de las caseínas en ovinos domésticos y salvajes. XVIII Jornadas sobre Producción Animal AIDA-ITEA, 2019, Zaragoza. Oral Communication.

- **Mármol-Sánchez, E.**, Quintanilla, R., Tibau, J., Cardoso, T. F. and Amills, M. Análisis de asociación de genes candidatos para fenotipos lipídicos en una población comercial duroc. XIX Reunión Nacional de Mejora Genética Animal, 2018, León. Oral Communication.

- Guan, D., **Mármol-Sánchez, E.**, Such, X., Landi, V. and Amills, M. Una perspectiva genómica sobre el origen de la variación genética de las caseínas caprinas. XIX Reunión Nacional de Mejora Genética Animal, 2018, León. Oral Communication.

- Cardoso, T. F., Quintanilla, R., Castelló, A., **Mármol-Sánchez, E.**, Ballester, M., Jordana, J. and Amills, M. Efecto de la ingestión de alimento sobre la expresión de genes circadianos en cinco tejidos porcinos. XIX Reunión Nacional de Mejora Genética Animal, 2018, León. Oral Communication.

- Ballester, M., Amills, M., González-Rodríguez, O., Cardoso, T. F., Pascual, M., **Mármol-Sánchez, E.,** Gil, M., Tibau, J. and Quintanilla, R. Effect of feed restriction on pig skeletal muscle transcriptome. 68th Annual Meeting of the European Federation of Animal Science (EAAP) 2017. Poster.

- **Mármol-Sánchez, E.,** Quintanilla, R., Cardoso, T. F., Tibau, J., González-Rodríguez, O., González-Prendes, R., Ballester, M. and Amills, M. Food intake promotes changes in microRNA muscle expression profile in pigs. 36th International Society for Animal Genetics Conference (ISAG) 2017, Dublin, Ireland. Poster.

- Cardoso, T. F., Quintanilla, R., Tibau, J., Gil, M., **Mármol-Sánchez, E.,** González-Rodríguez, O., González-Prendes, R. and Amills, M. Nutrient supply drives changes in the muscular expression of protein-encoding and non-coding RNA genes. 36Th International Society for Animal Genetics Conference (ISAG) 2017, Dublin, Ireland. Poster.

- **Mármol-Sánchez, E.,** Quintanilla, R., Cardoso, T. F., Tibau, J., González-Rodríguez, O., Ballester, M. and Amills, M. Efecto de la ingestión de alimento sobre la expresión muscular de miRNAs en porcino. XVII Jornadas sobre Producción Animal AIDA-ITEA, 2017, Zaragoza. Oral Communication.

- Cardoso, T. F., Quintanilla, R., Tibau, J., **Mármol-Sánchez, E.,** González-Rodríguez, O., González-Prendes, R., Ballester, M. and Amills, M. Expresión muscular de RNAs largos no-codificantes en respuesta a la ingestión de alimentos. XVII Jornadas sobre Producción Animal AIDA-ITEA, 2017, Zaragoza. Oral Communication.

- Ballester, M., González-Rodríguez, O., Amills, M., Cardoso, T. F., Pascual, M., **Mármol-Sánchez, E.,** Tibau, J. and Quintanilla, R. Efecto de la restricción alimentaria sobre el transcriptoma del músculo esquelético en porcino. XVII Jornadas sobre Producción Animal, 2017 AIDA-ITEA, Zaragoza. Oral Communication.

International Research Short Stays

- Two months Short Stay at the Department of Animal Genetics and Integrative Biology (GABI) at the National Institute for Agronomic Research (INRA), Jouy-en-Josas, France. 01/07/2019 – 31/08/2019. **Supervisor:** Dr. Dominique Rocha.

- Three months Short Stay at the Department of Veterinary Clinical and Animal Science, Animal Genetics, Bioinformatics and Breeding. Faculty of Health and Medical Sciences, University of Copenhagen, Denmark. 01/09/2018 – 30/11/2018. **Supervisor:** Dr. Susanna Cirera.

Both Research Short Stays were fully funded by awarded FPU International Short Stays open calls 2019 and 2018, respectively.

Abbreviations

A	Adenine
a*	Meat redness
ABCA1	ATP-binding cassette 1
ACACA	Acetyl-CoA carboxylase 1
ACADSB	Acyl-CoA dehydrogenase short/branched chain
Acc	Accuracy
ACLY	ATP citrate lyase
ADE	Allelic differential expression
ADM	Asian domestic
Agm	Adjusted geometric mean
AGO	Argonaute
AGO1	Argonaute I
AGO2	Argonaute II
AMPK	Adenosine monophosphate-activated protein kinase
ANOVA	Analysis of variance
ANXA5	Placental anticoagulant annexin 5
ARACNE	Algorithm for the reconstruction of accurate cellular networks
AREs	AU-rich sequence motifs
ARID5B	ARID domain-containing protein 5B
ARNTL	Aryl hydrocarbon receptor nuclear translocator like
ARNTL2	Aryl hydrocarbon receptor nuclear translocator like 2
ARRDC3	Arrestin domain-containing protein 3
ASS1	Argininosuccinate synthase 1
ATF3	Activating transcription factor 3
ATP1A2	ATPase Na^+/K^+ transporting α_2 subunit
ATP8B2	ATPase phospholipid transporting 8B2
AUC	Area under the curve
AUPR	Area under the precision-recall curve
AUROC	Area under the receiver operating characteristics curve

AWB	Asian wild boars
b*	Meat yellowness
BACH1	BTB domain and CNC homolog 1
BACH2	BTB domain and CNC homolog 2
BAM	Binary SAM
BBS9	Bardet Biedl syndrome 9
BCV	Biological coefficient of variation
BFTLR	Backfat thickness at last rib
BQSR	Base quality score recalibration
BP	Back propagation
C14:0	Myristic fatty acid
C16:0	Palmitic fatty acid
C16:1	Palmitoleic fatty acid
C18:0	Stearic fatty acid
C18:1	Oleic fatty acid
C18:1 n-7	Vaccenic fatty acid
C18:2	Linoleic fatty acid
C18:3	α-linolenic fatty acid
C20:0	Arachidic fatty acid
C20:1	Gadoleic fatty acid
C20:4	Arachidonic fatty acid
C3NET	Conservative causal core network
CASQ1	Calsequestrin 1
CBN	Categorical Bayesian network
cDNA	Complementary DNA
CE	Electric conductivity
CEBPA	CCAAT/enhancer-binding protein α
CEBPB	CCAAT/enhancer-binding protein β
CFLAR	Cellular FLICE-like inhibitory protein
ChoRES	Carbohydrates response elements
CHRND	Cholinergic receptor nicotinic δ subunit

CIART	Circadian associated repressor of transcription
CLOCK	Clock circadian regulator
COPA	Coatomer protein complex subunit *α*
CPM	Counts-per-million
CREB1	cAMP responsive element binding protein 1
CREB34L	cAMP-responsive element binding protein 3 like 4
CRTC2	CREB-regulated transcription coactivator 2
CRY2	Cryptochrome circadian regulator 2
CSNK1D	Casein kinases δ
CSNK1E	casein kinases ε
CTBPL2	C-terminal binding protein 2
CV	Coefficient of variation
DD	Differential dispersion
DE	Differential expression
DL	Deep learning
DEPP1	Fasting-induced gene protein
DPI	Data processing inequality
dr	Deviation rate
DROSHA	Drosha ribonuclease III
DSTN	Destrin/actin depolymerizing factor
DUOX2	Dual oxidase 2
DW	Differential wiring
E	Edge
EDM	European domestic
EGR1	Early growth factor 1
eQTL	Expression quantitative trait loci
ETS1	ETS proto-oncogene 1
EWB	European wild boar
F1	F-1 score
FA	Fatty acids
FABP3	Fatty acid binding protein 3

FABP4	Fatty acid binding protein 4
FABP7	Fatty acid-binding protein 7
FADS2	Fatty acid desaturase 2
FASN	Fatty acid synthase
FC	Fold change
FDR	False discovery rate
FGF2	Fibroblast growth factor 2
FOSL2	FOS-related antigen 2
FOXO1	Forkhead box O1
FSP-1	Fibroblast-specific protein-1
FSTL1	Follistatin-like 1
GBLUP	Genomic best linear unbiased prediction
GEMMA	Genome-wide efficient mixed model analysis
GEV	Gene expression variance
GEVB	Genomic breeding values
GGM	Graphical Gaussian model
GLUT1	Glucose transporter 1
GLUT4	Glucose transporter 4
GM	*Gluteus medius*
GO	Gene ontology
GPAT2	Glycerol-3 phosphate acyltransferase 2
GPR26	G protein-coupled receptor 26
GRB2	Growth factor receptor-bound 2
GRN	Gene regulatory network
GS	Gene significance
GWAS	Genome-wide association analysis
HADHA	Hydroxyacil-CoA dehydrogenase trifunctional multienzyme
HDL	High density lipoprotein
HMM	Hidden Markov model
IGF1R	Insulin like growth factor 1 receptor
IGF2	Insulin growth factor 2

IGSF8	Immunoglobulin superfamily member 8 gene
IMF	Intramuscular fat
IMPA1	Inositol monophosphatase 1
INDELs	Insertions and deletions
INS	Insulin
INSIG1	Insulin-induced 1
INSR	Insulin receptor
IRS1	Insulin receptor substrate 1
isomiRs	MiRNA isoforms
K	Hub score
KLF15	Kruppel-like factor 15
KNN	K-nearest neighbors
L*	Meat lightness
LD	*Longissimus dorsi*
LD (bis)	Linkage disequilibrium
LDL	Low density lipoprotein
LDLR	Low density lipoprotein receptor
LEP	Leptin
LEPR	Leptin receptor
lGBM	Light gradient boosting trees
lincRNAs	Long intergenic non-coding RNAs
lncRNAs	Long non-coding RNAs
LoF	Loss-of-function
MAD2L1	Mitotic arrest deficient 2-like 1
MAF	Minor allele frequency
MAS	Marker-assisted selection
Mb	Megabase
MC4R	Melanocortin 4 receptor
MES	Module eigengenes
MFE	Minimum free energy
MFEadj	Adjusted minimum free energy

MI	Mutual information
MIGA2	Mitoguardin 2
miRISC	MiRNA-induced silencing complex
miRNA*	MiRNA passenger strand
miRNAs	MicroRNAs
MKRN1	E3 ubiquitin ligase makorin ring finger protein 1
ML	Machine learning
MLXIPL	Carbohydrate-responsive element-binding protein
mRNAs	Messenger RNAs
mtRNAs	Mitochondrial RNAs
MUFA	Monounsaturated fatty acids
Mya	Million years ago
MYBPHL	Myosin binding protein H like
MYF6	Myogenic factor 6
MYOD1	Myoblast determination protein 1
MYOG	Myogenin
Myr	Million years
N	Neighborhood score
NADK	NAD kinase
NB	Naïve Bayes
NCOA3	Nuclear receptor coactivator 3
NEU3	Neuraminidase 3
NN	Neural network
NPAS2	Neuronal PAS domain protein 2
NR1D1	Rev-Erb-α
NR1D2	Rev-Erb-β
nt	Nucleotide
NUDT6	Nudix hydrolase 6
ORF	Open reading frame
OXSR1	Oxidative stress response kinase 1
p	Structural integrity score

PAR-CLIP	Photoactivatable-ribonucleoside-enhanced cross-linking immunoprecipitation
PCA	Principal components analysis
PCIT	Partial correlation coefficient with information theory
PCK	Protein kinase C
PCK1	Cytosolic phosphoenolpyruvate carboxykinase
PCR	Polymerase chain reaction
PHD	Pyruvate dehydrogenase
PDK4	Pyruvate dehydrogenase kinase 4
PEA15	Proliferation and apoptosis adaptor protein 15
PER1	Period circadian regulator 1
PER2	Period circadian regulator 2
PHKG1	Phosphorylase kinase catalytic subunit γ_1
PIC	Pig improvement company
PIF	Phenotype impact factor
PIK3CD	Phosphoinositide 3-kinase catalytic subunit δ
piRNA	Piwi-interacting RNA
PLA2G7	Phospholipase A2 Group VII
Pol-II	RNA polymerase II
POLR1B	RNA polymerase I subunit B
PPARD	Peroxisome proliferator activated receptor δ
PPARG	Peroxisome proliferator activated receptor γ
PPARGC1A	Peroxisome proliferator activated receptor γ coactivator 1 α
PPP1CC	Protein phosphatase 1 catalytic subunit γ
PR	Precision-recall
pre-miRNAs	Precursor miRNAs
pri-miRNAs	Primary miRNA transcripts
PRKAG3	Protein kinase AMP-activated non-catalytic subunit γ 3
PUFA	Polyunsaturated fatty acids
QTL	Quantitative trait loci
***q*-value**	Multiple testing corrected *P*-value
r	Pearson's correlation coefficient

RBBP8	RB binding protein 8 endonuclease
RBLP1	Retinaldehyde binding protein 1
RBP	RNA binding protein
RF	Random forests
RIF	Regulatory impact factor
RIN	RNA integrity number
RISC	RNA-induced silencing complex
Rlog	Regularized log$_2$
RMA	Robust multi-array average
RNA-seq	RNA sequencing
ROC	Receiver operating characteristics
RPKM	Reads per kilobase transcript per million mapped reads
RPM	Reads per million
RRM	RNA recognition motif
rRNAs	Ribosomal RNAs
RT	Reverse transcription
RT-qPCR	Quantitative real-time PCR
RXRG	Retinoid X receptor γ
RYR1	Ryanodine receptor 1
SAM	Sequence alignment map
SCAMP2	Secretory carrier membrane protein 2
SCD	Stearoyl-CoA desaturase
scRNA-seq	Single-cell RNA-seq
SD	Standard deviation
SE	Standard error
SE (bis)	Sensitivity
SEC24A	SEC24 homolog A COPII coat complex component
SELENOI	Selenoprotein I
SF3A3	Splicing factor 3A subunit 3
SFA	Saturated fatty acids
SFRP1	Secreted frizzled-related proteins 1

SFRP5	Secreted frizzled-related proteins 5
shRNAS	Short hairpin RNAs
SIFT	Sorting intolerant from tolerant score
SIGLEC10	Sialic acid binding Ig like lectin 10
SLC19A2	Thiamine transporter 1
SLC27A4	Fatty acid transport protein 4
sncRNAs	Small nuclear RNAs
snoRNAs	Small nucleolar RNAs
SNP	Single nucleotide polymorphism
SOM	Self-organizing maps
SP	Specificity
SPTBN4	Spectrin beta non-erythrocytic 4
SRA	Sequence read archive
SREBF1	Sterol regulatory element-binding protein 1
sRNA-seq	Small RNA sequencing
SRSF3	Serine and Arginine rich splicing factor 3
SSC1	*Sus scrofa* chromosome 1
SSC2	*Sus scrofa* chromosome 2
SSC3	*Sus scrofa* chromosome 3
SSC4	*Sus scrofa* chromosome 4
SSC14	*Sus scrofa* chromosome 14
STK36	Serine/Threonine kinase 36
SVM	Support vector machine
TADA2A	Transcriptional adapter-Ada2
TCA	Tricarboxylic acids
TFs	Transcription factors
TG	Triglycerides
TG (bis)	Transductive graphs
TGFBRAP1	Transforming growth factor *β* receptor-associated protein 1
THRSP	Thyroid hormone responsive protein
TIGRESS	Trustful inference of gene regulation with stability selection

TIM	Timeless
TMM	Trimmed mean of means
TOM	Topological overlapping matrix
TPM	Transcript per million
tRNAs	Transfer RNAs
TXNIP	Thioredoxin interacting protein
UFA	Unsaturated fatty acids
URB1	URB1 ribosome biogenesis homolog
V	Vertices
VCF	Variant calling format
WGCNA	Weighted gene correlation network analysis
WGS	Whole-genome sequencing
XGB	Extreme gradient boosting trees
XPO5	Exportin 5
ZNF142	Zinc finger 142
ZNF704	Zinc finger 704
α-SMA	α-smooth muscle actin
β	Power soft threshold
δ	Allele substitution effect
ΔD	Differential deviation
ω3	Omega-3 fatty acids
ω6	Omega-6 fatty acids

CHAPTER I. INTRODUCTION

1.1. Investigating the molecular basis of meat quality traits in pigs

1.1.1. The pig industry is highly technified

Pig industry is one of the most important sectors in meat production, together with poultry and beef. The development of an intensive and highly technified productive system has made possible the access of consumers to affordable and safe pig fresh meat and other processed products. With an observed annual growth of ~1.25% in meat production, spanning 2009-2018 (OECD-FAO Agricultural Outlook 2019-2028, agri-outlook.org), pig meat production has followed the increasing trend observed for the production and consumption of other types of meats and animal-derived products, although in relative terms, poultry, beef and sheep experienced higher percentual increases compared to pigs. In absolute numbers, pig meat production is expected to experience an increase of approximately 11 Mt during the next decade, with ~0.8% of additional growth per year, while the number of productive animals will only grow around 0.5% interannually, in accordance with data observed for the period comprising 2009-2018 (OECD-FAO Agricultural Outlook 2019-2028, agri-outlook.org, Figure 1). China will concentrate the majority of the predicted increases in pig production (42%), with two-thirds of forecasted growth coming from increasing production efficiency. Indeed, China is currently the most important producer of porcine-derived products. Other relevant producers are the USA, Brazil, Germany or Spain, where the porcine industry constitutes an important fraction of the food production system. On the other hand, most consumers of porcine-derived products are from European countries, followed by China, South Korea and USA. Consumers are increasingly favoring new standards in pig production other than affordability and offer, often related with animal welfare, such as traceability and high-quality standards for pig management as well as of the final marketed products (Thorslund et al., 2017; Xu et al., 2019).

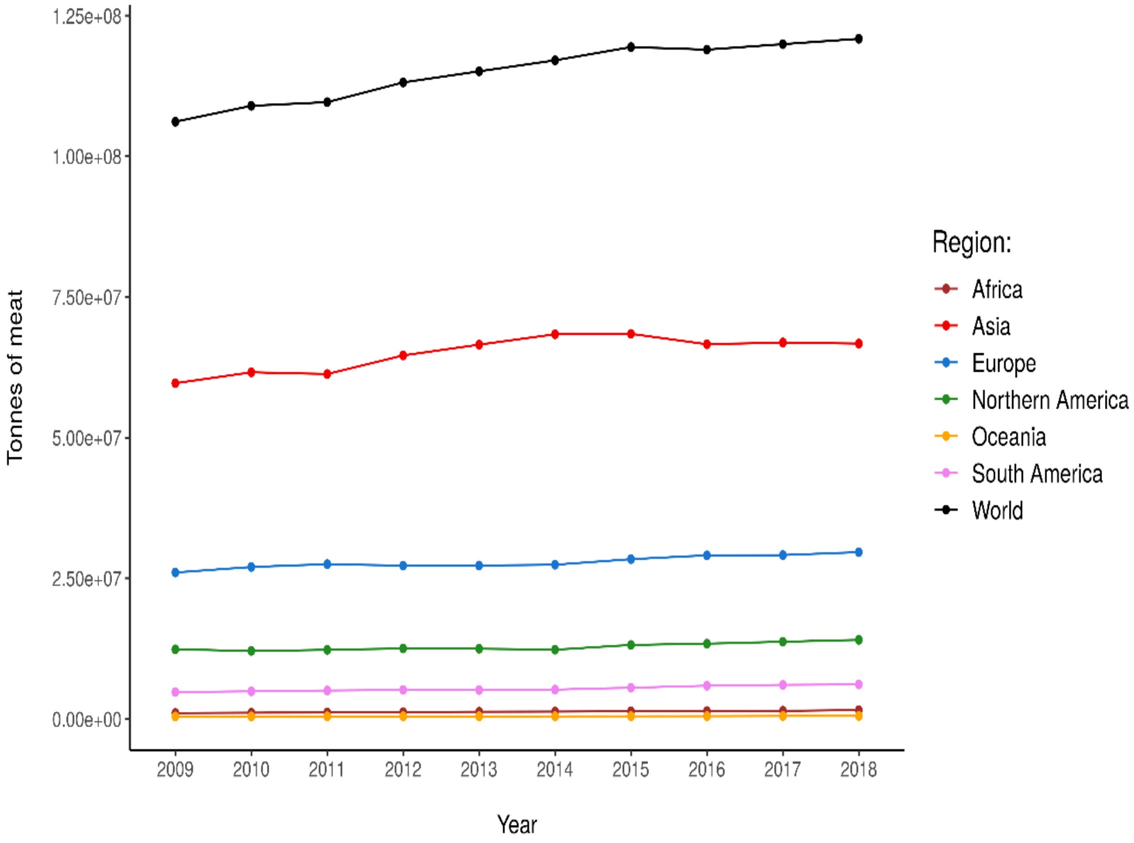


Figure 1: Consumption of pig meat during 2009-2018 in different geographic regions and expressed as the number of tonnes of meat produced by year (agri-outlook.org).

More specifically, pig industry is based on a cost-effective production system devoted to yield high-quality porcine products while meeting the highest standards in food security (Dekkers et al., 2011). This system often implies an intensive stratified scheme with three well-defined separated phases: 1) Nucleus farms, where purebred pig males (sires) and females (dams) are kept in health and management conditions following the highest standards and are subjected to intensive selection procedures. Dams are specifically selected for reproductive and maternal traits such as fertility, litter size or litter weaning weight, as well as for growth rate and meat leanness (Neeteson-van Nieuwenhoven et al., 2013). Conversely, the breeding sire line is primarily selected for growth rate, leanness, reduced mortality and meat quality traits

(Neeteson-van Nieuwenhoven et al., 2013). 2) Multiplier farms, where purebred selected lines are mated to generate a hybrid F_1 generation of sows that are sold to 3) commercial farms, where F_1 sows are inseminated with sperm from purebred sires from nucleus farms to produce the final crossbred F_2 generation of piglets that will complete a growth-finishing phase until slaughtering for commercial purposes. A schematic representation of the productive three-phase pyramid is shown in Figure 2.

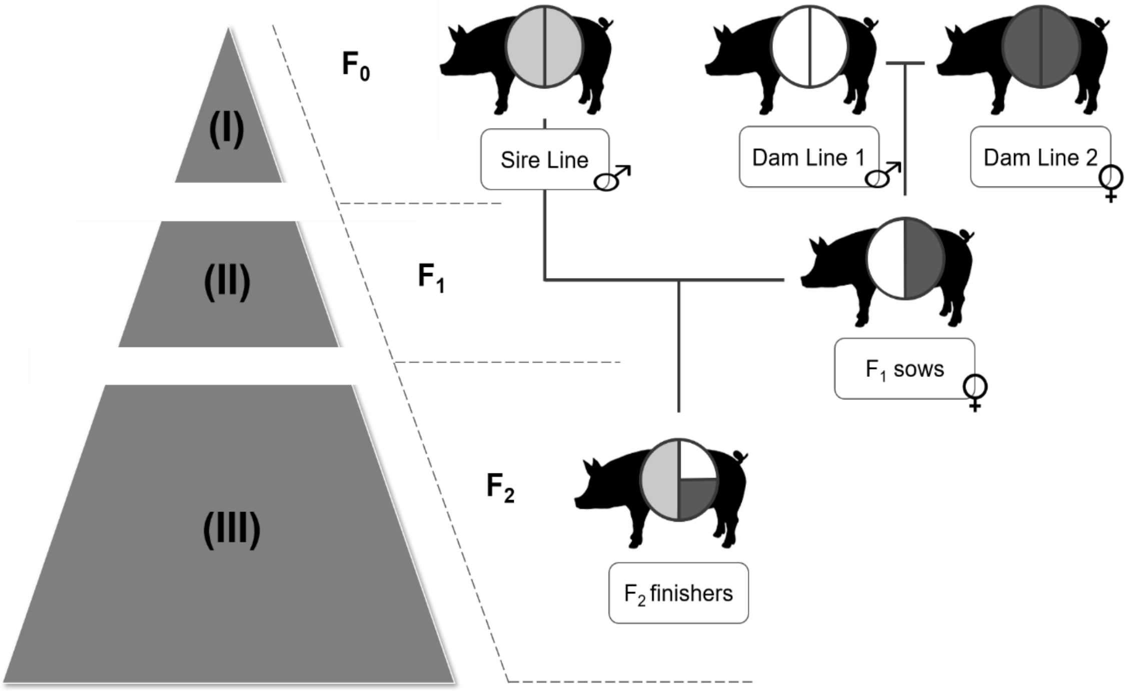


Figure 2: Breeding pyramid scheme. **(I)** Nucleus farm, where F_0 purebred animals are selected and subjected to intensive selection. **(II)** Multiplier farms, where purebred animals from different lines are crossed to generate F_1 animals (hybrid sows). **(III)** Production farms, where F_1 sows are inseminated by F_0 sires to generate F_2 commercial finishers that are bred and fattened until slaughter.

As a highly technified industry, breeders apply both classical BLUP and also genomic selection procedures in their selection nuclei in order to select breeders and improve the overall genetic and productive performance of pigs from the commercial stratum. Genomic selection relies on the estimation of the genomic breeding values (GEVB) of each individual of the population by using a large number of genotyped markers across the whole genome (Meuwissen et al., 2013). The additive effect of each marker is then estimated as a regression of the phenotype on the genotype information, using animals with both phenotypic and genotypic recordings and used to predict the GEVBs of the rest of individuals without the need to have phenotype data for the whole population (Samorè and Fontanesi, 2016). During the past decade, different methods for implementing genomic selection have been proposed: from Bayesian approaches (Gianola et al., 2009; Habier et al., 2011), to the widely used "single-step" genomic selection (Legarra et al., 2009), which can be seen as an extension of the genomic best linear unbiased prediction (GBLUP) method. In this approach, a genomic relationship matrix is used to account for family relationships (VanRaden, 2008), combining genomic relationships between genotyped animals with pedigree relationship matrices from non-genotyped animals.

Pig genomic selection has been capitalized by transnational companies such as Topigs Norsvin, Pig improvement company (PIC), Hypor or Monsanto, which maintain several selection nuclei under extensive phenotype recording. These companies use classical measurements of body conformation and other productive and reproductive traits, as well as more advanced phenotyping technologies. For instance, the Topigs Norsvin company (https://topigsnorsvin.com/) is using computed tomography scans to measure carcass traits and other phenotypes of interest. Additionally, these companies have also included massive genotyping procedures in their selection programs, making use of single nucleotide polymorphism (SNP) arrays to obtain genome-wide genotype information of pigs from the selection nuclei and perform genomic selection. However, GBLUP-based methods assume that quantitative traits are controlled by additive effects of a large number of genes and SNPs within them, explaining a small percentage of the observed phenotypic variance (Goddard, 2009). This assumption leads to suboptimal performances in the prediction of GEVBs because many quantitative traits are regulated by a certain number of genes with small additive effects (Hayes and Goddard, 2001). To overcome this situation, previous selection of SNPs based on genome-wide association studies (GWAS) to account for linkage disequilibrium, dominance, imprinting or breed-specific effects have been proposed (Su et al., 2012Costa et al., 2015; Esfandyari et al., 2015). Nevertheless, as the statistical power of GWAS heavily relies on the amount of available genotype data, the size of analyzed populations is of paramount importance for obtaining a comprehensive picture of genomic sites with functional effects on productive traits, both at coding and non-coding regions of the genome.

1.1.2. Uncovering the genetic basis of meat quality traits

Research about the genetic determinism of meat quality traits has focused on key phenotypes influencing the technological and organoleptic attributes of meat, including post-mortem pH, electric conductivity, water-holding capacity, drip loss, color, and intramuscular fat content and composition. Heritabilities for these traits are in general moderate to high, as reported by Rothschild et al. (2011) and Van Eenennam et al. (2014). Of course, heritability values of phenotypes depend on many factors including breed, animal cohorts and analyzed muscle tissues (Suzuki et al.,2006; Casellas et al., 2010; Gjerlaug-Enger et al., 2010; Ramayo-Caldas et al., 2012). Nevertheless, meat quality traits are expected to provide a good response to artificial selection.

In the nineties, the development of molecular markers that could be genotyped by polymerase chain reaction (PCR) made possible the development of marker-assisted selection (MAS) programs to identify quantitative trait loci (QTL), i.e. regions of the genome containing polymorphisms with causal effects on quantitative traits (Große-Brinkhaus et al., 2010). The first QTL studies (Andersson et al., 1994) were based on panels of 100-200 microsatellites, so their resolution was limited. However, they were extraordinarily useful to obtain a first glimpse about the genomic architecture of meat quality traits. It was clear that phenotypes of economic interest are polygenic and that genomic contributions to phenotypic variance are considerably heterogeneous across loci (Van Eenennaam et al., 2014). Since then, a great number of studies have identified meat quality QTL, mainly in intercrosses such as wild boar × Large White (Andersson et al., 1994), Landrace × Iberian (Pérez-Enciso et al., 2000; Revilla et al., 2014), Berkshire × Yorkshire (Malek et al., 2001), Japanese wild boar × Large White (Nii et al., 2005), Duroc × Landrace (Rohrer et al., 2006), Duroc × Pietrain (Liu et al., 2007; Große-Brinkhaus et al., 2010; Choi et al., 2011), Landrace × Korean pigs (Cho et al.,

2015) or white Duroc × Erhualian pigs (Guo et al., 2019). However, genome scans carried out in purebred populations evidenced that many of the QTL identified in intercrosses do not segregate in commercial populations (Evans et al., 2003; Vidal et al., 2005), thus limiting the applicability of such knowledge. Despite these limitations, a plethora of QTL have been mapped and thoroughly collected and summarized in ad-hoc databases like PigQTLdb (Hu et al., 2005). Up to date, the PigQTLdb database encompasses a total of 30,170 porcine QTL identified in a range of 687 different scientific publications and representing 688 porcine quantitative traits (release 40, https://www.animalgenome.org/cgi-bin/QTLdb/index). Although these initial studies allowed the identification of multiple QTL, their resolution to fine map the boundaries of such QTL were usually hindered by the small sizes of the investigated populations, which prevented mutations with weak effects to be detected. Moreover, the small number of known microsatellites markers also reduced the confidence and resolution with which QTL were mapped (Nagamine et al., 2003). Nonetheless, several causal mutations underlying changes in porcine productive traits were described using MAS approaches or by candidate gene search studies. One of the very early examples of candidate gene study led to the identification of the causal mutation of the halothane syndrome in porcine populations selected for carcass conformation and lean meat. In certain highly muscled pig breeds (e.g. Pietrain), the frequency of a missense variant (Arg615Cys) in the coding region of the ryanodine 1 (*RYR1*) gene was markedly high. This allele causes the dysregulation of the flow of Ca^{2+} from the sarcoplasmic reticulum to the cytosol of the myocyte under stressful conditions, favoring the development of a metabolic disbalance due to acidosis caused by energy exhaustion (Fujii et al., 1991). This syndrome results in pale, soft and exudative meats and subsequent economic loses for producers (MacLennan et al., 1990; Fujii et al., 1991).

With regard to QTL studies that were successful at identifying causal mutations, it is worth mentioning the one that demonstrated that an intronic mutation in the paternally imprinted insulin growth factor 2 (*IGF2*) gene influences skeletal and cardiac muscle mass development in pigs (Jeon et al., 1999). In a study by Milan et al. (2000), the precise location of a dominant mutation causing high glycogen content, low ultimate pH and decreased water-holding capacity in the skeletal muscle of Hampshire pigs was successfully mapped. Through the use of microsatellites, Milan et al. (2000) built a high-resolution linkage map which allowed the detection of a missense mutation (Arg200Gln) located at the protein kinase AMP-activated non-catalytic subunit γ 3 (*PRKAG3*) gene, which encodes a muscle-specific isoform of the regulatory γ subunit of the adenosine monophosphate-activated protein kinase (AMPK). Linking the observed Rendement Napole phenotype with the putative effects of the missense mutation in the PRKAG3 protein, the authors hypothesized that the replacement of Arginine by Glutamine could lead to an increase of AMPK basal activity and thus to an augmented glycogen content in the muscle (Milan et al., 2000). Besides, another missense variant in the melanocortin 4 receptor (*MC4R*) was found at high frequencies in Hampshire, Landrace and Duroc breeds, possibly as a consequence of selection for daily gain in these populations (Bruun et al., 2006). Indeed, the *MC4R* gene maps to a QTL associated with carcass fat/meat ratio and to another QTL affecting muscle gain (Houston et al., 2004), which would be in agreement with the role of *MC4R* gene as a key regulator of feed intake and energy homeostasis (Bruun et al., 2006). Polymorphisms in the leptin receptor (*LEPR*) and fatty acid binding protein 3 (*FABP3*) genes have been also associated with meat quality traits like intramuscular fat content, meat moisture, cholesterol content and flavor score, as well as with the expression of *LEPR* and *FABP3* transcripts (Li et al., 2010; Óvilo et al., 2010; Pérez-Montarelo et al., 2013).

To increase the power and resolution of these initial QTL mappings, researchers started to combine different sources of genotype information into meta-analyses (Tortereau et al., 2010; Rückert et al., 2012), an approach that contributed to further refine the chromosomal location of many described QTL while reducing their confidence intervals (Silva et al., 2011). Moreover, the advent of SNP arrays for genotyping studies in pigs (Ramos et al., 2009) greatly improved previous attempts for building detailed maps of genomic regions influencing meat quality traits. Jointly with the publication of the first sequenced genome assembly of the pig by Groenen et al. (2012), commercial SNP panels made possible to perform GWAS in pigs and paved the way for implementing genomic selection schemes in the pig industry. With more than 62K SNPs included in the Porcine SNP60 Beadchip (Ramos et al., 2009), GWAS approaches started to be applied with the aim of mapping pig QTL. The first study implementing the genome-wide scale mapping of QTL with the porcine chip was reported by Duijvesteijn et al. (2010). These authors described two regions in porcine chromosomes 1 and 6 that were associated with androsterone levels in a commercial Duroc population (Duijvesteijn et al., 2010). After this first report, many other authors have also applied these techniques to gain further insight into the genomic architecture of meat quality traits such as intramuscular fat content and composition (Puig-Oliveras et al., 2016; Ros-Freixedes et al., 2016; Sato et al., 2017; Zhang et al., 2016, 2019), meat pH (Davoli et al., 2019; Liu et al., 2019b) and meat color (Zhang et al., 2015; González-Prendes et al., 2017; Cho et al., 2019).

As previous QTL scans based on microsatellites, GWAS evidenced the highly complex genetic basis of meat quality traits. For instance, a total of 865 polymorphisms, clustered in 11 genome-wide significant loci across 9 different chromosomes were associated with 33 fatty acid phenotypes in five different porcine populations (Zhang et al., 2016). Furthermore, the authors discussed the role of relevant lipid-related genes mapping to QTL regions such as the fatty acid desaturase 2 (*FADS2*), the sterol regulatory element binding transcription factor 1 (*SREBF1*) or the phospholipase A2 Group VII (*PLA2G7*) loci. Many other studies focused on the genetics of fatty acids traits have been published (Ramayo-Caldas et al., 2012; Casellas et al., 2013; Muñoz et al., 2013; Yang et al., 2013; Sato et al., 2017). Lipid content and composition affect the technological properties of meat. For instance, unsaturated fatty acids tend to decrease the melting point of fat and polyunsaturated fatty acids are prone to oxidation, thus deteriorating meat flavor (Wood et al., 2008). Moreover, fatty acid composition has also effects on the nutritional quality of food (unsaturated fatty acids are healthier than the saturated ones), a feature that has prompted the identification of causal mutations with effects on fat composition phenotypes. In this regard, Estany et al. (2014) found 18 mutations located at the 5’end and 3’-UTR regions of the stearoyl-CoA desaturase (*SCD*) gene in Duroc pigs, from which they identified a T/C SNP in the 5’end of the *SCD* gene. This site was described as highly associated with the enhanced desaturation ratio of stearic (C18:0) vs oleic (C18:1) fatty acids, both in muscle and subcutaneous fat, but not in liver (Estany et al., 2014). Two years later, Ros-Freixedes and collaborators used the Porcine SNP60 Beadchip (Illumina) to carry out a GWAS for intramuscular fat content and composition in Duroc pigs (Ros-Freixedes et al., 2016). In this second study, the influence of SNPs located in the *SCD* gene on the desaturation of stearic to oleic acid was further confirmed.

Other traits such as pH, meat color or glycogen content have also been investigated at a genome-wide scale. In a study employing a Duroc sire line crossed with hybrid females (Landrace × Large White), Zhang and collaborators described several polymorphisms, associated with multiple pH and color-related measurements, which mapped to the Zinc finger 142 gene (*ZNF142*) or close to the protein kinase AMP-activated non-catalytic subunit γ 3 (*PRKAG3*) and the Serine/Threonine kinase 36 (*STK36*), among others (Zhang et al., 2015). More recently, González-Prendes et al. (2017) reported several QTL located at porcine chromosomes 3, 4, 5, 13 and 17, that were associated with post-mortem meat pH, electric conductivity or meat redness (a*), lightness (L*) and yellowness (b*) in the *gluteus medius* and *longissimus dorsi* muscles from a dedicated commercial Duroc line.

1.1.3. Expression QTL for analyzing gene regulation in pigs

A number of studies have been made in order to identify expression QTL (eQTL), i.e. regions of the pig genome containing polymorphisms with causal effects on the expression of genes. Microarrays and, more recently, the high-throughput sequencing of RNA transcripts (RNA-seq) have made possible to generate massive amounts of gene expression data, thus facilitating the detection of underlying variations in transcript expression. In this regard, expression profiles of RNA transcripts can be used as quantitative phenotypes and subjected to statistical association analyses, mirroring QTL detection (Ernst and Steibel, 2013). Such information can be very useful to understand fundamental processes related with gene regulation, as well as to interpret the results of GWAS studies. In this context, the co-localization of QTL and eQTL has been employed to generate working hypothesis regarding the gene and type of polymorphism explaining the QTL (Westra and Franke, 2014). Moreover, gene regulatory networks and key regulators or hub genes and putative causal variants can be inferred by integrating QTL and eQTL information (Nica and Dermitzakis, 2013). Two types of eQTL are usually defined, i.e. *cis*-eQTL which regulate the expression of a nearby locus (e.g. less than 1 Mb apart) and *trans*-eQTL, which act on distant loci. It is clear that this definition is quite arbitrary because the terms “nearby” and “distant” will greatly depend on the experimental model system and on the resolution of the genome scan. A schematic view of the differences between *cis*- and *trans*-eQTL is shown in Figure 3.

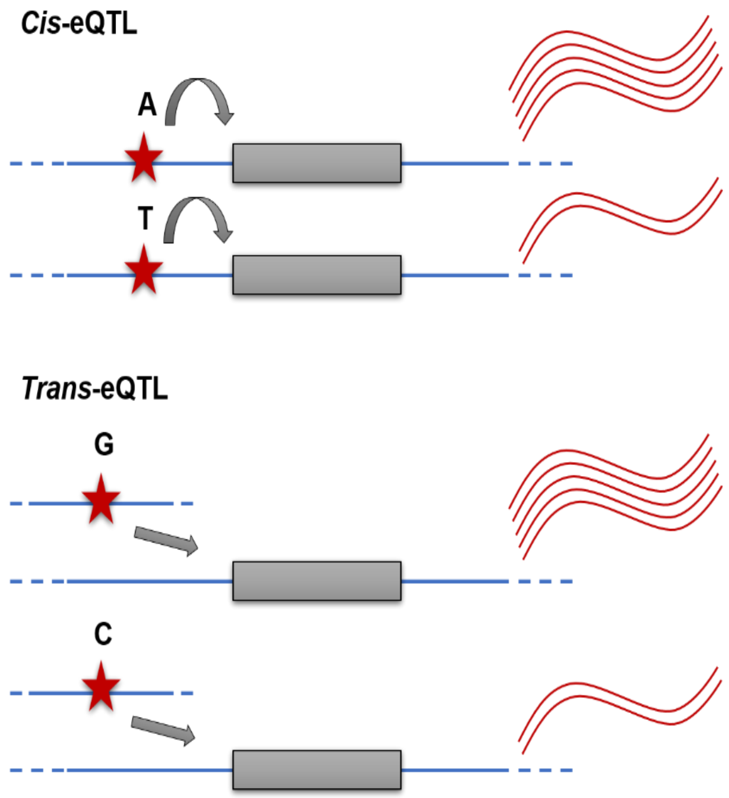


Figure 3: Representation of *cis*- and *trans*-eQTL affecting gene expression. Polymorphisms with *cis*-effects act over the expression of genes located at proximal distances (normally up to 1 Mb distance). Polymorphisms with *trans*-effect are located at distal positions (more than 1 Mb or even in other chromosomal regions).

The first study that used an eQTL approach in pigs was reported by Ponsuksili et al. (2008). In this study, the authors used Affimetrix GeneChip microarray expression data measured in *longissimus dorsi* skeletal muscle samples from 72 Duroc × Pietrain F_2 pigs to detect eQTL affecting meat water holding capacity. Further surveys made by the same authors investigated the existence of eQTL co-localizing with QTL associated with meat quality (Ponsuksili et al., 2010), fatty acids metabolism and fatness-related traits (Ponsuksili et al., 2011), plasma metabolites concentration (Ponsuksili et al., 2012) and muscle pH or electric conductivity (Ponsuksili et al., 2014). In this latter study, several SNPs located on porcine chromosomes 4 and 6 were associated with the expression of the inositol monophosphatase 1 (*IMPA1*), zinc finger 704 (*ZNF704*), oxidative stress response kinase 1 (*OXSR1*) or sialic acid binding Ig like lectin 10 (*SIGLEC10*) genes. Cánovas et al. (2012) also performed eQTL analyses using *gluteus medius* mRNA expression levels obtained with Affymetrix microarrays in 105 Duroc pigs selected for divergent fatness traits. These authors reported a predominance of *trans*-acting eQTL signals over *cis*-regulatory effects (Cánovas et al., 2012).

More recently, other authors have implemented eQTL analyses jointly with other sources of information to reconstruct gene regulatory networks with potential effects on pig productive traits. For instance, Peñagaricano et al. (2015) integrated QTL and eQTL data to infer gene regulatory networks including genes regulated by eQTL and focusing on genomic regions containing QTL for productive phenotypes.

In an additional work by González-Prendes et al. (2019a), the authors compared *cis*- and *trans*-eQTL signals between skeletal muscle and liver tissues and found a total of 76 and 28 genome-wide significant *cis*-eQTL in the *gluteus medius* skeletal muscle and liver tissue, respectively. Several of the eQTL-regulated genes in muscle might be involved in meat quality traits. From those eQTL identified by González-Prendes and collaborators in the liver, almost 43% (12) were also detected in the muscle (González-Prendes et al., 2019a). Although the number of shared *cis*-eQTL between these two tissues was relatively high, the reduced sample size limited the power of the study. Overall, the proportion of *cis*-eQTL shared across tissues in humans (Aguet et al., 2017) is similar to that obtained in pigs, and the main genetic mechanisms that regulate gene expression in pigs and humans are also expected to be similar (Pant et al., 2015). In this regard, Aguet et al. (2017) reported a sample size-dependent effect of the magnitude and significance of the effects mediated by *cis*- and *trans*-eQTL across tissues. While the number of identified *cis*-eQTL increased in tissues with smaller sample sizes, significance increased with sample size. This would suggest that a good strategy for augmenting the statistical power of experiments with a limited number of individuals could be the tissue-specific identification of eQTL and their joint-analysis with other publicly available data to infer their significance (Aguet et al., 2017). Equivalent results have been obtained in other meta-analyses: Võsa et al. (2018) conducted an extensive analysis of different sources of blood eQTL data and reported that 92% of the lead *cis*-eQTL SNPs map close (± 100 kb) to the gene they regulate. Moreover, the fine-mapping of the *cis*-effects increased with sample size, as significant leading SNPs were detected closer to the *cis*-eQTL associated gene (Võsa et al., 2018). Similarly, around 33% of traits were associated with some observed eQTL in *trans*, and many of them were related to *cis*-effects in transcription factors (TFs) that were themselves co-expressed with genes showing significant *trans*-eQTL signals. However, the correct mapping of *trans*-effects was still tightly linked to a sufficient sample size (Võsa et al., 2018). The apparent lack of significant shared *trans*-effects, even when considering large sample sizes, and their prominent tissue-specificity might be explained by the fact that most of *trans*-eQTL function as weak modulators of the expression of peripheral genes that, at the same time, affect the expression of core genes that are typically associated with *cis*-signals. This suggests that even when considering experiments with many individuals and different body tissues, the magnitude of missing heritabilities for gene expression phenotypes might be high as a consequence of weak *trans*-effects that are not detectable due to limited statistical power (Liu et al., 2019a).

1.1.4. Detection of deleterious variants in pigs

Natural selection has favored the spread of beneficial mutations, while removing the harmful ones due to their negative effects on fitness and offspring survivability. The load of deleterious mutations in a given population depends on multiple factors such as their level of harmfulness on fitness, dominance and epistatic effects, environmental interactions and linkage disequilibrium with adjacent polymorphic sites (Makino et al., 2018). Demography can also have huge effects on the spread and retention of deleterious mutations. For instance, it is well known that the reduction in the effective population size increases inbreeding and the chance of the emergence of homozygous genotypes for harmful recessive mutations. Indeed, weakly deleterious sites may persist through generations, thus contributing to debilitate population fitness (Bosse et al., 2019). Additionally, livestock populations have been selected for centuries, and more prominently during the past decades with classical breeding strategies and recently with novel techniques based on genomic selection. This has contributed to create almost closed breeding lines with small (several hundreds) or very small (some dozens) effective population sizes, which are therefore more prone to suffer from inbreeding depression (Charlesworth and Willis, 2009; González-Peña et al., 2015). This effect is mostly caused by an increased homozygosity in partially detrimental recessive alleles segregating in animal populations and a consequent reduction in fitness (Charlesworth and Willis, 2009). The frequencies of harmful or even lethal mutations can increase if they are linked to alleles with favorable effects on the traits that are selected for. Given the complex history of pig breeds (Bosse et al., 2014), the segregation of deleterious alleles in commercial pig populations has been the focus of several studies during the past years. A selection of studies characterizing loss-of-function mutations and their putative deleterious consequences on transcript structure and related phenotypes is shown in Table 1.

Table 1: List of loss-of-function mutations with harmful predicted consequences on productive traits in pigs.

SSC[a]	Position (Mb)	Gene	Type[b]	Trait	Reference
1	141.23	*DUOX2*	Missense	Hypothyroidism	(Cao et al., 2019)
3	43.95	*POLR1B*	Splice-region	Litter size	(Derks et al., 2019a)
6	48.75-50.25	*SPTBN4*	Deletion	Postnatal mortality	(Derks et al., 2019b)
6	54.88	*PNKP*	Missense	Litter size	(Derks et al., 2019a)
8	107-113.3	*MAD2L1*		Stillborn piglets	(Häggman and Uimari, 2017)
8	107-113.3	*FGF2*		Stillborn piglets	(Häggman and Uimari, 2017)
8	107-113.3	*NUDT6*		Stillborn piglets	(Häggman and Uimari, 2017)
8	107-113.3	*ANXA5*		Stillborn piglets	(Häggman and Uimari, 2017)
12	38.92	*TADA2A*	Splice-donor	Litter size	(Derks et al., 2019a)
13	195.98	*URB1*	Frameshift	Litter size	(Derks et al., 2019a)
18	39.2-40.1	*BBS9*	Deletion	Growth rate, loin depth, feed intake	(Derks et al., 2018)
18	39.1-40.1	*BMPER*	Deletion	Litter size, stillborn piglets, mummification	(Derks et al., 2018)

[a]SSC: Porcine chromosome; [b]Type: Predicted effect of the causal polymorphism.

Derks et al. (2018) analyzed the genetic basis of a recessive lethal haplotype on pig chromosome 18 showing reduced or missing homozygosity. A thorough study of this region in heterozygous carriers revealed a deletion of 212 kb partially spanning the Bardet Biedl syndrome 9 (*BBS9*) gene. After examining the expression patterns and exon structure of *BBS9* transcripts in several tissues, the authors found that the observed deletion induced the skipping of several exons. Such alteration introduced 11 novel amino acids immediately before a premature stop codon, thus generating a truncated non-functional BBS9 protein.

Furthermore, quantitative real-time PCR (RT-qPCR) analyses between wild type and heterozygous animals demonstrated a 50% lower expression of the *BBS9* transcripts in carriers compared with non-carrier individuals (Derks et al., 2018). Apart from the observed reduction in *BBS9* expression, association analyses with reproductive traits revealed that carriers also presented a decreased number (~20%) of born piglets compared with crosses involving pigs homozygous for the wild type allele, as well as with carrier × wild type matings. Besides, stillborn and mummified piglets were more prevalent in carrier × carrier matings, and the homozygous status for the deletion in several of the mummified individuals was also confirmed. These results contrasted with the relatively high frequency of the lethal mutation in the pig population, i.e. a 10.8% carrier frequency (5.4% allele frequency) was detected. When the authors examined the productive performance of carriers and non-carriers, increased growth rate, smaller loin depth, litters with reduced weight and increased feed intake were found in carrier animals. Such results would explain the maintenance of the lethal allele at non-negligible frequencies due to artificial selective pressure of selection programs aimed at improving growth rates. Other relevant examples of harmful deletions have also been reported by Derks and collaborators, like one producing a truncated form of the spectrin β non-erythrocytic 4 protein (SPTBN4) and resulting in postnatal mortality with homozygous piglets suffering severe myopathy, hind-limb paralysis and tremors (Derks et al., 2019b).

Analogously, other surveys have reported harmful recessive haplotypes causing reproductive problems in pigs. In a study by Häggman and Uimari (2017), the authors identified 26 putative lethal haplotypes spanning 12 chromosomes by estimating the deviation in recessive homozygosity according to the observed carrier proportion in a population of Finnish Yorkshire pigs. One haplotype located on chromosome 18 showed significant associations with increased numbers of stillborn piglets in first and later parities. This region harbored several interesting genes such as the mitotic arrest deficient 2-like 1 (*MAD2L1*) gene, which is involved in the regulation of meiosis and prevents aneuploidy events (Homer et al., 2005), the fibroblast growth factor 2 (*FGF2*) and its antisense transcript, nudix hydrolase 6 (*NUDT6*), which participates in the vascular reorganization of uterine and placental beds during pregnancy (Chrusciel et al., 2010), or the placental anticoagulant annexin 5 (*ANXA5*), which was linked to pregnancy losses in humans (Bogdanova et al., 2007). Moreover, another study reported sets of frameshift, splice-site and missense variants causing a complete loss-of-function of the affected genes and reduced litters in carrier × carrier matings. Essential genes for regulating DNA transcription and repair of DNA damage such as the transcriptional adapter-Ada2 (*TADA2A*), the RNA polymerase I subunit B (*POLR1B*), the URB1 ribosome biogenesis homolog (*URB1*) or the polynucleotide kinase 3'-phosphatase (*PNKP*) were affected by these deleterious variants, which generated either truncated proteins or proteins with amino acid substitutions compromising their function (Derks et al., 2019a). Similarly, missense variants with predicted harmful effects have also been reported for non-reproductive traits in pigs. For instance, an A/G mutation located at a splicing enhancer region of the dual oxidase 2 (*DUOX2*) gene was described to cause aberrant alternatively spliced mRNA isoforms, hence impairing the production of H_2O_2 and altering the synthesis of thyroid hormones in Chinese Bama pigs (Cao et al., 2019).

1.2. The era of transcriptomics and integrative analyses

1.2.1. An introduction to the analysis of transcriptomes

The advent of the next-generation sequencing (NGS) techniques allowed the high-throughput sequencing of the cell transcriptome with unprecedented resolution. The NGS of RNA transcripts (RNA-seq) paved the way for the systematic characterization of the transcriptomes of multiple tissues. In this regard, RNA-seq involves the bulk extraction of the RNA fraction from a tissue sample obtained from animals subjected to a specific experimental treatment (Wolf, 2013). This RNA preparation encompasses the whole fraction of RNA molecules present in the cell, where many different types of RNA transcripts exist. Indeed, the most abundant RNAs present in a typical metazoan cell corresponds to ribosomal RNAs (rRNAs), which accounts for ~80 % of the total amount of RNA molecules present within the cell (Wu et al., 2014b). The remaining RNAs correspond to other types of transcripts: either protein-coding RNAs (messenger RNAs or mRNAs) yielding ~4 % of the total RNA mass, or other types of RNA molecules (Wu et al., 2014b). This latter group of RNAs drives major cellular activities, such as chromosomal structure and organization, DNA replication and repair or transcriptional/post-transcriptional regulation, and it encompasses transfer RNAs (tRNAs), microRNAs (miRNAs), long non-coding RNAs (lncRNAs), circular RNAs (circRNAs), small nuclear RNAs (sncRNAs), small nucleolar RNAs (snoRNAs) and mitochondrial RNAs (mtRNAs), among few others. In section 1.3 we will discuss with more detail the biology of miRNAs, which is one of the main subjects of this thesis.

Prior to sequencing, the abundant rRNAs must be depleted with the goal of concentrating sequencing efforts on the remaining types of RNAs, (Wolf, 2013). Moreover, extraction protocols and sequencing techniques should also be adjusted for the correct capturing of particularly small RNA molecules like, for instance, miRNAs (Brown et al., 2018). Subsequently, researchers must decide whether to perform single-end or paired-end library preparation, which also requires PCR-based transcript amplification. Paired-end protocols are useful for initial transcriptome assembly or isoform detection, and are the preferred choice for common RNA-seq experiments aiming at sequencing the mRNA transcriptome, whereas single-end sequencing allows a better capture of the small size fraction of transcripts such as miRNAs (Wolf, 2013). In addition, strand-specific protocols should be considered when attempting to reach accurate sequencing of regions with overlapping transcription from both genomic strands (Borodina et al., 2011).

After library preparation, sequencing of the extracted RNA transcripts must be performed. A varied range of NGS platforms are commercially available nowadays, with Illumina, Pacific Biosciences, Oxford Nanopore and ThermoFisher Ion Torrent being among the most currently used ones (Levy and Myers, 2016). While some of them are based on optical light or fluorescence detection (Illumina and Pacific Biosciences), others use different methods of sequence detection such as changes in pH during a polymerization reaction via a solid-state sensor (ThermoFisher Ion Torrent) or modifications of an electrical field as the nucleic acids pass through a protein nanopore (Oxford Nanopore). Among them, the Illumina sequencing platform is probably one of the most used approaches, followed by the emerging Oxford Nanopore technique, which allows long read sequencing and hence a more accurate *de novo* assembly of genomes and a better characterization of large structural variations in the genome (Amarasinghe et al., 2020). During the course of the present thesis, Illumina platform was used for short-read sequencing of miRNAs, and is therefore further explained in Figure 4.

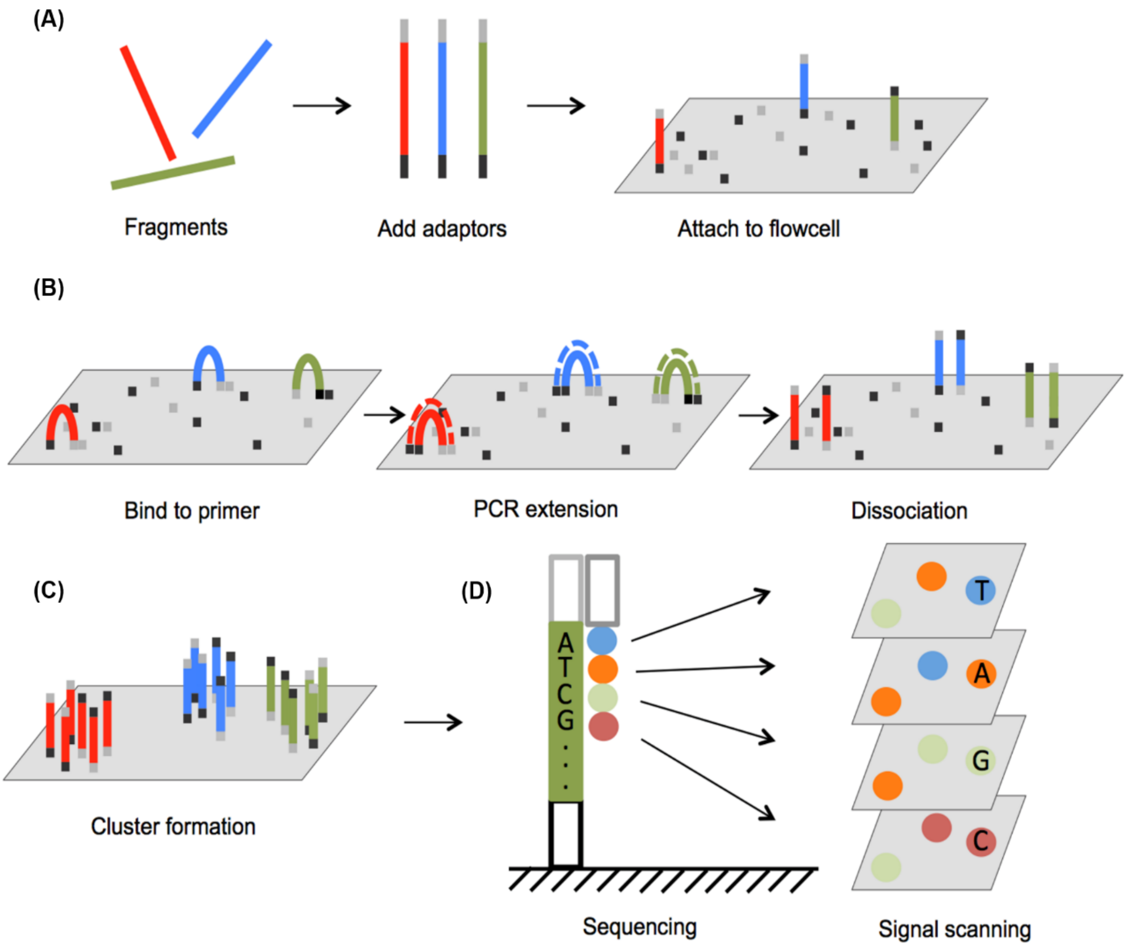


Figure 4: Illumina sequencing of short DNA transcript fragments after library preparation and PCR amplification. **(A)** Adaptors are annealed to the 5' and 3' ends of fragments and attached to primer primer-loaded flow cell. **(B)** Sequence fragments form bridge structures allowing PCR amplification and dissociation where the bridge amplification is repeated. **(C)** Through successive cycles of PCR bridge amplification, double stranded DNA sequences are denatured and attach to the flow cell to form sequence clusters. **(D)** When clonal amplification terminates, all formed reverse strands are washed away and primers are attached to the forward strands. A DNA polymerase then adds fluorescently tagged nucleotides complementary to the sequence. Only one base is added per round. After each cycle, the machine scans which nucleotide was added by using a color signal recording in order to reconstruct the read sequence.

Modified from http://www.3402bioinformaticsgroup.com/service/.

Once sequencing has been performed, the resulting reads are processed and typically stored in FASTA or FASTQ formatted files. Then, in order to process the sequenced data, a broad array of bioinformatic pipelines can be applied to retrieve the transcriptional information contained in the sequenced RNA reads (Wolf, 2013, Conesa et al., 2016). A schematic workflow of the successive steps for processing and analyzing RNA-seq data is depicted in Figure 5 and summarized hereunder:

1) Most commonly, the pipeline for processing FASTQ formatted reads involves a first step of quality check and filtering process of the raw reads, in which poorly determined nucleotides at the 3’ end of the reads are removed according to pre-established quality thresholds, and any remaining adapters used during sequencing are also filtered out in order to generate clean processed and quality-checked reads. Different dedicated software tools are available for the quality checking of the reads, such as FASTQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Common tools for adapter trimming are, for instance, Cutadapt (Martin, 2011), Trimmomatic (Bolger et al., 2014), fastp (Chen et al., 2018) or the FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/).

Although several distinct pipelines can be used to analyze RNA-seq data, here we will discuss one of the most widely used for the mapping and quantification of transcripts, differential expression (DE) analysis and pathway and gene ontology enrichment analyses (Figure 5).

2) After quality check and adapter trimming, clean reads are now ready to be aligned to a reference genome assembly (as long as it is available for the species of interest). In the event that no assembly can be used, *de novo* transcript assembly can be performed in order to obtain a putative representation of the transcriptome defined by the set of sequenced reads (Haas et al., 2013). Moreover, reference annotation-free mapping can also be performed to generate novel prediction of unannotated transcripts (Kim et al., 2019). Among the most used and cited tools for reference assembly-based alignment, it is worth mentioning Bowtie (Langmead et al., 2009), Bowtie 2 (Langmead and Salzberg, 2012), STAR (Dobin et al., 2012), HISAT2 (Kim et al., 2019), or BWA (http://bio-bwa.sourceforge.net/). The alignment of reads generates a representation of the mapping location of the reads towards the employed reference assembly in the format of sequence alignment map (SAM) files, or their binary version, BAM files.

3) Once the alignment files are generated, successfully mapped transcripts need to be quantified in order to determine the expression abundance of each gene. Genes showing a higher expression will be represented by more transcripts mapping to their genomic location compared with lowly expressed genes, which will gather less mapped reads. In this way, tools like featureCounts (Liao et al., 2014), StringTie (Pertea et al., 2015), HTSeq (Anders et al., 2015), Kallisto (Bray et al., 2016), Salmon (Patro et al., 2017) or StringTie2 (Kovaka et al., 2019) allow the quantification of the number of transcripts representing each annotated locus.

4) Before DE analyses, normalization of the count matrices must be performed in order to account for library size biases. Several normalization methods like the trimmed mean of means (TMM), quantile normalization, size factors, transcript per million mapped reads (TPM) or reads per kilobase transcript per million mapped reads (RPKM) have been compared and thoroughly reviewed by Abrams et al. (2019).

5) With quantified and normalized gene matrices, DE analyses between two or more defined contrasting conditions can be performed. Multiple tools are available for this purpose, which, in general, fit a probabilistic negative binomial distribution or linear additive response error models to the input quantification data in order to detect differences in average gene expression between two groups of samples. Dedicated tools for DE analysis are, for instance, edgeR (Robinson et al., 2010), DESeq2 (Love et al., 2014), NOISeq (Tarazona et al., 2015) or Sleuth (Pimentel et al., 2017). Through these analyses, researchers can obtain a representative list of genes showing significant expression differences (i.e. activation or repression) between contrasts after multiple testing correction (Benjamini and Hochberg, 1995; Benjamini and Yekutieli, 2001). Aside of canonical DE analyses based on contrasting mean gene expression differences by groups, alternative models targeting statistical differences on the variance of gene expression have also been recently proposed (Ran and Daye, 2017).

6) Finally, enrichment analyses are performed in order to obtain enriched gene ontologies or metabolic pathways that are overrepresented in the set of DE genes above a random distribution, typically by means of hypergeometric or Fisher’s exact tests. Available tools for such analyses are, to mention a few, ClueGO (Bindea et al., 2009), GOrilla (Eden et al., 2009), DAVID (Jiao et al., 2012), enrichR (Kuleshov et al., 2016), g:Profiler (Reimand et al., 2016) or PANTHER (Mi et al., 2019).

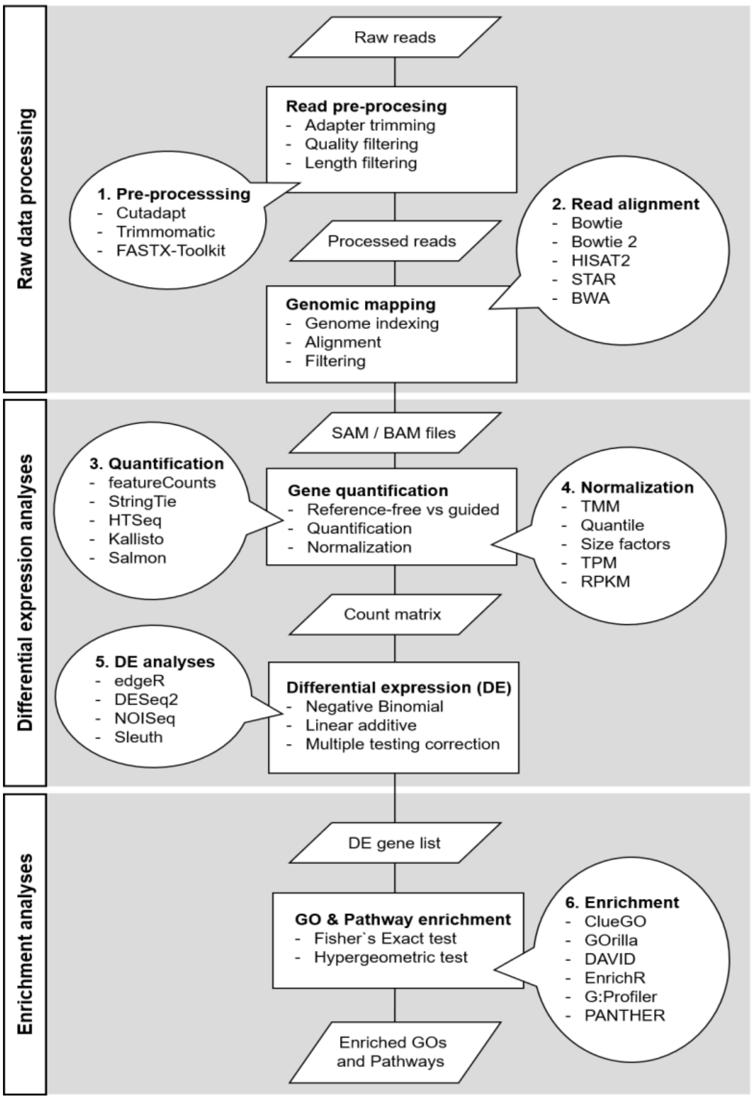


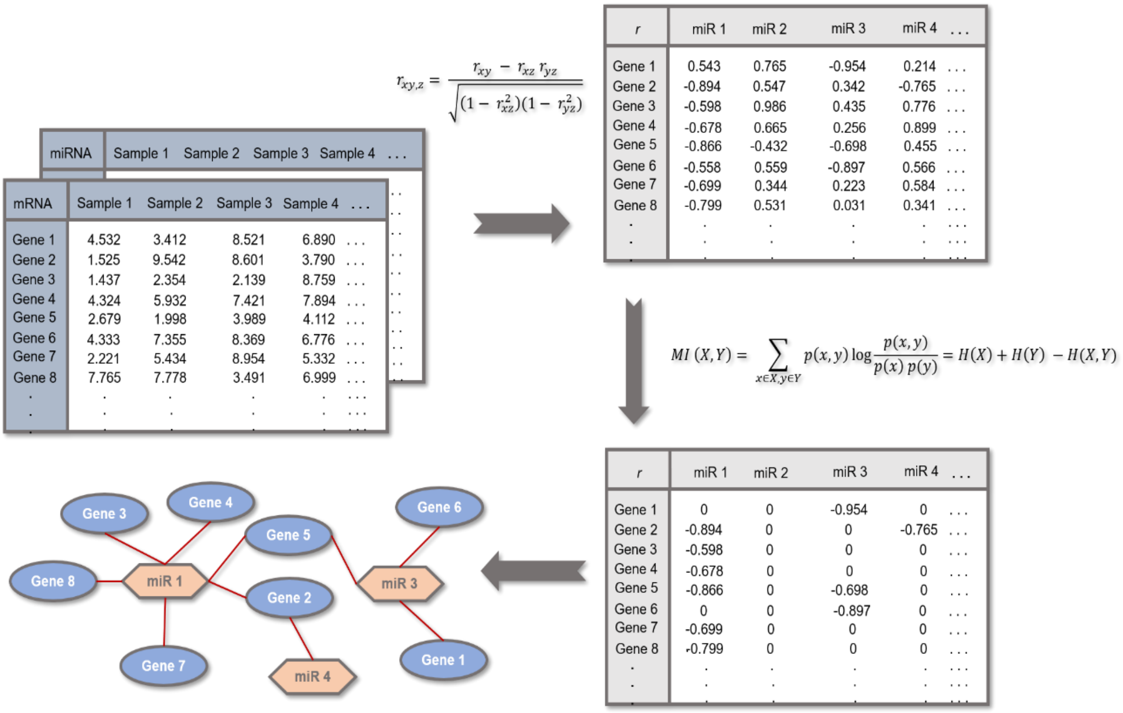
Figure 5: Schematic workflow for RNA-seq analysis of differentially expressed genes and enriched gene ontologies and metabolic pathways. **1)** Raw reads are quality checked and sequencing adaptors are trimmed. **2)** Clean reads are mapped to a reference genome assembly, if available. **3)** Gene counts are estimated based on read alignment and reference gene annotation. Alternatively, reference-free annotation can be performed for novel transcripts. **4)** Count matrices must be normalized prior to differential expression analyses. **5)** Different probabilistic distributions can be fitted to count data in order to perform differential expression analyses and multiple testing correction. **6)** Differentially expressed gene lists can be used to perform gene ontology and/or pathway enrichment analyses hence obtaining overrepresented ontologies and pathways.

1.2.2. Gene regulatory networks

Shortly after high-throughput sequencing techniques became a common methodology for many genomic projects, the amount of data generated made possible the integration of different sources of information to obtain a comprehensive hierarchy of the relationships among the different components that form part of the cell metabolism as a whole. This task has been accomplished by reconstructing gene-to-gene interactions, fundamentally based on transcriptomic information from NGS experiments, although non-transcriptomic data, such as DNA methylation patterns, histone marks, chromatin organization or post-translational modifications (e.g. phosphorylation or acetylation signals) have also been taken into consideration in several studies (Thompson et al., 2015). Overall, all these sources of data can be integrated to build a gene regulatory network (GRN), thus allowing a comprehensive representation of the multiple interactions that occur among RNA transcripts, proteins and other epigenetic factors (Figure 6). The inference of GRNs from different -omic data assumes that the variations in the expression of a given gene can be modeled as a function of one or more elements with which this gene interacts in a certain manner (Barbosa et al., 2018). Nevertheless, such assumption relies on multiple variables such as the experimental design, the sequencing depth, the type and techniques used for generating the information, confounding variables, intrinsic noise etc.

As discussed by Emmert-Streib et al. (2014), GRN inference is subjected to such a number of interacting variables that, virtually, no methodology has attained sufficient consensus to be applied as an all-purpose approach. It is therefore critical to correctly understand which are the main features of the data and the underlying hypothesis to be tested, in order to make good decisions about which approach to choose for reconstructing GRN.

The graphical representation of GRNs can be expressed as $G = \{V, E\}$, in which genes are considered as nodes or vertices (V) of the graph, while the edges (E) define the connections between genes in the form of expression correlations or any other metric. A connection is established when the interaction between two vertices is thought to exist or to be potentially meaningful. If the directionality of edges is known, i.e., the regulatory gene and its target are defined and differentiated, the graph is directed, whereas if no information regarding edge direction is included, the graph is undirected (Barbosa et al., 2018).



miRNA	Sample 1	Sample 2	Sample 3	Sample 4	. . .

mRNA	Sample 1	Sample 2	Sample 3	Sample 4	. . .
Gene 1	4.532	3.412	8.521	6.890	. . .
Gene 2	1.525	9.542	8.601	3.790	. . .
Gene 3	1.437	2.354	2.139	8.759	. . .
Gene 4	4.324	5.932	7.421	7.894	. . .
Gene 5	2.679	1.998	3.989	4.112	. . .
Gene 6	4.333	7.355	8.369	6.776	. . .
Gene 7	2.221	5.434	8.954	5.332	. . .
Gene 8	7.765	7.778	3.491	6.999	. . .
.	.	.	.	.	. . .

$$r_{xy,z} = \frac{r_{xy} - r_{xz}\, r_{yz}}{\sqrt{(1 - r_{xz}^2)(1 - r_{yz}^2)}}$$

r	miR 1	miR 2	miR 3	miR 4	. . .
Gene 1	0.543	0.765	-0.954	0.214	. . .
Gene 2	-0.894	0.547	0.342	-0.765	. . .
Gene 3	-0.598	0.986	0.435	0.776	. . .
Gene 4	-0.678	0.665	0.256	0.899	. . .
Gene 5	-0.866	-0.432	-0.698	0.455	. . .
Gene 6	-0.558	0.559	-0.897	0.566	. . .
Gene 7	-0.699	0.344	0.223	0.584	. . .
Gene 8	-0.799	0.531	0.031	0.341	. . .
.	.	.	.	.	. . .

$$MI\,(X,Y) = \sum_{x \in X, y \in Y} p(x,y) \log \frac{p(x,y)}{p(x)\,p(y)} = H(X) + H(Y) - H(X,Y)$$

r	miR 1	miR 2	miR 3	miR 4	. . .
Gene 1	0	0	-0.954	0	. . .
Gene 2	-0.894	0	0	-0.765	. . .
Gene 3	-0.598	0	0	0	. . .
Gene 4	-0.678	0	0	0	. . .
Gene 5	-0.866	0	-0.698	0	. . .
Gene 6	0	0	-0.897	0	. . .
Gene 7	-0.699	0	0	0	. . .
Gene 8	-0.799	0	0	0	. . .
.	.	.	.	.	. . .

Figure 6: Gene regulatory network (GRN) inference. From a series of -omic data representing gene expression measures of different types (e.g. mRNAs or miRNAs), ranking and prioritization algorithms can be applied to infer regulatory interactions among the different layers of input data.

The correct method used for inferring GRNs is highly dependent on the type, quality and quantity of data available. Several commonly used approaches for GRN inference are: 1) co-expression or correlation-based methods, 2) information-theoretic approaches, 3) regression-based algorithms and 4) Bayesian models. A summarized list of several representative methods for each of these categories is shown in Table 2.

Table 2: Summary of representative methods for Gene regulatory network inference.

Method	Type	Platform	Package	Reference
GeneNet	Correlation	R	GeneNet[b]	(Opgen-Rhein and Strimmer, 2007)
WGCNA	Correlation	R	WGCNA[c]	(Langfelder and Horvath, 2008)
MutRank	Correlation	R	NetBenchmark[d]	(Obayashi and Kinoshita, 2009)
RELNET	IT[a]			(Butte and Kohane, 2000)
ARACNE	IT[a]	R	MINET[e]	(Margolin et al., 2006)
CLR	IT[a]	R	MINET[e]	(Faith et al., 2007)
PCIT	IT[a]	R	PCIT[f]	(Reverter and Chan, 2008)
C3NET	IT[a]	R	c3net[g]	(Altay and Emmert-Streib, 2010)
BC3NET	IT[a]	R	bc3net[h]	(de Matos Simoes and Emmert-Streib, 2012)
MIDER	IT[a]	Matlab		(Villaverde et al., 2014)
PREMER	IT[a]	Matlab		(Villaverde et al., 2018)
Genie3	Regression	R	GENIE3[i]	(Huynh-Thu et al., 2010)
TIGRESS	Regression	R	Tigress[j]	(Haury et al., 2012)
GGM	Bayesian	R	GeneNet[b]	(Schäfer and Strimmer, 2005)
CBN	Bayesian	R	catnet[k]	(Balov, 2013)

[a] IT: Information Theory. [b] https://CRAN.R-project.org/package=GeneNet.

[c] https://CRAN.R-project.org/package=WGCNA.

[d] https://www.bioconductor.org/packages/release/bioc/html/netbenchmark.html (Bellot et al., 2015).

[e] https://www.bioconductor.org/packages/release/bioc/html/minet.html (Meyer et al., 2008).

[f] https://CRAN.R-project.org/package=PCIT (Watson-Haigh et al., 2010).

[g] https://CRAN.R-project.org/package=c3net.

[h] https://CRAN.R-project.org/package=bc3net. [i] https://bioconductor.org/packages/release/bioc/html/GENIE3.html.

[j] https://github.com/jpvert/tigress.

[k] https://CRAN.R-project.org/package=catnet.

Co-expression methods

These methods assume that similar expression profiles between genes are suggestive of underlying relationships that are revealed by measuring the patterns of co-expression. Co-expression can be explained by direct or indirect regulatory mechanisms and/or participation in common biological pathways. Co-expression networks are reconstructed by computing a similarity score reflecting pairwise interactions among genes. The simplest metrics would imply calculating correlation coefficients between the sets of gene expression data. If the observed correlation ranks above a predefined threshold, the interaction is considered as meaningful and conserved for building the GRN. Representative methods of such approach are, for instance, GeneNet (Opgen-Rhein and Strimmer, 2007), which relies on the conversion of the inferred correlation network into a partial correlation graph. In this context, the ranking of nodes is assigned by means of a multiple testing of the log-ratio of the standardized partial variances. The weighted gene correlation network analysis (WGCNA) proposed by Langfelder and Horvath (2008) is probably one of the most used and cited methods of this category. This procedure prioritizes relevant high co-expression relationships by raising the absolute value of the correlation to a pre-defined power threshold greater or equal to 1 ($\beta \geq 1$), which produces weighted adjacency matrices that are then transformed and clustered into co-expression modules formed by highly similar nodes. MutRank (Obayashi and Kinoshita, 2009) is a simple algorithm that ranks correlations between pairs of genes by considering a similarity score between genes, the reliability of which is measured as the geometric mean of the scores obtained between gene *i* and *j*, and vice versa.

Information-Theoretic approaches

These approaches use a generalization of the pairwise correlation coefficients known as mutual information (MI), which measures the degree of dependence between two given genes (Cover and Thomas, 2005). The RELNET algorithm (Butte and Kohane, 2000) is a simple method using MI between two nodes *i* and *j*. A connection is created if the computed MI_{ij} surpasses an established threshold. The algorithm for the reconstruction of accurate cellular networks (ARACNE) assumes that many expression similarities between pairs of genes might be the result of other weak indirect interactions. Thus, this method uses the data processing inequality (DPI) algorithm to remove the weakest edge, i.e. the one with lowest MI among each considered triplet of genes (Margolin et al., 2006). The CLR method (Faith et al., 2007) is an extension of the RELNET algorithm, which derives a normalized z-score of the MI between two nodes *i* and *j*, thus facilitating the removal of indirect connections. The partial correlation coefficient with information theory (PCIT) algorithm (Reverter and Chan, 2008) extracts all possible triplets of genes and applies DPI to remove indirect connections, combined with first-order partial correlations for weighting interactions. This approach aims to eliminate third indirect interactors for any given pairwise correlation of genes *i* and *j*. The conservative causal core network (C3NET) algorithm (Altay and Emmert-Streib, 2010) and the updated version BC3NET (de Matos Simoes and Emmert-Streib, 2012) incorporating a bagging procedure, first removes non-significant connections from pairwise MI estimates that do not surpass a pre-established threshold α. Subsequently, the most significant connection for each gene is selected. Non-parametric bootstrap is then applied to generate an ensemble of independent networks that are used to build a weighted network to determine the significance of each pairwise connection between genes. An entropy reduction step from MI distances was proposed by Villaverde et al. (2014) in the MIDER method. The goal was to discriminate direct from indirect interactions between genes in order to minimize the false positive rate. A parallelized version of this method (PREMER) was also developed to incorporate prior knowledge to the network inference (Villaverde et al., 2018).

Regression algorithms

Regression-based algorithms aim at finding the statistical relationship between two or more gene expression measures. In this framework, a change in a dependent variable can be modeled by considering changes in other independent variables. Examples of this method are, to mention a few, the Genie3 algorithm (Huynh-Thu et al., 2010), which uses random forest feature selection to solve a regression problem that consists of predicting the expression profile of a given target gene by means of the expression of other defined regulators. The trustful inference of gene regulation with stability selection (TIGRESS) algorithm (Haury et al., 2012) predicts the expression of a given g_i gene from the expression patterns of its predicted regulators. Through this approach, TIGRESS finds the minimum set of regulators able to predict the expression of the *i* target gene by scoring their potential involvement in the regression model through a likelihood approach.

Bayesian models

Multiple approaches based on Bayesian networks (Friedman et al., 2000) have been proposed. One of the most used in the inference of gene regulatory networks are graphical Gaussian models (GGM), which evaluate network structure by estimating the covariance matrix to detect conditionally dependent genes and apply multiple testing correction for a heuristic network search (Schäfer and Strimmer, 2005). Another relevant Bayesian method is the categorical Bayesian network (CBN), which learns the graph structure by a score-based procedure identifying significant differences in gene interactions between the two conditions being compared (Balov, 2013).

1.2.3. The analysis of pig transcriptomes

The FAANG consortium (Giuffra et al., 2019) is currently generating multiple sources of -omic data from different domestic species, while increasing the number of analyzed animals for gene expression data. The integration of RNA-seq transcriptomics, histone modification marks, DNA methylation patterns and chromatin spatial conformation and accessibility data is expected to provide an extensive knowledge about the genetic basis of phenotypes of economic interest in domestic species.

In pigs, many RNA-seq experiments have been performed during the past years, analyzing the landscape of RNA transcript expression in multiple porcine breeds, tissues, developmental stages and experimental conditions (Puig-Oliveras et al., 2014; Pérez-Montarelo et al., 2014; Pilcher et al., 2015; Ayuso et al., 2016; Cardoso et al., 2017a, 2017b; Horodyska et al., 2018; Ramayo-Caldas et al., 2018; Benítez et al., 2019). The majority of these studies are focused on the protein-coding mRNA fraction of the transcriptome, since it can provide a comprehensive representation of the metabolic state of the analyzed tissues in response to different stimuli. Multiple porcine tissues have been analyzed with RNA-seq technologies. For instance, Ayuso et al. (2016) investigated the mRNA expression profiles of two metabolically divergent skeletal muscles (*biceps femoris* and *longissimus dorsi*) in Iberian and Iberian × Duroc pigs. Differential expression analyses showed that the transcriptome of the *biceps femoris* muscle is enriched in pathways related with lipid metabolism and adipocyte differentiation, thus suggesting an active intramuscular fat deposition. Besides, Iberian pigs presented increased expression of glucose and lipid metabolism-related genes compared with Iberian × Duroc crosses, and enrichment analyses of the set of DE genes between both breeds revealed pathways involved in protein deposition and cellular growth (Ayuso et al., 2016). When comparing pigs in a fasting condition with others fed during variable period of times, multiple differentially expressed genes related with angiogenesis and oxidative stress, ribosomal proteins or the regulation of peripheral circadian rhythms were detected (Cardoso et al., 2017b). In a study performed by Benitez et al. (2020), the influence of diet and breed-specific genetic determinants was evaluated by RNA-seq of subcutaneous adipose tissue samples from the hams of Iberian and Duroc pigs. As expected, several genes involved in the regulation of lipid metabolism, such as leptin (*LEP*), cytosolic phosphoenolpyruvate carboxykinase (*PCK1*) or the retinoid X receptor γ (*RXRG*), were upregulated in Iberian pigs.

Besides, other upregulated genes in Duroc pigs were related with cell growth and insulin signaling such as the insulin growth factor 2 (*IGF2*), insulin (*INS*), insulin receptor (*INSR*) and insulin-induced 1 (*INSIG1*) genes (Benítez et al., 2019). These results are compatible with the existence of insulin resistance in fatty Iberian pigs, a breed which is prone to increased adiposity and obesity (Torres-Rovira et al., 2012). Moreover, the glucose transporter 4 (*GLUT4*) was upregulated in Iberian pigs, a finding that might be explained by leptin resistance compensatory mechanisms (Waller et al., 2011).

Although porcine non-coding RNAs have been much less studied than mRNAs, there is solid evidence that miRNAs have an important role in regulating fatty acids metabolism in the skeletal muscle and in adipose tissue, which highly influences the organoleptic properties of meat and its shelf life (Wood et al., 2008). Noteworthy, Li et al. (2011) were among the first authors to report putative functions for miRNAs expressed in porcine adipose tissue. According to their results, members of the ssc-miR-143, ssc-miR-103, ssc-let-7 and ssc-miR-148 families were abundantly expressed in the backfat tissue of Rongchang pigs. Moreover, several relevant lipid-related pathways such as MAPK, Wnt, TGF-β or insulin signaling were enriched in the set of predicted putative mRNA targets. The role of ssc-miR-122 in lipid metabolism was also revealed in a study using Göttingen minipigs fed with standard and high-cholesterol diets (Cirera et al., 2010). This research showed that the expression of ssc-miR-122 was reduced in pigs with high-cholesterol food intake (Cirera et al., 2010). In a similar investigation comparing lean and obese minipigs, Mentzel et al. (2015) reported ssc-miR-10b-5p, ssc-miR-143-3p, ssc-miR-26a-5p or ssc-miR-22-3p as highly expressed in subcutaneous adipose tissue, whereas ssc-miR-9-5p and ssc-miR-124a-3p were overexpressed in obese pigs compared with their lean counterparts. These two miRNAs have been associated with weight gain, insulin resistance and proinflammatory signaling (Blüher et al., 2007; Bazzoni et al., 2009; Grandjean et al., 2009). Many other recent studies have also analyzed the miRNA expression profiles of adipose tissues. For instance, the overexpression of ssc-miR-17-5p significantly reduced the transcript levels of the nuclear receptor coactivator 3 (*NCOA3*), the fatty acid binding protein 4 (*FABP4*) and the peroxisome proliferator activated receptor γ (*PPARG*) genes, which inhibit the differentiation of intramuscular pre-adipocytes (Han et al., 2017). The ssc-miR-146b, on the other hand, reduced glucose uptake in pre-adipocytes by targeting the insulin receptor substrate 1 (*IRS1*) and glucose transporter 4 (*GLUT4*) genes (Zhu et al., 2018).

In the liver, an organ of paramount importance in the integration and processing of nutrients after food intake, as well as in the maintenance of glucose homeostasis (Han et al., 2016), porcine miRNAs have also been described to play relevant regulatory roles. Mentzel et al. (2016) profiled the expression of several miRNAs and mRNAs in the liver from Göttingen minipigs, and found ssc-miR-34a and ssc-miR-1285 as highly upregulated in obese pigs fed *ad libitum*, whereas ssc-miR-181d was downregulated. Another recent study reported a total of 13 differentially expressed miRNAs in liver when contrasting pigs with low and high fat deposition, among which ssc-miR-451, ssc-miR-127 or ssc-miR34c were significantly downregulated in pigs with a high fat profile (Xing et al., 2019).

In the same study made by Mentzel et al. (2016), the authors also reported the upregulated expression of ssc-miR-215, ssc-miR-1285, ssc-miR-208b or ssc-miR-1 in the skeletal muscle of obese pigs. In an additional study comparing the miRNA expression profiles of *longissimus dorsi* skeletal muscle samples with divergent phenotypes for drip loss, Wei et al. (2018) reported ssc-miR-22-5p and ssc-miR-499 as highly expressed, and inferred potential GRNs from their putative mRNA targets. A miRNA-mediated differential muscle fiber development was described for Tongcheng and Yorkshire pig breeds, further validating, with luciferase assays, the regulation of the destrin/actin depolymerizing factor (*DSTN*) mRNA by ssc-miR-499-5p (Xi et al., 2018). Moreover, co-expression network integration of miRNA and mRNA sequencing data provided additional insights into the role of ssc-miR-499-5p as a key regulator of AMPK, mTOR and TGF-β signaling pathways (Xie et al., 2019).

In the light of these results, it is clear that non-coding RNAs, and especially, miRNAs, play important roles in regulating the expression of many metabolic pathways in multiple tissues. Nevertheless, the accumulated knowledge regarding their expression, regulation, function and variation across porcine tissues is still limited. Such phenomenon motivated our investigation of the putative roles that miRNAs might have in skeletal muscle metabolism in response to nutrient supply, as well as the analysis of their variability across different populations, in search of putative causal mutations that might be driving changes in the expression profiles of their targeted mRNAs.

1.3. MicroRNAs as key post-transcriptional regulators

1.3.1. An introduction to microRNA biology

MicroRNAs (miRNAs) are small endogenous post-transcriptional regulators of gene expression found in a wide range of eukaryotes. These non-coding RNA transcripts derive from the processing of primary miRNA spliced transcripts (pri-miRNAs) into hairpin-like precursors (pre-miRNAs), which are formed by a single fully base-paired stem with two arms containing both -5p and -3p mature miRNA sequences and an apical loop (Bartel, 2018), as depicted in Figure 7. Subsequently, these precursor transcripts are processed into functionally active mature miRNAs (~18-22 nucleotides long) in the cytoplasm.

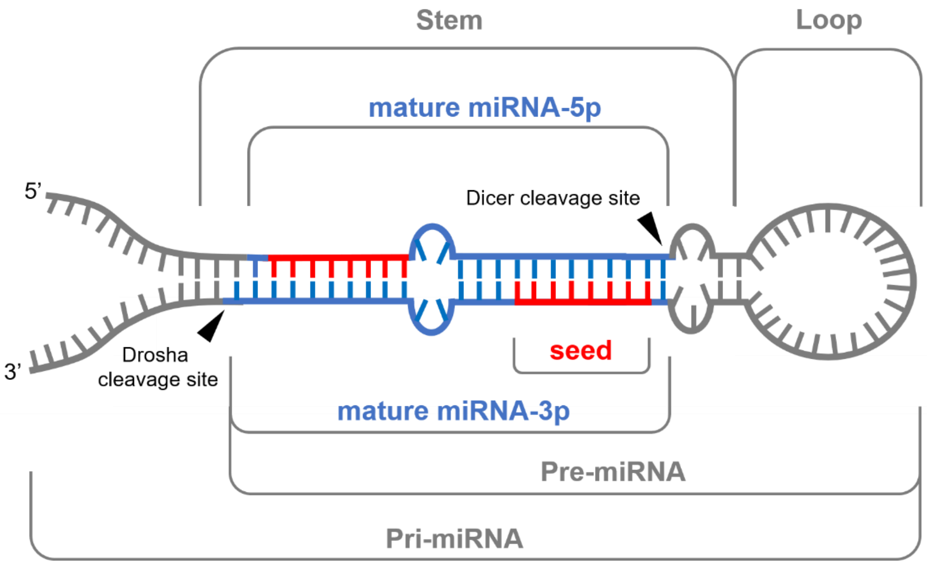


Figure 7: Secondary structure of the folded pri-miRNA containing the pre-miRNA and the -5p and -3p mature miRNAs. The pri-miRNA forms a hairpin-like structure with one apical loop and a central stem composed by two paired arms of the RNA sequence. The terminal 5' and 3' ends are unpaired. The basal junction at the end of the hairpin determines the Drosha cleavage site for generating the pre-miRNA. In the cytoplasm, Dicer recognizes the end of the central stem near the apical loop and cleaves the pre-miRNA to release the miRNA-miRNA* duplex with both -5p and -3p mature miRNAs that are loaded into the RISC complex.

Thousands of different miRNAs have been described since they were firstly reported in *C. elegans* (Lee et al., 1993). MiRNAs act as post-transcriptional regulators of targeted messenger RNAs (mRNAs) to which they bind by base complementarity between the seed region of mature miRNAs (2nd to 8th 5’ nucleotides) and short matching sequences in the 3’-UTR of targeted mRNAs (Friedman et al., 2009). It is well known that miRNAs can regulate the expression of hundreds of targeted mRNAs, thus modulating multiple biological pathways and contributing to fine-tune the expression of protein-coding transcripts. Indeed, many miRNA loss-of-function studies have reported the influence of miRNAs in several key biological processes such as development (Bhaskaran and Mohan, 2014), energy homeostasis (Dumortier et al., 2013), circadian clock regulation (Cheng et al., 2007) or lipid metabolism (Lynn, 2009). Moreover, they are also involved in the progression of numerous pathologies like obesity (Iacomino and Siani, 2017), diabetes (Feng et al., 2016), heart failure (Zhou et al., 2018) or cancer (Peng and Croce, 2016).

Early studies about the function of the lin-4 gene in *C. elegans* found that this transcript was not able to encode a protein (Lee et al., 1993). Instead, it had the ability to target the mRNA transcripts encoded by the *lin-14* gene in an antisense manner. Such finding led to the realization that lin-4 generated a short non-coding RNA transcript of ~22 nucleotides in length, and that its sequence had imperfect complementarity to conserved sites in the 3’-UTR of the targeted *lin-14* mRNA. Studies carried out by Reinhart et al. (2000) and focused on let-7, another non-coding gene, reinforced such vision about the function of these short non-coding RNAs as regulators of mRNA expression. Besides, the let-7 gene was reported to be highly conserved in a wide range of animal species (Pasquinelli et al., 2000). These initial results, along with other investigations reporting the existence of short non-coding RNA transcripts processed from hairpin-like precursors and displaying regulatory functions (Lagos-Quintana et al., 2001; Lau et al., 2001; Lee and Ambros, 2001), were essential steps to understand the role of small regulatory RNAs in the modulation of gene expression.

1.3.2. Biogenesis and function of microRNAs

Animal miRNA genes are typically transcribed by RNA Polymerase II (Pol-II) as pri-miRNAs (Lee et al., 2004). Transcription of pri-miRNAs results in non-coding RNAs harboring single miRNA sequences (Figure 8A). At some instances, multiple miRNAs are embedded in a single large polycistronic pri-miRNA, thus being transcribed simultaneously (Ameres and Zamore, 2013). Pri-miRNAs fold back on themselves to form long hairpin-like structures in the nucleus, with an apical loop, a long imperfect stem of ~33-35 nucleotides and a basal junction formed by the two 5' and 3' strands, followed by flanking single-stranded segments (Ha and Kim, 2014). This primary structure is then recognized by the Microprocessor machinery, which is a heterotrimeric complex composed by one molecule of the Drosha endonuclease and two DGCR8 (named Pasha in flies and nematodes) molecules (Nguyen et al., 2015). Drosha, which contains two RNase III domains, along with other cofactors, binds to the pri-miRNA transcripts via the recognition and positioning of the Drosha subunit in the basal region of the hairpin (Partin et al., 2017; Kwon et al., 2019). Subsequently, it cleaves the pri-miRNA, thus releasing a sliced stem-loop folded structure of ~60-80 nucleotides called pre-miRNA (Bartel, 2018). The Drosha processing mechanism mediates the canonical intergenic miRNA maturation (Figure 8B). Nevertheless, other alternative processing pathways exist. For instance, intronic miRNAs, the so-called mirtrons, are directly spliced from intronic segments by the spliceosome (Ruby et al., 2007). Other less prevalent pathways include endogenous short-hairpin RNAs (shRNAs) or chimeric miRNAs that are transcribed in tandem as part of other non-coding RNAs (Babiarz et al., 2008; Ender et al., 2008). Whatever the mechanism of synthesis might be, once generated, pre-miRNAs are transported to the cytoplasm (Figure 8C) by Exportin 5 (XPO5) and RAN-GTP (Lund et al., 2004).

In the cytoplasm, pre-miRNAs are recognized by another component of the processing machinery: Dicer (Zhang et al., 2004), a RNase-II protein that cleaves the apical loop region from the hairpin (Figure 8D), hence yielding a double-stranded short miRNA-miRNA* duplex of ~22 nucleotides (Hutvágner et al., 2001; Bartel, 2018). This miRNA-miRNA* duplex contains the mature guide miRNA coupled with its passenger strand (miRNA*), and an overhang of ~2 nucleotides in the 3' end of the sequence, previously created by the slicing action of Drosha. Once formed, the miRNA duplex is loaded into the RNA-induced silencing complex (RISC) to form the so-called miRISC complex. This complex contains an Argonaute (AGO) subunit protein that allocates the miRNA duplex with the aid of chaperon proteins (HSC70/HSP90). These cofactors induce Argonaute to adopt a high-energy open structure suitable for binding to the miRNA duplex (Iwasaki et al., 2010). Argonaute proteins contain four domains: PAZ (which binds to the 3' end of the miRNA-miRNA* duplex), Mid (which binds to the 5’-phosphate group of the miRNA), C-terminal PIWI (presumably exhibits endonucleolytic activity) and the N-terminal domain. Among the different existing paralogs of the Ago protein, only Ago2 shows the ability to elicit miRNA-induced mRNA repression (Liu et al., 2004).

After the loading of the miRNA duplex into Argonaute, the miRNA* passenger strand is removed via a structural reorganization of Argonaute into a more relaxed form and subsequently undergoes rapid exonuclease-mediated degradation (Kawamata and Tomari, 2010). Both strands of the initial duplex have the ability to act as guide or passenger strands (Griffiths-Jones et al., 2011). The choice of which strand is used as a guide miRNA depends on the thermodynamic stability of the miRNA ends. Indeed, the less stable 5’-end in the duplex is typically used as the guide strand (Khvorova et al., 2003). Besides, A- and U-residues are preferentially loaded in Ago compared with G- or C-residues at the 5’-end of the selected guide miRNA (Frank et al., 2010). Other factors can also influence the choice of the guide strand, such as sequence composition (Hu et al., 2009). When the functional miRISC complex is generated, the seed sequence of the miRNA guides the specificity of the pairing to short complementary sequences in the 3’-UTR of mRNAs (Figure 8E). The recognition by base-pair complementarity between the miRNA and the targeted mRNA often occurs through a perfect match. However, such interaction can be also mediated through subtle imperfect pairings (Chipman and Pasquinelli, 2019), which can be compensated or stabilized by supplementary pairings alongside the body of the miRNA (typically 13th-to-16th nucleotides in the mature miRNA).

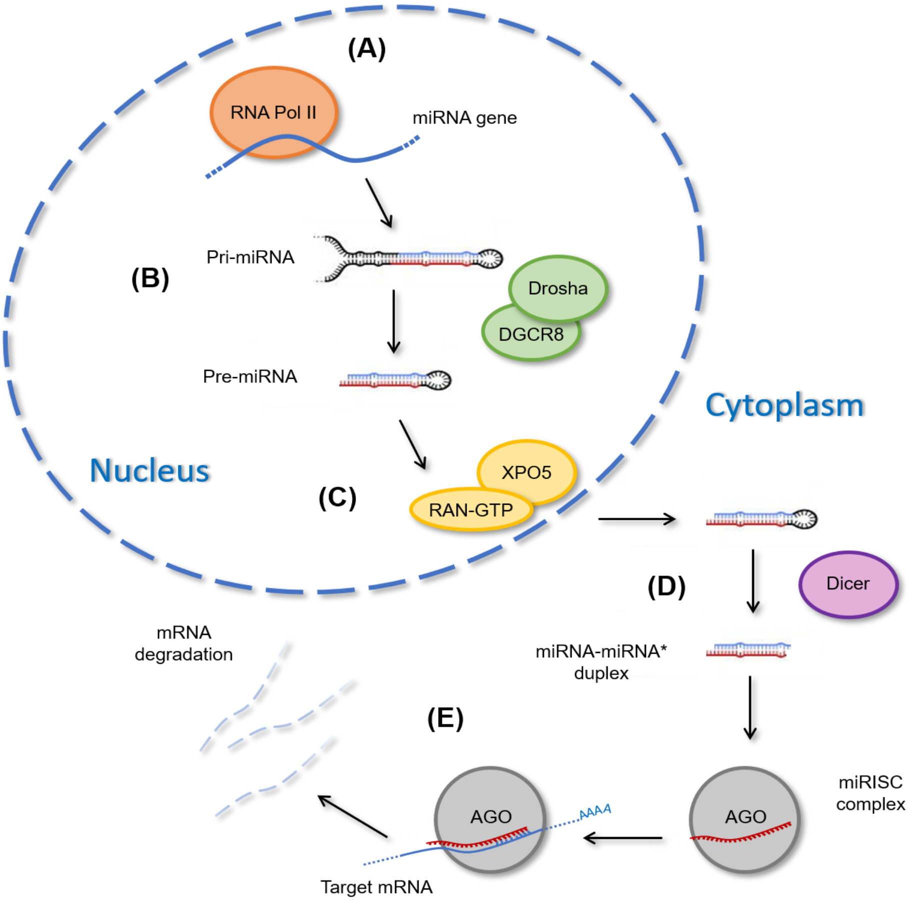


Figure 8: Biogenesis and function of miRNAs. **(A)** The miRNA gene is transcribed by RNA-Polymerase II (Pol II) into a primary miRNA transcript (pri-miRNA) which adopts a hairpin-like secondary structure. **(B)** The Microprocessor complex formed by Drosha and DGCR8 proteins bind to the pri-miRNA and cuts the edges of the stem to generate the precursor miRNA (pre-miRNA). **(C)** The pre-miRNA transcript is then transported from the nucleus to the cytoplasm by the action of Exportin 5 (XPO5) and RAN-GTP. **(D)** Once in the cytoplasm, the pre-miRNAs are processed by Dicer, which removes the apical loop to form the miRNA-miRNA* duplex. This miRNA duplex is then loaded into the Ago protein to form the miRISC complex, where the passenger miRNA* is degraded and the guide mature miRNA remains. **(E)** The active miRISC interacts with target mRNAs via base-pairing of the mRNA 3'-UTR with the mature miRNA seed (2nd to 8th 5' nucleotides) and elicits the degradation of the mRNA by shortening its poly(A) tail or, alternatively, by impeding its translation in the ribosomes.

Active target matches often occur via the pairing of 7 nucleotides corresponding to the miRNA seed (2nd to 8th 5’ nucleotides) and a complementary target site in the 3’-UTR of the mRNA (Lewis et al., 2005). Canonical pairing involves the more active 8mer matches and also 7mer-m8 matches, which can be defined as follows:

1) The 8mer interaction consists of 7 nucleotide pairings along the entire miRNA seed and the target 3’-UTR binding site, plus an additional interaction between the adenine (A) in the 3’ end of the mRNA which binds to a pocket inside the Argonaute complex, thus contributing to stabilize mRNA positioning (Schirle et al., 2015).

2) When nucleotides other than adenine are placed at position 1 of the target mRNA, the 7-nucleotide matching site is called 7mer-m8.

Alternatively, the 7mer interaction can take place when the miRNA seed pairing encompasses 5’ nucleotides 2st to 7th of the mature miRNA, provided the presence of an A nucleotide in the 3’ end of matching sequence from the target mRNA, analogously to the 8mer sites. This intermediate 7mer pairing is called 7mer-A1 (Bartel, 2018). Other additional and less common non-canonical pairings can also occur, like 6mer, 6mer-offset and 3’-compensatory pairings or CDS interactions (Chipman and Pasquinelli, 2019). A schematic representation of some canonical and non-canonical target sites is depicted in Figure 9.

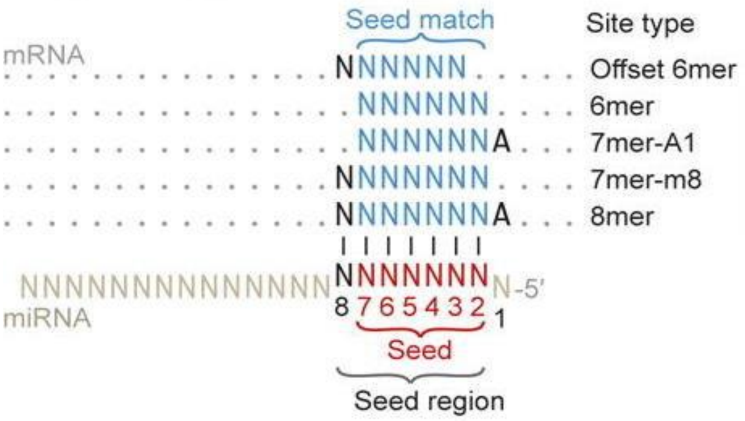


Figure 9: Canonical and non-canonical target sites between the miRNA seed region and short matching sequences in the 3’-UTR of mRNAs. Modified from Bartel (2018).

The miRNA-induced repression of targeted mRNAs occurs via the degradation of targeted mRNAs by poly(A) deadenylation-mediated decay and also by impeding their translation in the ribosomes (Jonas and Izaurralde, 2015). This repression mechanism requires the recruitment of the protein adaptor TNRC6 by Argonaute, which interacts with the poly(A)-binding protein PABC, in conjunction with deadenylases PAN2-PAN3 and the CCR4-NOT complex (Jonas and Izaurralde, 2015). The recruited deadenylases promote the shortening of the poly(A) tail of the targeted mRNA, triggering the destabilization of the mRNA through 5'-to-3' decay and decapping (Chen and Shyu, 2011; Jonas and Izaurralde, 2015). Besides, translation inhibition is also induced by TNRC6 and CCR4-NOT through the recruitment of the DDX6 helicase, a cofactor with inhibitory effects on translation (Ozgur et al., 2015).

Other than the widely reported post-transcriptional regulation of targeted mRNAs by the action of miRNAs, several alternative repressive mechanisms have been elucidated for miRNA transcripts. Eiring et al. (2010) described the interfering of miR-338 with RNA-binding proteins in the nucleus in a decoy manner, independently of its interaction with targeted mRNAs through the seed. Other authors identified miRNAs acting at the transcriptional level to inhibit the expression of targeted mRNAs by binding to the promoter regions of the corresponding genes, thus suggesting the existence of a miRNA-directed transcriptional gene-silencing effect (Kim et al., 2008a). Moreover, Vasudevan et al. (2007) proposed the upregulation of mRNA translation in the ribosomes by direct intervention of miRNAs recruiting AU-rich sequence motifs (AREs) and sets of associated proteins, a mechanism that was further supported by studies on cell cycle activation (Truesdell et al., 2012). Nevertheless, despite some studies describing non-canonical mechanisms of gene regulation mediated by miRNAs, they still remain poorly characterized and further research is needed to better document these non-canonical functions of miRNAs.

1.3.3. MicroRNA sequence motifs and isoforms

Given the extensive variety of RNA hairpins that can arise from genome transcription, the precise recognition of miRNA hairpins primarily relies on the correct orientation of Drosha for the pri-miRNA processing. Such event is fundamental for discriminating between true pri-miRNA sequences and other hairpin-like structures. Several studies have identified different processing motifs (Auyeung et al., 2013; Fang and Bartel, 2015; Roden et al., 2017) that allow the recognition of pri-miRNA hairpins by Drosha (Nguyen et al., 2015). The basal upstream UG motif is found in ~40% of studied miRNAs and helps recruiting Drosha to the basal junction of the pri-miRNA (Auyeung et al., 2013). Located at the start of the apical loop, the UGU motif interacts with the DGCR8 dimer of the Microprocessor, and indirectly prevents Drosha to bind the apical loop, facilitating its positioning at the basal junction (Auyeung et al., 2013; Nguyen et al., 2015). The downstream CNNC motif is one of the most commonly found sequences across all miRNA genes, with ~50-60% of annotated miRNAs presenting such motif (Auyeung et al., 2013). The CNNC sequence is recognized by the RNA-binding protein SRSF3, a molecule that promotes Microprocessor activity and hence the miRNA maturation process (Kim et al., 2018). The binding affinity of SRSF3 to the CNNC motif is influenced by the structural conformation of the pri-miRNA hairpin (Fernandez et al., 2017). Other additional motifs have been proposed, like basal GHG mismatches (Fang and Bartel, 2015), as well as bulge-depleted regions and stem length preference (Roden et al., 2017).

Additionally, the existence of several alternative processing pathways and post-transcriptional modifications has favored the emergence of multiple miRNA isoforms (isomiRs) derived from the same expressed miRNA gene. Inaccurate cleavage by Drosha or Dicer can generate variable 5’ and/or 3’ ends and different nucleotide overhangs within the miRNA duplex, which can lead to the generation of alternative seeds and guide strand switching in the miRISC complex (Neilsen et al., 2012). Other alternative mechanisms have also been reported, such as adenosine-to-inosine (A-to-I) RNA-editing (Kume et al., 2014) and 3’ shortening of the mature miRNA due to exonuclease activity (Kim et al., 2016). The existence of a broad repertoire of miRNA isoforms with individual specificities contributes to increase the array of mRNAs that can be targeted by a given miRNA without the need of novel miRNA gain or the fixation of polymorphic sites in miRNA regions.

1.3.4. Phylogenetic conservation

Early after the identification of the first miRNA genes (Lau et al., 2001; Lee and Ambros, 2001), the existence of orthologous miRNA sequences that were highly conserved across different species became obvious (Tanzer and Stadler, 2004; Peterson et al., 2009). The relatively low divergence of miRNAs across species led to their utilization as useful molecular markers for phylogenetic and evolutionary studies (Wheeler et al., 2009; Kenny et al., 2015). In this regard, it is worth mentioning the let-7 miRNA, which is extensively conserved across many different species (Pasquinelli et al., 2000). Studies of let-7 function demonstrated its key role in cell development and transition between larval and adult stages in *C. elegans* (Reinhart et al., 2000), as well as in the regulation of metamorphosis in *D. melanogaster* (Sempere et al., 2003). In vertebrates, let-7 has also been identified as one of the main regulators of embryonic development in zebrafish (Chen et al., 2005), as well as of numerous tissue-specific signaling pathways during cell division and differentiation (Schulman et al., 2005; Watanabe et al., 2005). Another relevant case is exemplified by miRNA 1 (miR-1), one of the most important and highly expressed miRNAs in the muscle tissue and intimately related with cardiac and skeletal muscle development (Chen et al., 2006). Both in mammals and in *D. melanogaster,* the expression of this miRNA is regulated by the myocyte enhancer factor 2 (Mef-2) and myoblast determination protein 1 (MYOD1), which suggests that the regulation of miR-1 expression and function is evolutionarily conserved in vertebrates and insects (Kalsotra et al., 2014).

More recently, the evolutionary dynamics of miRNAs in domestic animals was surveyed by Penso-Dolfin et. al (2018). These authors thoroughly analyzed, in cow, dog, horse, pig and rabbit, the patterns of gains and losses in miRNA loci and their corresponding predicted target sites. Moreover, they described duplication events as a relevant mechanism for miRNA evolution and they also reported that young emerging miRNA families have a more restricted tissue expression profile than other more conserved old miRNA families (Penso-Dolfin et al., 2018). Besides, 3'-UTR sites targeted by conserved miRNA families across multiple species also showed reduced variation rates. Other species-specific miRNA target sites, conversely, did not evolve under such strong evolutionary constraints (Penso-Dolfin et al., 2018). Another recent study by Simkin et al. (2020) further described three differentiated groups of miRNAs in terms of gain and loss events in their predicted 3'-UTR target sites (Simkin et al., 2020).

While ancient and widely conserved miRNAs, like let-7, presented signatures of strong purifying selection in their 3'-UTR target sites, other less conserved miRNAs did not show evidence of gaining or loosing target sites. Besides, a reduced number of miRNAs like miR-146 evidenced a rapid turnover, gaining and loosing target sites with neutral, or even faster than neutral rates (Simkin et al., 2020).

1.3.5. MicroRNA annotation

The annotation of miRNAs is a fundamental step towards elucidating their crucial roles in regulating and fine-tuning of many relevant biological pathways. Regarding Metazoa, the amount of knowledge about miRNA identity, location, structure and function has been accumulating during the past two decades at a fast pace, and in well-studied model organisms such as humans or mice, comprehensive catalogues of annotated miRNAs have been built. The majority of studies focused on miRNA biology have used these two organisms as experimental models, while other species have been much less studied. This phenomenon is particularly obvious when analyzing the data set stored in one of the most accessed and cited miRNA databases, miRBase (Kozomara et al., 2019, https://www.mirbase.org). This database has become one of the reference sources of scientific information regarding miRNA annotation across species. In its last release (v.22, March 2018), miRBase hosted miRNA sequences from 271 different organisms, with a total of 38,589 hairpins representing pre-miRNA sequences, which encode a total of 48,860 mature miRNAs. Despite these impressive numbers, some species such as human or mice are predominantly represented in the database, while other relevant organisms are missing or poorly annotated. In Figure 10 we can observe that several livestock species such as pig, goats or sheep have reduced numbers of annotated miRNAs. In contrast, cow and chicken rank at third and fourth positions (immediately after human and mice), respectively.

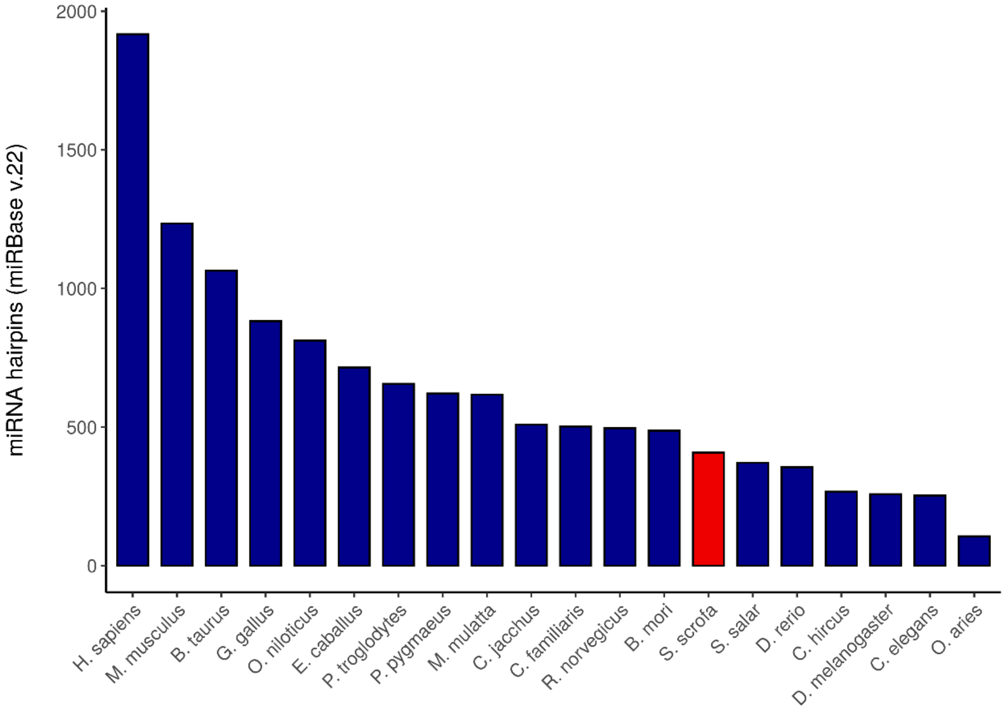


Figure 10: Number of miRNA hairpins annotated in the miRBase database v.22 in humans, primates, model organisms and domestic species such as cow (*B. Taurus*), chicken (*G. gallus*), pig (*S. scrofa*), goat (*C. hircus*) or sheep (*O. aries*). Highlighted in red is the number of annotated miRNA hairpins (408) in pigs.

This high heterogeneity across species, not only in the numbers of annotated miRNAs but also in the quality of the annotation, is a problem of high concern for scientists working in the miRNA field. Indeed, miRNA prediction algorithms obtain worse performance metrics when using miRNA sequences from miRBase, compared with other more curated databases (Saçar et al., 2013). These problems have motivated the construction of other alternative and more curated databases with stricter annotation rules and, accordingly, a much more reduced set of available miRNA sequences and organisms (Backes et al., 2018; Fromm et al., 2019). Moreover, to overcome the high heterogeneity and redundancy in miRNA annotation, several authors have proposed homogeneous and well-structured criteria for naming novel and orthologous miRNA sequences (Ambros et al., 2003; Budak et al., 2015; Fromm et al., 2015; Desvignes et al., 2019).

Only 388 miRNA loci are available in the last annotation release of the pig assembly (Sscrofa11.1) in the Ensembl database (https://www.esnembl.org). From these, 370 miRNA genes are in chromosomal locations, whereas 18 are located in scaffolds. The number of annotated miRNA genes in the Ensembl database closely resembles that obtained when screening porcine miRNAs in the miRBase database (Kozomara et al., 2019, https://www.mirbase.org), where, as previously shown in Figure 10, 408 porcine miRNAs are available. This subtle difference might be due to the presence of doubtful or poorly annotated porcine miRNAs in miRBase, a caveat that has been also reported for other species (Saçar et al., 2013).

The first set of porcine miRNAs were reported by Wernersson et al. (2005). At the same time, Sawera et al. (2005) also used a homology-based search to describe the miRNA-17-92 cluster and the expression profile of several miRNAs mapping to this cluster across different tissues. Subsequent surveys further described additional sets of porcine miRNAs using homology-based search (Kim et al., 2006), as well as cDNA cloning and sequencing techniques (Kim et al., 2008b; McDaneld et al., 2009; Cho et al., 2010; Cirera et al., 2010; Xie et al., 2011). Microarray hybridization profiling (Huang et al., 2008; Podolska et al., 2011) and *de novo* assembly and sequencing (McDaneld et al., 2012) were also used. Following these initial studies, and shortly after the release of the first porcine genome assembly by Groenen et al. (2012), researchers started to use RNA-seq technologies to characterize the functions, interactions and expression profiles of non-coding RNAs in different porcine tissues, developmental stages and experimental conditions. Nevertheless, our understanding about the extent of the porcine miRNA repertoire is still limited, as evidenced by the reduced amount of annotated miRNA loci in the porcine genome. Hence, there is still room for expanding the set of porcine miRNAs by means of state-of-the-art prediction techniques using homology-based search, as well as sequencing of miRNA expression profiles in porcine tissues.

1.3.6. MicroRNA gene prediction

Early approaches to detect miRNAs

Initial methods for miRNA discovery and annotation relied on laborious low-throughput procedures that required the isolation, cloning (Bentwich et al., 2005) and *in situ* hybridization of the miRNA transcripts (Nelson et al., 2006), followed by Sanger sequencing of the detected RNA molecules. This molecular research was backed up by computational methods such as comparative alignment scanning for other overlapping non-miRNA sequences, homology-based comparison among species or the *in silico* prediction of miRNA secondary structures based on the search of hairpin-like foldings by means of RNA-folding algorithms like UNAFold (Markham and Zuker, 2008) or RNAfold (Lorenz et al., 2011).

Following these early efforts, researchers focused their attention on the high evolutionary conservation of miRNA genes with the aim of using such feature to detect miRNA genes in poorly annotated species. In this regard, several methods using homology-based search approaches were published, like miRscan (Lim et al., 2003), miRSeeker (Lai et al., 2003), RNAmicro (Hertel and Stadler, 2006) or miROrtho (Gerlach et al., 2009). However, these approaches heavily relied on the 3′ and 5′ stem conservation of the candidate miRNAs (Lim et al., 2003), as well as in shared nucleotide patterns between the reference set of annotated miRNAs to which they are compared (Lai et al., 2003). Hence, it soon became evident that homology-based methods were only applicable to the detection of miRNAs exceptionally well-conserved across species, thus failing to detect species-specific miRNA loci or miRNAs with some degree of divergence among species.

Prediction of miRNAs from small RNA-seq data

Following the development of NGS technologies, specific protocols were also designed for the targeted sequencing of small RNAs (sRNA-seq), as canonical RNA-seq methods for mRNA capture tended to misrepresent the small RNA fraction. One of the first published papers using NGS technologies targeting miRNAs was authored by Bar et al. (2008). In this work, the authors sequenced the small RNA fraction of human embryonic stem cells, and profiled the expression of 1) 191 human miRNAs annotated in previous studies, 2) 56 miRNAs not previously annotated in humans but with conserved orthologs in other species, and 3) 13 novel miRNA candidates without orthologs in other species. For miRNA prediction, Bar et al. (2008) established a series of rule-based criteria for miRNA characterization: 1) lack of overlap with other coding and non-coding annotated elements, 2) perfect alignment towards the reference genome, 3) sufficient read alignment to mature miRNAs within the hairpin, 4) minimized free energy of the folding compared with other non-miRNA sequences (Bonnet et al., 2004) and 5) shared 5’ end among clustered reads (derived from the high conservation of the miRNA seed region).

These studies were essential to implement *in silico* tools for predicting miRNAs from sRNA-seq experimental data. The miRDeep algorithm (Friedländer et al., 2008) was one of the first end-to-end approaches incorporating rule-based methods for reconstructing miRNA hairpin candidates from sequencing data. Friedländer and collaborators designed a rule-based algorithm able to integrate sequence data from high-throughput analyses and extract miRNA hairpin candidates based on a combination of parameters, alike to those mentioned before, which were summarized in a score quantifying the probability of the hairpin to be a true miRNA. After this groundbreaking work, similar methods were developed, like miReap (https://sourceforge.net/projects/mireap/), MIReNA (Mathelier and Carbone, 2010), miRSeqNovel (Qian et al., 2012), sRNAbench (Barturen et al., 2014), miRdentify (Hansen et al., 2014) or miRNAFold (Tav et al., 2016). Updated versions of these tools, such as miRDeep2 (Friedländer et al., 2012), miRDeep* (An et al., 2013) and sRNAtoolbox (Rueda et al., 2015; Aparicio-Puerta et al., 2019) were also developed.

Besides, other comprehensive methods for sRNA-seq analysis also incorporated miRNA prediction tools in their pipelines, with miRDeep (Friedländer et al., 2008), sRNAbench (Barturen et al., 2014) and miReap being the most used ones. Additional examples of these general-purpose pipelines are wapRNA (Zhao et al., 2011), omiRas (Müller et al., 2013), miRTools2.0 (Wu et al., 2013), iMiR (Giurato et al., 2013), eRNA (Yuan et al., 2014), Cap-miRSeq (Sun et al., 2014), miARma-Seq (Andrés-León et al., 2016), the UEA sRNA workbench (Beckers et al., 2017) or the sRNAtoolbox (Aparicio-Puerta et al., 2019).

Rule-based methods vs machine learning approaches

In recent years, machine learning (ML) approaches have been implemented for miRNA prediction, detecting and discriminating miRNA hairpin structures from other types of non-coding RNAs by using statistical learning models instead of deterministic rules. Nevertheless, both ML and rule-based methods can benefit from sRNA-seq data to uncover novel miRNA transcripts (Figure 11). Different tools have addressed the problem of correctly classifying miRNAs by training several ML algorithms: Hidden Markov models (HMM), support vector machine (SVM), random forest (RF), naïve Bayes (NB) or neural networks (NN), among others, all of them with inherent strengths and caveats (Stegmayer et al., 2018).

One of the first examples of ML applied to miRNA prediction was proMiR (Nam et al., 2005) and its updated version proMiR II (Nam et al., 2006). These methods implemented an HMM approach with probabilistic co-learning based on conserved sequences, their secondary structures and other characteristic features like the G/C ratio, conservation score or the free energy of the candidate sequences. Other tools making use of HMM are, to mention a few, miRRim (Terai et al., 2007), HHMMiR (Kadri et al., 2009) and SSCprofiler (Oulas et al., 2009). With regard to SVM models, there are numerous methods implementing this type of learning approach, as early studies on miRNA prediction showed promising results in term of its accuracy and performance. The first SVM-based tool was reported by Xue et al. (2005), who used a specific triplet-encoding of the miRNA sequence to obtain a vector representation of the nucleotides in the hairpin. Methods developed subsequently were also based on different feature calculations, parameters and kernels from SVM algorithms. Examples of these tools are, for instance, miRAbela (Sewer et al., 2005), RNAmicro (Hertel and Stadler, 2006), miRFinder (Huang et al., 2007), miRCoS (Sheng et al., 2007), miROrtho (Gerlach et al., 2009), microPred (Batuwita and Palade, 2009), micro-ProcessorSVM (Helvik et al., 2007), miRenSVM (Ding et al., 2010), miRPara (Wu et al., 2011), BosFinder (Sadeghi et al., 2014), YamiPred (Kleftogiannis et al., 2015), miRNA-dis (Liu et al., 2015), miRBoost (Tran et al., 2015), iMiRNA-SSF (Chen et al., 2016a) or miRge 2.0 (Lu et al., 2018). Tools based on RF algorithms are miPred (Jiang et al., 2007), miRanalyzer (Hackenberg et al., 2009), miReader (Jha and Shankar, 2013), HuntMi (Gudyś et al., 2013), miRClassify (Zou et al., 2014) and Mirnovo (Vitsios et al., 2017). The BayesmiRNAfind (Yousef et al., 2006) and MatureBayes (Gkirtzou et al., 2010) tools, in contrast, use a classification method based on NB algorithm. Examples of the use of NN structures are miRANN (Rahman et al., 2012), a method using back-propagation on neural networks (Jiang et al., 2016), DP-miRNA (Thomas et al., 2017) or DeepMir (Tang and Sun, 2019). Other tools aimed to detect miRNAs have applied ensemble methods combining several ML algorithms, like izMiR (Saçar et al., 2017) or a recent work by Saçar et al. (2019) using 3D representations of RNA secondary structures.

These methods generally include a first step of selecting a positive set of hairpin sequences to contrast with a negative set of other non-coding RNAs or pseudo-miRNA sequences and subsequent extraction of a set of representative miRNA features. Finally, the ML algorithm is trained based on these features in order to build a classifier for miRNA prediction. In Table 3, the extensive range of positive and negative data sets used for ML models training is displayed.

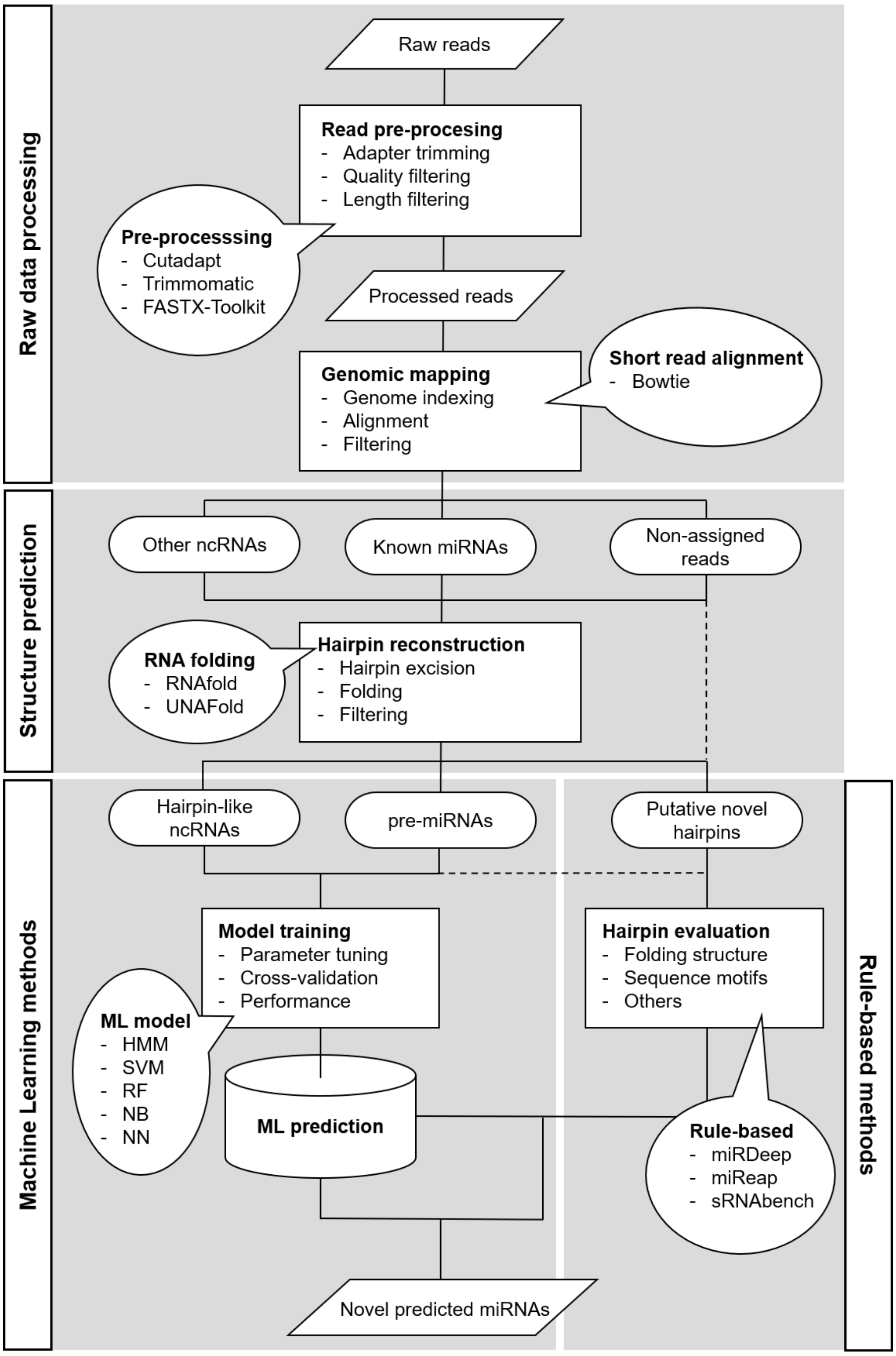


Figure 11: Schematic representation of the miRNA prediction workflow from sRNA-seq data. After pre-processing and filtering, reads are mapped and putative hairpin structures are reconstructed. Known pre-miRNAs and other hairpin-like non-coding RNAs can be used for training machine learning-based classifiers. Putative novel hairpins are then embedded into rule-based methods or as input for trained classifiers. Both approaches allow to obtain a list of novel predicted miRNA candidates.

Table 3: List of machine learning algorithms for microRNA gene prediction and the positive and negative data sets used for model training.

Model[a]	Tool	Positive data	Negative data	Reference
HMM	ProMiR	*H. sapiens* pre-miRNA hairpins	Random pseudo-hairpins from *H. sapiens*	(Nam et al., 2005)
	miRRim	*H. sapiens* conserved miRNAs +/- 50 bp	Randomly chosen conserved, moderately conserved and non-conserved regions in H. sapiens	(Terai et al., 2007)
	HHMMiR	*H. sapiens* pre-miRNA hairpins from miRBase v10	CDS and random regions from *H. sapiens*	(Kadri et al., 2009)
	SSCprofiler	*H. sapiens* pre-miRNA hairpins from miRBase v12	3'-UTRs from *H. sapiens*	(Oulas et al., 2009)
SVM	Triplet-SVM	*H. sapiens* pre-miRNA hairpins from miRBase v5	Random pseudo hairpins from *H. sapiens*	(Xue et al., 2005)
	mirAbela	MiRNA genes from Rfam	Random sequences isolated from tRNA, rRNA and mRNA genes	(Sewer et al., 2005)
	RNAmicro	Metazoan miRNAs from miRBase v6	Random shuffled sequences from positive data and tRNAs	(Hertel and Stadler, 2006)
	miRFinder	Pre-miRNA hairpins in human, mouse, pig, cattle, dog and sheep from miRBase v8.2	Random sequences from *H. sapiens* and *M. musculus*	(Huang et al., 2007)
	miRCoS	*M. musculus* pre-miRNA hairpins from miRBase v9.1	Random hairpins that did not pass filterings for positive data set	(Sheng et al., 2007)
	miROrtho	Metazoan miRNAs from miRBase	Randomly chosen hairpins from non-miRNA genes	(Gerlach et al., 2009)
	microPred	*H. sapiens* pre-miRNA hairpins from miRBase	Non-redundant *H. sapiens* pseudo hairpins from RefSeq	(Batuwita and Palade, 2009)
	micro-ProcessorSVM	*H. sapiens* pre-miRNA hairpins from miRBase v8	*H. sapiens* non-coding RNAs from Ensembl v37	(Helvik et al., 2007)
	miRenSVM	*H. sapiens* and *A. gambiae* pre-miRNA hairpins from miRBase v12.2	*H. sapiens* and *A. gambiae* 3'-UTRs and other non-coding sequences from 3'-UTRdb v22 and Rfam9.1	(Ding et al., 2010)
	miRPara	Metazoan miRNAs from miRBase v13	Sequences identical to positive data with mature miRNAs shifted to random starts	(Wu et al., 2011)
	BosFinder	*B. taurus* pre-miRNA hairpins from miRBase v20	Non-redundant *B. taurus* pseudo hairpins and other non-coding sequences from RefSeq	(Sadeghi et al., 2014)
	YamiPred	*H. sapiens* pre-miRNA hairpins from miRBase	Random pseudo hairpins from *H. sapiens*	(Kleftogiannis et al., 2015)
	miRNA-dis	Metazoan miRNAs from miRBase v20	Random pseudo hairpins from *H. sapiens*	(Liu et al., 2015)
	miRBoost	Pre-miRNA hairpins from eukaryotic genomes with at least 100 annotated miRNAs in miRBase v18	Exonic regions and other non-coding sequences from fRNAdb, NONCODE and sonRNA-LBME-db	(Tran et al., 2015)
	iMiRNA-SSF	Non-redundant *H. sapiens* pre-miRNA hairpins	Random pseudo hairpins from *H. sapiens*	(Chen et al., 2016a)

	miRge 2.0	*H. sapiens* and *M. musculus* pre-miRNA hairpins expressed in several tissues from miRGeneDB	Other expressed transcripts mapping to tRNA, snoRNA, rRNA or mRNA loci	(Lu et al., 2018)
RF	miPred	*H. sapiens* pre-miRNA hairpins from miRBase v8.2	Random pseudo hairpins from *H. sapiens*	(Jiang et al., 2007)
	miRanalyzer	Pre-miRNA hairpins in *H. sapiens*, *C. elegans* and *R. norvegicus* from miRBase v12	Random hairpins from non-miRNA genes	(Hackenberg et al., 2009)
	miReader	Read sequences from sRNA-seq data set mapping to miRNA loci	Read sequences from sRNA-seq data set mapping to non-miRNA loci	(Jha and Shankar, 2013)
	HuntMi	Pre-miRNAs from miRBase v17	Non-coding RNAs and mRNAs from 10 animal and 7 plant species, as well as 29 viruses	(Gudyś et al., 2013)
	miRClassify	Pre-miRNAs from miRBase v19	Random pseudo hairpins	(Zou et al., 2014)
	Mirnovo	Read sequences from sRNA-seq data set mapping to miRNA loci	Read sequences from sRNA-seq data set mapping to non-miRNA loci	(Vitsios et al., 2017)
NB	BayesmiRNAfind	Pre-miRNA hairpins in *C. elegans* and *M. musculus* genomes	Other non-coding sequences in *C. elegans* and *M. musculus* from UCSC	(Yousef et al., 2006)
	MatureBayes	Experimentally verified pre-miRNA hairpins in *H. sapiens* and *M. musculus* from miRBase v10	Random shuffled sequences from positive data and tRNAs	(Gkirtzou et al., 2010)
NN	miRANN	*H. sapiens* pre-miRNA hairpins from miRBase v18	Random pseudo hairpins from CDS regions in *H. sapiens* assembly	(Rahman et al., 2012)
BP	BP-miRNA	Pre-miRNA hairpins from *A. lyrata*	Other non-miRNA hairpins from *A. lyrata*	(Jiang et al., 2016)
DL	DP-miRNA	*H. sapiens* pre-miRNA hairpins from miRBase v18	Random pseudo hairpins from *H. sapiens*	(Thomas et al., 2017)
	DeepMir	Non-redundant pre-miRNA hairpins from Rfam		(Tang and Sun, 2019)
Ensemble	izMiR	Pre-miRNA hairpin data sets in human, mouse, chicken and zebrafish from miRGeneDB and miRBase	Random pseudo hairpins and CDS sequences	(Saçar et al., 2017)
TG	miRNAss	Pre-miRNA hairpins in *H. sapiens* and *A. thaliana*	Random pseudo hairpins extracted from genome assemblies	(Yones et al., 2018)
SOM	DeepSOM	Pre-miRNA hairpins from miRBase v17		(Stegmayer et al., 2017)

[a]HMM: Hidden Markov model; SVM: Support vector machine; RF: Random forest; NB: Naïve Bayes; NN: Neural networks; BP: Back-propagation SVM; DL: Deep learning; Ensemble: Ensemble from various ML models; TG: Transductive graphs; SOM: Self-organizing maps.

Despite the plethora of available tools, the majority of domestic species, still lack a complete and reliable set of annotated miRNAs in their genomes, as shown in Figure 10. This circumstance narrows the possibilities of gene expression profiling and carrying out functional studies for determining miRNA-mRNA interactions, hence limiting the outcome of experiments aiming at disentangling the miRNA-dependent regulation of biological pathways. Unfortunately, many of the tools developed for miRNA prediction are just focused on reference species such as humans or mice (Ding et al., 2010; Chen et al., 2016), hindering the training of species-specific or updated prediction models. Indeed, the prediction of miRNA genes has generally emerged as a species-dependent problem, and better results are often obtained when using species-specific training data sets (Lopes et al., 2016). Another pitfall is that a number of miRNA prediction tools are based on web servers and thus they are limited to test a limited number of sequences at a time (Liu et al., 2015; Tran et al., 2015; Chen et al., 2016a; Tav et al., 2016). Other miRNA prediction packages have become deprecated, being no longer available for downloading or online testing (Ding et al., 2010; Zou et al., 2014) or, even worse, they are not available for usage (Jiang et al., 2016; Saçar, 2019). Awareness of these limitations led to the development of other alternative methods taking advantage of unlabeled sequences in unsupervised schemes. Clustering techniques like graph-based transductive methods (Yones et al., 2018) or self-organizing maps (SOM), as implemented in deepSOM (Stegmayer et al., 2017) are examples of this unsupervised approach.

1.3.7. Prediction of mRNAs targeted by microRNAs

The role of miRNAs as post-transcriptional regulators has motivated the optimization of tools to predict and identify mRNAs targeted by miRNAs. Although several types of miRNA binding sites have been reported (Bartel, 2018), only canonical targets involving 7mer/8mer interactions seem to elicit observable and reproducible effects at the experimental level (Agarwal et al., 2015). Other alternative non-canonical targets might appear at marginal levels or correspond to very specific miRNA-mRNA interactions (Agarwal et al., 2018). To achieve a reliable prediction of miRNA targets, a useful approach is to directly search for sequence complementarity between the miRNA seed and short pairing regions within the 3’-UTR of mRNA transcripts. Moreover, restricting this simple sequence search to sites that are conserved among species might confer an additional strenght to the successful identification of the target sites. Indeed, conserved miRNA-mRNA interactions have a good functional correspondence with knockdown experiments involving miRNA loss, while inconsistent results have been obtained in the case of less conserved target sites (Baek et al., 2008).

To overcome these limitations, numerous tools and pipelines for miRNA target prediction have been developed, each applying a mixture of different approaches to accurately predict miRNA-mRNA interactions. In Table 4, the most commonly used miRNA target prediction tools are presented. One of the first released tools for miRNA target prediction was miRanda (John et al., 2004), a web-server software (microRNA.org) that integrates sequence complementarity and free energy of the miRNA-mRNA duplex to identify miRNA binding sites. Similarly, RNAhybrid (Krüger and Rehmsmeier, 2006) incorporates the free energy of the duplex. Other methods are based on the pattern-based recognition between miRNAs and 3’-UTRs. Examples of this approach are rna22 (Miranda et al., 2006) or PITA (Kertesz et al., 2007).

Probably, the most cited, used and accessed tool for miRNA target prediction is the TargetScan software (Agarwal et al., 2015, 2018). This software is lodged in a web-based site (http://www.targetscan.org/vert_72/) and predicts miRNA target sites according to their conservation across species and the type of interaction. Besides, several additional variables, such as the free energy of the miRNA-mRNA duplex, 3’ compensatory pairing, and the local AU context or the pairing position within the 3’-UTR, are used for improving the accurary of the prediction. All these parameters are then summarized for ranking the most probable targets. Other tools for miRNA target prediction have incorporated ML algorithms to their pipelines. Examples of ML-based approaches for miRNA target prediction are miRDB (Wang, 2008; Liu and Wang, 2019), which uses a SVM model trained on the basis of commonly used sequence-based and experimental features from functionally confirmed miRNA-mRNA interactions, or the DIANA-microT-CDS tool (Maragkakis et al., 2009; Paraskevopoulou et al., 2013), which incorporates experimental features based on photoactivatable-ribonucleoside-enhanced cross-linking immunoprecipitation (PAR-CLIP) data. More recently, deep-learning methods based on neural networks schemes have also been applied for this task, like the miRAW tool (Pla et al., 2018), as well as others SOM-based tools like MiRNATIP (Fiannaca et al., 2016).

Table 4: Relevant bioinformatic tools for microRNA target prediction.

Tool	Type	Organism[a]	URL	Reference
miRanda	database/web server/software	h, m, r, d, c	http://www.microrna.org/microrna/home.do	(John et al., 2004)
RNAhybrid	software		https://directory.fsf.org/wiki/RNAhybrid	(Krüger and Rehmsmeier, 2006)
rna22	database/web server	h, m, d, c	https://cm.jefferson.edu/rna22/Interactive/	(Miranda et al., 2006)
PITA	database/web server/software	h, m, d, c	https://genie.weizmann.ac.il/pubs/mir07/index.html	(Kertesz et al., 2007)
TargetScan	database/web server/software	h, m, r, p, b, f, g, x, o, ma, z	http://www.targetscan.org/vert_72/	(Agarwal et al., 2015, 2018)
miRDB	database/web server	h, m, r, f, g	http://mirdb.org/	(Wang, 2008; Liu and Wang, 2019)
DIANA-microT-CDS	database/web server/software	h	http://diana.imis.athena-innovation.gr/DianaTools/	(Maragkakis et al., 2009)
miRAW	software		https://bitbucket.org/bipous/workspace/projects/MIRAW	(Pla et al., 2018)
MiRNATIP	database	h, c	http://tblab.pa.icar.cnr.it/public/miRNATIP/1.0/	(Fiannaca et al., 2016)

[a] Acronyms for detailed organisms are: h (*H. sapiens*), m (*M. musculus*), r, (*R. norvegicus*), d (*D. melanogaster*), c (*C. elegans*), p (*P. troglodytes*), b (*B. taurus*), f (*C. familiaris*), g (*G. gallus*), x (*X. laevis*), o (*M. domestica*), ma (*M. mulatta*) and z (*D. rerio*).

CHAPTER II. OBJECTIVES

This Ph.D. thesis was carried out with data generated under the framework of projects “Study of traits related with pigs lipid metabolism and pork quality by means of integral analyses of high density genotyping and gene expression data” (grant number: AGL2010-22208-C02-02) and “Genomic physiology of intramuscular fat storage in pigs” (grant number: AGL2013-48742-C2-1-R). In the AGL2010-22208-C02-02 project, a number of QTL for meat quality traits were identified in a Duroc population through a GWAS approach and the genomes of the five founders of this population were sequenced. The goals of the thesis linked with this project were:

1. To identify candidate polymorphisms in the sequenced QTL regions and exploring their association with porcine meat quality traits.
2. To detect polymorphisms with potential deleterious effects in the five sequenced genomes and determining their association with viability and production traits.

In the AGL2013-48742-C2-1-R project, we sequenced the skeletal muscle transcriptomes of fasted and fed sows and we investigated the differential expression of mRNA genes in these two physiological stages. In this thesis, we have put a special emphasis on investigating the potential role of microRNAs in the determinism of meat quality traits through several complementary approaches. The specific goals of the thesis linked to project AGL2013-48742-C2-1-R were:

1. To determine whether the variability of a number of mRNA encoding genes differentially expressed in fasted and fed sows is associated with meat quality traits.
2. To characterize the patterns of variability of porcine microRNA genes and to investigate the association of such variation with mRNA expression and meat quality phenotypes.
3. To improve the yet-poorly annotated porcine microRNAome by developing a dedicated prediction software tool able to detect novel and annotated microRNA sequences from transcriptomic data.
4. To elucidate the patterns of microRNA expression in the skeletal muscle of fasted and fed sows and to integrate such information with the profiles of expression of mRNAs and long non-coding RNAs in order to obtain a comprehensive perspective about the consequences of nutrition on skeletal muscle metabolism.

CHAPTER III. PUBLICATIONS

ANIMAL GENETICS
Immunogenetics, Molecular Genetics and Functional Genomics

An association analysis for 14 candidate genes mapping to meat quality quantitative trait loci in a Duroc pig population reveals that the *ATP1A2* genotype is highly associated with muscle electric conductivity

Mármol-Sánchez, E.[1], Quintanilla, R.[2], Jordana, J.[3] and Amills, M.[1,3*]

[1]Department of Animal Genetics, Centre for Research in Agricultural Genomics (CRAG), CSIC-IRTA-UAB-UB, Campus de la Universitat Autònoma de Barcelona, Bellaterra, Spain. [2]Animal Breeding and Genetics Programme, Institute for Research and Technology in Food and Agriculture (IRTA), Torre Marimon, Caldes de Montbui, Spain. [3]Departament de Ciència Animal i dels Aliments, Universitat Autònoma de Barcelona, Bellaterra, Spain.

[*]Corresponding author: Marcel Amills. marcel.amills@uab.cat

Anim. Genet. (2020) 51, 95-100. doi: 10.1111/age.12864.

Abstract

In previous GWAS carried out in a Duroc commercial line (Lipgen population), we detected on pig chromosomes 3, 4 and 14 several QTL for *gluteus medius* muscle redness (GM a*), electric conductivity in the *longissimus dorsi* muscle (LD CE) and vaccenic acid content in the LD muscle (LD C18:1 n-7), respectively. We have genotyped, in the Lipgen population, 19 SNPs mapping to 14 genes located within these QTL. Subsequently, association analyses have been performed. After correction for multiple testing, two SNPs in the *TGFBRAP1* (rs321173745) and *SELENOI* (rs330820437) genes were associated with GM a*, whereas *ACADSB* (rs81449951) and *GPR26* (rs343087568) genotypes displayed significant associations with LD vaccenic content. Moreover, the polymorphisms located at the *ATP1A2* (rs344748241), *ATP8B2* (rs81382410) and *CREB3L4* (rs321278469 and rs330133789) genes showed significant associations with LD CE. We made a second round of association analyses including the SNPs mentioned above as well as other SNPs located in the chromosomes to which they map. After performing a correction for multiple testing, the only association that remained significant at the chromosome-wide level was that between the *ATP1A2* genotype and LD CE. From a functional point of view, this association is meaningful because this locus encodes a subunit of the Na^+ /K^+ -ATPase responsible for maintaining an electrochemical gradient across the plasma membrane.

Keywords: Na^+/K^+-ATPase; pig; single nucleotide polymorphism.

Meat quality traits are of paramount importance for the pig industry because they determine, to a great extent, consumer acceptance and financial profit. Once pigs are slaughtered, there is a decline of the pH of the skeletal muscle owing to the production of lactic acid through anaerobic glycolysis (Rosenvold & Andersen 2003). The rate of muscle acidification has a strong effect on meat color and water-holding capacity. In this way, a low ultimate pH (5.3–5.4) is associated with pale, soft and exudative meat, as well as with an increased electrical conductivity (CE) and elevated drip and cooking losses (Lee *et al*. 2000; Rosenvold & Andersen 2003). In contrast, a high ultimate pH (6.3 or higher) results in dark, firm and dry meat with a high water-holding capacity and a lowered CE (Lee *et al*. 2000; Kim *et al*. 2016). Adverse effects on meat quality are influenced by both genetic and environmental factors.

Recessive and dominant genotypes in the porcine ryanodine receptor 1 (*RYR1*) and the protein kinase AMP-activated non-catalytic subunit *γ* 3 (*PRKAG3*) genes, respectively, are strong predisposing factors to the occurrence of pale, soft and exudative meats (Fujii *et al*. 1991; Milan *et al*. 2000). On the other hand, there are multiple factors related to pig management and transportation (pre-slaughter stress), stunning method at slaughter, carcass chilling and pelvic suspension of carcasses that influence pork quality (Rosenvold & Andersen 2003). Another important parameter that determines meat quality is intramuscular fat (IMF) composition. In this regard, fatty acid composition can have important consequences on the oxidative stability of meat during processing and retail display as well as on fat firmness (Wood *et al*. 2008).

In previous GWAS, we identified several genomic regions containing QTL for meat Minolta *a** value (redness), CE (González-Prendes *et al*. 2017) and IMF composition (González-Prendes *et al*. 2019) traits measured in the *longissimus dorsi* (LD) and *gluteus medius* (GM) muscle samples of 350 Duroc barrows (Lipgen population). Details about the rearing of the Lipgen pigs can be found in Gallardo *et al*. (2009), whereas a thorough description of QTL mapping methods is reported in González-Prendes *et al*. (2017). The measurement of CE was done 24 h after slaughter using a Pork Quality Meter (PQM-I INTEK GmbH, Aichach, Germany), and Minolta *a** value was determined with a Minolta Chroma-Meter CR-200 (Konica Minolta, Osaka, Japan) equipment at the same time point. Muscle fatty acid composition was measured as previously described by Quintanilla *et al*. (2011). In the current work, we have selected 14 candidate genes located within QTL regions for GM *a** on SSC3, LD CE on SSC4, and LD vaccenic content on SSC14 (Table 1). These genes were as follows: phosphorylase kinase catalytic subunit *γ* 1 (*PHKG1*), transforming growth factor *β* receptor-associated protein 1 (*TGFBRAP1*), selenoprotein I (*SELENOI*), hydroxyacil-CoA dehydrogenase trifunctional multienzyme (*HADHA*), coatomer protein complex subunit *α* (*COPA*), proliferation and apoptosis adaptor protein 15 (*PEA15*), calsequestrin 1 (*CASQ1*), ATPase Na$^+$/K$^+$ transporting α_2 subunit (*ATP1A2*), ATPase phospholipid transporting 8B2 (*ATP8B2*), cAMP-responsive element binding protein 3 like 4 (*CREB3L4*), CREB-regulated transcription coactivator 2 (*CRTC2*), acyl-CoA dehydrogenase short/branched chain (*ACADSB*), G protein-coupled receptor 26 (*GPR26*) and C-terminal binding protein 2 (*CTBP2*).

Table 1: An association analysis between 19 SNPs mapping to 14 candidate genes and meat quality traits recorded in a Duroc pig population (significant associations are shown in bold)[a].

Gene	SNP	Type	Trait	*P*-value	*q*-value	*P*-value*	*q*-value*	δ ± SE	A_1	MAF
PHKG1	rs697732005 (3:16.830 Mb)	Splice region variant (G/A)	**GM a***	0.88661	0.88661	0.68325	0.96577	-0.02 (0.142)	A	0.3443
TGFBRAP1	**rs321173745 (3:49.516 Mb)**	**Missense variant (A/G)**		**0.00361**	**0.00902**	**0.03108**	0.6722	0.549 (0.186)	G	0.1875
SELENOI	**rs330820437 (3:112.635 Mb)**	**Missense variant (A/G)**		**0.00039**	**0.00196**	**0.01307**	0.51778	0.643 (0.181)	G	0.1757
HADHA	rs81215086 ((3:112.794 Mb)	Missense variant (G/A)		0.53993	0.67491	0.62966	0.96577	-0.102 (0.169)	A	0.2899
	rs344578723 (3:112.796 Mb)	Missense variant (G/A)		0.53466	0.67491	0.6798	0.96577	-0.104 (0.169)	A	0.2866
COPA	rs340853721 (4:90.163 Mb)	Splice region variant (T/C)	**LD CE**	0.90735	0.95684	0.79005	0.99942	0.014 (0.091)	T	0.4351
	rs333099339 (4:90.183 Mb)	Splice region variant (T/C)		0.87813	0.95684	0.88586	0.99942	0.017 (0.090)	T	0.4381
	rs80949931 (4:90.186 Mb)	Missense variant (A/G)		0.95684	0.95684	0.6899	0.99942	-0.002 (0.091)	A	0.4335
PEA15	rs329681990 (4:90.266 Mb)	Splice region variant (G/A)		0.85666	0.95684	0.58021	0.99942	-0.014 (0.091)	G	0.433
CASQ1	rs334946278 (4:90.280 Mb)	Splice region variant (G/A)		0.95267	0.95684	0.9224	0.99942	0.005 (0.104)	A	0.1304
ATP1A2	**rs344748241 (4:90.356 Mb)**	**Splice region variant (G/A)**		**6.52E-03**	**7.17E-02**	**0.00006**	**0.02518**	-0.325 (0.066)	G	0.497
ATP8B2	**rs81382410 (4:95.435 Mb)**	**Splice region variant (T/C)**		**0.00285**	**0.01565**	**0.00256**	0.21113	-0.233 (0.077)	T	0.3345
CREB3L4	rs329686514 (4:95.717 Mb)	Missense variant (C/T)		0.08043	0.17695	0.22592	0.97957	-0.155 (0.088)	T	0.3063
	rs321278469 (4:95.717 Mb)	**Missense variant (C/A)**		**0.00639**	**0.01757**	**0.00554**	0.30475	-0.228 0.083)	C	0.3084
	rs330133789 (4:95.721 Mb)	**Missense variant (G/A)**		**0.00493**	**0.01757**	**0.01769**	0.57188	0.254 (0.075)	A	0.3373

CRTC2	rs330198768 (4:95.740 Mb)	Intron variant (C/T)		0.32931	0.60373	0.5631	0.99942	-0.083 (0.085)	T	0.3687
ACADSB	rs81449951 (14:132.588 Mb)	Missense variant (C/A)	**LD (C18:1) n-7**	**0.04036**	0.08073	**0.0424837**	0.8322423	0.093 (0.045)	A	0.2109
GPR26	**rs343087568** **(14:133.182 Mb)**	**Splice region variant (A/G)**		**0.00333**	**0.01334**	0.1269422	0.9956111	-0.096 (0.032)	G	0.4632
CTBP2	rs339956077 (14:134.334 Mb)	Splice region variant (G/A)		0.88166	0.88166	0.1269422	0.9956111	0.007 (0.046)	A	0.2094

[a] The *P*-value and the *q*-value terms define the statistical significance of the association analysis before and after correcting for multiple testing with a false discovery rate approach, respectively. The correction for multiple testing took into account the number of selected candidate SNPs mapping to each one of the SSC3 GM *a** (5 SNPs), SSC4 CE (11 SNPs) and SSC14 LD (C18:1) *n* − 7 (3 SNPs) QTL. The *P*-value* and the *q*-value* terms define the statistical significance of the chromosome-wide association analysis before and after correcting for multiple testing with a false discovery rate approach, respectively. In this case, the correction for multiple testing took into account the number of markers in the porcine SNP60 BeadChip mapping to pig chromosomes SSC3 (3123 SNPs), SSC4 (3899 SNPs) and SSC14 (4203 SNPs). Other terms that need to be defined are: *δ*, estimated allele substitution effect and its standard error (SE); A_1, minor allele; MAF, minor allele frequency; GM *a**, Minolta *a** value (redness) in the *gluteus medius* muscle; LD CE, electric conductivity in the *longissimus dorsi* muscle; and LD (C18:1) *n* − 7, vaccenic acid content in the *longissimus dorsi* muscle.

Genes were selected based on bibliographic information about their biological functions that suggested that they could be involved in the determinism of meat quality. Based on available RNA-seq (Cardoso *et al*. 2017) and whole-genome data (our unpublished results), we called 19 SNPs mapping to these 14 genes using the GATK Best Practices workflow for SNP calling (https://software.broadinstitute.org/gatk/best-practices/workflow?xml:id=11145), in accordance with protocols reported by Mármol-Sánchez *et al*. (2019). Nineteen SNPs were finally selected because the SnpEff software predicted that they might have functional effects (Cingolani *et al*. 2012), as reported in Table S1. The 19 selected SNPs (Table 1) were genotyped at the Servei Veterinari de Genètica Molecular of the Universitat Autònoma de Barcelona (http://sct.uab.cat/svgm/en) using a QuantStudio 12K Flex Real-Time PCR System (Thermo Fisher Scientific, Barcelona, Spain). Association analyses between SNPs and phenotypes were performed with the genome-wide efficient mixed model association (GEMMA) software (Zhou & Stephens 2012). The following statistical model was used:

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

where y is the vector of phenotypic observations for every individual, α corresponds to a vector including the intercept plus the fixed effects, that is batch effect with four categories (all traits), and farm origin effect with three categories (all traits). The α vector also contains the regression coefficients of the following covariates: (i) carcass weight at slaughterhouse for meat quality traits; and (ii) IMF content in the LD muscle for LD fatty acid composition. W is the incidence matrix relating phenotypes with the corresponding effects; x is the vector of the genotypes corresponding to the set of selected polymorphisms; δ is the allele substitution effect for each polymorphism; u is a vector of random individual effects with an *n*-dimensional multivariate normal distribution MVN$_n$ $(0, \lambda\ \tau^{-1} K)$, where τ^{-1} is the variance of the residual errors, λ is the ratio between the two variance components and K is a known relatedness matrix derived from the SNPs; and $\boldsymbol{\varepsilon}$ is the vector of residual errors. Results were corrected for multiple testing using the false discovery rate method reported by Benjamini & Hochberg (1995). The correction for multiple testing took into account the number of candidate selected SNPs mapping to each one of the SSC3 GM *a** (5 SNPs), SSC4 LD CE (11 SNPs) and SSC14 LD (C18:1) $n - 7$ (3 SNPs) QTL.

Performance of association analyses with the methodology described above revealed the existence of several associations that remained significant even after correction for multiple testing. We found, for instance, an association between GM Minolta *a** value and missense mutations in the *TGFBRAP1* and *SELENOI* genes, which map to two different GM *a** QTL on SSC3 (Table 1). The inactivation of the *TGFBRAP1* gene results in the suppression of aerobic glycolysis and increased levels of mitochondrial respiration and fatty acid oxidation (Yoshida *et al*. 2013), whereas *SELENOI* encodes a selenoprotein that is fundamental for the synthesis of phosphatidylethanolamine, a molecule with important effects on the oxidation of lipid membranes, oxidative phosphorylation and mitochondrial morphology (Tasseva *et al*. 2013; Poyton *et al*. 2016). We have also detected significant associations between LD CE and SNPs in the *ATP1A2*, *ATP8B2* and *CREB3L4* genes, which map to SSC4 LD CE QTL covering two regions spanning 85.6–91 and 95.2–97.8 Mb. These findings are suggestive because the *ATP1A2* gene, the one showing the most significant association, is preferentially expressed in the skeletal and heart muscles and brain and it encodes the α_2 subunit of the ion pump Na^+/K^+ ATPase (Clausen *et al*. 2017). Noteworthy, Na^+/K^+-ATPases provide the energy necessary for the maintenance of Na^+ and K^+ electrochemical gradients across the plasma membrane by hydrolyzing ATP (Clausen *et al*. 2017; Sampedro Castañeda *et al*. 2018). These gradients are essential for the preservation of the resting membrane potential as well as for the generation of electrical impulses in the skeletal muscle and nervous system (Clausen *et al*. 2017; Sampedro Castañeda *et al*. 2018). The ATP8B2 protein is also an ATPase with flippase activity toward phosphatidyl choline, a key component of phospholipid membranes with important effects on the functioning of the sarcoendoplasmic reticulum Ca^{2+}ATPase pumps (Fajardo *et al*. 2018; Shin & Takatsu 2018), whereas CREB3L4 is a transmembrane bZip transcription factor involved in the modulation of endoplasmic reticulum stress (Kim *et al*. 2014). Our association analysis also revealed the existence of significant associations between the phenotypic variation of LD vaccenic (C18:1 $n - 7$) content and SSC14 SNPs located in the *ACADSB* gene, which catalyzes the oxidation of branched-chain fatty acids (Porta *et al*. 2019), and the *GPR26* gene, whose inactivation leads to hyperphagia, glucose intolerance, hyperinsulinemia, dyslipidemia and obesity in mice (Chen *et al*. 2012).

We made a second round of association analyses in which the SNPs that previously showed evidence of statistical significance were compared against the whole sets of the porcine SNP60 BeadChip SNPs co-localizing to the same chromosome (chromosome-wide analysis), that is 3,123 SNPs on SSC3, 3,899 SNPs on SSC4 and 4,203 SNPs on SSC14. These 11,225 SNPs were obtained from previously published porcine SNP60 BeadChip data reported by González-Prendes *et al*. (2017). In this case, the correction for multiple testing took into account the number of SNPs mentioned above for each one of the three chromosomes under analysis, that is 3,128, 3,910 and 4,206 independent tests were taken into consideration when performing association analyses for pig chromosomes SSC3, SSC4 and SSC14, respectively. Interestingly, the rs344748241 SNP in the *ATP1A2* gene was the only one that surpassed the chromosome-wide threshold of significance (q-value < 0.05; Table 1, Figure 1). Noteworthy, this SNP was not significant when we made an association analysis at the genome-wide level (data not shown). Additionally, we used the *LD* function of the *GASTON* R package (version 1.5.5; Perdry *et al*. 2019) to evaluate the presence of linkage disequilibrium among the SNP markers that showed significant associations with LD CE after correction for multiple testing at the chromosome-wide level (Figure S1). The amount of linkage disequilibrium was expressed as r^2 in accordance with the definition of Hill & Robertson (1968). As shown in Figure S1, we observed a high degree of linkage disequilibrium between the rs344748241 (*ATP1A2* gene) and the rs80782100 (*IGSF8* gene) markers. It is noteworthy that the rs80782100 SNP, which maps to an intronic position within the immunoglobulin superfamily member 8 gene, displays the highest association with the LD CE phenotype, as described in González-Prendes *et al*. (2017).

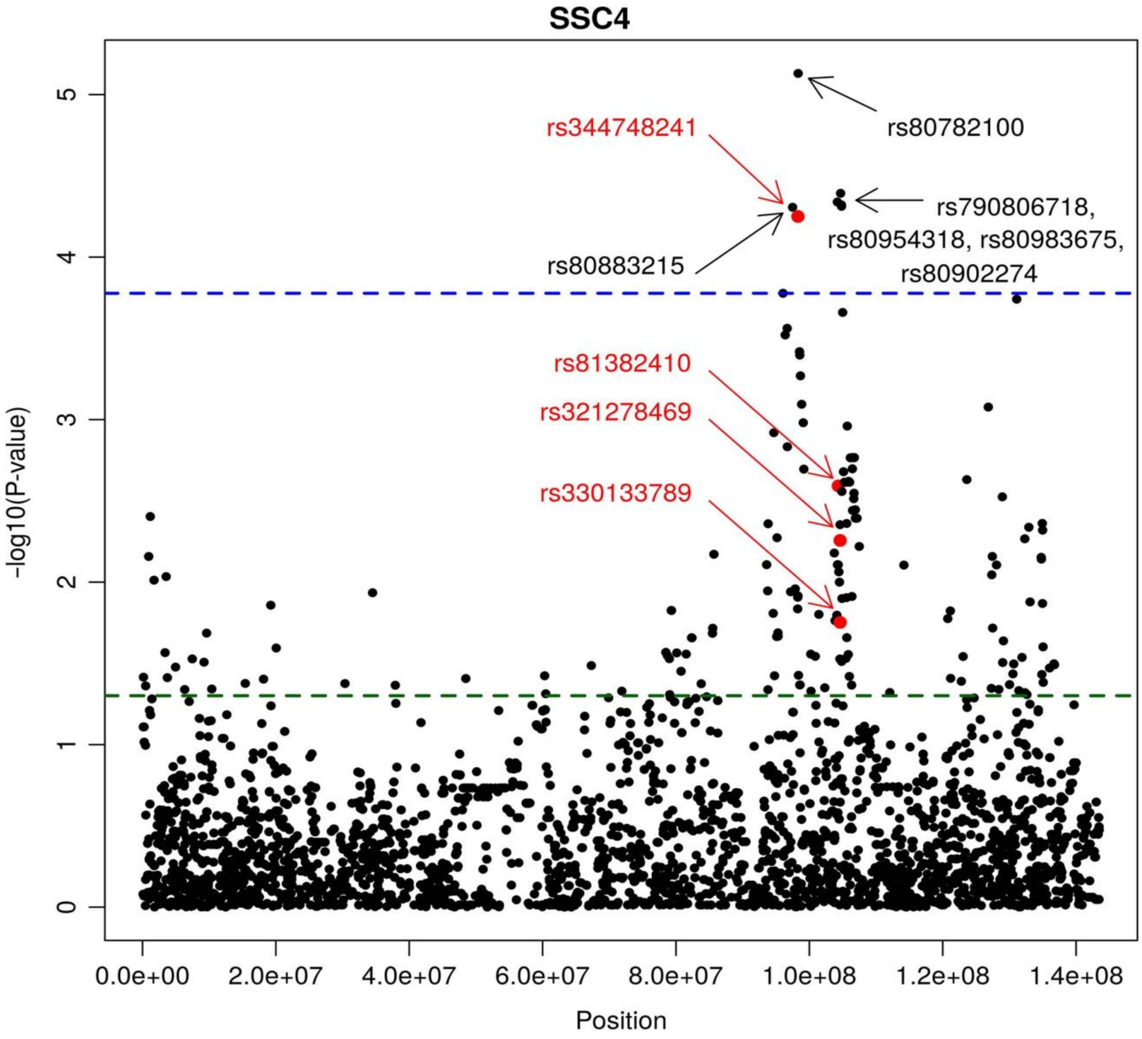


Figure 1: Manhattan plot depicting associations between electrical conductivity in the *longissimus dorsi* muscle and the genotypes of markers in the *ATP1A2* (rs344748241), *ATP8B2* (rs81382410) and *CREB3L4* (rs321278469 and rs330133789) loci plus 3,899 additional SNPs mapping to pig chromosome 4 (SSC4). The positions of these three genes are: SSC4, 90.292–90.371 Mb (*ATP1A2*); SSC4, 95.426 – 95.446 Mb (*ATP8B2*); and SSC4, 95.714–95.723 Mb (*CREB3L4*). The green line represents the nominal *P*-value of significance, whereas the blue line indicates the *P*-value of significance after correcting for multiple testing with a false discovery rate approach (*q*-value). The rs344748241 SNP in the *ATP1A2* gene is located 23 kb away from the peak of the LD CE QTL, that is ALGA0026686 (rs80782100; 4:90.378 Mb) SNP, as reported by González-Prendes *et al*. (2017).

As previously discussed, we consider that the *ATP1A2* gene is a strong positional and functional candidate to explain the CE QTL found on SSC4 because Na^+, K^+ ATPases are fundamental to inducing an electrochemical gradient across the plasma membrane of cells (Suhail 2010), and their kinetics are modulated by the extracellular pH (Salonikidis *et al*. 2000), a parameter that also displays strong effects on muscle electrical conductivity. In pigs, the *ATP1A2* gene has been sequenced (Henriksen *et al*. 2013) and its polymorphisms have been associated with fat cut percentage (Fontanesi *et al*. 2012). A next step would be to re-sequence the whole gene in Lipgen pigs with alternative genotypes (QQ vs. qq) for the LD CE QTL on SSC4, to build a complete catalog of SNPs with potential effects on protein activity and expression and to investigate their association with CE in the Lipgen population. Subsequently, functional tests should be applied to ascertain whether any of the mutations in the pig *ATP1A2* gene with highly significant *q*-values also have causal effects on muscle conductivity.

Supplementary Information

Supplementary Table 1: Additional information about selected SNP and their potential impact and deleteriousness (SIFT).

Supplementary Figure 1: Graph depicting the magnitude of linkage disequilibrium among SNPs that showed significant associations with *longissimus dorsi* electric conductivity after correction for multiple testing at the chromosome-wide level. Here, the amount of linkage disequilibrium is expressed as r^2 as defined by Will & Robertson (1968) and such parameter was calculated with the *LD* function of *gaston* R package.

Acknowledgements

The authors are indebted to Selecció Batallé S.A. for providing the animal material used in this study. We gratefully acknowledge J. Reixach and J. Soler (IRTA) for their collaboration in the experimental protocols. This research was partially funded with grants AGL2013-48742-C2-1-R and AGL2013-48742-C2-2-R awarded by the Spanish Ministry of Economy and Competitiveness and grant 2014 SGR 1528 from the Agency for Management of

University and Research Grants of the Generalitat de Catalunya. Emilio Mármol-Sánchez was funded with an FPU Ph.D. grant from the Spanish Ministry of Education (FPU15/01733). We also acknowledge the support of the Spanish Ministry of Economy and Competitiveness for the Center of Excellence Severo Ochoa 2016–2019 (SEV-2015-0533) grant awarded to the Center for Research in Agricultural Genomics. Thanks also to the CERCA Programme of the Generalitat de Catalunya.

Conflict of interest

The authors declare that they have no conflict of interest.

Data Availability

The 11,225 SNPs included in this study were obtained from published porcine SNP60 BeadChip data reported by González-Prendes *et al*. (2017), which can be accessed at the Figshare public repository (https://figshare.com/s/2e636697009360986794).

References

Benjamini Y. & Hochberg Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B* 57, 289-300.

Cardoso T.F., Cánovas A., Canela-Xandri O., González-Prendes R., Amills M. & Quintanilla R. (2017) RNA-seq based detection of differentially expressed genes in the skeletal muscle of Duroc pigs with distinct lipid profiles. *Scientific Reports* 7, 40005.

Chen D., Liu X., Zhang W. & Shi Y. (2012) Targeted inactivation of GPR26 leads to hyperphagia and adiposity by activating AMPK in the hypothalamus. *PLoS One* 7, e40764.

Cingolani P., Platts A., Wang Ie L., Coon M., Nguyen T., Wang L., Land S.J., Lu X. & Ruden D.M. (2012) A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly* 6, 80-92.

Clausen M.V., Hilbers F. & Poulsen H. (2017) The structure and function of the Na,K-ATPase isoforms in health and disease. *Frontiers in Physiology* 8, 371.

Fajardo V.A., Mikhaeil J.S., Leveille C.F., Tupling A.R. & LeBlanc P.J. (2018) Elevated whole muscle phosphatidylcholine: phosphatidylethanolamine ratio coincides with reduced SERCA activity in murine overloaded plantaris muscles. *Lipids in Health and Disease* 17, 47.

Fontanesi L., Galimberti G., Calò D.G. *et al.* (2012) Identification and association analysis of several hundred single nucleotide polymorphisms within candidate genes for back fat thickness in Italian Large White pigs using a selective genotyping approach. *Journal of Animal Sci*ence 90, 2450-64.

Fujii J., Otsu K., Zorzato F., de Leon S., Khana V.K., Weiler J.E., O' Brien P.J. & McLennan D.H. (1991) Identification of a mutation in porcine ryanodine receptor associated with malignant hyperthermia. *Science* 253, 448-451.

Gallardo D., Quintanilla R., Varona L., Díaz I., Ramírez O., Pena R.N., & Amills M. (2009) Polymorphism of the pig *acetyl-coenzyme A carboxylase α* gene is associated with fatty acid composition in a Duroc commercial line. *Animal Genetics* 40, 410-419.

González-Prendes R., Quintanilla R., Cánovas A., Manunza A., Figueiredo Cardoso T., Jordana J., Noguera J.L., Pena R.N. & Amills M. (2017) Joint QTL mapping and gene expression analysis identify positional candidate genes influencing pork quality traits. *Scientific Reports* 7, 39830.

González-Prendes R., Quintanilla R., Mármol-Sánchez E. *et al.* (2019) Comparing the mRNA expression profile and the genetic determinism of intramuscular fat traits in the porcine *gluteus medius* and *longissimus dorsi* muscles. *BMC Genomics* 20, 170.

Henriksen C., Kjaer-Sorensen K., Einholm A.P., Madsen L.B., Momeni J., Bendixen C., Oxvig C., Vilsen B. & Larsen K. (2013) Molecular cloning and characterization of porcine Na^+/K^+-ATPase isoforms α_1, α_2, α_3 and the ATP1A3 promoter. *PLoS One* 8, e79127.

Hill W.G. & Robertson A. (1968) Linkage disequilibrium in finite populations. *Theoretical and Applied Genetics* 38, 226-31.

Kim T.H., Jo S.H., Choi H., Park J.M., Kim M.Y., Nojima H., Kim J.W. & Ahn Y.H. (2014) Identification of *Creb3l4* as an essential negative regulator of adipogenesis. *Cell Death & Disease* 5, e1527.

Kim T.W., Kim C.W., Yang M.R., No G.R., Kim S.W. & Kim I.S. (2016) Pork quality traits according to postmortem pH and temperature in Berkshire. *Korean Journal for Food Science of Animal Resources* 36, 29-36.

Lee S., Norman J.M., Gunasekaran S., van Laack R.L., Kim B.C. & Kauffman R.G. (2000) Use of electrical conductivity to predict water-holding capacity in post-rigor pork. *Meat Science* 55, 385-9.

Mármol-Sánchez E., Quintanilla R., Cardoso T.F., Jordana J. & Amills M. (2019) Polymorphisms of the cryptochrome 2 and mitoguardin 2 genes are associated with the variation of lipid-related traits in Duroc pigs. *Scientific Reports* 9, 9025.

Milan D., Jeon J.T., Looft C. *et al.* (2000) A mutation in *PRKAG3 a*ssociated with excess glycogen content in pig skeletal muscle. *Science* 288, 1248-1251.

Perdry H., Dandine-Roullard C., Bandyopadhyay D. & Kettner L. (2019) gaston: Genetic Data Handling (QC, GRM, LD, PCA) & Linear Mixed Models. R package version 1.5.5. https://CRAN.R-project.org/package=gaston

Porta F., Chiesa N., Martinelli D. & Spada M. (2019) Clinical, biochemical and molecular spectrum of short/branched-chain acyl-CoA dehydrogenase deficiency: two new cases and review of literature. *Journal of Pediatric Endocrinology and Metabolism* 32, 101-108.

Poyton M.F., Sendecki A.M., Cong X. & Cremer P.S. (2016) Cu^{2+} binds to phosphatidylethanolamine and increases oxidation in lipid membranes. *Journal of the American Chemical Society* 138, 1584-90.

Quintanilla R., Pena R.N., Gallardo D., Cánovas A., Ramírez O., Díaz I., Noguera J.L. & Amills M. (2011) Porcine intramuscular fat content and composition are regulated by quantitative trait loci with muscle-specific effects. *Journal of Animal Science* 89, 2963-71.

Rosenvold K. & Andersen H.J. (2003) Factors of significance for pork quality-a review. *Meat Science* 64, 219-37.

Salonikidis P.S., Kirichenko S.N., Tatjanenko L.V., Schwarz W. & Vasilets L.A. (2000) Extracellular pH modulates kinetics of the Na^+,K^+-ATPase. *Biochimica et Biophysica Acta (BBA) - Biomembranes* 1509, 496-504.

Sampedro Castañeda M., Zanoteli E., Scalco R.S. *et al.* (2018) A novel *ATP1A2* mutation in a patient with hypokalaemic periodic paralysis and CNS symptoms. *Brain* 141, 3308-3318.

Shin H.W. & Takatsu H. (2018) Substrates of P4-ATPases: beyond aminophospholipids (phosphatidylserine and phosphatidylethanolamine). *The FASEB Journal* 3, 3087-3096.

Suhail M. (2010) Na^+, K^+-ATPase: Ubiquitous multifunctional transmembrane protein and its relevance to various pathophysiological conditions. *Journal of Clinical Medicine Res*earch 2, 1-17.

Tasseva G., Bai H.D., Davidescu M., Haromy A., Michelakis E. & Vance J.E. (2013) Phosphatidylethanolamine deficiency in mammalian mitochondria impairs oxidative phosphorylation and alters mitochondrial morphology. *Journal of Biological Chemistry* 288, 4158-4173.

Wood J.D., Enser M., Fisher A.V., Nute G.R., Sheard P.R., Richardson R.I., Hughes S.I. & Whittington F.M. (2008) Fat deposition, fatty acid composition and meat quality: A review. *Meat Science* 78, 343-58.

Yoshida S., Tsutsumi S., Muhlebach G. *et al.* (2013) Molecular chaperone TRAP1 regulates a metabolic switch between mitochondrial respiration and aerobic glycolysis. *Proceedings of the National Academy of Sciences of the United States of America* 110, E1604-12.

Zhou X. & Stephens M. (2012) Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics* 44, 821-824

ANIMAL GENETICS
Immunogenetics, Molecular Genetics and Functional Genomics

Detection of homozygous genotypes for a putatively lethal recessive mutation in the porcine argininosuccinate synthase 1 (*ASS1*) gene

Mármol-Sánchez, E.[1], Luigi-Sierra, M. G.[1], Quintanilla, R.[2] and Amills, M.[1,3*]

[1]Department of Animal Genetics, Centre for Research in Agricultural Genomics (CRAG), CSIC-IRTA-UAB-UB, Campus de la Universitat Autònoma de Barcelona, Bellaterra, Spain. [2]Animal Breeding and Genetics Programme, Institute for Research and Technology in Food and Agriculture (IRTA), Torre Marimon, Caldes de Montbui, Spain. [3]Departament de Ciència Animal i dels Aliments, Universitat Autònoma de Barcelona, Bellaterra, Spain.

[*]Corresponding author: Marcel Amills. marcel.amills@uab.cat

Anim. Genet. (2020) 51, 106-110. doi: 10.1111/age.12877.

Abstract

The sequencing of the pig genome revealed the existence of homozygous individuals for a nonsense mutation in the argininosuccinate synthase 1 (*ASS1*) gene (rs81212146, c.944T>A, L315X). Paradoxically, an AA homozygous genotype for this polymorphism is expected to abolish the function of the ASS1 enzyme that participates in the urea cycle, leading to citrullinemia, hyperammonemia, coma and death. Sequencing of five Duroc boars that sired a population of 350 Duroc barrows revealed the segregation of the c.944T>A polymorphism, so we aimed to investigate its phenotypic consequences. Genotyping of this mutation in the 350 Duroc barrows revealed the existence of seven individuals homozygous (AA) for the nonsense mutation. These AA pigs had a normal weight despite the fact that mild citrullinemia often involves impaired growth. Sequencing of the region surrounding the mutation in TT, TA and AA individuals revealed that the A substitution in the second position of the codon (c.944T>A) is in complete linkage disequilibrium with a C replacement (c.943T>C) in the first position of the codon. This second mutation would compensate for the potentially damaging effect of the c.944T>A replacement. In fact, this is the most probable reason why pigs with homozygous AA genotypes at the 944 site of the *ASS1* coding region are alive. Our results illustrate the complexities of predicting the consequences of nonsense mutations on gene function and phenotypes, not only because of annotation issues but also owing to the existence of genetic mechanisms that sometimes limit the penetrance of highly harmful mutations.

Keywords: citrullinemia; nonsense mutation; pig; premature stop codon; single nucleotide polymorphism.

The sequencing of the pig genome led to the discovery of 157 nonsense mutations mapping to 142 genes, and 11 of them were reported to have pathological effects in humans (Groenen *et al.* 2012). Although most of these 11 damaging nonsense variants were found in a heterozygous state, two mutations mapping to the argininosuccinate synthase 1 (*ASS1,* rs81212146, c.944T>A, L315X) and to the RB binding protein 8 endonuclease (*RBBP8*) genes displayed homozygous genotypes. The inactivation of the *RBBP8* gene causes embryonic lethality (Polato *et al.* 2014), but it should be noticed that the current release of the

Ensembl database (https://www.ensembl.org) does not report any stop gained mutation in the porcine *RBBP8* gene. With regard to the ASS1 enzyme, its inactivation leads to the disruption of the urea cycle and to citrullinemia, a disease characterized by increased ammonia levels in blood, stupor, convulsions, coma and death (Endo *et al.* 2004). Groenen *et al.* (2012) argued that homozygosity for the nonsense *ASS1* mutation might be associated with a milder form of citrullinemia. However, the mild course of this disease is usually, but not always, explained by mutations causing only a partial abolishment of the function of the ASS1 enzyme (Häberle *et al.* 2003). In some cases, the mild form of the disease involves the development of symptoms such as poor growth, liver failure, cerebral infarction or spasticity, whereas in other occasions patients remain asymptomatic (Häberle *et al.* 2003).

Whole-genome sequencing (WGS) (our unpublished data) of the five Duroc boars that sired a purebred population of 350 Duroc barrows (Gallardo *et al.* 2008, 2009) revealed the segregation of the rs81212146 *ASS1* nonsense polymorphism, providing an opportunity to investigate its phenotypic effects. Indeed, the consequences of this polymorphism were predicted by NCBI automated computational analysis, but no experiment was made to assess the accuracy of such a prediction. Using a QuantStudio 12 K flex Real-Time PCR System available at the Servei Veterinari de Genètica Molecular at the Universitat Autònoma de Barcelona (http://sct.uab.cat/svgm/en), we genotyped the 350 offspring of the five boars with a dedicated TaqMan Open Array multiplex assay. In total, 323 pigs were successfully genotyped for the rs81212146 polymorphism, which led to the identification of 239 TT, 77 TA and 7 AA pigs, hence confirming the existence of homozygous individuals for this mutation in the population under study. As one of the potential symptoms of mild citrullinemia is retarded growth, we inspected the final weight of the AA pigs compared with their TT and TA counterparts. Live weights were measured before slaughtering and carcass weights were also collected after evisceration at the abattoir. The average live weights at 190 days of TT, TA and AA pigs were 122.55 ± 12.18, 121.26 ± 16.66 and 119.92 ± 21.91 kg respectively. Moreover, carcass weights of TT, TA and AA pigs were 94.47 ± 10.18, 95.09 ± 11.78 and 93.67 ± 11.67 kg respectively (Figure S1). An analysis of variance (ANOVA) performed with the *aov* R function and contrasting *ASS1* genotypic means for both live (*P*-value = 0.724) and carcass (*P*-value = 0.893) weights did not reveal any significant difference. In summary, we did not find evidence of a significantly decreased weight, before or after slaughter, in AA pigs.

In order to further investigate the potential consequences of the rs81212146 polymorphism, we sequenced the region of the *ASS1* gene containing the putative nonsense mutation by making use of both genomic DNA and complementary DNA (cDNA) as templates. A total of 16 liver samples belonging to each of five TT and AA and six TA animals were selected at random. Genomic DNA extraction was performed by digestion of 30 mg of liver tissue in 0.5 ml lysis buffer (50 mM Tris–HCl, pH 8; 20 mᴍ EDTA, pH 8; 2% SDS) plus 15 µl (1 µg/µl) proteinase K and incubated overnight at 56 °C. Subsequently, 500 µl of the lysate was deproteinized with 0.5 ml of a mixture of phenol–chloroform–isoamyl alcohol (25:24:1). The resulting supernatant was mixed with 1 ml ice-cold pure ethanol plus 50 µl NaCl (2 M) and centrifuged for 30 min at maximum speed. The DNA pellets obtained in this way were washed with 500 µl of ethanol 70% and resuspended in 50 µl of ultrapure water. We also extracted RNA from the same selected liver samples corresponding to TT, AA and TA pigs. In brief, liver samples were pulverized in liquid nitrogen with a mortar and a pestle, homogenized and submerged in 1 ml of TRI Reagent (Thermo Fisher Scientific, Barcelona, Spain). Total RNA was then purified with the RiboPure kit (Ambion, Austin, TX, USA) in accordance with the instructions of the manufacturer. The concentration and purity of DNA and RNA samples were assessed with a NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific, Barcelona, Spain). A Bioanalyzer-2100 equipment (Agilent Technologies Inc., Santa Clara, CA) was employed for determining RNA integrity (RIN) with the Agilent RNA 6000 Nano Kit (Agilent Technologies Inc., Santa Clara, CA). All RNA samples had RIN values > 7. The average RIN values of RNA preparations corresponding to TT, TA and AA pigs were 7.46, 7.24 and 7.52 respectively. Reverse transcription (RT) was carried out with the High-Capacity cDNA Reverse Transcription Kit (Thermo Fisher Scientific, Barcelona, Spain). Each reverse transcription reaction contained 2 µl 10× RT Buffer, 0.8 µL 25× dNTP Mix (100 mM), 2 µl 10× RT Random Primers, 1 µl MultiScribe Reverse Transcriptase (50 U/µl) and 10 µl total RNA (~100 ng/µl). Ultrapure water was added until a final volume of 20 µl was reached. The RT thermal profile included an incubation step at 25 °C for 10 min, followed by 120 min at 37 °C and an inactivation step at 95 °C for 5 min.

Genomic DNA and cDNA samples were then subjected to PCR amplification. Primers (Table S1) were designed with the Primer3 software (Untergasser *et al.* 2012) to span contiguous exon–intron and exon–exon junctions for genomic DNA and cDNA amplicons respectively. Expected sizes were 278 and 221 bp for PCR products amplified from genomic

DNA and cDNA templates respectively. The relative positions of genotyped rs81212146 polymorphism in genomic and cDNA amplicons are depicted in Figure S2. Amplification reactions contained 2 µl of 10× PCR buffer, 0.2 µl dNTPs (25 mM), 0.6 µl of each primer (10 µM), 2 µl of $MgCl_2$ (25 mM), 2.5 µl of genomic DNA (10 ng/µl) or 2.5 µl of a 5-fold dilution of the RT-reaction, and 0.2 µl Amplitaq Gold DNA Polymerase (5 U/µl) (Thermo Fisher Scientific, Barcelona, Spain). Ultrapure water was added until a 20 µl final volume was reached. The thermal profile included a denaturation step at 95 °C for 10 min, followed by 35 cycles of denaturation at 95 °C for 1 min, annealing at 60 °C for 1 min and extension at 72 °C for 1 min, plus a final extension step at 72 °C for 7 min. Amplicons with the expected size were purified with the ExoSAP-IT PCR Clean-up kit (Thermo Fisher Scientific, Barcelona, Spain). They were subsequently sequenced with the BigDye Terminator Cycle Sequencing Kit v1.1 (Applied Biosystems, Foster City, CA, USA) and with primers listed in Table S1. Sequencing reactions were electrophoresed in an ABI 3730 DNA analyzer (Applied Biosystems, Foster City, CA, USA). The MEGA software version 6.0 (Tamura *et al.* 2013) was employed to visualize the results of the sequencing experiments. Partial *ASS1* sequences obtained from genomic DNA (accession numbers: MN296492–MN296493) and cDNA (MN296494–MN296495) were submitted to the Genbank database.

The predicted consequence of the replacement of T by A at the second position of codon 315 would be the introduction of a premature stop codon (TTG>TAG), completely abolishing the function of the ASS1 enzyme. However, sequencing of *ASS1* DNA and cDNA amplicons revealed that the A allele in the second position of the codon is linked to a C replacement (rs81212145, c.943T>C) in the first position of the codon (Figure 1, Figure S3), leading to the generation of a benign missense (L315Q) mutation. This second polymorphism is expected to compensate for the potentially damaging effect of the c.944T>A replacement. As revealed by the PolyPhen-2 algorithm (Adzhubei *et al.* 2010), the substitution of leucine (TTG) by glutamine (CAG) is predicted to be tolerated (PolyPhen-2 score = 0.012). Indeed, homozygosity for the TAG codon at position 315 should be lethal in pigs, and in consequence, it might have been strongly selected against. Interestingly, all sequenced animals displaying an AA genotype for the second position of codon 315 were also homozygous CC for the first position (rs81212145), i.e. all of them were CAG for codon 315, suggesting the existence of complete linkage disequilibrium (LD) between both polymorphisms. By using a previously generated liver microarray dataset from the same

Duroc population analyzed herewith (Manunza *et al.* 2014), we compared the levels of *ASS1* mRNA expression between two c.944T>A genotypes, i.e. TA (N = 18) vs TT (N = 67). A *t*-test analysis performed with the *t.test* R function did not reveal any significant difference in *ASS1* mRNA expression between these two genotypes (*P*-value = 0.346), suggesting that the c.944T>A polymorphism does not have any effect on the transcriptional rate of the *ASS1* gene.

In order to estimate the co-association between the two mutations in the first and second positions of codon 315, 120 WGS belonging to European and Asian domestic pigs and wild boars were retrieved from the NCBI Sequence Read Archive (SRA, https://www.ncbi.nlm.nih.gov/sra). Detailed information about these WGSs is available in Table S2. All raw SRA files were converted into FASTQ format using the fastq-dump 2.8.2 tool from the SRA-TOOLKIT package (https://www.ncbi.nlm.nih.gov/sra/docs/toolkitsoft). The FASTQ files were subsequently filtered for any sequencing adaptors with the Trimmomatic version 0.36 software (Bolger *et al.* 2014). Paired-end filtered sequences were then aligned to the porcine reference genome (Sscrofa11.1, Warr *et al.* 2019) with the BWA MEM algorithm (Li 2013). Alignment files were sorted and binarized and PCR duplicates were marked and removed with the PICARD tool (https://broadinstitute.github.io/picard). INDEL realignment and base recalibration were performed and the HaplotypeCaller function from the GATK 3.8 tool (McKenna *et al.*, 2010) with default parameters was used to generate variant call format (VCF) files. Hard filtering was applied according to GATK best practices (https://software.broadinstitute.org/gatk/best-practices/). The rs81212145 and rs81212146 contiguous polymorphisms were retrieved and their co-segregation in European and Asian domestic pigs as well as in European and Asian wild boars was investigated by estimating the r^2 coefficient, which defines the amount of LD between two markers (Hill & Robertson 1968).

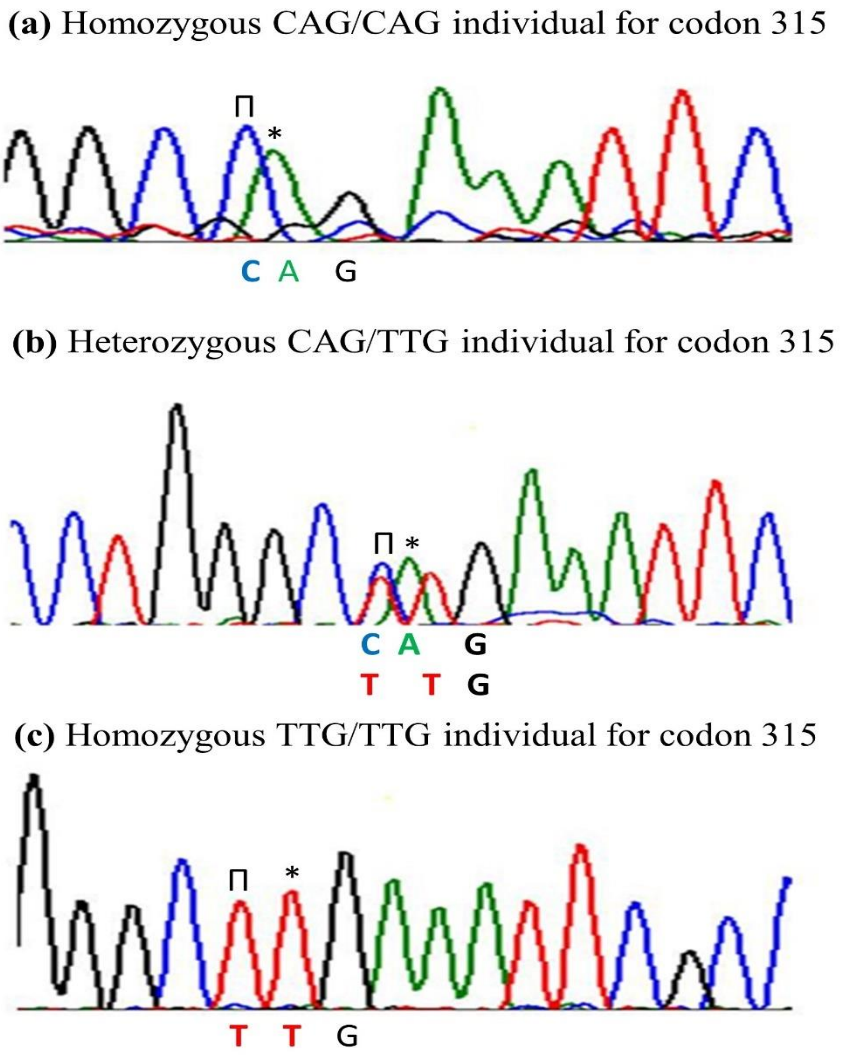


Figure 1: Sequencing of codon 315 of the porcine *ASS1* gene and its surrounding region using genomic DNA as a template. The upper (a), central (b) and lower (c) electropherograms display the three codon 315 genotypes (CAG/CAG CAG/TTG and TTG/TTG) detected by Sanger sequencing in a sample of 16 pigs. The c.943T>C and c.944T>A polymorphisms are indicated with the (Π) and (*) symbols, respectively.

This analysis supported the notion that rs81212145 and rs81212146 polymorphisms are in complete LD (Table 1), implying the existence of two potential sequences, CAG and TTG, at codon 315. In contrast, the TAG sequence, which would have severe deleterious effects on ASS1 enzyme activity, was not detected in our WGS dataset. The frequency of the CAG haplotype was much higher in pigs and wild boars from Asia than in those with a European origin (Table 1). This result is probably due to the high genetic divergence between Asian and European pigs, which separated 1 million years ago (Frantz *et al.* 2015).

Table 1: Frequency of the codon 315 CAG *ASS1* haplotype in 120 sequenced pigs and wild boars (NCBI Sequence Read Archive) and measurement of the r^2 coefficient between polymorphisms c.943T>C and c.944T>A.

Parameter	**European domestic pigs (N = 40)**	**European wild boars (N = 20)**	**Asian domestic pigs (N = 40)**	**Asian wild boars (N = 20)**
Missing[1]	0.375	1	0.05	0.15
CAG frequency[2]	0.08	-	0.89	0.47
r^2 LD[3]	1	-	1	1

[1] Percentage of pigs with missing genotypes for codon 315 of the *ASS1* gene; [2] CAG haplotype frequency; [3] r^2 LD: magnitude of the linkage disequilibrium between polymorphisms c.943T>C and c.944T>A expressed as r^2 (Hill & Robertson 1968).

There is an increasing interest in characterizing nonsense mutations associated with lethality because they can have a negative effect on the profitability of pig farms. For instance, Derks *et al.* (2017) analyzed, with an 80K SNP array, 24,000 pigs from commercial farms and found 35 haplotypes with complete absence or depletion of homozygous genotypes and showing adverse effects on reproduction traits. Moreover, Derks *et al.* (2019) detected five relatively frequent recessive lethal haplotypes in two commercial Norwegian Landrace and Duroc purebred populations which cause important reductions (15.1–21.6%) in litter size owing to the embryonic death of homozygous individuals. Interestingly, these recessive lethal haplotypes increase litter size in crossbred individuals owing to a positive heterotic effect on fertility.

The results of our study reflect the difficulties of predicting the outcome of putative loss-of-function mutations, either because problems in their correct annotation (Narasimhan *et al.* 2016) or owing to the existence of mechanisms of genetic compensation that prevent lethality. Indeed, the rs81212145 and rs81212146 SNPs are annotated as synonymous and stop gained substitutions in the Sscrofa10.2 and Sscrofa11.1 assemblies of the pig genome respectively, but according to our analyses they should be jointly considered as a dinucleotide polymorphism in codon 315 with a missense effect. With regard to genetic compensation, an analysis of 589,306 human genomes led to the identification of 13 individuals with homozygous (autosomal recessive disease) or heterozygous (autosomal dominant disease) genotypes for eight severe Mendelian childhood diseases (Chen *et al.* 2016). These individuals should have manifested serious clinical symptoms before the age of 18 years, but apparently they were perfectly healthy (Chen *et al.* 2016). The only explanation for such a paradoxical result is that there are mechanisms at play that decrease the penetrance of nonsense mutations, including suppressor mutations able to change the sequence of the affected codon or to induce splicing events eliminating the exon carrying the nonsense mutation (MacArthur *et al.* 2012). Alternatively, the readthrough of the premature stop codon during ribosomal translation might also prevent its truncating effect on protein synthesis (Rausell *et al.* 2014). In conclusion, our data indicate that the c.944T>A *ASS1* mutation probably does not have pathological consequences on pigs owing to the existence of an adjacent mutation that prevents the formation of a premature stop codon. The considerable amount of deleterious variation segregating in domestic animals (Makino *et al.* 2018) offers an unparalleled opportunity to explore the effects of loss-of-function mutations on phenotypes of economic interest, as well as to elucidate the genetic mechanisms that, on some occasions, counteract their harmful consequences.

Supplementary Information

Supplementary Table 1: Primers employed in the PCR amplification and partial sequencing of the porcine argininosuccinate synthase 1 (*ASS1*) gene.

Supplementary Table 2: List of the porcine WGSs used in the current study and genotype of codon 315 in Asian domestic (ADM), Asian wild boar (AWB), European domestic (EDM) and European wild boar (EWB) pigs.

Supplementary Figure 1: Boxplots depicting the distribution of (a) live weight and (b) carcass weight for pigs with TT (N = 239), TA (N = 77) and AA (N = 7) rs81212146 genotypes.

Supplementary Figure 2: Primer binding regions (highlighted in bold) in the *ASS1* amplicons generated from (a) genomic DNA and (b) cDNA.

Supplementary Figure 3: Sequencing of codon 315 of the porcine *ASS1* gene and its surrounding region using cDNA as a template.

Acknowledgements

The authors are indebted to Selección Batallé S.A. for providing the animal material and IRTA collaborators for participating in experimental protocols and animal management. Part of the research presented in this publication was funded by grants AGL2013-48742-C2-1-R and AGL2013-48742-C2-2-R awarded by the Spanish Ministry of Economy and Competitivity. We also acknowledge the support of the Spanish Ministry of Economy and Competitivity for the Center of Excellence Severo Ochoa 2016–2019 (SEV-2015-0533) grant awarded to the Centre for Research in Agricultural Genomics. Emilio Mármol-Sánchez was funded by an FPU Ph.D. grant from the Spanish Ministry of Education (FPU15/01733). María Gracia Luigi-Sierra was funded by an FPI Ph.D. grant (BES-C-2017-0024) awarded by the Spanish Ministry of Economy and Competitivity. Thanks are also due to the CERCA Programme of the Generalitat de Catalunya.

Conflict of interest

The authors declare that they have no conflict of interest.

References

Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., Kondrashov, A. S. & Sunyaev, S. R. (2010) A method and server for predicting damaging missense mutations. *Nature Methods* 7, 248-49.

Bolger, A. M., Lohse, M. & Usadel, B. (2014) Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114-2120.

Chen, R. Shi, L., Hakenberg, J. *et al*. (2016) Analysis of 589,306 genomes identifies individuals resilient to severe Mendelian childhood diseases. *Nature Biotechnology* 34, 531–538.

Derks, M. F. L., Megens, H.-J., Bosse, M., Lopes, M. S., Harlizius, B. & Groenen, M. A. M. (2017) A systematic survey to identify lethal recessive variation in highly managed pig populations. *BMC Genomics* 18, 858.

Derks, M. F. L., Gjuvsland, A. B., Bosse, M. *et al.* (2019) Loss of function mutations in essential genes cause embryonic lethality in pigs. *PLoS Genetics* 15, e1008055.

Endo, F., Matsuura, T., Yanagita, K. & Matsuda, I. (2004) Clinical manifestations of inborn errors of the urea cycle and related metabolic disorders during childhood. *The Journal of Nutrition* 134, 1605S–1609S.

Frantz, L. A. F., Madsen, O., Megens, H.-J., Schraiber, J. G., Paudel, Y., Bosse, M., Crooijmans, R. P. M. A., Larson, G. & Groenen M. A. M. (2015) Evolution of Tibetan wild boars. *Nature Genetics* 47, 188–189.

Gallardo, D., Pena, R. N., Amills, M., Varona, L., Ramírez, O., Reixach, J., Díaz, I., Tibau, J., Soler, J., Prat-Cuffi, J. M., Noguera, J. L. & Quintanilla, R. (2008) Mapping of quantitative trait loci for cholesterol, LDL, HDL, and triglyceride serum concentrations in pigs. *Physiological Genomics* 35, 199–209.

Gallardo, D., Quintanilla, R., Varona, L., Díaz, I., Ramírez, O., Pena, R. N. & Amills, M. (2009) Polymorphism of the pig *acetyl-coenzyme A carboxylase α* gene is associated with fatty acid composition in a Duroc commercial line. *Animal Genetics* 40, 410–417.

Groenen, M. A. M., Archibald, A. L., Uenishi, H. *et al.* (2012) Analyses of pig genomes provide insight into porcine demography and evolution. *Nature* 491, 393–398.

Häberle J., Pauli S., Schmidt E., Schulze-Eilfing, B., Berning, C. & Koch, H. G. (2003) Mild citrullinemia in Caucasians is an allelic variant of argininosuccinate synthetase deficiency (citrullinemia type 1). *Molecular Genetics and Metabolism* 80, 302-306.

Hill, W. G. & Robertson, A. (1968) Linkage disequilibrium in finite populations. *Theoretical and Applied Genetics* 38, 226-31.

Li, H. (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997v2[q-bio.GN]

MacArthur, D. G., Balasubramanian, S., Frankish, A. *et al.* (2012) A systematic survey of loss-of-function variants in human protein-coding genes. *Science* 335, 823–8.

Makino, T., Rubin, C.-J., Carneiro, M., Axelsson, E., Andersson, E. & Webster, M. T. (2018) Elevated proportions of deleterious genetic variation in domestic animals and plants. *Genome Biology and Evolution* 10, 276–290.

Manunza, A., Casellas, J., Quintanilla, R. *et al.* (2014) A genome-wide association analysis for porcine serum lipid traits reveals the existence of age-specific genetic determinants. *BMC Genomics* 15, 758.

McKenna, A., Hanna, M., Banks, E. *et al.* (2010) The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* 20, 1297-1303.

Narasimhan, V. M., Xue, Y. & Tyler-Smith, C. (2016) Human knockout carriers: dead, diseased, healthy or improved? *Trends in Molecular Medicine* 22, 341–351.

Polato, F., Callen, E., Wong, N. *et al.* (2014) CtIP-mediated resection is essential for viability and can operate independently of *BRCA1*. *The Journal of Experimental Medicine* 211, 1027-36.

Rausell, A., Mohammadi, P., McLaren, P. J., Bartha, I., Xenarios, I., Fellay, J. & Telenti, A. (2014) Analysis of stop-gain and frameshift variants in human innate immunity genes. *PLoS Computational Biology* 10, e1003757.

Tamura K., Stecher G., Peterson D., Filipski, A. & Kumar, S. (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution* 30, 2725-2729.

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M. & Rozen, S. G. (2012) Primer3—new capabilities and interfaces. *Nucleic Acids Research* 40, e115–e115.

Warr, A., Affara, N., Aken, B. *et al.* (2019) An improved pig reference genome sequence to enable pig genetics and genomics research. bioRxiv 668921.

Polymorphisms of the cryptochrome 2 and mitoguardin 2 genes are associated with the variation of lipid-related traits in Duroc pigs

Mármol-Sánchez, E.[1], Quintanilla, R.[2], Cardoso, T. F.[1,3], Jordana Vidal, J.[4] and Amills, M.[1,4,*]

[1]Department of Animal Genetics, Centre for Research in Agricultural Genomics (CRAG), CSIC-IRTA-UAB-UB, Campus de la Universitat Autònoma de Barcelona, Bellaterra, Spain. [2]Animal Breeding and Genetics Programme, Institute for Research and Technology in Food and Agriculture (IRTA), Torre Marimon, Caldes de Montbui, Spain. [3]CAPES Foundation, Ministry of Education of Brazil, Brasilia, D. F., Brazil. [4]Departament de Ciència Animal i dels Aliments, Universitat Autònoma de Barcelona, Bellaterra, Spain.

[*]Corresponding author: Marcel Amills. marcel.amills@uab.cat

Sci. Rep. (2019) 9, 9025. doi: 10.1038/s41598-019-45108-z.

Abstract

The genetic factors determining the phenotypic variation of porcine fatness phenotypes are still largely unknown. We investigated whether the polymorphism of eight genes (*MIGA2*, *CRY2*, *NPAS2*, *CIART*, *ARNTL2*, *PER1*, *PER2* and *PCK1*), which display differential expression in the skeletal muscle of fasted and fed sows, is associated with the variation of lipid and mRNA expression phenotypes in Duroc pigs. The performance of an association analysis with the GEMMA software demonstrated that the rs330779504 SNP in the *MIGA2* gene is associated with LDL concentration at 190 days (LDL_2, corrected *P*-value = 0.057). Moreover, the rs320439526 SNP of the *CRY2* gene displayed a significant association with stearic acid content in the *longissimus dorsi* muscle (LD C18:0, corrected *P*-value = 0.015). Both SNPs were also associated with the mRNA levels of the corresponding genes in the *gluteus medius* skeletal muscle. From a biological perspective these results are meaningful because *MIGA2* protein plays an essential role in mitochondrial fusion, a process tightly connected with the energy status of the cell, while *CRY2* is a fundamental component of the circadian clock. However, inclusion of these two SNPs in chromosome-wide association analyses demonstrated that they are not located at the peaks of significance for the two traits under study (LDL_2 for rs330779504 and LD C18:0 for rs320439526), thus implying that these two SNPs do not have causal effects.

Introduction

The genome-wide analysis of gene expression data obtained from RNA-seq experiments can provide valuable information in order to understand the biology of production phenotypes and how they are genetically regulated. Cardoso *et al*.[1] compared the muscle transcriptomic profiles of Duroc sows before and after feeding and, in doing so, they demonstrated that the ingestion of food is associated with changes in the mRNA levels of several circadian genes including the cryptochrome circadian regulator 2 (*CRY2*), neuronal PAS domain protein 2 (*NPAS2*), circadian associated repressor of transcription (*CIART*), aryl hydrocarbon receptor nuclear translocator like 2 (*ARNTL2*), period circadian regulator 1 (*PER1*) and period circadian regulator 2 (*PER2*). The identification of circadian clock regulator genes is particularly relevant because they have been broadly reported as major contributors to lipid metabolism and energy homeostasis[2,3,4,5,6,7], driving changes in the expression of multiple transcripts and modulating cell response to different stimuli such as food intake[5,8,9]. Two other interesting genes identified by Cardoso *et al*.[1] as differentially expressed before and after eating were mitoguardin 2 (*MIGA2*), which regulates mitochondrial fusion[10], a process tightly connected with energy homeostasis[11], and phosphoenolpyruvate carboxykinase 1 (*PCK1*), an enzyme fundamental for the maintenance of glucose and lipid levels[12].

The expression of the eight genes mentioned above (*ARNTL2*, *CIART*, *CRY2*, *NPAS2*, *PER1*, *PER2*, *PCK1* and *MIGA2*) is affected by food intake and there is ample evidence that they have a key role in carbohydrate and lipid metabolism[8,10,13,14,15]. The main hypothesis that we aim to test in the current work is whether the variability of these eight genes is associated with lipid phenotypes recorded in a Duroc pig population denominated as Lipgen (Supplementary Table 1). To achieve this goal, we have first identified a total of 20 polymorphisms (Table 1) in these eight genes by using a previously published RNA-seq data set corresponding to 52 pigs from the Lipgen population[16]. These 20 SNPs have been genotyped in 345 pigs from the Lipgen population with available records for a broad array of lipid traits listed in Supplementary Table 1, *i.e*. serum lipid concentrations[17,18], *longissimus dorsi* (LD) and *gluteus medius* (GM) muscle fatty acid composition[19] and backfat thickness. Subsequently, those SNPs showing significant associations (after correction for multiple testing) with a given lipid trait, have been further studied by investigating if they are associated with gene expression as well as by performing chromosome-wide association analyses based on Porcine SNP60 BeadChip data. The liver and GM muscle mRNA expression data sets[18,20] and the Porcine SNP60 BeadChip genotypes[18,21] used for this purpose were generated in previous studies (Supplementary Table 2).

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

Gene	Sscrofa.11.1					
	SSC	Start	End	Strand	SNP	Effect
MIGA2	1	269338125	269362328	+	rs322533788	Splice region
					rs80923452	Splice region
					rs80832336	Splice region
					rs330779504	Splice region
CRY2	2	16584431	16620669	-	rs320439526	5'-UTR
NPAS2	3	53295744	53418480	-	rs335603631	Missense
CIART	4	98797378	98814553	-	rs322666984	Missense
ARNTL2	5	46330522	46482280	-	rs326158774	Splice region
PER1	12	53341329	53376723	-	rs699427837	Missense
					rs345340955	Splice region
					rs81436952	Missense
PER2	15	137793421	137836268	-	rs344440225	Splice region
					rs329662925	Missense
					rs324793161	Missense
					rs80910874	Missense
					rs325502974	Missense
PCK1	17	57930356	57938817	+	rs343196765	Missense
					rs331782052	Missense
					rs345064848	Splice region
					rs320568163	Missense

Results

Association analyses for lipid traits

Previous data sets employed for making the association analyses with a wide variety of lipid-related traits are listed in Supplementary Table 2. Performance of association analyses between the 20 selected SNPs and the phenotypes listed in Supplementary Table 1 allowed us to identify several associations that were significant at the nominal level (Table 2). Three SNPs in the *PER1* gene were associated with LD and GM C18:3, and there was also an association between the *CIART* genotype and backfat thickness. Two SNPs in the *PCK1* gene were associated with LD C17:0, and *CRY2* and *MIGA2* genotypes showed associations with several serum lipid and fatty acid composition traits. These results were consistent with the relevant role of the genes under study on metabolism and energy homeostasis. However, only two associations remained significant after correction for multiple testing. The serum concentration of low-density-lipoproteins (LDL) measured at ~190 days was significantly associated with the rs330779504 SNP (Table 2), a splice region variant located in the beginning of intron 14 (1:269.360 Mb) of the mitoguardin 2 gene (*MIGA2*). Pigs inheriting the A-allele showed an increased LDL cholesterol concentration (Figure 1A), with homozygous AA animals having a higher median blood LDL concentration (69.35 mg/dL) than GA (61.75 mg/dL) and GG (58.40 mg/dL) individuals. Kruskal-Wallis ranking test for differences in median LDL concentrations yielded a *P*-value of 5.14E-03 (Supplementary Table 3), thus supporting the existence of significant differences among the three rs330779504 genotypes. Besides, this *MIGA2* polymorphism also displayed an additive effect on palmitic acid content in LD muscle, total serum cholesterol concentration at ~190 days of age and the ratio between omega-6 and omega-3 desaturation in LD, but only at the nominal *P*-value level of significance (Table 2). The proportion of variance in LDL cholesterol concentration explained by rs330779504 genotype was 2.16% (SE = 0.03%).

The other association that remained significant after correction for multiple testing was that between rs320439526 genotype and stearic acid content (C18:0) of the LD muscle (Table 2). This polymorphism is located in the 5′ end of the *CRY2* gene, and it was annotated as having a putative stop gain effect in the former *Sus scrofa* assembly record (Sscrofa10.2). This led us to select it due to the high impact effect that the inactivation of this gene could have on the regulation of circadian clock rhythms and many other relevant metabolic processes. However, when interrogated in the last available assembly release for the porcine genome (Sscrofa11.1), this variant appeared to be located in the 5′-UTR of the *CRY2* gene. The Kruskal-Wallis ranking test for differences in median C18:0 content in the LD muscle yielded a *P*-value of 5.71E-03 (Supplementary Table 3), with homozygous TT pigs having a higher median stearic acid content (12.52%) than their CT (11.54%) and CC (11.30%) counterparts (Figure 1C). The proportion of variance in stearic acid content in LD muscle explained by the rs320439526 genotype was 8.87% (SE = 0.04%).

Table 2: Polymorphisms significantly associated with lipid-related traits[a].

Gene	SNP	Type	Trait	*P*-value	*q*-value	δ ± SE	A$_1$	MAF
MIGA2^1	**rs330779504 (1:269.360 Mb)**	**Splice region variant (G/A)**	**LDL$_2$**	**2.71E-03**	**5.69E-02**	**6.26 (2.01)**	**A**	**0.236**
			LD (C16:0)	7.89E-03	1.66E-01	-0.43 (0.14)		
			TotalCholest$_2$	3.06E-02	2.73E-01	5.86 (2.53)		
			LDFAn6/FAn3	2.91E-02	6.10E-01	-1.00 (0.42)		
	rs80832336 (1:269.359 Mb)	Splice region variant (C/T)	LD (C16:0)	7.06E-02	3.63E-01	-0.24 (0.13)	T	0.381
			TotalCholest$_2$	2.69E-02	2.73E-01	5.03 (2.13)		
	rs322533788 (1:269.341 Mb)	Splice region variant (T/C)	GM (C10:0)	3.77E-02	4.32E-01	-0.01 (0.01)	C	0.093
			GM (C20:0)	5.46E-03	1.15E-01	-0.04 (0.01)		
CRY2^1	**rs320439526 (2:16.620 Mb)**	**5'-UTR variant (C/T)**	**LD (C18:0)**	**7.04E-04**	**1.48E-02**	**0.39 (0.12)**	**T**	**0.353**
			TotalCholest$_2$	4.22E-02	2.73E-01	-4.74 (2.23)		
			LDUFA	3.10E-02	3.82E-01	-0.43 (0.20)		
			LDSFA	3.69E-02	3.75E-01	0.42 (0.20)		
			LD (C16:1)	2.05E-02	4.31E-01	-0.12 (0.05)		
CIART	rs322666984 (4:98.801 Mb)	Missense variant (G/C)	BFT$_1$	4.07E-03	8.54E-02	-3.20 (1.12)	C	0.214
PER1	rs81436952 (12:53.368 Mb)	Missense variant (C/T)	LD (C18:3)	1.16E-02	1.22E-01	0.04 (0.02)	C	0.058
			GM (C18:3)	4.11E-02	2.96E-01	0.04 (0.02)		
	rs699427837 (12:53.365 Mb)	Missense variant (A/G)	LD (C18:3)	9.65E-03	1.22E-01	0.04 (0.02)	G	0.059
			GM (C18:3)	4.22E-02	2.96E-01	0.04 (0.02)		
	rs345340955 (12:53.368 Mb)	Splice region variant (A/T)	LD (C18:3)	2.33E-02	1.62E-01	0.04 (0.02)	T	0.054
			GM (C18:3)	1.33E-02	2.96E-01	0.05 (0.02)		
PCK1	rs320568163 (17:57.936 Mb)	Missense variant (A/G)	LD (C17:0)	2.23E-02	2.70E-01	-0.02 (0.01)	G	0.144
	rs331782052 (17:57.933 Mb)	Missense variant (A/G)	LD (C17:0)	2.57E-02	2.70E-01	-0.02 (0.01)	G	0.138

[a] SNPs in bold show associations that remained significant after correction for multiple testing; *q*-value: *q*-value calculated with the false-discovery rate (FDR) method; δ: estimated allele substitution effect and standard error (SE); A$_1$: minor allele, MAF: Minor allele frequency; LD: *longissimus dorsi* muscle, GM: *gluteus medius* muscle; trait acronyms are defined in Supplementary Table 1.

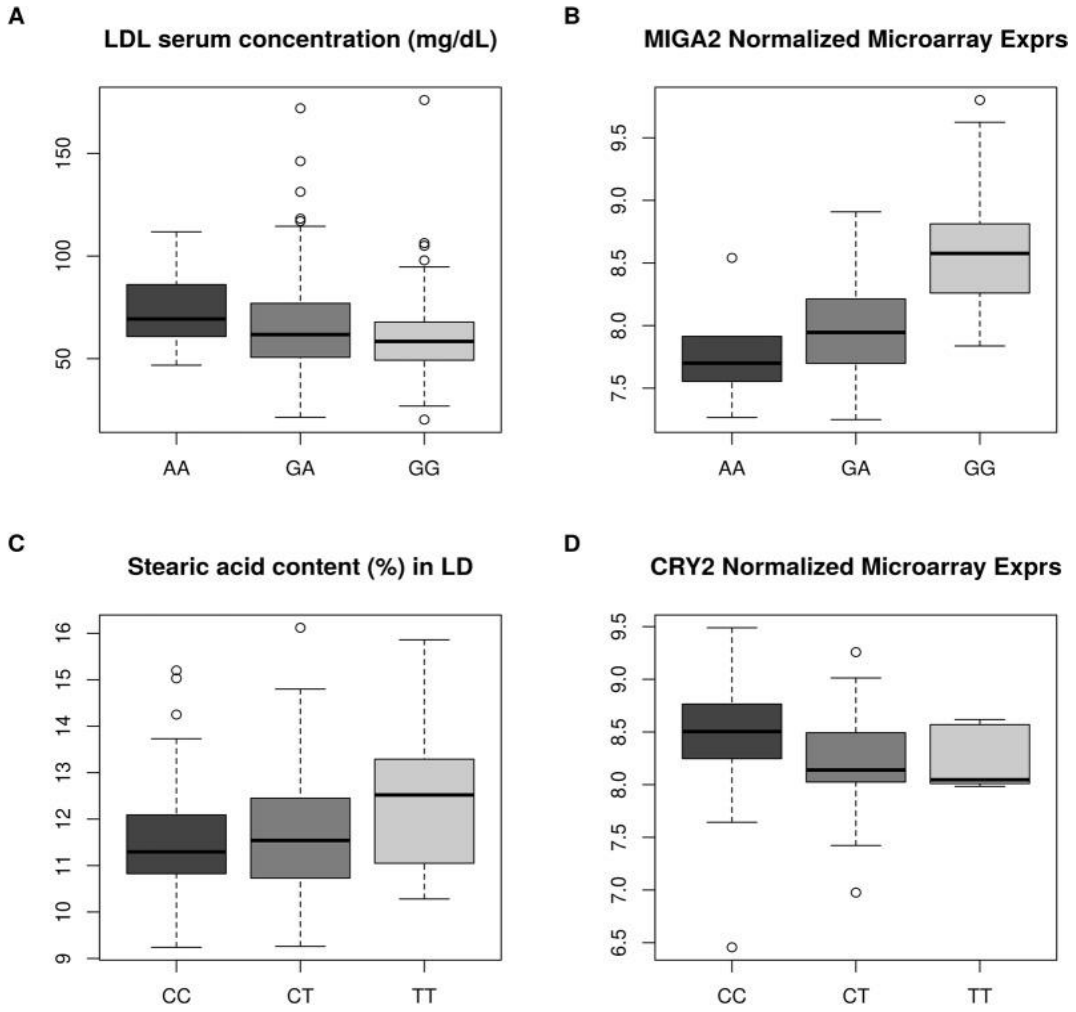


Figure 1: (**A**) Boxplots depicting the median and the distribution of serum low density lipoprotein concentrations at ~190 days for each one of the three rs330779504 genotypes: GG (N = 191), GA (N = 125) and AA (N = 16). (**B**) Boxplots depicting the median and the distribution of *MIGA2* mRNA expression levels in the *gluteus medius* skeletal muscle for each one of the three rs330779504 genotypes: GG (N = 48), GA (N = 33) and AA (N = 6). (**C**) Boxplots depicting the median and the distribution of stearic acid (C18:0) content in LD skeletal muscle for each one of the three rs320439526 genotypes: CC (N = 135), CT (N = 161) and TT (N = 37). **(D)** Boxplots depicting the median and the distribution of *CRY2* mRNA expression levels in the *gluteus medius* skeletal muscle for each one of the three rs320439526 genotypes: CC (N = 37), CT (N = 45) and TT (N = 6).

Polymorphisms in the *MIGA2* and *CRY2* genes are associated with mRNA expression levels

To gain new insights into the molecular basis of the two significant associations found (Table 2), we investigated whether the rs330779504 and the rs320439526 SNPs are associated with the mRNA expression of the *MIGA2* and *CRY2* genes, respectively. Previously reported hepatic and GM muscle microarray data sets[18,20] were employed for this purpose (Supplementary Table 2). Analysis with the GEMMA software revealed a significant association between the rs330779504 polymorphism and *MIGA2* mRNA expression levels in the GM muscle (Table 3). Pigs inheriting the A-allele of the rs330779504 polymorphism showed a reduced *MIGA2* mRNA expression (Figure 1B). Performance of a test based on the analysis of variance (ANOVA) confirmed the existence of statistically significant differences among genotypes (Supplementary Table 4). Moreover, a weak but significant association between the SNP rs330779504 and one of the probes defining liver *MIGA2* mRNA expression was also found (Table 3). With regard to the *CRY2* gene, when we performed an association analysis with the GEMMA software, the rs320439526 5′-UTR variant happened to be significantly associated with the expression of the corresponding gene in the GM muscle (Table 3). When we compared the *CRY2* mRNA levels corresponding to each one of the three rs320439526 genotypes (Figure 1D) by using an ANOVA test, we found differences that almost reached significance (Supplementary Table 4).

Inclusion of significant SNPs in a chromosome-wide association analysis

After demonstrating that in the Lipgen population the rs330779504 (*MIGA2*) and rs320439526 (*CRY2*) SNPs are associated with serum LDL concentration at ~190 days and LD C18:0, respectively, we aimed to investigate whether other SNP markers located in the vicinity of rs330779504 and rs320439526 display associations with these two traits with a higher level of significance than those observed for rs330779504 and rs320439526. To achieve this goal, we merged the rs330779504 SNP with 7,188 SNPs mapping to pig chromosome 1 (SSC1) and the rs320439526 SNP with 3,684 SNPs mapping to SSC2. The SSC1 and SSC2 SNP data were extracted from Porcine SNP60 BeadChip genotyping data reported by Manunza *et al.*[18] and González-Prendes *et al.*[21] in the Lipgen population (Supplementary Table 2). The associations between the markers rs330779504 (*MIGA2*) and rs320439526 (*CRY2*) with LDL serum concentration at ~190 days of age and with stearic acid content in LD, respectively, were only detected at the nominal level (Figure 2). Indeed, we did not find any significant association at the chromosome-wide level when correcting for multiple testing with the false discovery rate (FDR) approach[22] (Figure 2).

Table 3: Associations between *MIGA2* and *CRY2* genotypes and the mRNA levels of the corresponding genes estimated with microarrays in *gluteus medius* skeletal muscle and liver samples from Duroc pigs; δ: estimated allele substitution effect and standard error (SE); A_1: minor allele, MAF: Minor allele frequency; GM: *gluteus medius* skeletal muscle.

Gene	SNP	Type	Probe	Tissue	*P*-value	δ ± SE	A_1	MAF
MIGA2	rs330779504 (1:269.360 Mb)	Splice region variant (G/A)	Ssc.19153.2.A1_at	GM	8.11E-06	-0.39 (0.08)	A	0.236
				LIVER	5.99E-01	0.05 (0.09)		
			Ssc.19153.1.S1_at	GM	2.78E-07	-0.53 (0.09)		
				LIVER	2.60E-02	0.16 (0.07)		
CRY2	rs320439526 (2:16.620 Mb)	5'-UTR variant (C/T)	Ssc.26267.1.S1_at	GM	3.01E-02	-0.19 (0.09)	T	0.353

Discussion

One of the main goals of our study was to evaluate whether the polymorphisms of six genes with critical roles in the regulation of the circadian rhythms (*CRY2*, *NPAS2*, *CIART*, *ARNTL2*, *PER1*, *PER2*) are associated with the variation of lipid traits recorded in 345 Duroc pigs (Lipgen population). Indeed, circadian clock genes have been broadly reported as major contributors to the regulation of lipid metabolism and maintenance of energy homeostasis[3,4,7], driving changes in the expression of multiple transcripts and thus causing differences in protein and enzymatic activity across the day-night cycle[23,24]. We have found multiple associations between the variability of the pig circadian genes and fatty acid composition traits, but the majority of them were only significant at the nominal level (Table 2). There are reports that indicate that there is a close relationship between the activity of circadian genes and fatty acid synthesis. For instance, fatty acid elongation is under circadian control because the cyclic acetylation of acetyl-CoA synthetase 1 by the SIRT1 deacetylase modulates the intracellular concentration of acetyl-CoA[25]. Moreover, alteration of circadian genes can potentially influence liver lipid metabolism in mice[26]. With regard to the circadian genes, the only association that remained significant after correction for multiple testing was that between the rs320439526 SNP of the *CRY2* gene and C18:0 content in the *longissimus dorsi* muscle (Table 2, Supplementary Table 3). This association is particularly interesting because it has been demonstrated that the *CRY1/2* genes can repress the peroxisome proliferator-activated receptor δ (*PPARD*) transcription factor, which has a fundamental role in lipid metabolism[15]. Indeed, the inhibition of *PPARD* by *CRY1/2* is expected to decrease the rates of fatty acids transport and oxidation in the skeletal muscle[15]. Besides, the *CRY2* gene has multiple effects on lipid metabolism. In response to a high-fat diet, CRY-deficient mice showed an increased body weight gain despite less feed consumption compared with wild-type animals, as a result of the activation of lipogenic pathways combined with increased insulin secretion and lipid storage[27], thus leading to obesity propensity when CRY regulatory function was disrupted. We have also shown that the rs320439526 SNP of the *CRY2* gene is associated with the expression of the *CRY2* mRNA in the *gluteus medius* muscle (Table 3, Supplementary Table 4), suggesting that this polymorphism, or a nearby mutation, has regulatory effects on the transcriptional rate of the *CRY2* gene. This polymorphism maps to the 5′-UTR of the *CRY2* gene, a region that can have broad transcriptional effects on gene expression by interacting with RNA-binding proteins[28]. However, the Ensembl annotation of the rs320439526 SNP does not predict any functional effect, so we favor the hypothesis that this SNP is linked to another mutation with regulatory effects on *CRY2* mRNA levels. The chromosome-wide (SSC2) association analysis depicted in Figure 2 clearly shows that rs320439526 is not the marker displaying the most significant association with LD C18:0, thus indicating that the associations detected in our study have been produced by the existence of linkage disequilibrium between the rs320439526 SNP and a causal mutation yet to be found.

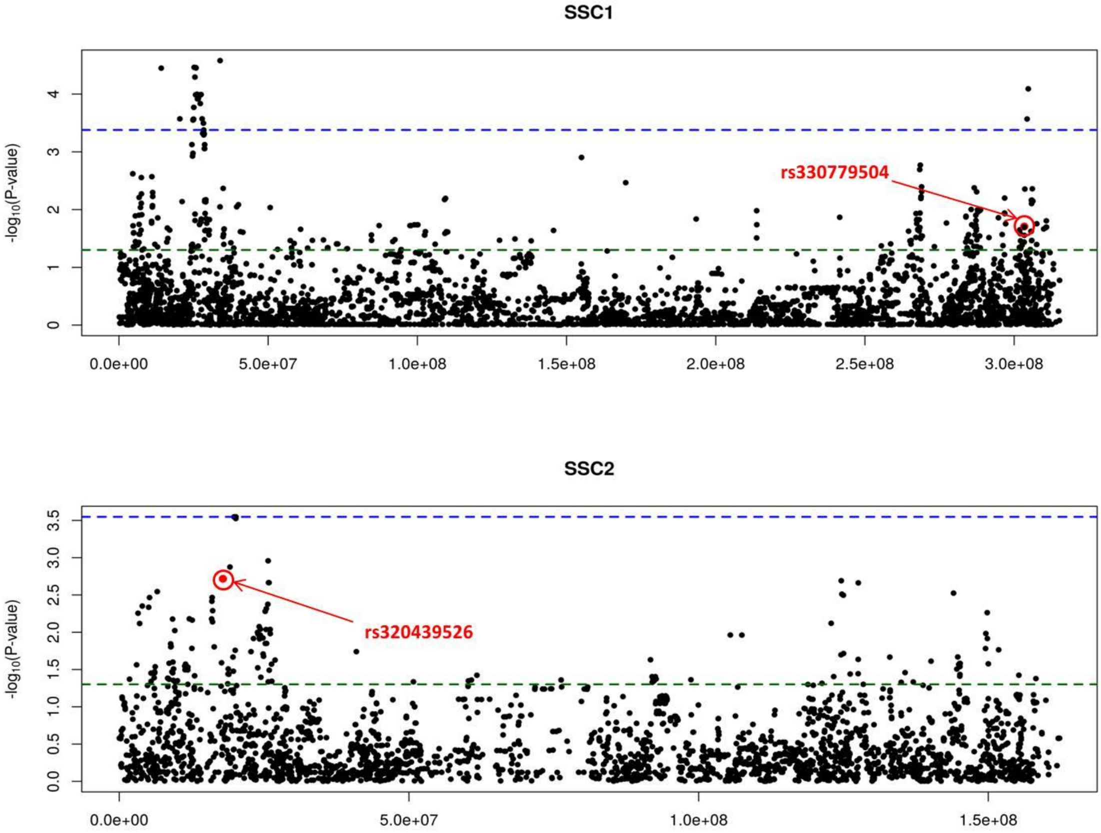


Figure 2: (**A**) Manhattan plot depicting the association of SNP rs330779504 and 7,188 additional SNPs mapping to pig chromosome 1 (SSC1) with serum low density lipoprotein concentration at ~190 days of age recorded in 345 Duroc pigs (Lipgen population). (**B**) Manhattan plot depicting the association of SNP rs320439526 and 3,684 additional SNPs mapping to pig chromosome 2 (SSC2) with stearic acid content in the *longissimus dorsi* muscle. The green line represents the nominal *P*-value of significance, while the blue line indicates the *P*-value of significance after correcting for multiple testing with an FDR test.

Another association that remained significant after correction for multiple testing was that between the *MIGA2* rs330779504 SNP and serum LDL concentrations at ~190 days (Table 2, Supplementary Table 3). Moreover, this SNP was also associated with *MIGA2* mRNA expression in the GM muscle and liver tissues (Table 3, Supplementary Table 4). The *MIGA2* gene, also known as *FAM73B*, and its homolog *MIGA1* (*FAM73A*) encode proteins localized to the outer membrane of mitochondria as membrane-integrated proteins and they have been previously associated with reduced body weight in mice[29] and variations in backfat thickness in pigs[30]. In a study performed by Zhang *et al.*[10], it was reported that MIGA1/2 proteins stabilize the dimeric complex formed by active MitoPLD, thus facilitating mitochondrial fusion[31]. Interestingly, the dynamics of mitochondrial fusion and fission is tightly related with the energy demand of cells. Indeed, nutrient abundance and starvation are associated with an increased frequency of fission and fusion events, respectively[27,32]. Besides, the capacity to produce ATP in response to changes in energy demand and supply is modulated by mitochondrial morphology[33]. A recent study reported that mitochondrial fusion induced by leptin could have important effects on the hepatic lipid accumulation[34], but to the best of our knowledge it is currently unknown whether mitochondrial fusion/fission has any effect on cholesterol and lipoprotein metabolism. Noteworthy, the chromosome-wide analysis pictured in Figure 2 evidenced that the association observed between the *MIGA2* rs330779504 marker and serum LDL levels at ~190 days is probably not causal, as there are some other neighboring SNPs that show more significant associations with this trait.

Conclusions

In this work, we wanted to test whether the variability of six circadian genes (*ARNTL2*, *CIART*, *CRY2*, *NPAS2*, *PER1* and *PER2*) and two additional genes (*MIGA2* and *PCK1*) with key roles in energy homeostasis is associated with a set of lipid phenotypes recorded in Duroc pigs (Lipgen population). We have observed multiple associations between the variation of circadian genes and muscle fatty acid composition, but only that between the rs320439526 SNP of the *CRY2* gene and LD C18:0 content remained significant after correction for multiple testing. We have also detected a significant association between the rs330779504 SNP of the *MIGA2* gene and LDL concentration at 190 days. In the light of the results of the chromosome-wide analyses, we conclude that none of these two associations are causal.

Methods

Ethics approval

Animal care and management procedures were performed following Spanish national guidelines for the Good Experimental Practices and they were approved by the Ethical Committee of the Institut de Recerca i Tecnologia Agroalimentàries (IRTA).

Animal material and phenotype recording

As previously reported by Gallardo *et al.*[17,35] a total of 345 Duroc barrows belonging to 5 half-sib families and distributed in 4 fattening batches were selected from a commercial pig line, devoted to high quality meat production. This line is characterized by its high content of intramuscular fat, a feature that results in the improvement of meat juiciness and taste, hence conferring a better consumer acceptance[36]. Pigs were bred under intensive conditions of feeding and handling, and slaughtered when they reached 122 kg of live weight (~190 days of age). Phenotypic measures for different traits (Supplementary Table 1) were recorded during the productive cycle or after slaughtering: Triglycerides (TG), total cholesterol (TotalCholest), high-density lipoprotein (HDL) and low-density lipoprotein (LDL) serum concentrations at ~45 and ~190 days of age as reported by Gallardo *et al.*[17], whereas intramuscular fat content in the LD and GM muscles and fatty acid composition for LD and GM were determined as described by Quintanilla *et al.*[19].

Selection and genotyping of twenty SNPs mapping to eight candidate genes

Based on the results reported by Cardoso *et al.*[1], we took into consideration eight genes that showed differential expression before and after feeding and that, moreover, play important roles in metabolism and circadian clock regulation (Table 1). The variability of these 8 genes was characterized by using, as a source of information, RNA-seq data results from 52 Duroc pigs retrieved from the same population analyzed herewith (Supplementary Table 2). Single nucleotide polymorphisms within selected genes were retrieved among variant calling results from sequences generated by Cardoso *et al*.[16].

Variant discovery analyses were performed by following the *GATK Best Practices workflow for SNP calling*:

(https://software.broadinstitute.org/gatk/documentation/article.php?id=3891). Briefly, after read mapping, sequences were split into exon segments and intronic overhanging sequences hard-clipped. Mapping qualities were reassigned by using the SplitNCigarReads GATK tool (https://software.broadinstitute.org/gatk), and the Haplotype Caller tool (https://software.broadinstitute.org/gatk) was used to detect SNPs for each analyzed sample (N = 52). Variant effect prediction on detected polymorphisms was estimated by using the SnpEff software[37] and those that showed potential functional or regulatory effects (*i.e.* high impact, missense, splice site regions, 5′-UTR) within selected genes were kept for genotyping. Moreover, we also selected 3 SNPs showing potential functional or regulatory effects (rs322533788, rs335603631 and rs326158774) that were retrieved from the Ensembl database (https://www.ensembl.org). A total of 20 selected SNPs and their flanking sequences (60 nucleotides upstream and downstream), were submitted to the Custom TaqMan Assay Design Tool website (https://www5.appliedbiosystems.com /tools/cadt/; Life Technologies) to ascertain if they were amenable to genotyping in a TaqMan Open Array multiplex assay platform. Genotyping was performed at the Servei Veterinari de Genètica Molecular at the Universitat Autònoma de Barcelona (http://sct.uab.cat/svgm/en) by using a QuantStudio 12 K flex Real-Time PCR System (ThermoFisher Scientific).

Association analyses between twenty selected SNPs and porcine lipid-related traits

The PLINK software[38] was used for processing genotyped data. Association analyses between genotyped polymorphisms and phenotypes were performed with the genome-wide efficient mixed model association (GEMMA) software[39]. This package uses a mixed model approach to account for population stratification and relatedness by calculating a genomic kinship matrix with SNPs genotypes as random effects and provides an exact test of significance. We implemented a univariate mixed model as follows:

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

where y is the vector of phenotypic observations for every individual; α corresponds to a vector including the intercept plus the fixed effects, *i.e.* batch effect with 4 categories (all traits), farm origin effect with 3 categories (all traits), data of extraction with 2 categories within batch (only for TotalCholest, TG, HDL and LDL serum concentration, that were measured at approximately 45 and 190 days). The α vector also contains the regression coefficients of the following covariates: live weight at slaughterhouse for TotalCholest, TG, HDL and LDL serum concentrations, and IMF content in LD and GM for LD and GM fatty acid composition respectively; W is the incidence matrix relating phenotypes with the corresponding effects; x is the vector of the genotypes corresponding to the set of selected polymorphisms; δ is the allele substitution effect for each polymorphism; u is a vector of random individual effects with a n-dimensional multivariate normal distribution MVN_n (0, $\lambda\ \tau^{-1}$ K), where τ^{-1} is the variance of the residual errors, λ is the ratio between the two variance components and K is a known relatedness matrix derived from the SNPs; and $\boldsymbol{\varepsilon}$ is the vector of residual errors.

The association between lipid-related traits and the twenty analyzed polymorphisms was assessed on the basis of the estimated allele substitution effects. The significance of these effects was established by implementing a correction for multiple testing using the FDR method reported by Benjamini and Hochberg[22]. Moreover, we compared the phenotypic medians corresponding to each one of the three possible genotypes by applying the non-parametric Kruskal-Wallis test, due to the non-normal data distribution of lipid phenotypes under study.

Association analyses between the rs330779504 and rs320439526 polymorphisms and the expression of the genes that contain them

Gluteus medius skeletal muscle and liver samples were collected from 103 Duroc pigs belonging to the Lipgen population. Samples were retrieved after slaughtering and immediately frozen at −80 °C in liquid nitrogen. Total RNA was isolated from GM samples by using the TRIzol method[40] and the RiboPure kit (Ambion, Austin, TX) following manufacturer's recommendations. Transcriptomic mRNA expression profiles were then assessed by hybridization to the GeneChip Porcine arrays (Affymetrix Inc., Santa Clara, CA), as previously reported by Cánovas *et al.*[20]. Expression data corresponding to GM muscle and liver samples are deposited in NCBI's Gene Expression Omnibus[41] and can be accessed through GEO Series accession number GSE115484 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE115484). Data pre-processing, background correction, normalization and log-transformation of expression values between samples were carried out by computing a Robust Multi-array Average (RMA) as described by Irizarry *et al.*[42].

The correspondence between genes and microarray expressed probes was assessed with the Biomart database available at Ensembl repositories (https://www.ensembl.org/biomart/martview/). Expression levels for selected genes were then extracted from microarray samples for both GM muscle and liver tissues and used as continuous variables in association analyses, following the same statistical model previously described for phenotype records and correcting for batch (4 categories), farm of origin (3 categories) and laboratory (2 categories) as fixed effects. Moreover, we compared the phenotypic means corresponding to each one of the three possible genotypes by applying an ANOVA test.

Inclusion of the *MIGA2* rs330779504 and *CRY2* rs320439526 SNPs in a chromosome-wide association analysis

As previously described by Manunza *et al.*[18] and González-Prendes *et al.*[21], the population employed in the current experiment was typed with the Porcine SNP60 BeadChip (Illumina, San Diego, CA) which contains probes for 62,163 SNPs (Supplementary Table 2). The GenomeStudio software (Illumina) was used for quality control analyses, as reported by Manunza *et al.*[18]. The PLINK software[38] was used for removing SNPs that (a) did not map to autosomal chromosomes, (b) had minor allele frequency (MAF) <0.05, (c) with rate of missing genotypes > 0.05, (d) major departures from the Hardy-Weinberg equilibrium (*P*-value = 0.001), (e) had a GenCall score < 0.15, (f) had a call rate < 0.95, or (g) that could not be mapped to the pig reference genome. A total of 36,710 SNPs were finally retrieved after filtering and merged with genotyping data corresponding to the rs330779504 and the rs320439526 SNPs. Association analyses were performed with the GEMMA software[39] as described before, and multiple testing correction was implemented with the FDR method[22] by establishing a chromosome-wide threshold of significance.

Supplementary Information

Supplementary Table 1: Analyzed phenotypic traits for 345 Duroc pigs belonging to the Lipgen population. LD: *longissimus dorsi* skeletal muscle; GM: *gluteus medius* skeletal muscle.

Supplementary Table 2: Sources of information for data described in the current work.

Supplementary Table 3: Phenotypic distribution by *MIGA2* and *CRY2* genotypes. Median ± SE: Median values ± standard error for LDL cholesterol serum concentration (mg/dL) and stearic acid content (%) in the *longissimus dorsi* skeletal muscle. KW *P*-value: Nominal *P*-value obtained with the Kruskal-Wallis test.

Supplementary Table 4: Normalized probe expression values for *MIGA2* and *CRY2* genotypes. Mean ± SE: Mean values ± standard error for the estimated normalized probe expression values; ANOVA *P*-value: Nominal *P*-value obtained with an ANOVA test.

Acknowledgements

The authors are indebted to Selección Batallé S.A. for providing the animal material. We gratefully acknowledge to J. Reixach (Selecció Batallé), J. Soler (IRTA), C. Millan (IRTA), A. Quintana (IRTA) and A. Rossell (IRTA) for their collaboration in the experimental protocols and pig management. We also want to thank MINECO for the Center of Excellence Severo Ochoa 2016–2019 (SEV-2015-0533) grant awarded to the Centre for Research in Agricultural Genomics (CRAG), and the CERCA Programme of the Generalitat de Catalunya for their support. Part of the research presented in this publication was funded by grants AGL2013–48742-C2–1-R and AGL2013–48742-C2–2-R awarded by the Spanish Ministry of

Economy and Competitivity, and grant 2014 SGR 1528 from the Agency for Management of University and Research Grants of the Generalitat de Catalunya. Tainã Figueiredo Cardoso was funded with a fellowship from the CAPES Foundation-Coordination of Improvement of Higher Education, Ministry of Education of the Federal Government of Brazil. Emilio Mármol-Sánchez was funded with an FPU pre-doctoral fellowship from the Spanish Ministry of Education (FPU15/01733).

Author Contributions

M.A., R.Q. and J.J. designed the experiment. R.Q. generated the animal material and collected the phenotypic and microarray data. E.M.S. and M.A. selected the SNPs to be genotyped. E.M.S. did all bioinformatic and statistical analyses. T.F.C. contributed to the analysis of gene expression data. E.M.S. and M.A. wrote the paper. All authors read and approved the content of the manuscript.

Data Availability

Expression data corresponding to GM muscle and liver samples are deposited at NCBI's Gene Expression Omnibus and are accessible through GEO Series accession number GSE115484: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE115484. Genotypes and phenotypes for the 345 Duroc pigs (Lipgen population) have been deposited in the Figshare public repository (https://figshare.com/s/2e636697009360986794).

References

1. Cardoso TF, et al. Nutrient supply affects the mRNA expression profile of the porcine skeletal muscle. BMC Genomics. 2017; 18:603.

2. Froy O. The relationship between nutrition and circadian rhythms in mammals. Front. Neuroendocrinol. 2007; 28:61–71.

3. Green CB, Takahashi JS, Bass J. The Meter of Metabolism. Cell. 2008; 134:728–742.

4. Laposky AD, Bass J, Kohsaka A, Turek FW. Sleep and circadian rhythms: Key components in the regulation of energy metabolism. FEBS Lett. 2008; 582:142–151.

5. Froy O, Miskin R. Effect of feeding regimens on circadian rhythms: implications for aging and longevity. Aging (Albany. NY). 2010; 2:7–27.

6. Paschos GK. Circadian clocks, feeding time, and metabolic homeostasis. Front. Pharmacol. 2015; 6:112.

7. McGinnis GR, Young ME. Circadian regulation of metabolic homeostasis: causes and consequences. Nat. Sci. Sleep. 2016; 8:163–80.

8. Patel SA, Velingkaar N, Makwana K, Chaudhari A, Kondratov R. Calorie restriction regulates circadian clock gene expression through BMAL1 dependent and independent mechanisms. Sci. Rep. 2016; 6:25970.

9. Chaudhari A, Gupta R, Makwana K, Kondratov R. Circadian clocks, diets and aging. Nutr. Heal. aging. 2017; 4:101–112.

10. Zhang Y, et al. Mitoguardin regulates mitochondrial fusion through MitoPLD and is required for neuronal homeostasis. Mol. Cell. 2016; 61:111–24.

11. Westermann B. Bioenergetic role of mitochondrial fusion and fission. Biochim. Biophys. Acta - Bioenerg. 2012; 1817:1833–1838.

12. Millward CA, et al. Phosphoenolpyruvate carboxykinase (*Pck1*) helps regulate the triglyceride/fatty acid cycle and development of insulin resistance in mice. J. Lipid Res. 2010; 51:1452–1463.

13. Grimaldi B, et al. PER2 controls lipid metabolism by direct regulation of PPARγ Cell Metab. 2010; 12:509–20.

14. Machicao F, et al. Glucose-raising polymorphisms in the human Clock gene Cryptochrome 2 (CRY2) affect hepatic lipid content. PLoS One. 2016; 11:e0145563.

15. Jordan SD, et al. CRY1/2 Selectively repress PPARδ and limit exercise capacity. Cell Metab. 2017; 26:243–255.e6.

16. Cardoso TF, et al. RNA-seq based detection of differentially expressed genes in the skeletal muscle of Duroc pigs with distinct lipid profiles. Sci. Rep. 2017; 7:40005.

17. Gallardo D, et al. Mapping of quantitative trait loci for cholesterol, LDL, HDL, and triglyceride serum concentrations in pigs. Physiol. Genomics. 2008; 35:199–209.

18. Manunza A, et al. A genome-wide association analysis for porcine serum lipid traits reveals the existence of age-specific genetic determinants. BMC Genomics. 2014; 15:758.

19. Quintanilla R, et al. Porcine intramuscular fat content and composition are regulated by quantitative trait loci with muscle-specific effects. J. Anim. Sci. 2011; 89:2963–71.

20. Cánovas A, et al. Segregation of regulatory polymorphisms with effects on the *gluteus medius* transcriptome in a purebred pig population. PLoS One. 2012; 7:e35583.

21. González-Prendes R, et al. Joint QTL mapping and gene expression analysis identify positional candidate genes influencing pork quality traits. Sci. Rep. 2017; 7:39830.

22. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B. 1995; 57:289–300.

23. Ruiter M, et al. The daily rhythm in plasma glucagon concentrations in the rat is modulated by the biological clock and by feeding behavior. Diabetes. 2003; 52:1709–15.

24. Kumar Jha P, Challet E, Kalsbeek A. Circadian rhythms in glucose and lipid metabolism in nocturnal and diurnal mammals. Mol. Cell. Endocrinol. 2015; 418:74–88.

25. Sahar S, et al. Circadian control of fatty acid elongation by SIRT1 protein-mediated deacetylation of acetyl-coenzyme A synthetase 1. J. Biol. Chem. 2014; 289:6091–6097.

26. Green CB, et al. Loss of Nocturnin, a circadian deadenylase, confers resistance to hepatic steatosis and diet-induced obesity. Proc. Natl. Acad. Sci. 2007; 104:9888–9893.

27. Putti R, Sica R, Migliaccio V, Lionetti L. Diet impact on mitochondrial bioenergetics and dynamics. Front. Physiol. 2015; 6:109.

28. Mignone F, Gissi C, Liuni S, Pesole G. Untranslated regions of mRNAs. Genome Biol. 2002; 3:reviews0004.

29. Bassett JHD, et al. Rapid-throughput skeletal phenotyping of 100 knockout mice identifies 9 new genes that determine bone strength. PLoS Genet. 2012; 8:e1002858.

30. Lee K-T, et al. Neuronal genes for subcutaneous fat thickness in human and pig are identified by local genomic sequencing and combined SNP association study. PLoS One. 2011; 6:e16356.

31. Choi S-Y, et al. A common lipid links Mfn-mediated mitochondrial fusion and SNARE-regulated exocytosis. Nat. Cell Biol. 2006; 8:1255–1262.

32. Vamecq J, et al. Mitochondrial dysfunction and lipid homeostasis. Curr. Drug Metab. 2012; 13:1388–1400.

33. Schrepfer E, Scorrano L. Mitofusins, from mitochondria to metabolism. Mol. Cell. 2016; 61:683–694.

34. Hsu W-H, Lee B-H, Pan T-M. Leptin-induced mitochondrial fusion mediates hepatic lipid accumulation. Int. J. Obes. 2015; 39:1750–6.

35. Gallardo D, et al. Polymorphism of the pig *acetyl-coenzyme A carboxylase α* gene is associated with fatty acid composition in a Duroc commercial line. Anim. Genet. 2009; 40:410–417.

36. Wood JD, et al. Fat deposition, fatty acid composition and meat quality: A review. Meat Sci. 2008; 78:343–358.

37. Cingolani P, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. Fly. 2012; 6:80–92.

38. Purcell S, et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. Am. J. Hum. Genet. 2007; 81:559–575.

39. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. Nat. Genet. 2012; 44:821–824.

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

41. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. 2002; 30:207–10.

42. Irizarry RA, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics. 2003; 4:249–264

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

Mármol-Sánchez, E.[1], Guan, D.[1], Quintanilla, R.[2], Tonda, R.[3] and Amills, M.[1,4]

[1]Centre for Research in Agricultural Genomics (CRAG), CSIC-IRTA-UAB-UB, Universitat Autònoma de Barcelona, 08193 Bellaterra, Spain. [2]Animal Breeding and Genetics Program, Institute for Research and Technology in Food and Agriculture (IRTA), Torre Marimon, 08140 Caldes de Montbui, Spain. [3]CNAG-CRG, Centre for Genomic Regulation (CRG), Barcelona Institute of Science and Technology (BIST), Barcelona, Spain. [4]Departament de Ciència Animal i dels Aliments, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain.

[*]Corresponding author: Marcel Amills. marcel.amills@uab.cat

Genet. Sel. Evol. Under Review. doi: 10.1101/2020.04.17.038315.

Abstract

Background

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

Results

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

Conclusions

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

mapping within and outside the seed region could have important regulatory effects on gene expression.

Background

Mature microRNA transcripts (miRNAs) are short (~22 nt) non-coding RNAs which play an essential role in the regulation of gene expression [1]. During the biogenesis of miRNAs, one strand of the miRNA duplex binds to the guide-strand channel of an Argonaute protein forming a miRNA-induced silencing complex (miRISC) with the ability of repressing mRNA expression through binding to specific 3'-UTR target sites [2]. In postembryonic cells, this repressor mechanism mainly acts by destabilizing the mRNA through decapping and poly(A)-tail shortening [3,4] and less often by hindering translation [2]. The binding of the miRNA to its 3'-UTR target site depends critically on the sequence of the seed region, which encompasses nucleotides (nt) 2nd to 8th from the 5'end of the miRNA and interacts with the target site through Watson-Crick pairing [1]. Polymorphisms located within the seed region have the potential to cause strong effects on target recognition, due to the emergence of novel target sites complementary to the mutated seed and the ablation of canonical wild-type miRNA-mRNA interactions [5,6]. Additionally, polymorphisms affecting miRNA binding sites in the 3'-UTR of targeted mRNAs can also notably influence the miRNA-dependent expression of these genes [5,7,8]. Nevertheless, imperfect seed matches can still be compensated by nucleotides 13th to 16th of the miRNA, thus providing additional anchor pairing to the seed region [2]. Other sites relevant for miRNA processing and function are the basal UG, flanking CNNC and apical UGU motifs [9], as well as a mismatched GHG motif (Roden et al., 2017), all of which contribute to facilitate miRNA processing [2].

Saunders et al. (2007) [5] investigated the variability of 474 human miRNA genes and found that SNP density within such loci (~1.3 SNPs per miRNA) is lower than that of surrounding regions (~3 SNPs). Moreover, they found that ~90% of human miRNA genes do not contain polymorphisms, and the majority of SNPs mapping to miRNAs are outside the seed region, thus evidencing that the variability of this critical functional motif evolves under strong selective constraints [5]. Indeed, polymorphisms in the first 14 nucleotides of the mature miRNA, and particularly those within the seed region, might abrogate the binding of the miRNA to its 3’-UTR targets, leading to an extensive rewiring of the miRNA-mediated regulatory network and, in some instances, to detrimental consequences [11]. Mammalian miRNA knockouts often display abnormal phenotypes, reduced viability and clinical disorders [2], although functional redundancy among miRNA family members might mitigate to some extent the severity of such manifestations [2]. Polymorphisms within miRNA loci laying outside the seed region can also affect the processing and stability of miRNAs during their maturation and loading into the functional silencing complex [12,13]. Moreover, Saunders et al. (2007) [5] showed that a broad array of predicted miRNA target sites in the 3’-UTR of mRNAs are polymorphic, a finding that suggests that purifying selection on these regions is less intense than in miRNA genes [11].

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

Methods

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

Retrieval of Porcine Whole-Genome Sequences

Whole-genome sequences from a total of 120 wild and domestic pigs (*Sus scrofa*) were retrieved from the NCBI sequence read archive (SRA) database (https://www.ncbi.nlm.nih.gov/sra). Detailed information about each sequenced porcine sample is available at Additional file 1: Table S1. The 120 selected genome sequences were classified into Asian domestic pigs (ADM, N = 40), Asian wild boars (AWB, N = 20), European domestic pigs (EDM, N = 40) and European wild boars (EWB, N = 20). More specifically, the European and Asian domestic individuals encompassed a selection of different representative porcine breeds, i.e. Meishan, Tongchen, Jinhua, Rongchan, Wuzhishan, Tibetan, Sichuan, Hetao, Minzhu, Bamaixang and Laiwu for ADM population, and Pietrain, Mangalitza, Iberian, Duroc, American Yucatan (from America, but with a European origin), Yorkshire, Landrace, Hampshire and Large-white for EDM population, each of them including ~1-7 individuals, depending on the availability of good-quality whole-genome sequenced samples. Regarding European and Asian wild boars, samples were selected according to their location of origin, spanning a broad proportion of Europe and the Far East (Additional file 1: Table S1). The European pool contained one sample from the Near East. Raw data in SRA format were downloaded from SRA public repositories and converted into fastq format by using the fastq-dump 2.8.2 tool available in the SRA-toolkit package (ncbi.github.io/sra-tools/).

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

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

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

Single nucleotide polymorphisms (SNPs) mapping to annotated porcine miRNA loci (N = 370) were retrieved by making use of the curated Sscrofa11.1 annotation for miRNA regions available at miRCarta v1.1 database [22]. Additionally, annotated mature miRNA loci (N = 409) inside miRNA genes (N = 370) were retrieved and SNPs within miRNA genes residing in mature and seed regions (2nd to 8th positions at 5' end in the mature miRNA) were identified. Moreover, we expanded our analysis to putative miRNA target sites in the 3'-UTR of protein-coding mRNAs. To this end, we downloaded the current annotation of the 3'-UTR of porcine mRNA transcripts from Ensembl repositories (http://www.ensembl.org/info/data/ftp/index.html) and interrogated the corresponding set of sequences against seed regions of miRNAs available at miRBase database [23]. Seed sequences were reverse-complemented and searched along the 3'-UTR sequences of mRNA genes by making use of the SeqKit toolkit [24]. All 7mer-m8 canonical seed pairing (2nd to 8th 5' nts perfect match) as well as 8mer seed pairing (adding an additional adenine to the 1st position of the matched region within the mRNA 3'-UTR, which is used as a complementary anchor to the Argonaute miRISC complex), were identified for each miRNA seed and the corresponding putative miRNA-mRNA target pairs were defined. Subsequently, the genomic location of both 7mer-m8 and 8mer matching regions were determined, and SNP variants residing in such targeted sites were retrieved. Principal components analyses (PCA) based on autosomal whole-genome SNPs, as well as on autosomal SNPs inside miRNAs genes, 8mer and 7mer-m8 3'-UTR sites and whole 3'-UTRs were performed with the smartPCA software [25]. When implementing the whole-genome PCA, polymorphisms complying with the following parameters were retained: 1) minimum allele frequency (MAF) > 0.05, 2) Hardy-Weinberg equilibrium exact test *P*-value > 0.001. In contrast, in the PCA based on miRNA SNPs no filtering was performed and all retrieved autosomal miRNA-related polymorphic sites were included.

Identifying signatures of selection in miRNA genes

Whole-genome SNP data sets generated for the four ADM, AWB, EDM and EWB populations were filtered on the basis of the following parameters: 1) MAF > 0.01, 2) Hardy-Weinberg equilibrium exact test *P*-value > 0.001, and 3) missing rate < 0.9. Only bi-allelic

SNPs were considered. The Weir and Cockerham's F_{ST} statistic [26] was employed for evaluating the presence of signatures of selection between the four defined contrasts (i.e., ADM vs AWB, EDM vs EWB, ADM vs EDM and AWB vs EWB). In this regard, the F_{ST} estimate can be defined as a measure of population differentiation due to changes in genetic structure based on the variance of allele frequency. The wcFst calculator implemented in the Genotype Association Toolkit (GPAT++, https://github.com/vcflib/vcflib/) was used to estimate the magnitude of F_{ST} in each of the defined pairwise contrasts. Statistical significance was determined with a permutation test (1,000 permutations). An empirical *P*-value ≤ 0.05 was set as the cutoff of significance.

Frequency and distribution of SNPs in miRNA genes

To assess the patterns of SNP distribution across miRNA-related loci, polymorphic sites mapping to miRNA genes were classified according to their location, i.e., SNPs inside the seed region of mature miRNAs were flagged as "*seed*", whereas the rest of SNPs located in mature miRNAs but outside the seed were classified as "*mature*". The remaining SNPs inside miRNA loci that did not belong to any of the two previous categories were assigned to the "*precursor*" class. Allele frequencies for reference and alternative alleles were separately calculated for the whole set of porcine samples (N = 120), as well as for each defined porcine population separately (Additional file 2: Table S2).

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

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

where N_{snp} is the total number of miRNA SNPs (N = 285), as well as the number of detected SNPs in each of the defined regions, (i.e. 221, 52 and 12 polymorphisms in precursor, mature and seed regions, respectively); N_r is the total nucleotide length of all miRNA loci (31,145 nts), as well as the total number of nucleotides in miRNA loci belonging to either "*precursor*" (22,229 bp), "*mature*" (6,053 bp) or "*seed*" (2,863 bp) regions. These calculations would yield the number of SNPs per bp for each category. We decided to adjust this estimate to a window of 100 bp, and this is why N_{snp} is multiplied by 100 in the above formula. Furthermore, the average SNP density at the whole-genome level was compared with the SNP density within miRNA genes. To this end, we retrieved all detected SNPs in our whole-genome sequenced dataset (N = 120) and applied the SNP density formula considering the length of Sscrofa11.1 whole-genome assembly (~2.48 Gb) available at Ensembl repositories (http://www.ensembl.org/info/data/ftp/index.html).

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

Whole-Genome sequencing of five Duroc pigs

In 2003, five Duroc boars were selected and used as founders of a half-sib population of purebred Duroc pigs devoted to the production of high-quality cured ham and phenotypic and genotypic recording for subsequent analyses. In the current work, we aimed to characterize the variability of these five individuals by whole-genome sequencing, identify SNPs located in miRNA genes and investigate their association with mRNA levels and lipid phenotypes recorded in their offspring (Lipgen population, N = 350). Extraction of DNA was performed for the five Duroc founders and genome sequencing was carried at the Centro Nacional de Análisis Genómico (CNAG, Barcelona, Spain). Paired-end multiplex libraries were prepared according to the instructions of the manufacturer with the KAPA PE Library Preparation kit (Kapa Biosystems, Wilmington, MA). Libraries were loaded to Illumina flow-cells for cluster generation prior to producing 100 bp paired-end reads on a HiSeq2000 instrument following the Illumina protocol. Base calling and quality control analyses were performed with the Illumina RTA sequence analysis pipeline according to the instructions of the manufacturer. Quality-checked filtered reads were mapped to the *Sus scrofa* genome version 11.1 and processed for SNP calling according to GATK best practices recommendations and the protocol implemented for Asian/European wild boars and domestic pigs.

Description of the Duroc Lipgen population and phenotype recording

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

Total DNA was extracted from each sample following Vidal et al. 2005 [32]. A total of 345 DNA samples from the initial set of 350 pigs were successfully obtained and processed for further genotyping. Total RNA was extracted from a number of GM (N = 89 pigs) and liver (N = 87 pigs) tissue samples obtained from the Lipgen population following the acid/phenol RNA extraction method [33] implemented in the Ribopure isolation kit (Ambion, Austin, TX). Expression mRNA profiles were then characterized by means of hybridization to the GeneChip porcine arrays (Affymetrix Inc., Santa Clara, CA), as reported by Cánovas et al. 2010 [29]. Further details about tissue collection, sample selection, RNA isolation and microarray hybridization procedures can be found in [29]. Microarray data pre-processing, background correction, normalization and $\log_2$-transformation of expression estimates were performed with a robust multi-array average (RMA) method per probe [34]. The *mas5calls* function from *affy* R package [34] was then applied to infer probe intensity significance level in order to detect gene expression above background noise. This function applies a Wilcoxon signed rank-based gene expression presence/absence detection algorithm for labeling expressed probes in each sample. Control probes and those with expression levels below the detection threshold in more than 50% of samples were discarded from further analyses.

Besides gene expression phenotypes, we also obtained lipid-related phenotypes in the Duroc population. Backfat thickness measured between 3rd and 4th ribs and at last rib, ham fat thickness, as well as fatty acids (FA) composition phenotypes from GM and LD skeletal muscle samples (N = 345) were determined as described in [28,30]. Briefly, intramuscular fatty acid (IMF) percentages in GM and LD muscles were calculated with the Near Infrared Transmittance technique (NIT, Infratec 1625, Tecator Hoganas, Sweden), while muscle cholesterol measurements were inferred following Cayuela et al. 2003 [35]. A gas chromatography of methyl esters protocol was used to determine muscle fatty acids composition of saturated (SFA), unsaturated (UFA), monounsaturated (MUFA) and polyunsaturated (PUFA) fatty acids. Weight (kg), backfat and ham fat thickness (mm) were measured on a regular basis prior and after slaughter. Mean and standard deviation values for the Lipgen population (N = 345) and further details about measured phenotypes are described in Additional file 3: Table S3.

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

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

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

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

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

Association analyses between miRNA SNPs and mRNA expression and lipid phenotypes

Genotype data from 15 miRNA SNPs (Table 1) were processed with the PLINK software [36] in order to generate formatted files for subsequent analyses. The genome-wide efficient mixed-model association (GEMMA) software [37] was used to implement association analyses between genotyped SNPs and lipid phenotypes and microarray expression data in GM and liver tissues. The following univariate mixed model was used:

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

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

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

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

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

miRNA structural inference with RNAfold

The potential effects of SNPs on pri-miRNA structural conformation as well as on miRNA stability and processing availability were predicted with the RNAfold tool from the ViennaRNA Package 2.0 [41] based on the Boltzmann-weighted centroid structure ensemble of the RNA sequence [42].

Results

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

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

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

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

When we analyzed population structure based on whole-genome autosomal 3’-UTR SNPs (N = 709,343 SNPs, Figure 1C), 3’-UTR 7mer-m8 site SNPs (N = 107,196 SNPs, Figure 1D) and 3’-UTR 8mer site SNPs (N = 33,511 SNPs, Figure 1E), genetic differentiation between Asian and European populations was evident, in close concordance with results shown in Figure 1A and 1B. However, we also detected a pronounced differentiation between domestic pigs and wild boars, and this observation was particularly true for Asian pigs and wild boars when considering variants in 3’-UTR regions, either at the full 3’-UTR sequence (Figure 1C) or at miRNA target sites (Figure 1D-E).

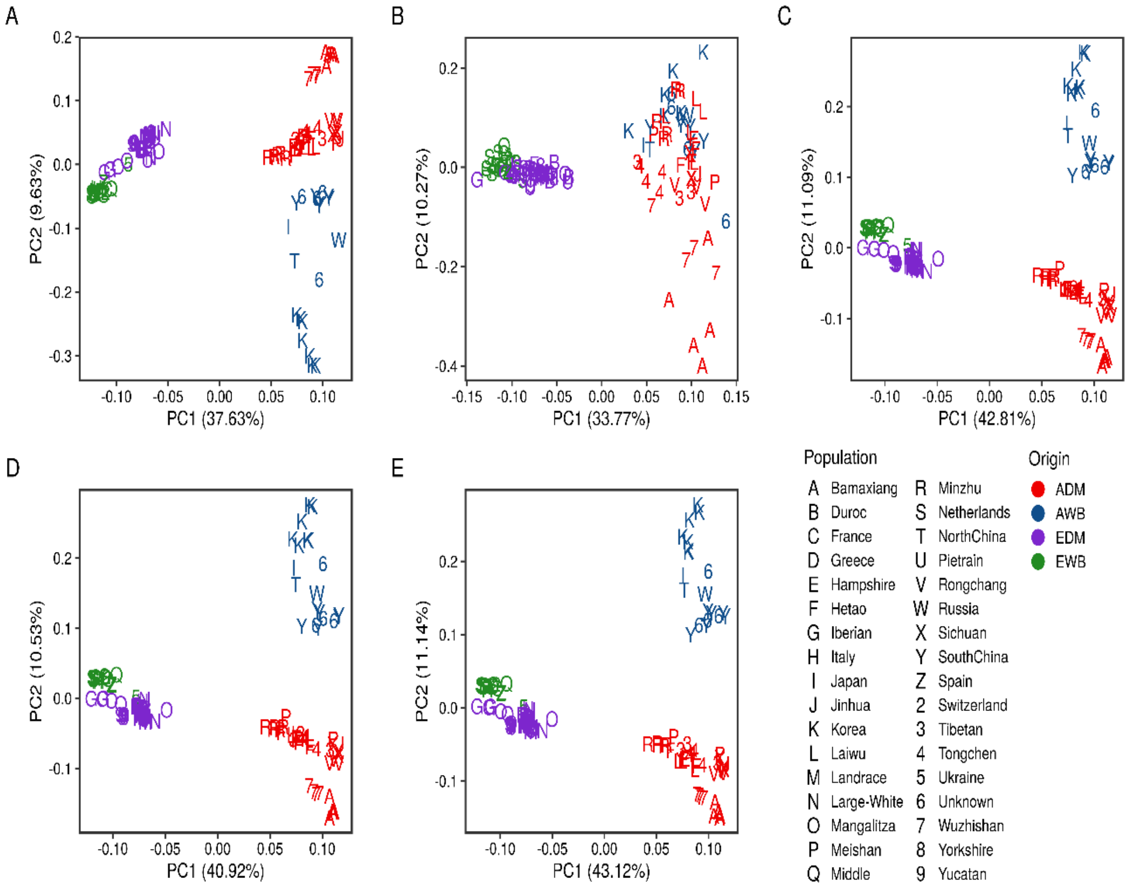


Figure 1: Principal component analysis plots based on SNPs mapping to: (**A**) the whole-genome, (**B**) miRNA genes, (**C**) Full 3’-UTRs, (**D**) 3’-UTR 7mer-m8 sites and (**E**) 3’-UTR 8mer sites, respectively.

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

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

Moreover, the observed SNP density adjusted to 100 bp for miRNA precursor, mature and seed regions consistently followed the order *precursor* > *mature* > *seed* when we considered all whole-genome sequenced pigs (N = 120). Indeed, ~1 SNP per 100 bp was detected in precursor regions, whereas ~0.86 and ~0.42 SNPs per 100 bp were observed in the mature and seed regions, respectively (Figure 3A). These results implied a slight difference in SNP density between precursor and mature regions, while for seed regions, which are critical determinants of miRNA-mRNA interaction, the observed SNP density was ~2.4 fold lower than in precursor regions. This differential distribution of the SNP density across miRNA regions (*precursor* > *mature miRNA* > *seed*) was also observed in each of the analyses performed in the ADM, EDM, AWB and EWB groups (Figure 3A). With regard to variants located within mature miRNAs (N = 64), both inside (N = 12) and outside seed regions (N = 52), their observed distribution within the whole body of the mature miRNA (~22 nts) showed a characteristic pattern (Figure 3B): among all the detected SNPs, the 1st position of the mature miRNA 5' end, which binds to the MID domain of the Argonaute protein in the miRISC complex, exhibited a SNP density of ~0.49 SNPs per 100 bp. Such scarcity in polymorphic sites was also observed when considering the next 2nd to 8th positions in the mature miRNA sequence (seed region), where up to ~0.73 SNPs per 100 bp were observed in the 6th position of the mature miRNA, and an average of ~0.42 SNPs per 100 bp where found among all analyzed miRNA seeds. In contrast, the interval comprising positions 9th to 12th showed an average SNP density of ~0.98 SNPs per 100 bp. Regarding the next 13th to 18th positions of the mature miRNA, which roughly corresponds to a functional region providing additional anchor pairing to the seed region, we observed a decreased SNP density, more prominent at positions 16th to 17th (Figure 3B). Finally, an increased SNP density was found at positions 19th to 22nd.

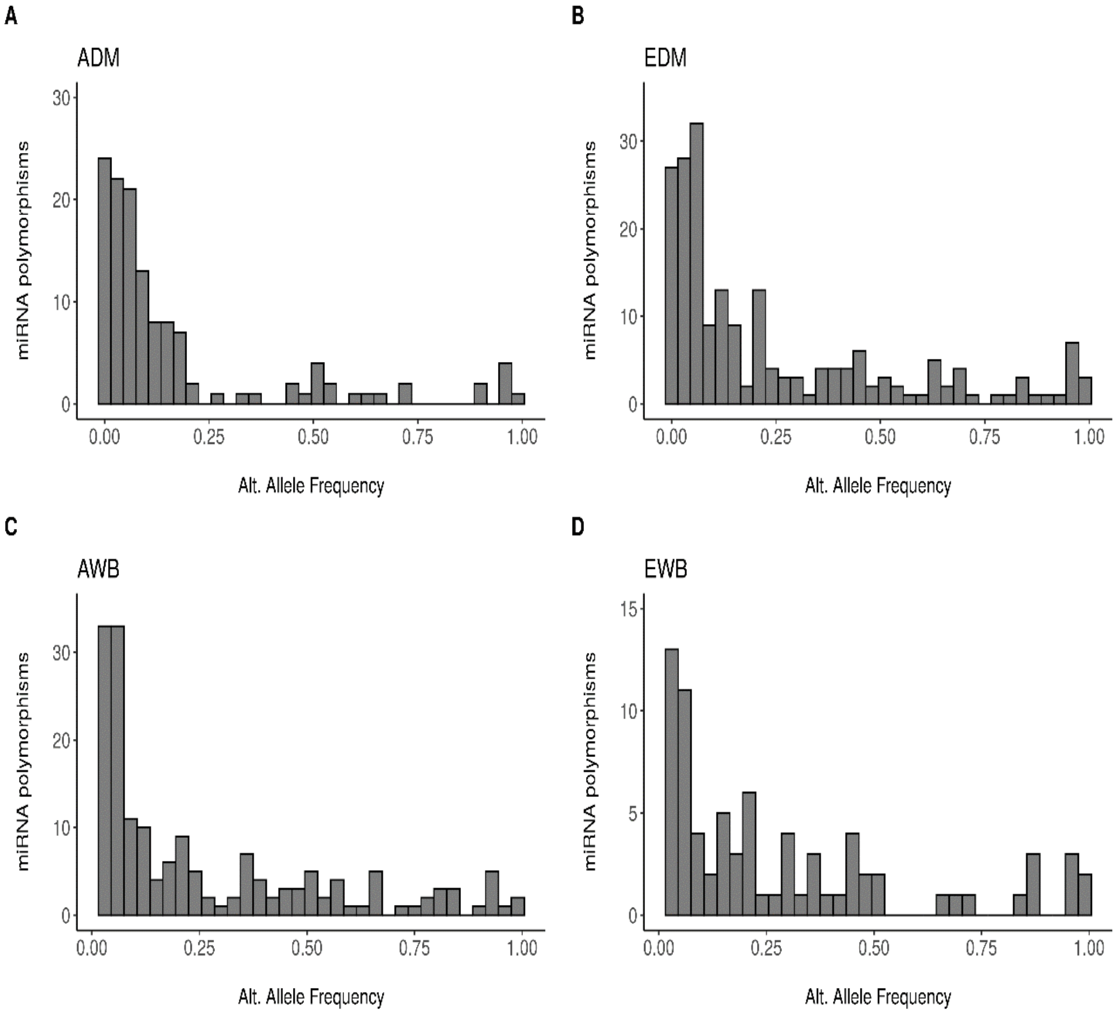


Figure 2: Alternative allele frequency distribution of polymorphisms located at miRNA loci in (**A**) Asian domestic pigs (ADM), (**B**) European domestic (EDM) pigs, (**C**) Asian wild boars (AWB) and (**D**) European wild boars (EWB).

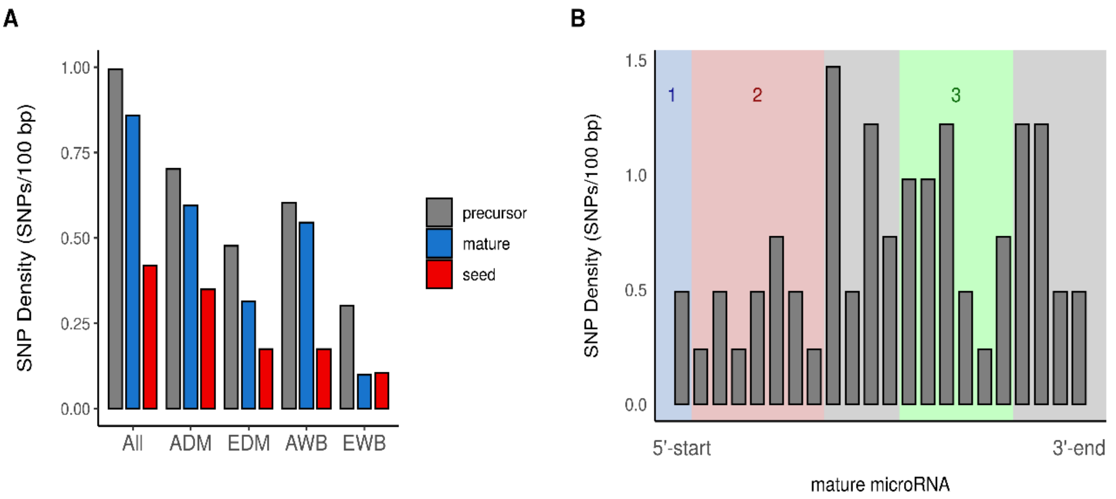


Figure 3: (**A**) SNP density per 100 bp for each analyzed miRNA region considering the full set of 120 whole-genome sequenced porcine samples as well as each of the ADM, EDM, AWB and EWB populations. (**B**) SNP density across mature miRNA regions. 1: anchor (1st 5' end position), 2: seed (2nd to 8th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

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

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

Statistics of the whole-genome sequencing of five Duroc boars

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

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

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

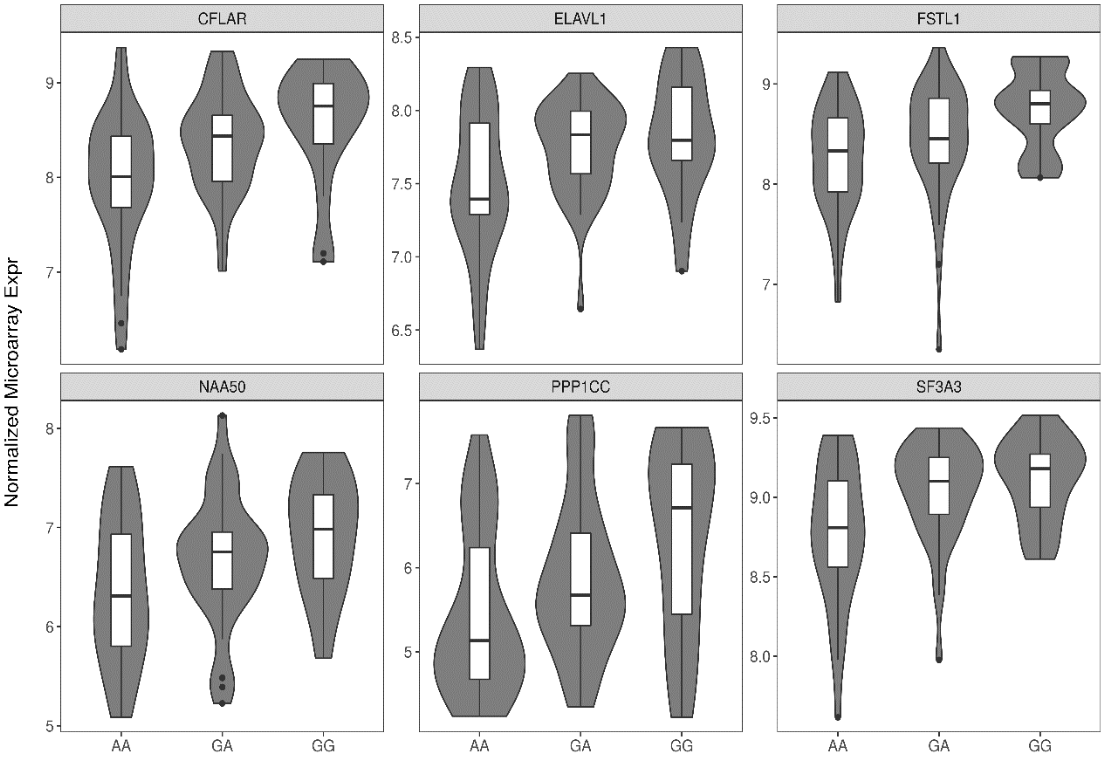


Figure 4: Hepatic mRNA expression levels of the *CFLAR*, *ELAVL1*, *FSTL1*, *NAA50*, *PPP1CC* and *SF3A3* genes according to the genotype of the rs319154814 apical loop polymorphism in the ssc-miR-326 gene. The number of individuals representing each genotype were: GG (N = 17), GA (N = 37) and AA (N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

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

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

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

We further aimed to predict the potential consequences of the rs319154814 SNP on the secondary structure of the ssc-miR-326 transcript by means of the RNAfold centroid-based algorithm [41]. According to structural folding inference (Figure 5), the presence of the rs319154814 polymorphism would have an impact on the steric forces contributing to the adequate base-pairing selection of the hairpin, thus generating an unpaired bulge with a considerable size at the base of the pri-miRNA folded structure generated immediately after transcription (Figure 5). More importantly, the accessibility of the two observed contiguous CNNC motifs (located at positions -16/-21 from the miRNA gene boundaries and inside the originated bulge at the base of the pri-miRNA) would be facilitated by the open unpaired bulge, hence improving its recognition by the Serine-rich splicing factor 3 (SRSF3) protein, a relevant miRNA processing factor that binds to the CNNC motifs of the pri-miRNA sequence and contributes to the correct positioning of Drosha endonuclease for miRNA maturation.

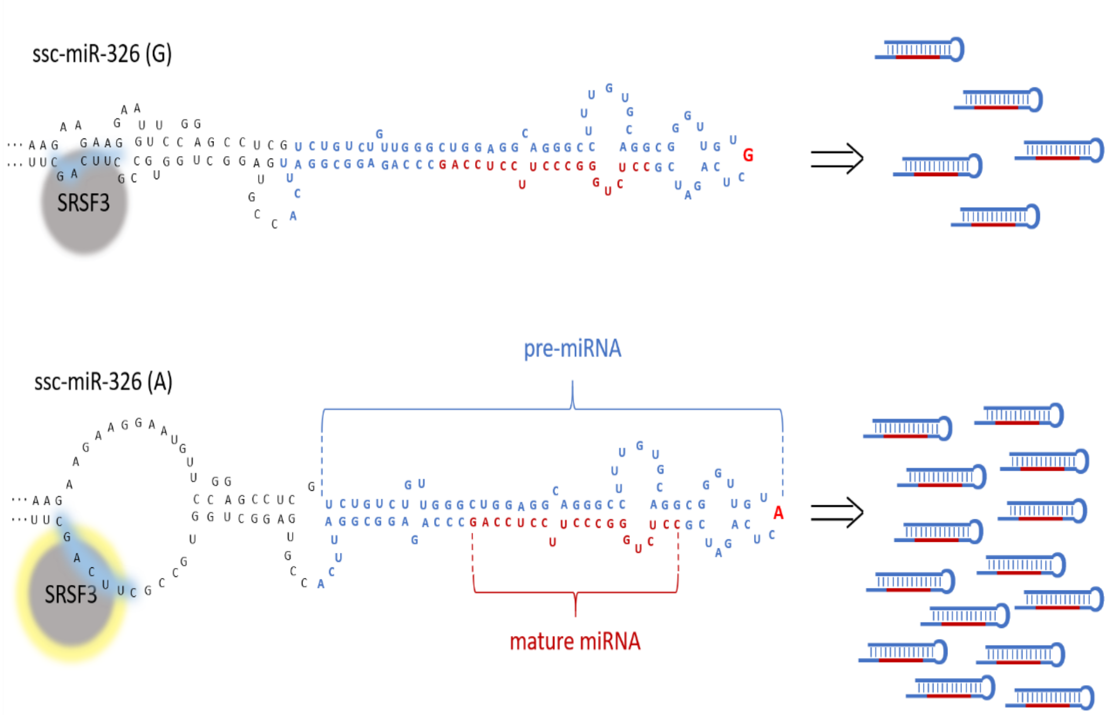


Figure 5: Secondary structural folding of the initial ssc-miR-326 primary miRNA transcript (pri-miRNA) predicted with the RNAfold centroid-based algorithm. The rs319154814 (G/A) polymorphism in the apical loop of the ssc-miR-326 is thought to produce a reorganization to the base-pairing selection of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

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

Polymorphisms in the seed of mature miRNAs might alter the repertoire of their potential targets by abolishing existing miRNA-mRNA interactions and creating new ones. Only one SNP (rs322514450, G/A) located in the seed region of the ssc-miR-9792-5p segregated in our Duroc population. After genotyping, the observed allele frequency for the alternative mutated allele was of 19.89% (Table 1). In this case, we considered as ssc-miR-9792-5p mRNA targets those probes corresponding to mRNAs with predicted 7mer-m8 binding sites (we did not consider experimental validation in humans due to a lack of information in Tarbase v.8 database, Additional file 9: Table S6). Significant associations between this SNP and the expression levels of targeted mRNAs were detected (Table 2). The levels of the retinaldehyde binding protein 1 (*RBLP1*) mRNA in the liver and of the nudix hydrolase 6 (*NUDT6*) mRNA in the GM muscle were significantly associated with rs322514450 genotypes (Table 2). Moreover, we observed a decreased expression of *RBLP1* and *NUTD6* mRNAs in pigs homozygous for the mutated allele (Figure 6).

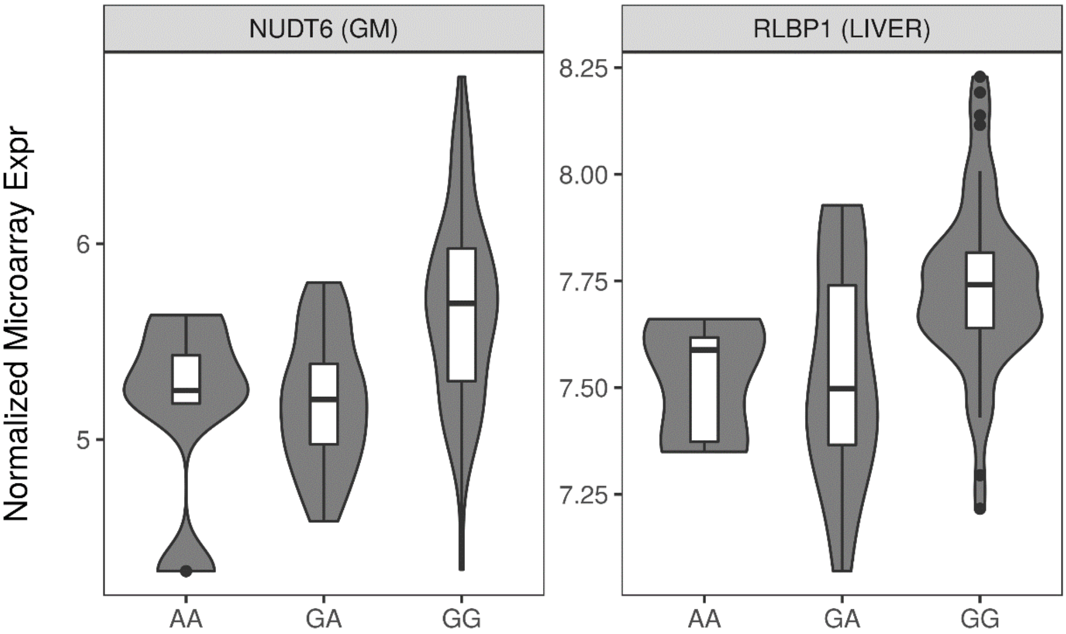


Figure 6: *Gluteus medius* (GM) skeletal muscle and hepatic mRNA expression levels of the *NUDT6* and *RBLP1* genes according to the genotype of the rs322514450 polymorphism located in the seed of the ssc-miR-9792-5p gene. The number of pigs representing each genotype were: GG (N = 56 in GM, 58 in liver),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

Polymorphisms outside the mature microRNA region are associated with mRNA expression

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

Porcine lipid phenotypes are associated with the genotypes of miRNA genes

We also evaluated the association between miRNA SNPs and several lipid-related phenotypes recorded in the Lipgen population (Additional file 3: Table S3). Only the rs319154814 variant inside ssc-miR-326 was significantly associated (*q*-value < 0.1) with lipid traits (Table 3, Additional file 10: Table S7). More specifically, we found significant associations with the myristic acid (C14:0) content in both LD and GM muscles, as well as with the gadoleic acid (C20:1) content and the ratio between PUFA and MUFA in the LD muscle (Table 3, Additional file 10: Table S7). We also observed several associations between the rs319154814 SNP and fatty acid composition traits but they were significant only at the nominal level (*P*-value < 0.01), as shown in Table 3. Other apical loop SNPs like (1) rs81349391 at ssc-miR-339 and (2) rs344472188 at ssc-miR-130a were significantly associated at the nominal level (*P*-value < 0.01) with palmitic acid (C16:0) content and SFA and UFA proportion in the GM muscle, as well as with backfat thickness. Besides, a SNP located in the precursor 3' stem of ssc-miR-30d showed a nominally significant association with the content of arachidic acid (C20:0) (Table 3). Other relevant examples of significant associations at the nominal level were, for instance, those between the rs322514450 (G/A) polymorphism in the seed of the ssc-miR-9792-5p and several fatty acid composition phenotypes in the LD muscle, such as palmitic (C16:0), palmitoleic (C16:1), linoleic (C18:2), α-linolenic (C18:3) and arachidonic acids (C20:4) content. Besides, the rs334590580 (T/C) SNP located at the precursor stem region of ssc-miR-335 was associated with palmitic and arachidic acids content in GM tissue (Table 3).

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

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

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

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

Discussion

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

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

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

We have also found that the degree of population differentiation between Asian domestic pigs and Asian wild boars increases when PCAs are built on the basis of SNPs located in the 3’-UTR (709,343 SNPs), 3’-UTR 7mer-m8 sites (107,196 SNPs) or 3’-UTR 8mer (33,511 SNPs) sites (Figs. 1C-E). The potential effects of 3’-UTR SNPs are the modulation of mRNA expression, secondary structure, stability, localization, translation, and binding to miRNAs and RNA-binding proteins [44], so in general they are not expected to have drastic consequences on gene function. In humans, only a small fraction (3.7%) of the polymorphisms associated with phenotypes reside in 3’-UTRs [44]. Thus, it is reasonable to assume that the intensity of purifying selection is lower in 3’-UTRs than in protein-coding regions, meaning that 3’-UTRs evolve faster and accumulate a larger fraction of recent polymorphisms contributing to population differentiation. In this regard, Bachtiar et al. (2019) [45] reported that ~95% of population-differentiated polymorphisms reside in non-genic regions, compared to the proportion of all SNPs (58%) found in non-genic regions.

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

The percentage of SNPs located within precursor miRNAs that were shared by all four populations was approximately 18.55%, while the percentages of SNPs found exclusively in ADM (13.57%), AWB (12.67%), EDM (10.40%) and EWB (5.43%) populations were lower (Additional file 5: Figure S2A). Moreover, the examination of Figs. S2B and S2C indicates that in mature miRNAs and, to a lesser extent, in seed regions, the percentages of group-specific SNPs exceed that of SNPs shared by all four populations, being ADM pigs the population which displays a higher percentage of exclusive SNPs. By using a data set of 133 porcine whole-genome sequences, Bianco and coworkers [46] computed the percentages of total SNPs that were private to ADM (4.8%), AWB (23.4%), EDM (12.6%) and EWB (2.6%) populations, as well as the fraction of SNPs shared by all four of them (11.7%). In close similarity with our results, the majority of SNPs (> 80%) were shared by two or more populations, and EWB pigs was the population that showed a lower percentage of private SNPs. It should be noticed, however, that the number of genomes used in our study and in the one of Bianco et al. (2015) [46] is moderate (~120-130 genomes), so these estimates might change if additional whole-genome sequences were incorporated to the investigated data sets. The sharing of SNPs between the four populations under consideration could be due to common ancestry as well as to the occurrence of an extensive gene flow between (1) domestic vs wild pigs, and (2) European vs Asian populations [47]. Reproductive isolation between wild and domestic pigs was disrupted to some extent during and after domestication in Asia and Europe [47]. Even in present times, significant levels of domestic pig introgression have been reported in northwest European [48], Sardinian [49] and Romanian [50] wild boars. On the other hand, there are abundant evidences of domestic pig exchanges between Europe and Asia [47,51,52]. The high frequency of Asian mitochondrial variants in European pigs has been interpreted in the light of the massive importation of Chinese sows into England in the

18th-19th centuries with the goal of increasing fatness and reproductive efficiency [51-53]. Conversely, the importation of improved European breeds into Asia explains the admixture of certain Chinese breeds with European blood [54]. In close agreement with our results, Bianco et al. (2015) identified Asian wild boars as the population with a higher percentage of private SNPs [46]. As previously said, the most likely explanation for this result is the Asian origin of *Sus scrofa* and the occurrence of a strong founder effect in Europe [15,43].

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

We have found that, in general, miRNA loci have a substantially lower SNP density than that of the global pig genome (2.5-fold reduction), a result that is concordant with data presented by Saunders et al. (2007) [5]. This reduction in SNP density was stronger in the seeds compared with mature and precursor regions (Figure 3A). The low variability of miRNA genes, a feature that was particularly evident in the seed region, is probably due to the intense effects of purifying selection. Indeed, the importance of the miRNA seeds is revealed by the high conservation of their sequence across species [17,55,56], as this sequence ultimately determines the success of the miRNA-mRNA interactions [2]. In our study, a total of 221, 52 and 12 SNPs were found in precursor, mature and seed regions within miRNA loci, respectively (Additional files 2 and 5: Table S2, Figure S2). Gong et al. (2012) [57] described the existence of 40% polymorphic miRNAs in the human genome but only 16% of them displayed more than one SNP. In a more recent study, He et al. (2018) [58] reported 1,879 SNPs in 1,226 (43.6%) human miRNA seed regions, and 97.5% of these polymorphisms had frequencies below 5%, results in accordance with the overall frequency distribution of miRNA SNPs in the European and Asian populations analyzed in the current work (Figure 2). He et al. (2018) also demonstrated that 1,587, 749, 340, 102, 31, and 4 miRNAs harbored zero, one, two, three, four, and five SNPs, respectively, in their seed regions, reflecting that mutations in this critical functional region are not well tolerated [58]. This distribution is similar to the one that we have observed in domestic pigs and wild boars, with 81, 31, 11, 9 and 5 miRNAs harboring one, two, three, four and five SNPs (Additional file 4: Figure S1). Only 4 and 2 miRNAs showed a total of seven and ten polymorphisms within their sequences.

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

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

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

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

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

No experimentally confirmed targets were available for the porcine ssc-miR-9272-5p, which contains the rs322514450 SNP in its seed. If we just consider mRNA targets on the basis of the 7mer-m8 criterion (without requiring experimental confirmation in humans), two significant associations in GM and liver tissues were detected for the *NUDT6* and *RBLP1* genes, respectively (Table 2, Additional file 9: Table S6). The *NUDT6* gene overlaps the fibroblast growth factor 2 (*FGF2*) gene in an antisense manner, and is thought to negatively regulate its expression [62]. The *FGF2* gene is involved in myoblast proliferation and is a potent inhibitor of skeletal muscle cells differentiation through the activation of the phosphorylation cascade triggered by the protein kinase C (PCK) [63]. Both *NUDT6* and *RBLP1* transcripts showed a reduced expression in pigs homozygous for the mutated A-allele (Figure 6).

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

Regarding other polymorphisms located in the miRNA precursor region, the rs333787816 (T/C) polymorphism within the ssc-miR-23a showed several significant associations, with *AGO1* and *AGO2* transcripts among them, two essential components of the miRNA-mediated cell metabolism regulation [65,66]. Besides, the hepatic expression of *MKRN1* transcript, whose depletion promotes glucose usage and reduces lipid accumulation via AMPK stabilization and activation [67], was also associated with the rs335924546 SNP at ssc-miR-1224 (Table 2).

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

Although the apical region of the miRNA does not have a function as critical as the seed, polymorphisms located in this particular region can have relevant effects on the structural conformation of the pre-miRNA hairpin [13]. In other words, SNPs in the apical region can modify the efficiency with which the Drosha processing machinery, mediating the initial slicing of the hairpin, is recruited. We found particularly interesting that pigs homozygous for the derived A-allele of the rs319154814 SNP showed a consistent downregulation of the hepatic mRNA expression of the *PPP1CC*, *CFLAR*, *FSTL1*, *SF3A3*, *NAA50* and *ELAVL1* genes (Figure 4), thus suggesting that this variant may increase the repressive activity of ssc-miR-326. We might hypothesize that the rs319154814 A-allele in the apical region of ssc-miR-326 could enhance the processing and the expression of this miRNA. However, a functional test will be required to demonstrate the hypothesis outlined in Figure 5, where we indicate the potential mechanism by which the rs319154814 SNP might affect the processing of the miRNA. The A allele is predicted to produce a change in the organization of the hairpin that allows the creation of a bulge around the CNNC processing motif. Such specific region, located at the basis of the pri-miRNA hairpin, is recognized by the SRSF3 protein, which induces Drosha to correctly slice both stems of the pri-miRNA, thus generating the pre-miRNA hairpin [76,77] that will be subsequently exported to the cytoplasm and further processed by Dicer into the mature miRNA transcript [2].

The aforementioned downstream CNNC motif, jointly with the basal UG motif, was first reported by Auyeung and collaborators (2013). These authors described both sequence motifs located around the miRNA precursor hairpins and assessed their contribution to the miRNA maturation process [9]. Additional surveys further reported other motifs affecting the expression of miRNAs [10] and analyzed the structural specifications of RNA-protein interactions between miRNAs and protein subunits in the Microprocessor complex [77-79], as well as the influence of such processing motifs in miRNA prediction analyses [80]. Particularly relevant was the study by Fernandez et al. (2017) [13], where they described a mutation in the apical loop of hsa-miR-30c (G/A) that creates a steric disruption of the pri-miRNA folding structure of the hairpin, hence creating a bulge around the CNNC motif that facilitates the SRSF3 factor accessibility to the RNA sequence. The SRSF3 protein incorporates an RNA-recognition motif (RRM), broadly conserved across many bilaterian animals, which recognizes a degenerate CNNC motif in a base-specific manner [81]. Interestingly, the rs319154814 (G/A) polymorphism detected in porcine ssc-miR-326 might have structural consequences similar to those described for the hsa-miR-30c apical loop variant [13]. This interpretation is further corroborated by structural analyses of the ssc-miR-326 hairpin with and without incorporating the rs319154814 variant (Figure 5). Such similarity could hence be compatible with a conserved mechanism for enhanced miRNA processing via restructuration of the hairpin through sequence modifications.

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

The only miRNA SNP showing significant associations (*q*-value < 0.1) with lipid-related phenotypes was, once again, the rs319154814 in the apical loop of ssc-miR-326. As shown in Table 3, this SNP was associated with myristic content in the GM and LD muscles of Duroc pigs. To the best of our knowledge, no direct effect of ssc-miR-326 on the metabolism of myristic fatty acid has been described so far, but there are reports suggesting that several of the targets of this miRNA might be involved in diverse carbohydrate and lipid metabolism pathways. For instance, increased expression of miR-326 has been detected in type 1 diabetic patients with ongoing islet autoimmunity [82] and there are evidences that this miRNA represses PKM2, an enzyme that catalyzes the final and limiting step in glycolysis [83]. Moreover, targets of miR-326 are enriched in pathways related with sphingolipid metabolism and arachidonic acid metabolism [71]. In this regard, the *PPP1CC* transcript, one of the predicted targets of miR-326, encodes a subunit of protein phosphatase-1 which activates acetyl-CoA carboxylase α and 6-phosphofructo2-kinase/fructose-2,6-bisphosphatase, the main regulators of fatty acid synthesis and glycolysis, respectively [72]. Protein phosphatase-1 also activates lipogenic transcription factors such as sterol regulatory element-binding protein 1 (SREBF1), carbohydrate-responsive element-binding protein (MLXIPL) and, moreover, it dephosphorylates the DNA-dependent protein kinase encoded by the *PRKDC* gene, which is another main determinant of hepatic lipogenesis [72]. In summary, a potential effect of the rs319154814 SNP in the synthesis or degradation of myristic acid can be envisaged, but this hypothesis still needs to be confirmed at the functional level.

Conclusions

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

Supplementary Information

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

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

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

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

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

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

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

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

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

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

Ethics approval

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

Consent of publication

Not applicable.

Availability of data materials

Microarray expression data used in the current study were deposited in the Gene Expression Omnibus (GEO) public repository and are accessible through GEO Series Accession Number GSE115484. Phenotypic and genotypic data sets generated and analyzed during the current study have been deposited in the Figshare public repository available at https://figshare.com/projects/SNPs_miRNA/78690. Whole-genome sequencing dataset from 5 Duroc boars is available at the sequence read archive (SRA) database (BioProject: PRJNA626370).

Competing interests

The authors declare that they have no competing interests.

Funding

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

Authors' contributions

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

References

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

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

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

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

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

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

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

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

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

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

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

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

13. Fernandez N, Cordiner RA, Young RS, Hug N, MacIas S, Cáceres JF. Genetic variation and RNA structure regulate microRNA biogenesis. Nat Commun. 2017; 8:15114.

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

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

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

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

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

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

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

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

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

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

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

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

2006; 38:904–9.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

48. Goedbloed DJ, Megens HJ, Van Hooft P, Herrero-Medrano JM, Lutz W, Alexandri P, et al. Genome-wide single nucleotide polymorphism analysis reveals recent genetic

introgression from domestic pigs into Northwest European wild boar populations. Mol Ecol. 2013; 856–66.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

71. Liu X, Song B, Li S, Wang N, Yang H. Identification and functional analysis of the risk

microRNAs associated with cerebral low-grade glioma prognosis. Mol Med Rep. 2017; 16:1173–9.

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

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

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

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

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

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

78. Kwon SC, Baek SC, Choi YG, Yang J, Lee Y suk, Woo JS, et al. Molecular basis for the single-nucleotide precision of primary microRNA processing. Mol Cell. 2019; 73:505-518.e5.

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

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

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

82. Sebastiani G, Grieco FA, Spagnuolo I, Galleri L, Cataldo D, Dotta F. Increased expression of microRNA miR-326 in type 1 diabetic patients with ongoing islet

autoimmunity. Diabetes Metab Res Rev. 2011; 27:862–6.

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

Discovery and annotation of novel microRNAs in the porcine genome by using a semi-supervised transductive learning approach

Mármol-Sánchez, E.[1*], Cirera, S.[2], Quintanilla, R.[3], Pla, A.[4] and Amills, M.[1,5]

[1]Department of Animal Genetics, Centre for Research in Agricultural Genomics (CRAG), CSIC-IRTA-UAB-UB, Campus de la Universitat Autònoma de Barcelona, Bellaterra, Spain. [2]Department of Veterinary and Animal Sciences, Faculty of Health and Medical Sciences, University of Copenhagen, Grønnegårdsvej 3, 2nd Floor, 1870 Frederiksberg C, Denmark. [3]Animal Breeding and Genetics Programme, Institute for Research and Technology in Food and Agriculture (IRTA), Torre Marimon, Caldes de Montbui, Spain. [4]Department of Medical Genetics, University of Oslo and Oslo University Hospital, Oslo, Norway. [5]Departament de Ciència Animal i dels Aliments, Universitat Autònoma de Barcelona, Bellaterra, Spain.

[*]Corresponding author: Emilio Mármol-Sánchez. emilio.marmol@cragenomica.es

Genomics (2020) 112, 2107-2118. doi: 10.1016/j.ygeno.2019.12.005.

Highlights

- Motif search improved pre-miRNA reconstruction from mature microRNA sequences.
- Semi-supervised methods outperformed canonical supervised classification algorithms.
- The presence of multiple isomiRs in the porcine muscle miRNA repertoire was uncovered.
- A total of 47 novel microRNA genes were identified in the porcine genome.
- RT-qPCR analyses allowed us to confirm the existence of three novel porcine microRNAs.

Abstract

Despite the broad variety of available microRNA (miRNA) prediction tools, their application to the discovery and annotation of novel miRNA genes in domestic species is still limited. In this study we designed a comprehensive pipeline (eMIRNA) for miRNA identification in the yet poorly annotated porcine genome and demonstrated the usefulness of implementing a motif search positional refinement strategy for the accurate determination of precursor miRNA boundaries. The small RNA fraction from *gluteus medius* skeletal muscle of 48 Duroc gilts was sequenced and used for the prediction of novel miRNA loci. Additionally, we selected the human miRNA annotation for a homology-based search of porcine miRNAs with orthologous genes in the human genome. A total of 20 novel expressed miRNAs were identified in the porcine muscle transcriptome and 27 additional novel porcine miRNAs were also detected by homology-based search using the human miRNA annotation. The existence of three selected novel miRNAs (ssc-miR-483, ssc-miR484 and ssc-miR-200a) was further confirmed by reverse transcription quantitative real-time PCR analyses in the muscle and liver tissues of Göttingen minipigs. In summary, the eMIRNA pipeline presented in the current work allowed us to expand the catalogue of porcine miRNAs and showed better performance than other commonly used miRNA prediction approaches. More importantly, the flexibility of our pipeline makes possible its application in other yet poorly annotated non-model species.

Keywords: MicroRNA discovery; Motif search; Porcine skeletal muscle; Semi-supervised learning; Small RNA-seq.

Introduction

The accurate annotation of a comprehensive set of miRNAs in different species has been challenging since the first genome assemblies were published, although an ever-increasing amount of knowledge about miRNA diversity across species has been accumulating during the past years, being available in public databases [1-3]. Despite these advances, many commonly studied domestic species still lack a complete and reliable set of annotated miRNAs in their genomes [1].

The computational prediction of miRNAs in sequenced genomes initially relied on the strong conservation of mature miRNA sequences across closely related species [4,5], taking advantage of homology-based comparisons between well annotated genome assemblies and other poorly annotated organisms [6-8]. Other approaches focused on rule-based classification, integrating other sources of information such as sequencing data or structural features to identify novel miRNAs [9-12]. More recently, several machine learning (ML) approaches have been proposed for miRNA prediction. Different tools have addressed the problem of correctly classifying miRNAs by training ML algorithms with a set of positive (annotated miRNAs) and negative (other non-miRNA sequences) data sets. [13-16]. Nevertheless, despite the broad array of available tools for novel miRNA identification, their application to the discovery and annotation of novel miRNAs in domestic species is still limited [17-25]. Moreover, the majority of miRNA surveys carried out in domestic species do not generally take into account several issues regarding miRNA genes prediction that have recently emerged. For instance, the set of positive training annotated miRNAs often include misclassified sequences [26,27], whereas the negative class is sometimes not clearly defined, i.e. different types of sequences have been used as negative data sets (coding regions, pseudo-hairpins, non-coding hairpins or artificial randomized miRNA sequences). Despite some efforts [28], obtaining a truly representative negative class is still challenging and few approaches have critically addressed this important issue [29-31]. Besides, miRNAs are thought to encompass a small percentage of the total non-coding transcriptomic repertoire, with thousands of other non-miRNA hairpin-like RNA molecules that represent a major fraction of it. This circumstance contributes to create a high class-imbalance between positive and negative sequences. Different approaches have dealt with such phenomenon [32], but recent studies have shown that commonly used techniques for solving the high-class imbalance problem in microRNA prediction may not be suited to a real-case classification scenario [15].

In this study we present eMIRNA, a bioinformatics pipeline for miRNA discovery and annotation in sequenced genomes. The proposed pipeline implements a semi-supervised transductive learning approach to predict and annotate novel microRNAs in the porcine genome, overcoming several of the drawbacks outlined above. In order to validate the performance of our pipeline in a real-case scenario, we have applied it to the analysis of a data set comprising the small RNA fraction of *gluteus medius* skeletal muscle from a population of 48 Duroc gilts [33,34]. Furthermore, making use of the better annotated *H. sapiens* miRNAome, an additional set of novel porcine miRNA genes were identified based on a homology-based search approach. Finally, some of the identified novel porcine miRNA candidates were independently validated in a Göttingen minipig population, investigating their expression in skeletal muscle and liver tissues.

Materials and methods

A detailed flow chart depicting all steps described in the eMIRNA pipeline is shown in Figure 1. Additional instructions and modular scripts needed for the implementation of eMIRNA are available at: https://github.com/emarmolsanchez/eMIRNA/.

Positive and negative training data sets

To define the corresponding positive (annotated miRNAs) data set required for novel miRNA prediction, two approaches were considered:

1) The annotated pre-miRNA coordinates in Sscrofa11.1 genome assembly were obtained from Ensembl repositories, release version 97 (http://www.ensembl.org/info/data/ftp/index.html), and the corresponding sequences were extracted from the pig reference genome by using the BEDTools suite v2.27.0 software [35].

miRNA loci located in scaffolds were removed from further analyses, resulting in a total of 484 annotated porcine miRNA genes. Sequence repeats from pre-miRNA duplicated elements were removed from the retrieved positive data set by using the CD-HIT Suite [36] with a 0.9 sequence identity cut-off value (i.e. sequences showing a similarity ≥ 90% to each other were removed and only unique representative pre-miRNA candidates were retained). Moreover, to avoid the inclusion of miss-annotated miRNA loci, an additional filtering based on secondary structure folding was applied. To this end, the RNAfold tool from the ViennaRNA Package 2.0 [37] was used to select sequences with canonical pre-miRNA hairpin secondary structures (stem-loop conformation with one single terminal loop and two stems). Sequences that failed to comply with required folding structure pre-requisites were removed.

2) In the second approach, the curated miRNA annotation for Sscrofa11.1 available in the miRCarta database [2] was retrieved, and the same pre-filtering criteria based on sequence identity and secondary structure employed in the analysis of the Ensembl data set were applied. The miRCarta database [2] integrates one of the most comprehensive and curated databases for miRNA annotation and functional activity, aiming to overcome the limitations of other widely used miRNA databases such as miRBase [1].

Regarding the negative data set (other hairpin-like sequences), two different data sources were used. First, the annotated non-coding transcripts in Ensembl repositories were retrieved and non-miRNA sequences were retained. Analogously to what was implemented for the positive data set, identity by sequence and secondary structure pre-filters were applied, and non-miRNA non-coding hairpin-like unique sequences were obtained. Only sequences ranging from 50 up to 150 nucleotides (nt) were retained, thus removing hairpin-like long non-coding RNAs from the negative data set. Additionally, a set of unlabeled sequences within the porcine reference genome (Sscrofa11.1) were generated by extracting candidate pre-miRNA-like sequences from random blocks of 1 Megabase (Mb) in each of the chromosomes of the porcine assembly with the *HextractoR* package [38], and the previously described pre-filters for the negative class were subsequently applied.

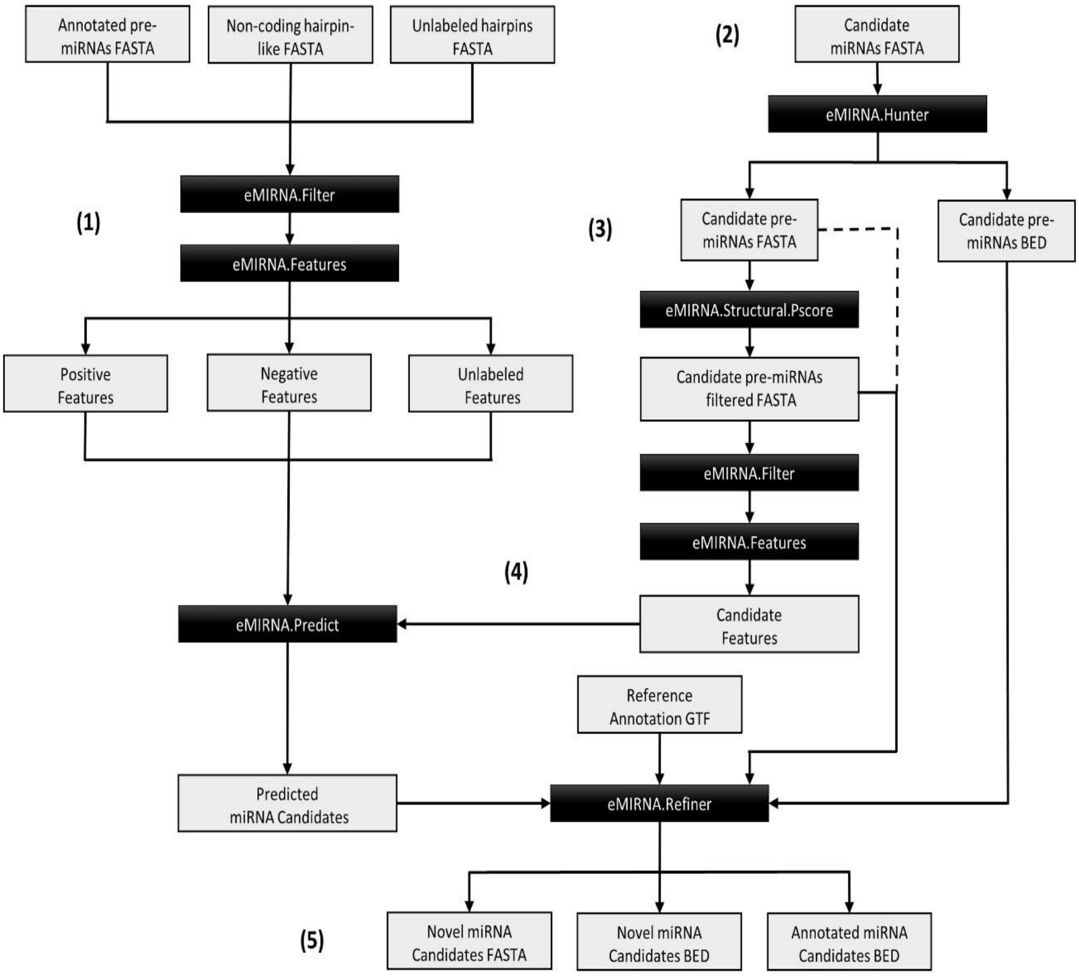


Figure 1: eMIRNA pipeline scheme for homology-based miRNA prediction using data from closely related species and *de novo* miRNA prediction from small RNA-seq data. (**1**) Positive, negative and unlabeled data are filtered based on size and secondary folding structure and a set of features is extracted for each sequence. (**2**) Mature miRNA sequences from small RNA-seq data or related model species are mapped against the selected genome assembly and elongated to reconstruct putative pre-miRNA candidates. (**3**) Candidate precursors are filtered based on size and secondary folding structure and a set of features is extracted for each candidate sequence. Optionally, sequences showing unstable secondary structure are removed. (**4**) Candidate sequences are embedded in the semi-supervised transductive classifier and a list of putative miRNAs is predicted. (**5**) Predicted miRNAs are either assigned to already annotated miRNA loci in the provided reference assembly or classified as putative novel miRNAs genes.

Obtaining putative miRNA candidate sequences from the porcine genome

In order to test our method with pig transcriptomic data, a small RNA-seq data set was generated by sequencing the muscle transcriptome of 48 gilts used in two previous studies [33,34]. Upon collection, muscle samples were individually submerged in RNAlater and snap-frozen in liquid nitrogen. Samples were pulverized and homogenized in 1 ml of TRI Reagent (Thermo Fisher Scientific, Barcelona, Spain). Total RNA was isolated with the RiboPure kit (Ambion, Austin, TX). A Nanodrop ND-100 spectrophotometer (Thermo Fisher Scientific, Barcelona, Spain) was used to assess RNA concentration and quality. RNA integrity expressed in RNA Integrity Number (RIN) units was measured with a Bionalyzer-2100 equipment (Agilent Technologies Inc., Santa Clara, CA). High quality RNA samples were then submitted to Sistemas Genómicos S.L. (https://www.sistemasgenomicos.com) for small RNA sequencing. Library preparation for each individual sample was carried out with the TruSeq Small RNA Sample Preparation Kit (Illumina Inc., USA) and small RNA libraries were single-end sequenced (1 × 50 bp) in a HiSeq 2500 platform (Illumina Inc., CA).

FASTQ sequence files were subjected to a quality control check as reported by Cardoso et al. [33]. After preliminary quality-based filtering, sequencing adaptors were trimmed with the Cutadapt software [39] and an acceptance sequence window of 15–30 nt per read was established. Processed FASTQ files from all sequenced samples (N = 48) were pooled and collapsed to unique FASTA sequences with the FASTQ collapser tool from FASTX-toolkit (http://hannonlab.cshl.edu/fastx_toolkit/). Unique FASTA sequences represented by >10 reads-per-million (RPM) were considered to be significantly expressed above the background noise [40], and thus selected for further analyses (File S1). The CD-HIT Suite [36] was employed to build sequence clusters with >0.9 sequence identity.

Furthermore, the human mature miRNA coordinates were obtained from Ensembl repositories and the corresponding sequences were retrieved from the GRCh38.p12 assembly. Pre-filtering based on sequence identity was applied and a set of non-redundant human mature miRNAs was generated for homology-based search in the Sscrofa11.1 porcine assembly (File S2).

Pre-miRNA reconstruction by sequence elongation and motif search

Once putative mature miRNA candidate sequences from the small RNA-seq data set and the human mature miRNA sequences were retrieved, they were aligned against the porcine reference assembly (Sscrofa11.1) with the Bowtie aligner [41] and the following specifications for short reads: 1) allowing 2 mismatches within the entire aligned sequence with respect to the reference assembly, 2) removing reads with >50 putative mapping sites and 3) reporting first single best stratum alignment (*bowtie -n 2 -l 25 -m 50 -k 1 --best --strata*). Reported alignment genome positions for successfully mapped putative mature miRNAs were elongated upstream and downstream, thus ensuring an adequate pre-miRNA reconstruction. As no prior knowledge about the 3p or 5p identity of putative mature miRNA sequences was available for porcine small RNA-seq data, two candidate pre-miRNA structures were generated for each expressed sequence. The same procedure was applied to human mature miRNAs when 3p or 5p identity was not specified. Candidate sequences that were aligned and extracted from overlapping regions corresponding to other annotated non-miRNA non-coding loci were discarded from further analyses.

Elongation patterns were based on previously reported pre-miRNA favored size, with a stem length of ~35 ± 3 nt and an apical loop ≥10 nt [42,43]. With these specifications, we established two upstream and three downstream elongation pattern combinations: 1) from the starting genome position of each aligned sequence, 15 and 30 nt were added upstream, beginning from each mature miRNA sequence start position. 2) Additionally, 60, 70 and 80 nt were added from each miRNA end position, resulting in the following elongation pattern combinations for each candidate sequence: 15/60, 30/60, 15/70, 30/70, 15/80 and 30/80 added nt (i.e. we generated a total of 12 putative elongated pre-miRNA candidates per each aligned sequence). Besides, the presence of flanking microprocessor motifs was assessed for positionally correcting the elongated pre-miRNA candidate sequences. Downstream CNNC and upstream UG motifs were assessed within the 30/60, 30/70 and 30/80 elongated candidates for each sequence, as described in [44], whereas downstream mismatched GHG and upstream CHC motifs were searched in 15/60, 15/70 and 15/80 candidates [42].

To determine the most prevalent positional range of flanking processing motifs surrounding pre-miRNA sequences in the porcine genome, 30 and 15 nt were added at the flanking positions of annotated porcine pre-miRNAs available at the curated miRCarta database [2]. The presence of CNNC and UG motifs within flanking ±30 nt, as well as GHG and CHC motifs within ±15 nt was hence assessed. According to positional results (Figure 2A), the CNNC and UG flanking motifs appeared more prominently located 18 nt after miRNA gene ending and 12 nt before miRNA starting points, respectively. Therefore, when downstream

CNNC or upstream UG motifs were found within ±30 nt flanking windows along pre-miRNA candidates, −18 and +12 nt positions were added from CNNC and UG motifs location, respectively, so as to establish accurate miRNA genes boundaries determined by the microprocessor machinery. In the event that none of the aforementioned motifs within flanking upstream and/or downstream defined regions were found, the original elongated pre-miRNA candidates with no motif-based positional refinement were kept.

Selecting putative pre-miRNA candidate sequences based on structural integrity

To better assess the optimal elongation pattern for each candidate sequence, the structural stability of the 12 pre-miRNA candidates per single sequence was determined based on the *randfold* algorithm [45]. This approach assumes the estimated minimum free energy (MFE) of the folded pre-miRNA hairpin to be consistently lower than that of other random sequences resembling hairpin-like folded structures [45]. Based on this property of pre-miRNA sequences, we implemented a Monte Carlo randomization test to select the most stable hairpin, i.e. those having the least folding minimum free energy (MFE) values among the 12 previously generated candidates during pre-miRNA elongation reconstruction for each of the analyzed sequences. To this end, we generated a total of 100 randomized sequences per candidate by shuffling their nucleotide distribution while maintaining k-let counts [46]. The corresponding MFE values for each shuffled and original hairpin-folded sequences were calculated with the RNAfold tool [37] and the structural integrity score (*p*) was defined as:

$$p = \frac{R}{N + 1}$$

where R is the number of randomized sequences having an MFE value equal or smaller than that of the MFE value of the original sequence and N is the number of generated iterations (100 in this study).

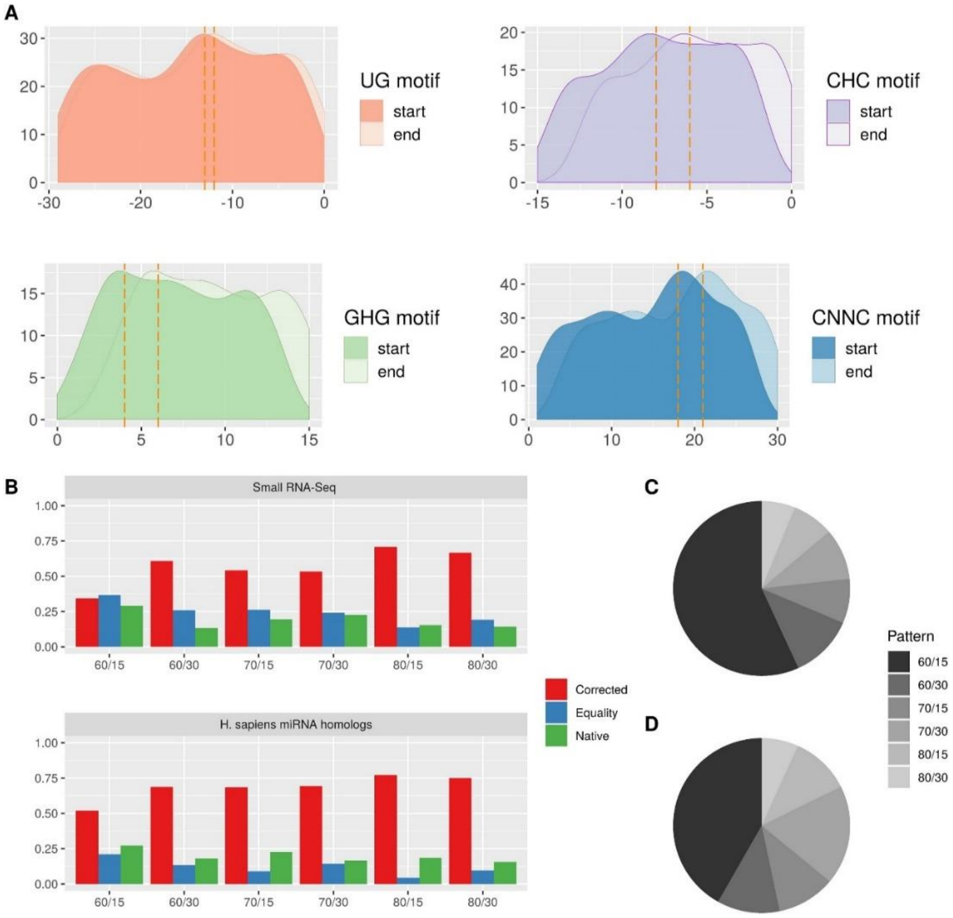


Figure 2: Processing motifs distribution and structural stability metrics. (**A**) Positional distribution of upstream and downstream motifs across annotated pre-miRNA boundaries in the porcine genome. (**B**) Proportion of candidate sequences for each elongation pattern showing the most stable folding structure according to randfold *p* score. The proportion of sequences for which the structural stability was higher in motif corrected candidates or, conversely, in non-corrected (native) candidates are shown as red and green bars, respectively. The proportion of sequences for which the structural stability was equivalent between motif corrected and native candidates were labeled as equally stable (blue). (**C**) Proportion of selected pre-miRNA candidates detected in the porcine *gluteus medius* muscle small RNA-seq data and (**D**) Proportion of selected pre-miRNA candidates detected through a *H. sapiens* homology-based miRNA search strategy, according to the most structurally stable elongation pattern tested. If two or more pre-miRNA sequences showed equivalent stability, the shortest motif-corrected candidate was selected.

Subsequently, the candidate sequence showing the higher structural integrity (i.e. the one showing the smallest *p* score) among all 12 generated pre-miRNA candidates per sequence was selected. The proportion of the most structurally stable sequences for each elongation pattern is shown in Figure 2B. When two or more sequences had equal *p* scores (i.e. they had equivalent structural stability irrespective of the elongation pattern) the reconstructed candidates belonging to the motif-corrected (if available) and shortest elongation pattern were retained. The proportion of each elongation pattern selected as the most structurally stable among all 12 tested patterns from expression-based and homology-based data is shown in Figure 2C and D, respectively.

Candidates classification with semi-supervised transductive learning

After defining training and candidate data sets, we selected a total of 100 features representing structural and statistical properties from each pre-defined sequence. These extracted features have been previously reported in other state-of-the-art methods and thoroughly reviewed in [47]. A complete list of all used features is shown in Table 1.

For pre-miRNA classification, the *miRNAss* algorithm proposed by Yones et al. [31] was applied. This method implements a semi-supervised transductive learning scheme by using well defined labeled cases, either positives (annotated pre-miRNAs) or negatives (comprising other annotated non-coding hairpin-like sequences and unlabeled cases with unknown hairpins), in order to draw a graph-based representation of each sequence based on input features. Each node in the graph represents a sequence, whereas the corresponding edges account for the expected similarities among them. In order to accurately represent the spatial distribution and connections of each node, the feature importance is obtained by applying the Relief-F algorithm [48,49], where k-nearest predictors are weighted based on conditional dependencies among all the considered features and the response vector of labels. This algorithm penalizes those predictor features giving different values to k-neighbors from the same label class and vice versa. After graph construction, a prediction score is assigned to each sequence node [31].

Table 1: List of calculated features extracted from candidate hairpins.

Sequence Features	**Symbol**	**Number of variables**
Triplet Elements by SVM-Triplet	T1 … T32	32
Sequence Length	Length	1
G+C/Length	GC	1
A+U/G+C	AU.GCr	1
A, U, G, C/Length	Ar, Ur, Gr, Cr	4
Dinucleotide/Length	Aar, GGr, CCr …	16
Secondary Structure metrics	**Symbol**	**Number of variables**
Hairpin loop Length	Hl	1
5’ and 3’ Stems Length	Steml5, Steml3	2
Basepairs in Secondary Structure	BP	1
Matches in 5’ and 3’ Stems	BP5, BP3	2
Mismatches in 5’ and 3’ Stems	Mism5, Mism3	2
Bulges in 5’ and 3’ Stems	B5, B3	2
Bulges in 5’ and 3’ Stems of types 1 to 7 mismatches	BN1.5, BN1.3 …	14
A-U, G-C and G-U basepairs	Aup, GCp, Gup	3
Structural Statistics	**Symbol**	**Number of variables**
Minimum Free Energy	MFE	1
Ensemble and Centroid Free Energy	EFE, CFE	2
Centroid Distance to Ensemble	CDE	1
Maximum Expected Accuracy	MEA, MEAFE	2
BP/Length	BPP	1
MFE Ensemble Frequency	Efreq	1
Ensemble Diversity	ED	1
MFE/Length, EFE/Length and CDE/Length	MFEadj, EFEadj, Dadj	3
Shannon Entropy/Length	Seadj	1
MFE-EFE/Length	DiffMFE.EFE	1
MFEadj/GC and MFEadj/BP	MFEadj.GC, MFEadj.BP	2
MEAFE/Length and ED/Length	MEAFEadj, Edadj	2

Sscrofa11.1 pre-miRNA sequences from Ensembl and miRCarta databases were evaluated and different imbalance ratios between positive (taken as reference) and negative data sets were applied to assess the performance of the classification algorithm for miRNA discovery in the porcine genome (i.e. 1:1, 1:2, 1:10, 1:20, 1:40, 1:60, 1:80, 1:100, 1:150 and 1:200 imbalance ratios were considered). Labeled sequences comprised annotated pre-miRNAs (+1) as positive sequences, while other non-coding hairpin-like transcripts (−1) were considered as negative. Genome-wide randomly extracted hairpins were assigned as unlabeled cases (0) within the negative data set.

Testing subsets were randomly assigned from all proposed imbalanced training data set combinations using a 0.25 ratio. The performance of the classification algorithm for miRNA identification was assessed with a total of 100 random Monte Carlo iterations and average performance measures based on sensitivity (SE), specificity (SP), accuracy (Acc), F-1 score (F1) and adjusted geometric-mean (Agm) [50] were estimated (Figure 3A). Furthermore, we evaluated the performance for each imbalance scenario by computing the corresponding receiver operating characteristics (ROC) curves and the precision-recall (PR) curves. PR curves can be more informative than ROC curves for highly imbalanced data sets [51]. ROC and PR curves as well as the corresponding Areas under the curve (AUC) estimates are shown in Figure S1 and Table S1. The ability of the algorithm to correctly classify the list of Ensembl and miRCarta annotated porcine miRNAs was also assessed by incorporating the positive data set as unlabeled candidate sequences during the classification process in each of the defined imbalance scenarios. Results for annotated porcine miRNAs assignment are shown in Table S2.

Finally, the reconstructed expressed candidate sequences from the porcine small RNA-seq data and *H. sapiens* homologous miRNAs detected in the porcine genome were used for identifying putative novel miRNAs. For this purpose, annotated pre-miRNAs from the Ensembl database were used as positive class and other hairpin-like sequences were considered as either negative or unlabeled sequences. Candidates classification was implemented with all previously proposed imbalance ratios. In order to reduce the false positive rate (i.e. reducing the misclassification of non-miRNA short hairpins as true miRNA candidates), the Ensembl miRNA data set was defined as the positive class, due to its higher overall reported specificity (Figure 3A and B). Prediction of novel miRNA candidates was carried independently with every defined imbalance ratio. Only candidates consistently reported as putative miRNAs in all imbalance scenarios were kept in order to minimize the number of false positive miRNA candidates, albeit probably at the expense of increasing the false negative rate.

Besides, for homology-based predicted novel pre-miRNA candidates, we calculated the proportion of shared neighboring genes (setting a 2 Mb window before and after each annotated human miRNA detected in the porcine genome) present in both *S. scrofa* and *H. sapiens* assemblies and expressed as a Neighborhood Score (N):

$$N = \frac{G_r \cap G_i}{G_r}$$

where Gr is the number of orthologous genes within the 4 Mb window in the model species (*H. sapiens*) and Gi is the number of genes within the same window in the species of interest (*S. scrofa*). Only homology-based novel pre-miRNA candidates with N > 0.1 were considered for further analyses, based on the assumption that microRNAs residing in genomic regions with surrounding and/or host genes phylogenetically conserved across species are more prone to be integrated in biologically relevant transcriptional networks [52].

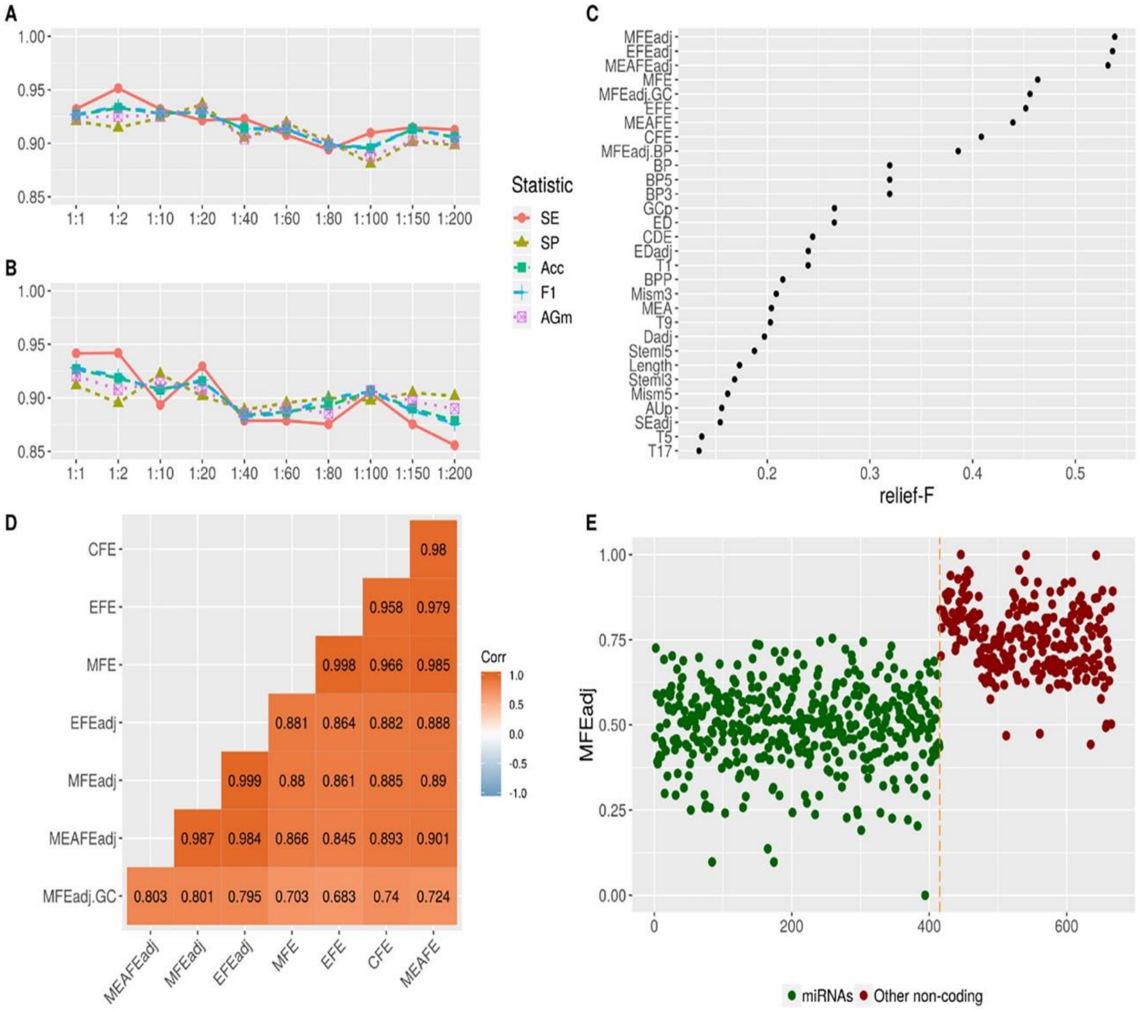


Figure 3: Classification performance and feature importance statistics. Performance metrics for sensitivity (SE), specificity (SP), accuracy (Acc), F1-score (F1) and adjusted geometric-mean (Agm) across incremental imbalance-ratios by using positive miRNAs from (**A**) Ensembl and (**B**) miRCarta databases. (**C**) Thirty most discriminant features according to the relief-F algorithm. (**D**) Pearson’s correlation coefficient among the seven most discriminant features associated with secondary structure stability metrics. (**E**) Comparison of the folding structure stability between annotated miRNAs and other hairpin-like non-coding RNA sequences present in the porcine genome. Stability is expressed as the scaled minimum free energy of the folded hairpins adjusted by sequence length (MFEadj).

Benchmarking for miRNA prediction performance

One of the most cited and used prediction miRNA algorithms is miRDeep. This tool was developed by Friedländer et al. [53], and further improvements were made in subsequent updates [11,54]. This algorithm implements a series of heuristics to compute a score for each miRNA candidate expressing the log-odds probability of a sequence being a true miRNA gene against the probability of being a miRNA-like pseudo-hairpin [53]. In order to benchmark the eMIRNA pipeline compared with the widely used miRDeep approach, we used the miRDeep2 algorithm [54] to identify novel and annotated miRNAs by using the same small RNA-seq data set employed for *de novo* miRNA identification with the eMIRNA pipeline. To ensure a fair comparison, the arf alignment file needed for running the miRDeep2 software was generated from the eMIRNA alignment pipeline using the bowtie tool (*bowtie -n 2 -l 25 -m 50 -k 1 --best --strata*) on pre-filtered expressed small RNA sequences generated in this study. After running the miRDeep2 algorithm, both novel and already annotated pre-miRNA candidates were compared with those obtained with the eMIRNA pipeline. The positional accuracy of the annotated pre-miRNA candidates concurrently identified with both approaches was then determined using the Ensembl annotation available for the Sscrofa11.1 assembly. To further determine which of the two approaches provided a better positional annotation of predicted miRNAs, the deviation rate (dr) of each miRNA gene commonly detected was calculated for both eMIRNA and miRDeep2, expressed as the average number of upstream and downstream overhanging nucleotides compared with the latest porcine miRNA Ensembl annotation (v97). The differential deviation estimate (ΔD) was assessed separately for each predicted pre-miRNA candidate, as follows:

$$\Delta D = eMIRNA_{dr} - miRDeep2_{dr}$$

Additionally, the performance statistics of the semi-supervised transductive learning method [31] implemented in the eMIRNA pipeline was compared with other canonical widely used state-of-the-art supervised ML approaches for miRNA prediction, such as support vector machine (SVM), random forest (RF), k-nearest neighbors (KNN), naïve Bayes (NB), extreme gradient boosting trees (XGB) and light gradient boosting trees (lGBM). Only labeled positive and negative data sets were used for comparison between semi-supervised and supervised algorithms. Training and testing subsets were randomly generated with a 0.25 ratio for testing data and commonly used with all the proposed methods. No imbalance correcting procedure was applied. The comparative performance of these tools was assessed on the basis of SE, SP, F1-score, ROC and PR curves obtained for each algorithm implementation. SVM, RF, KNN and NB algorithms were trained allowing 10 iterations for parameter tuning and a 10-fold cross-validation scheme, using built-in functions included in the *caret* R package [55]. The *xgboost* [56] and *lightgbm* (https://github.com/microsoft/LightGBM/tree/master/R-package) R packages with default parameters were employed for the training of XGB and lGBM classifiers, respectively.

Experimental confirmation of novel identified porcine miRNAs through the RT-qPCR analysis of an independent Göttingen minipig population

In order to investigate the existence of several of the novel putative predicted miRNAs in the porcine genome, three well established orthologous novel miRNA candidates detected by homology-based search and not previously annotated in the Sscrofa11.1 assembly were selected (hsa-miR-483-3p, hsa-miR-484-5p and hsa-miR-200a-3p). The existence of miRNA genes orthologous to hsa-miR-483-3p and hsa-miR-484-5p was supported by the identification of the corresponding expressed mature miRNA sequences in our small RNA-seq data set. Transcripts corresponding to hsa-miR-200a-3p were detected at very low expression levels (RPM < 10) in the porcine skeletal muscle transcriptomic data, so they were not considered as biologically relevant or functionally active in our experimental conditions. *Longissimus dorsi* muscle and liver RNA samples were collected from an independent Göttingen minipig population [57]. A total of 7 extracted RNA samples from muscle and liver tissues were randomly selected and cDNA synthesis was carried out as reported by Balcells et al. [58]. Primers for the qPCR amplification of miRNAs were designed with the miRprimer software [59] according to described protocols [60] and they are indicated in Table S3.

MiRspecific qPCR was performed on a MX3005P machine (Stratagene, USA). Briefly, 1 μl of cDNA diluted 8 fold, 5 μl of 2× QuantiFast SYBR Green PCR master mix (Qiagen, Germany) and 250 nM of each primer (Table S3) were mixed in a final volume of 10 μl. Cycling conditions were: 95 °C for 5 min followed by 40 cycles of 95 °C for 10 s and 60 °C for 30 s. Melting curve analyses (60 °C to 99 °C) were performed after completing amplification reaction to ensure the specificity of the assays. Data were processed with the MxPro qPCR associated software. Assays were considered successful when: 1) the melting curve was specific (1 single peak) and 2) the samples had Cq values <33 cycles (i.e. sufficiently expressed to be considered biologically functional). Finally, amplified products for muscle and liver samples were visually inspected by electrophoresis in a 3% agarose gel.

Results

Motif-based positional refinement enhances structural stability of pre-miRNA candidates

We have evaluated the usefulness of previously reported flanking motifs that enhance pre-miRNA processing [42,44] as possible novel determinants for pre-miRNA reconstruction from mature sequences. The presence of UG and CHC motifs in upstream flanking regions as well as of downstream CNNC and GHG motifs was assessed in the curated porcine miRNA annotation available in the miRCarta database [2] (Figure 2A). Consistent with data reported by Fang et al. [42] and Auyeung et al. [44], the most common flanking upstream positions for UG and CHC motifs from the 5′ start of the porcine pre-miRNA genes were −13/−12 and −7/−5, respectively, whereas for downstream CNNC and GHG motifs, the most common position from the 3’end of the pre-miRNA genes were +18/+21 and +4/+6 (Figure 2A).

Moreover, we determined the percentage of annotated porcine miRNAs that were surrounded by the aforementioned processing motifs, allowing ±2 nt of positional variation from their corresponding expected sites. From a total of 328 confidently annotated porcine pre-miRNAs in the miRCarta database [2], CNNC, UG, GHG and CHC flanking motifs were found in 53.05%, 42.68%, 30.79% and 33.54% of the sequences, respectively. The high frequency of the CNNC motif agrees well with its key role in the correct Drosha ribonuclease III (DROSHA) positioning through the recruitment of Serine and Arginine rich splicing factor 3 (SRSF3) at the basal junction of the processed pri-miRNA [61]. The proportion of the three other flanking motifs were also consistent with previously reported surveys [42,44].

To further elucidate the contribution of each motif to better delineate the boundaries of pri-miRNA processing, we compared the structural stability (i.e. the estimated *p* score of the hairpin secondary structure with the randfold approach [45]) for every pre-miRNA candidate in each of the 12 generated elongation patterns per sequence (15/60, 30/60, 15/70, 30/70, 15/80 and 30/80, with and without taking into account motif search positional refinement). As depicted in Figure 2B, predictions of candidate miRNA sequences based on positional information obtained through processing motif search showed a consistently increased structural stability compared with non-positionally corrected original sequences. This phenomenon was less evident for shorter elongation patterns, where the structural stability of the positionally corrected hairpins resembled that of non-corrected candidates (Figure 2B). In certain cases, both approaches resulted in equally stable secondary structures. Furthermore, shorter elongation patterns appeared to be more favored than their longer counterparts, showing higher overall structural stability both in small RNA-seq and homology-based derived candidate sequences (Figure 2C and D). This result highlights that the preferred length for pre-miRNA processed transcripts would be approximately in the range of 80 to 90 nt, with few cases showing longer stable hairpin structures. Interestingly, this favored pre-miRNA length interval coincides with that reported by Roden et al. [43], who determined a preferred 2× stem length of 35 nt and a terminal loop of ~10 nt, accounting for a total pre-miRNA sequence length of ~80 nt. Indeed, the average length of annotated pre-miRNAs in the porcine genome after filtering for secondary structure and sequence similarity was 84.63 nt, also in accordance with results obtained after selecting the most structurally stable elongation pattern from all generated candidates per sequence.

Classifier performance and feature importance

For assessing the performance of transductive semi-supervised miRNA classification on the porcine transcriptome, Ensembl and miRCarta positive pre-filtered porcine miRNA data sets (415 Ensembl and 244 miRCarta non-redundant hairpin-like stable annotated miRNAs) were tested against selected non-coding hairpin-like sequences (252 annotated non-coding hairpin-like RNAs other than miRNAs) and different imbalance ratios were applied by incorporating genome-wide randomly extracted hairpins (unlabeled). Overall, SE and SP obtained with the Ensembl miRNA data set (Figure 3A) were slightly better than those inferred for the miRCarta data set (Figure 3B). Ensembl average SE and SP were 0.9199 and 0.9101 respectively, whereas results obtained with the miRCarta data set were slightly worse (SE = 0.8975, SP = 0.9019). Optimal performance was achieved by using a balanced ratio between positive and negative classes, with a slightly descending trend in the classifier performance when increasing the imbalance ratio (Figure 3A and B), a result that was also observed when analyzing the ROC and PR curves (Figure S1). When we compared the performance of the semi-supervised approach vs that of other supervised algorithms, the *miRNAss* algorithm [31] implemented in the eMIRNA pipeline outperformed the rest of supervised approaches, with the exception of lGBM, which showed similar performance results (Table 2). SP, as well as AUROC and AUPR estimates obtained with the *miRNAss* method [31] showed its high ability to discard false positives miRNA candidates, at the cost of a lower SE (Table 2). Additionally, after evaluating the ability of the algorithm to correctly identify the annotated porcine miRNA loci in all defined imbalance scenarios, a total of 399 (89.92%) and 213 (87.30%) annotated miRNAs were consistently classified as miRNA sequences using Ensembl (415) and miRCarta (244) positive databases, respectively.

Table 2: Comparative benchmarking between the semi-supervised transductive learning approach employed by the *miRNAss* algorithm and other state-of-the-art supervised algorithms (i.e. SVM: Support vector machine, RF: Random forest, KNN: k-Nearest neighbors, NB: Naïve Bayes, XGB: Extreme gradient boosting and lGBM: Light gradient boosting tree) for miRNA classification. Only labeled positive and negative data sets were used for training.

Statistic	SVM	RF	KNN	NB	XGB	lGBM	miRNAss
SE	0.932	0.932	0.9223	0.9126	0.9515	0.9223	0.8835
SP	0.8413	0.9524	0.9524	0.9683	0.9365	0.9048	0.9683
F-1	0.9187	0.9505	0.9453	0.9447	0.9561	0.9314	0.9226
AUROC	0.6428	0.7246	0.5757	0.4291	0.7063	0.9781	0.9783
AUPR	0.7222	0.8489	0.6751	0.5818	0.8509	0.9873	0.987

SE: Sensitivity; SP: Specificity; F-1: F-score measure of the harmonic mean of the precision and recall; AUROC: Area under the receiver operating characteristics (ROC) curve; AUPR: Area under the precision-recall curve.

The improved performance achieved with the Ensembl data set was expected because Ensembl annotation includes a more diverse and complete miRNA catalogue (415) than miRCarta (244). However, these differences are probably due to a stricter miRNA annotation procedure in the case of miRCarta database [2], which only includes manually curated bona fide miRNA genes. Nevertheless, the slight increase in overall performance observed in the Ensembl miRNA data set evidenced that even when reducing the set of positive sequences to a more stringent annotation, as that available in the miRCarta database [2], the ability of the eMIRNA pipeline to accurately distinguish miRNA sequences from other non-miRNA hairpins remained almost unaltered.

Besides, we determined the importance of the set of calculated features for classifying the miRNA candidates with the relief-F algorithm [48,49]. The estimated importance of the 30 most discriminant features is depicted in Figure 3C. The estimated impact of each feature on the accuracy of miRNA is shown in Table S4. Structural stability-related features accounted for the most important variables for classifying miRNAs correctly (MFEadj, EFEadj, MFE, EFE, MEAFE, MFEadj.GC and CFE). All of these parameters represented different hairpin structure folding statistics and they were highly intercorrelated (Figure 3D). The discriminant power of structural stability features is better exemplified in Figure 3E, where Ensembl annotated pre-miRNA sequences had an overall higher structural stability (i.e. lower MFEadj values) compared with that of other non-coding hairpin-like RNA sequences. These results clearly show the outmost importance of the structural folding configuration in order to discriminate true miRNA candidates from other hairpin-like sequences, hence supporting the need of a careful determination of pre-miRNA boundaries.

Novel porcine miRNA identified in the muscle transcriptome and by homology-based search

After microRNA identification from the porcine small RNA-seq data set, a total of 1,403 reconstructed pre-miRNA candidates from expressed transcripts were successfully identified as putative novel miRNAs in the porcine *gluteus medius* transcriptome, which corresponded to 160 unique miRNA loci after assigning clustered isomiRs to consensus single miRNA genes. Among these, 140 consensus candidates (87.5%) overlapped already annotated miRNAs in the porcine genome, whereas the 20 remaining ones (12.5%) were classified as novel miRNA candidates.

Regarding homology-based search miRNA discovery in the porcine assembly (Sscrofa11.1), a total of 310 annotated human miRNAs had orthologous miRNA genes in the porcine genome. The already annotated miRNAs in the porcine genome comprised 281 (90.64%) of the 310 homologous miRNAs detected with eMIRNA (File S3), and the 29 (N > 0.1) remaining candidates were classified as novel non-previously annotated homologous miRNAs in the porcine assembly (Table 3). The miR-483 and miR-484 genes were also identified as novel expressed miRNA candidates in the *gluteus medius* muscle transcriptome generated in our small RNA-seq experiment. A complete list of the novel miRNA candidates obtained with *de novo* and homology-based approaches is shown in Table 3. The full list of detected miRNAs that had been already annotated and all isomiRs associated with novel miRNA sequences can be found in File S3. The existence of multiple isoform candidates for single predicted miRNA loci, either displaying polymorphisms within the mature miRNA sequence or corresponding to 5′ or 3′-trimming variations (File S3), evidenced the wide variety of isomiR sequences expressed at significant levels in our *gluteus medius* muscle transcriptomic data set.

Table 3: Novel porcine miRNA genes predicted through a homology-based comparison with human miRNA annotation and on the basis of data generated by sequencing small RNAs expressed in the *gluteus medius* muscle of Duroc pigs.

Chr	Start	End	Strand	ID	N
1	191218572	191218651	+	miR-3529	0.33
1	268816970	268817050	+	miR-219b	0.92
2	32718	32792	+	miR-6743	0.82
2	1473428	1473495	-	miR-483	0.84
2	1474436	1474513	-	3229-4643	-
2	40104336	40104403	-	1325-14520	-
2	134660802	134660897	-	1323-14559	-
3	7180536	7180603	-	miR-484	0.1
3	40421320	40421409	+	427-63874	-
3	40772345	40772445	+	176-178526	-
4	22195784	22195880	+	2340-6855	-
5	3397056	3397130	-	1111-18619	-
5	17410008	17410122	+	1794-9841	-

5	95548384	95548458	+	miR-3059	1
6	56426941	564267012	-	miR-520e	0.3
6	63490755	63490822	+	miR-200a	0.6
8	1205684	1205760	-	miR-4800	0.85
9	52087075	52087155	+	1864-9314	-
9	114528009	114528076	+	miR-3120	0.7
10	27079413	27079489	-	miR-24-1	0.79
11	1824995	1825062	+	504-51258	-
11	49808356	49808431	-	miR-3665	0.86
12	1538011	1538119	+	337-84973	-
12	1601453	1601506	-	miR-3065	0.82
12	18989584	18989651	+	399-69074	-
12	45088806	45088863	+	miR-451b	0.78
12	45597382	45597459	+	miR-4523	0.81
12	46211527	46211594	-	miR-3184	0.61
12	48162620	48162704	-	miR-132	0.84
12	56201226	56201300	-	518-49963	-
13	30242047	30242114	+	772-29980	-
13	33152284	33152383	+	miR-4787	0.83
13	197168804	197168901	+	miR-6501	0.97
14	87673881	87673954	+	3552-4147	-
14	109233945	109234032	-	miR-3085	0.95
14	122706280	122706361	+	miR-6715a	0.96
14	122706285	122706353	-	miR-6715b	0.96
14	127016706	127016794	-	miR-9851	0.83
14	140979533	140979627	+	3525-4198	-
15	128165751	128165827	-	miR-5702	0.86
17	61915309	61915376	+	1544-12001	-
X	41793240	41793315	+	451-58980	-
X	43716471	43716538	+	miR-502	0.73
X	59551153	59551220	+	miR-374c	0.8
X	94122543	94122610	+	miR-1264	0.83
X	96979691	96979765	+	miR-1277	0.68
X	124724889	124724956	-	miR-718	0.89

Chr: Chromosome; N: Neighborhood score.

The eMIRNA pipeline accurately recalls miRNA loci

The same *gluteus medius* skeletal muscle transcriptomic data from the small RNA-seq experiment employed for de novo miRNA discovery with the eMIRNA pipeline was used for running the miRDeep2 algorithm [54]. A total of 148 transcripts belonging to 134 unique annotated miRNA loci were identified with miRDeep2. These numbers were slightly smaller than the 140 annotated porcine miRNAs recovered as expressed transcripts by the eMIRNA pipeline. Among these, 126 annotated miRNAs (85.14%) were consistently recovered with eMIRNA and miRDeep2, 14 (9.46%) were only reported by eMIRNA, and 8 (5.41%) were exclusively predicted by miRDeep2 (Table S5).

Regarding novel candidates, miRDeep2 was able to recover a total of 11 putative novel candidates belonging to 10 unique loci (Table S6). Seven of these candidates displayed an estimated probability of being a true positive miRNA above 19% (miRDeep2 score ≥ 4, Table S6). Noteworthy, two of the putatively true miRNAs detected by miRDeep2 spanned other previously annotated non-coding RNAs in the porcine assembly and were hence considered as miRNA-like false positives (Table S6). Among the 5 remaining candidates, 4 of them (miR-193a, miR-26a, miR-106b and miR-17) spanned other already annotated miRNAs in the porcine assembly and were thus wrongly classified as novel miRNAs by miRDeep2. The remaining candidate corresponded to miR-483, which had already been identified with the eMIRNA pipeline (Table 3, Table S6).

When comparing the accuracy of miRNA loci boundaries determined by the eMIRNA pipeline and miRDeep2, the eMIRNA approach demonstrated an overall better capability to accurately assign miRNA boundaries according to data from porcine miRNA loci annotated in the Ensembl database. A total of 103 out of 126 (81.74%) annotated miRNA genes detected by both eMIRNA and miRDeep2 showed reduced ΔD values (Table S7). This result implies that genomic positions of miRNA precursors predicted with the eMIRNA pipeline were more concordant with the annotation of the Sscrofa11.1 assembly than those predicted with miRDeep2. This outcome illustrates the effectiveness of motif search positional correction for reconstructing pre-miRNA candidates with a higher reliability than the fixed elongation patterns strategy used by miRDeep2 [54]. Three of the miRNA candidates showed no differences in positional accuracy between both approaches, while the positions of the remaining sequences (15.87%) were more accurately predicted with miRDeep2 (Table S7).

Experimental confirmation of the existence of three novel miRNAs in the muscle and liver tissues of Göttingen minipigs

The RT-qPCR analyses allowed us to detect the expression of the novel ssc-miR-483, ssc-miR-484 and ssc-miR-200a candidates in both *longissimus dorsi* skeletal muscle and liver tissues (Figure S2A and B) retrieved from Göttingen minipigs. Both ssc-miR-483 and ssc-miR-484 were also detected as consistently expressed in the skeletal muscle of Duroc gilts from our small RNA-seq experiment. The ssc-miR-200a was also detected in our generated data set but at very low expression levels. Nevertheless, its expression was further confirmed independently by RT-qPCR analyses. Amplification profiles and melting curves for the three novel miRNA candidates detected by RT-qPCR are shown in File S4.

Discussion

In the discovery of novel miRNA genes, one essential issue is the generation of pre-miRNA sequence candidates, given that the majority of miRNA prediction tools are based on feature extraction from the well-defined pre-miRNA hairpin structure [62]. At the cellular level, the most abundant and stable miRNA transcripts are the mature miRNA forms. Indeed, precursor stages, such as pri or pre-miRNAs, are much less abundant and have shorter half-lives than mature miRNAs [63,64]. Therefore, the accurate definition of pre-miRNA boundaries reconstructed from mature miRNAs is a crucial issue in order to predict folding structure and minimum free energy (MFE) estimates in a robust manner.

Noteworthy, the majority of state-of-the-art methods for miRNA prediction are solely focused on the miRNA classification of predefined candidate sequences. Moreover, many of them do not contemplate the generation of such candidates for the identification of unannotated miRNAs. On the contrary, they rely on well-known hairpins or on sets of manually curated candidate sequences that are embedded in their prediction pipelines [30,31,65-72].

Several other algorithms take advantage of the automated generation of hairpin candidates, adopting fixed defined elongation patterns in order to reconstruct pre-miRNA candidates from mature miRNA sequences [9,11,73,74]. However, fixed assumptions about elongation patterns do not take into consideration the expected variable length of pre-miRNA loci, and tend to generate candidate sequences that, despite harboring mature miRNAs, might have unreliable boundaries. This may lead to inaccuracies in the folding prediction and thus to an augmentation of the false negative rate. Even worse, non-miRNA hairpin-like sequences strongly resembling pre-miRNAs may be generated through the blind elongation of short sequences, which could result in the emergence of false positive candidates. This situation is particularly critical when analyzing the reliability of miRNA annotation in public databases [27,75,76]. Other approaches have also adopted a multiple hairpin candidate search for each query sequence to further select those showing a higher structural stability [77-79]. By using this strategy, we explored the influence of flanking processing motifs on the accurate determination of the length and boundaries of pre-miRNA candidates. By doing so, we have demonstrated that the inclusion of processing motif search criteria for the estimation of pre-miRNA boundaries resulted in an improved ability to better assess the optimal candidate sequences to be used for miRNA prediction.

Compared with miRDeep2 [54], the eMIRNA pipeline showed an improved ability to better assess the already annotated miRNA loci boundaries after pre-miRNA sequence reconstruction. However, the presence of embedded processing motifs within the boundaries of miRNA genes is not a universal feature, with a non-negligible amount of miRNA loci lacking the well-known CNNC and UG motifs [44], as well as the CHC and GHG mismatches [42] in their proximal surroundings. Additional work is needed to better characterize other processing motifs or structural determinants that may also contribute to miRNA maturation.

In contrast with pre-existing supervised methods for miRNA discovery, few semi-supervised methods have been developed for such purpose [31,80]. From a biological perspective, the scarce miRNA annotation typically found in non-model species poses a great challenge when attempting to predict novel miRNA loci uniquely based on labeled data. This happens because the amount of unknown non-miRNA sequences with hairpin-like secondary structures is expected to be hundreds of times larger than the number of confidently annotated miRNAs to be used for training supervised algorithms. Despite the fact that good performance statistics may be obtained after classifier training, supervised algorithms heavily depend on the existence of an extensive miRNA annotation. Indeed, the ability of such classifiers to detect unannotated miRNA sequences is mainly driven by the amount and diversity of positive and negative instances used for learning training.

On the contrary, semi-supervised transductive approaches [31] are able to overcome such limitation by incorporating unlabeled cases to the training process, with the aim of increasing the variability of the data used for target sequences classification. In fact, allowing the classifier to check hundreds or thousands of unknown unlabeled sequences has proven to increase the validity of microRNA prediction over other methods solely based on labeled data [31], a result that was also verified when comparing the semi-supervised approach used in this study with other broadly reported supervised methods (Table 2). This strategy is particularly reliable when few positive data are available and the annotated negative data set only represent a small proportion of the whole non-miRNA class. Besides, in classification problems where the negative class is expected to be dozens or hundreds of times larger than the positive class, the accurate identification of false positives is crucial. Indeed, such scenario is completely applicable to miRNAs, where thousands of non-miRNA sequences exist compared with the few hundreds of reliably annotated miRNA genes, and the annotation of negative hairpin-like sequences only represents a small proportion of the whole non-miRNA class.

After miRNA prediction, the detection of multiple isoforms for each single predicted miRNA loci evidenced the existence of a broad array of isomiR sequences expressed at significant levels in our *gluteus medius* muscle transcriptomic data set (File S3). Previous studies have highlighted the importance of isomiRs in expanding the biological diversity of miRNA function [81-84]. Like canonical miRNAs, isomiRs are also evolutionary conserved [81]. Both 5′ and 3′ miRNA isoforms can be generated either from alternative processing sites of DROSHA and Dicer [43,85] or from post-transcriptional modifications, influencing miRNA half-lives as well as their interactions with RNA-binding proteins (RBPs) [86,87].

More recently, other integrative approaches have addressed the detection of isomiRs and the potential functional influence that subtle modifications in the 3′ and 5′ boundaries of mature miRNA sequences might have on target recognition [88-91]. Other studies have also reported 5′ alternative processing events in a large number of miRNAs, contributing to the expansion of their target repertoire at a higher rate than previously thought [92]. Despite these promising results, the biological implications of miRNA alternative processing events leading to the generation of isomiRs are still poorly understood and further research is needed in order to exclude potential biases in isomiR quantification and functional validation, as variations in 3′

or 5′ ends of mature miRNAs can strongly affect the reliability of stem-loop qPCR amplification protocols [93].

One potential limitation of our study is that 17 of the novel miRNAs predicted with eMIRNA and based on muscle transcriptomic data have not been further investigated in order to confirm their existence by RT-qPCR, so their experimental validation is still pending. Indeed, we only investigated 3 out of 20 predicted novel porcine miRNAs. Noteworthy, the three selected miRNAs were successfully confirmed as bona fide miRNAs by RT-qPCR thus suggesting that eMIRNA predictions are accurate.

Among the three validated miRNAs, it is worth mentioning miR-483, which has been functionally associated with cell growth regulation [94] as well as with insulin resistance and metabolic syndrome susceptibility likely due to its strong implication in the regulation of glucose metabolism [95,96]. Additionally, the expression of miR-483, whose coding sequence maps to the second intron of the insulin growth factor 2 (*IGF2*) gene, has been tightly associated with an enhancement of *IGF2* gene expression. This is achieved through the binding of miR-483 to transcription factors in a positive feed-back loop [97], although other authors have questioned such dependence [98]. Other relevant successfully profiled miRNAs were ssc-miR-200a and ssc-miR-484. The miR-200a gene has been mainly reported as a regulator of cell growth and differentiation through targeting several protein-encoding transcripts like the growth factor receptor-bound 2 (*GRB2*), α-smooth muscle actin (*α-SMA*) or the fibroblast-specific protein-1 (*FSP-1*), thus hampering the endothelial-mesenchymal transition [99]. Furthermore, miR-484 has been associated with the inhibition of Fis1-mediated mitochondrial fission and apoptosis signaling [100].

Conclusions

In this study we have implemented an end-to-end pipeline that may facilitate the identification of novel miRNAs in the porcine genome. We have tested the eMIRNA pipeline by following a homology-based approach making use of the well annotated human microRNA transcriptome. Besides, we have analyzed the presence of non-annotated miRNAs in the porcine genome using data from a small RNA-seq experiment comprising muscle samples from 48 Duroc gilts. We have also taken into consideration several issues that are critical to robustly predict miRNA genes, such as the accurate reconstruction of candidate pre-miRNAs, the correct definition of negative training data sets and the evaluation of the high class-imbalance phenomenon, which is not fully addressed in many miRNA-prediction studies. In parallel, we have established hard-threshold filtering steps to keep false positive predictions at a minimum. We have also demonstrated the usefulness of positional refinement through flanking motif search to better determine the boundaries of pre-miRNA hairpin-like candidate sequences. The expression of several of the novel miRNAs described in this work was further confirmed by RT-qPCR analyses. In the light of these results, we believe that the eMIRNA pipeline will facilitate the discovery and annotation of novel miRNAs, thus broadening the miRNA catalogue of non-model species with yet poorly annotated genome assemblies.

Supplementary Information

Figure S1: (A) Receiver operating characteristics (ROC) and **(B)** precision-recall (PR) curves computed for each pre-defined imbalance scenario using porcine Ensembl annotation for positive (miRNAs) and negative (other hairpin-like non-coding RNAs) data sets.

Figure S2: RT-qPCR results of selected novel miRNAs. Successfully profiled novel miRNAs in **(A)** the *longissimus dorsi* skeletal muscle and **(B)** liver tissues from 7 Göttingen minipigs.

File S1: FASTA file of collapsed expressed sequences (RPM > 10) used in the *de novo* discovery of miRNAs expressed in the porcine *gluteus medius* skeletal muscle.

File S2: Non-redundant annotated mature miRNA sequences obtained from the *H. sapiens* GRCh38.p12 genome assembly used as a reference in the homology-based search of novel miRNAs in the current release of the porcine genome (Sscrofa11.1).

File S3: List of already annotated miRNAs and all isomiRs detected as expressed (RPM > 10) in the porcine *gluteus medius* skeletal muscle.

File S4: Amplification profiles and melting curves for the three novel miRNA candidates subjected to confirmation by RT-qPCR analyses.

Table S1: Area under the curve (AUC) computed for each pre-defined imbalance scenario using Ensembl annotation for positive and negative data sets.

Table S2: True positive ratio of porcine miRNA loci annotated in the Ensembl and miRCarta databases and identified by the eMIRNA pipeline in all considered imbalance scenarios.

Table S3: Mature miRNAs and primers used for RT-qPCR confirmation of selected novel miRNA candidates.

Table S4: Feature importance according to the relief-F algorithm.

Table S5: Previously annotated miRNAs genes that are correctly classified as miRNAs by eMIRNA and miRDeep2.

Table S6: miRDeep2 algorithm results for miRNA prediction using the *gluteus medius* muscle small RNA-seq data generated in the present study.

Table S7: Deviation rates (dr) and Differential deviation (ΔD) estimates for miRNA genomic positional prediction with eMIRNA and miRDeep2.

Support

The research presented in this publication was funded by grants AGL2013-48742-C2-1-R and AGL2013-48742-C2-2-R awarded by the Spanish Ministry of Economy and Competitivity. E. Mármol-Sánchez was funded with a Ph.D. fellowship FPU15/01733 awarded by the Spanish Ministry of Education and Culture (MECD).

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

The authors would like to thank the Department of Veterinary Animal Sciences in the Faculty of Health and Medical Sciences of the University of Copenhagen for providing their facilities and materials for RT-qPCR experiments. We would also like to express our gratitude to Dr. Caroline M. Junker Mentzel for kindly providing RNA samples for qPCR analyses. We also acknowledge Selección Batallé S.A. for providing animal material and the Spanish Ministry of Economy and Competitivity for the Center of Excellence Severo Ochoa 2016-2019 (SEV-2015-0533) grant awarded to the Center for Research in Agricultural Genomics (CRAG). Thanks also to the CERCA Programme of the Generalitat the Catalunya for their support.

References

[1] A. Kozomara, M. Birgaoanu, S. Griffiths-Jones, miRBase: from microRNA sequences to function, Nucleic Acids Res. 47 (2019) D155–D162.

[2] C. Backes, T. Fehlmann, F. Kern, T. Kehl, H.-P. Lenhof, E. Meese, A. Keller, miRCarta: a central repository for collecting miRNA candidates, Nucleic Acids Res. 46 (2018) D160–D167.

[3] B. From, D. Domanska, L. Høye, V. Ovchinnikov, W. Kang, E. Aparicio-Puerta, M. Johansen, K. Flatmark, A. Mathelier, E. Hovig, M. Hackenberg, M.R. Friedländer, K.J. Peterson, MirGeneDB2.0: the metazoan microRNA complement, Nucleic Acids Res. (2019) gkz885.

[4] J. Meunier, F. Lemoine, M. Soumillon, A. Liechti, M. Weier, K. Guschanski, H. Hu, P. Khaitovich, H. Kaessmann, Birth and expression evolution of mammalian microRNA genes, Genome Res. 23 (2013) 34–45.

[5] M. Warnefors, A. Liechti, J. Halbert, D. Valloton, H. Kaessmann, Conserved microRNA editing in mammalian evolution, development and disease, Genome Biol. 15 (2014) R83.

[6] L.P. Lim, N.C. Lau, E.G. Weinstein, A. Abdelhakim, S. Yekta, M.W. Rhoades, C.B. Burge, D.P. Bartel, The microRNAs of Caenorhabditis elegans, Genes Dev. 17 (2003) 991–1008.

[7] E.C. Lai, P. Tomancak, R.W. Williams, G.M. Rubin, Computational identification of Drosophila microRNA genes, Genome Biol. 4 (2003) R42.

[8] X. Wang, J. Zhang, F. Li, J. Gu, T. He, X. Zhang, Y. Li, MicroRNA identification based on sequence and structure alignment, Bioinformatics. 21 (2005) 3610–3614.

[9] A. Mathelier, A. Carbone, MIReNA: finding microRNAs with high accuracy and no learning at genome scale and from deep sequencing data, Bioinformatics. 26 (2010) 2226–2234.

[10] K. Qian, E. Auvinen, D. Greco, P. Auvinen, miRSeqNovel: An R based workflow for analyzing miRNA sequencing data, Mol. Cell. Probes 26 (2012) 208–211.

[11] J. An, J. Lai, M.L. Lehman, C.C. Nelson, MiRDeep*: An integrated application tool for miRNA identification from RNA sequencing data, Nucleic Acids Res. 41 (2013) 727–737.

[12] T.B. Hansen, M.T. Venø, J. Kjems, C.K. Damgaard, miRdentify: high stringency miRNA predictor identifies several novel animal miRNAs, Nucleic Acids Res. 42 (2014) e124.

[13] D. Kleftogiannis, A. Korfiati, K. Theofilatos, S. Likothanassis, A. Tsakalidis, S. Mavroudi, Where we stand, where we are moving: surveying computational techniques for identifying miRNA genes and uncovering their regulatory role, J. Biomed. Inform. 46 (2013) 563–573.

[14] M. Bortolomeazzi, E. Gaffo, S. Bortoluzzi, A survey of software tools for microRNA discovery and characterization using RNA-seq, Brief. Bioinform. 20 (2017) 918–930.

[15] G. Stegmayer, L.E. Di Persia, M. Rubiolo, M. Gerard, M. Pividori, C. Yones, L.A. Bugnon, T. Rodriguez, J. Raad, D.H. Milone, Predicting novel microRNA: a comprehensive comparison of machine learning approaches, Brief. Bioinform. (2018) bby037.

[16] A. Rajendiran, A. Chatterjee, A. Pan, Computational approaches and related tools to identify microRNAs in a species: a bird's eye view, Interdiscip. Sci. Comput. Life Sci. 10 (2018) 616–635.

[17] J.-E. Long, H.-X. Chen, Identification and characteristics of cattle microRNAs by homology searching and small RNA cloning, Biochem. Genet. 47 (2009) 329–343.

[18] Z. Wang, K. He, Q. Wang, Y. Yang, Y. Pan, The prediction of the porcine premicroRNAs in genome-wide based on support vector machine (SVM) and homology searching, BMC Genomics 13 (2012) 729.

[19] X. Hou, Z. Tang, H. Liu, N. Wang, H. Ju, K. Li, Discovery of microRNAs associated with myogenesis by deep sequencing of serial developmental skeletal muscles in pigs, PLoS One 7 (2012) e52123.

[20] C. Yuan, X. Wang, R. Geng, X. He, L. Qu, Y. Chen, Discovery of cashmere goat (Capra hircus) microRNAs in skin and hair follicles by Solexa sequencing, BMC Genomics 14 (2013) 511.

[21] J. Sun, M. Li, Z. Li, J. Xue, X. Lan, C. Zhang, C. Lei, H. Chen, Identification and profiling of conserved and novel microRNAs from Chinese Qinchuan bovine longissimus thoracis, BMC Genomics 14 (2013) 42.

[22] T. Buza, M. Arick, H. Wang, D.G. Peterson, Computational prediction of disease microRNAs in domestic animals, BMC Res. Notes. 7 (2014) 403.

[23] B. Sadeghi, H. Ahmadi, S. Azimzadeh-Jamalkandi, M.R. Nassiri, A. Masoudi-Nejad, BosFinder: a novel pre-microRNA gene prediction algorithm in Bos taurus, Anim. Genet. 45 (2014) 479–484.

[24] J. Wu, H. Zhu, W. Song, M. Li, C. Liu, N. Li, F. Tang, H. Mu, M. Liao, X. Li, W. Guan, X. Li, J. Hua, Identification of conservative microRNAs in Saanen dairy goat testis through deep sequencing, Reprod. Domest. Anim. 49 (2014) 32–40.

[25] Z. Li, H. Wang, L. Chen, L. Wang, X. Liu, C. Ru, A. Song, Identification and characterization of novel and differentially expressed microRNAs in peripheral blood from healthy and mastitis Holstein cattle by deep sequencing, Anim. Genet. 45 (2014) 20–27.

[26] D.M.D. Saçar, H. Hamzeiy, J. Allmer, Can miRBase provide positive data for machine learning for the detection of miRNA hairpins? J. Integr. Bioinform. 10 (2013) 1–11.

[27] N. Ludwig, M. Becker, T. Schumann, T. Speer, T. Fehlmann, A. Keller, E. Meese, Bias in recent miRBase annotations potentially associated with RNA quality issues, Sci. Rep. 7 (2017) 5162.

[28] L. Wei, M. Liao, Y. Gao, R. Ji, Z. He, Q. Zou, Improved and promising identification of human microRNAs by incorporating a high-quality negative set, IEEE/ACM Trans. Comput. Biol. Bioinforma. 11 (2014) 192–201.

[29] M. Yousef, J. Allmer, W. Khalifa, Accurate plant microRNA prediction can be achieved using sequence motif features, J. Intell. Learn. Syst. Appl. 8 (2016) 9–22.

[30] G. Stegmayer, C. Yones, L. Kamenetzky, D.H. Milone, High class-imbalance in premiRNA prediction: a novel approach based on deepSOM, IEEE/ACM Trans. Comput. Biol. Bioinforma. 14 (2017) 1316–1326.

[31] C. Yones, G. Stegmayer, D.H. Milone, C. Sahinalp, Genome-wide pre-miRNA discovery from few labeled examples, Bioinformatics. 34 (2018) 541–549.

[32] Y. Wang, X. Li, B. Tao, Improving classification of mature microRNA by solving class imbalance problem, Sci. Rep. 6 (2016) 25941.

[33] T.F. Cardoso, R. Quintanilla, J. Tibau, M. Gil, E. Mármol-Sánchez, O. González-Rodríguez, R. González-Prendes, M. Amills, Nutrient supply affects the mRNA expression profile of the porcine skeletal muscle, BMC Genomics 18 (2017) 603.

[34] M. Ballester, M. Amills, O. González-Rodríguez, T.F. Cardoso, M. Pascual, R. González-Prendes, N. Panella-Riera, I. Díaz, J. Tibau, R. Quintanilla, Role of AMPK signaling pathway during compensatory growth in pigs, BMC Genomics 19 (2018) 682.

[35] A.R. Quinlan, I.M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features, Bioinformatics. 26 (2010) 841–842.

[36] Y. Huang, B. Niu, Y. Gao, L. Fu, W. Li, CD-HIT suite: a web server for clustering and comparing biological sequences, Bioinformatics. 26 (2010) 680–682.

[37] R. Lorenz, S.H. Bernhart, C. Höner zu Siederdissen, H. Tafer, C. Flamm, P.F. Stadler, I.L. Hofacker, ViennaRNA Package 2.0, Algorithms Mol. Biol. 6 (2011) 26.

[38] C. Yones, HextractoR: Integrated tool for hairpin extraction of RNA sequences, R Package Version 1.3, 2018 https://cran.r-project.org/package=HextractoR.

[39] M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads, EMBnet.journal. 17 (2011) 10.

[40] Y. Lu, A.S. Baras, M.K. Halushka, miRge 2.0 for comprehensive analysis of microRNA sequencing data, BMC Bioinforma. 19 (2018) 275.

[41] B. Langmead, C. Trapnell, M. Pop, S. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome, Genome Biol. 10 (2009) R25.

[42] W. Fang, D.P. Bartel, The menu of features that define primary microRNAs and enable de novo design of microRNA genes, Mol. Cell 60 (2015) 131–145.

[43] C. Roden, J. Gaillard, S. Kanoria, W. Rennie, S. Barish, J. Cheng, W. Pan, J. Liu, C. Cotsapas, Y. Ding, J. Lu, Novel determinants of mammalian primary microRNA processing revealed by systematic evaluation of hairpin-containing transcripts and human genetic variation, Genome Res. 27 (2017) 374–384.

[44] V.C. Auyeung, I. Ulitsky, S.E. McGeary, D.P. Bartel, Beyond secondary structure: primary-sequence determinants license pri-miRNA hairpins for processing, Cell. 152 (2013) 844–858.

[45] E. Bonnet, J. Wuyts, P. Rouze, Y. Van de Peer, Evidence that microRNA precursors, unlike other non-coding RNAs, have lower folding free energies than random sequences, Bioinformatics. 20 (2004) 2911–2917.

[46] M. Jiang, J. Anderson, J. Gillespie, M. Mayne, uShuffle: a useful tool for shuffling biological sequences while preserving the k-let counts, BMC Bioinforma. 9 (2008) 192.

[47] I. Lopes, A. Schliep, A.C.L.F. de Carvalho, The discriminant power of RNA features for pre-miRNA recognition, BMC Bioinforma. 15 (2014) 124.

[48] I. Kononenko, E. Šimec, M. Robnik-Šikonja, Overcoming the myopia of inductive learning algorithms with RELIEFF, Appl. Intell. 7 (1997) 39–55.

[49] M. Robnik-Šikonja, I. Kononenko, Theoretical and empirical analysis of ReliefF and RReliefF, Mach. Learn. 53 (2003) 23–69, https://doi.org/10.1023/ A:1025667309714.

[50] R. Batuwita, V. Palade, Adjusted geometric-mean: a novel performance measure for imbalanced bioinformatics data sets learning, J. Bioinforma. Comput. Biol. 10 (2012) 1250003.

[51] J. Davis, M. Goadrich, The relationship between precision-recall and ROC curves, ACM Int. Conf. Proceeding Ser. (2006) 233–240.

[52] G.S. França, M.D. Vibranovski, P.A.F. Galante, Host gene constraints and genomic context impact the expression and evolution of human microRNAs, Nat. Commun. 7 (2016) 11438.

[53] M.R. Friedländer, W. Chen, C. Adamidi, J. Maaskola, R. Einspanier, S. Knespel, N. Rajewsky, Discovering microRNAs from deep sequencing data using miRDeep, Nat. Biotechnol. 26 (2008) 407–415.

[54] M.R. Friedländer, S.D. MacKowiak, N. Li, W. Chen, N. Rajewsky, MiRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades, Nucleic Acids Res. 40 (2012) 37–52.

[55] M. Kuhn, Building predictive models in R using the caret package, J. Stat. Softw. 28 (2008) 1–26.

[56] T. Chen, C. Guestrin, XGBoost: A scalable tree boosting system, Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., 2016, pp. 785–794.

[57] C.M.J. Mentzel, C. Anthon, M.J. Jacobsen, P. Karlskov-Mortensen, C.S. Bruun, C.B. Jørgensen, J. Gorodkin, S. Cirera, M. Fredholm, Gender and obesity specific microRNA expression in adipose tissue from lean and obese pigs, PLoS One 10 (2015) e0131650.

[58] I. Balcells, S. Cirera, P.K. Busk, Specific and sensitive quantitative RT-PCR of miRNAs with DNA primers, BMC Biotechnol. 11 (2011) 70.

[59] P.K. Busk, A tool for design of primers for microRNA-specific quantitative RT-qPCR, BMC Bioinforma. 15 (2014) 29.

[60] S. Cirera, P.K. Busk, Quantification of miRNAs by a simple and specific qPCR method, Methods Mol. Biol. (2014) 73–81.

[61] K. Kim, T. Duc Nguyen, S. Li, T. Anh Nguyen, SRSF3 recruits DROSHA to the basal junction of primary microRNAs, RNA. 24 (2018) 892–898.

[62] D.P. Bartel, Metazoan microRNAs, Cell. 173 (2018) 20–51.

[63] L. Gan, B. Denecke, Profiling pre-microRNA and mature microRNA expressions using a single microarray and avoiding separate sample preparation, Microarrays. 2 (2013) 24–33.

[64] Y. Guo, J. Liu, S.J. Elfenbein, Y. Ma, M. Zhong, C. Qiu, Y. Ding, J. Lu, Characterization of the mammalian miRNA turnover landscape, Nucleic Acids Res. 43 (2015) 2326–2341.

[65] C. Xue, F. Li, T. He, G.-P. Liu, Y. Li, X. Zhang, Classification of real and pseudo microRNA precursors using local structure-sequence features and support vector machine, BMC Bioinforma. 6 (2005) 310.

[66] P. Jiang, H. Wu, W. Wang, W. Ma, X. Sun, Z. Lu, MiPred: classification of real and pseudo microRNA precursors using random forest prediction model with combined features, Nucleic Acids Res. 35 (2007) W339–W344.

[67] R. Batuwita, V. Palade, microPred: effective classification of pre-miRNAs for human miRNA gene prediction, Bioinformatics. 25 (2009) 989–995.

[68] Y. Wu, B. Wei, H. Liu, T. Li, S. Rayner, MiRPara: a SVM-based software tool for prediction of most probable microRNA coding regions in genome scale sequences, BMC Bioinforma. 12 (2011) 107.

[69] A. Gudyś, M.W. Szcześniak, M. Sikora, I. Makałowska, HuntMi: an efficient and taxon-specific approach in pre-miRNA identification, BMC Bioinforma. 14 (2013) 83.

[70] Q. Zou, Y. Mao, L. Hu, Y. Wu, Z. Ji, miRClassify: an advanced web server for miRNA family classification and annotation, Comput. Biol. Med. 45 (2014) 157–160.

[71] D. Kleftogiannis, K. Theofilatos, S. Likothanassis, S. Mavroudi, YamiPred: a novel evolutionary method for predicting pre-miRNAs and selecting relevant features, IEEE/ACM Trans. Comput. Biol. Bioinforma. 12 (2015) 1183–1192.

[72] D.M.D. Saçar, J. Baumbach, J. Allmer, On the performance of pre-microRNA detection algorithms, Nat. Commun. 8 (2017) 330.

[73] D.M. Vitsios, E. Kentepozidou, L. Quintais, E. Benito-Gutiérrez, S. van Dongen, M.P. Davis, A.J. Enright, Mirnovo: genome-free prediction of microRNAs from small RNA sequencing data and single-cells using decision forests, Nucleic Acids Res. 45 (2017) e177.

[74] R.J. Peace, M. Sheikh Hassani, J.R. Green, miPIE: NGS-based prediction of miRNA using integrated evidence, Sci. Rep. 9 (2019) 1548.

[75] M.J. Axtell, B.C. Meyers, Revisiting criteria for plant microRNA annotation in the era of big data, Plant Cell 30 (2018) 272–284.

[76] J. Alles, T. Fehlmann, U. Fischer, C. Backes, V. Galata, M. Minet, M. Hart, M. AbuHalima, F.A. Grässer, H.-P. Lenhof, A. Keller, E. Meese, An estimate of the total number of true human miRNAs, Nucleic Acids Res. 47 (2019) 3353–3364.

[77] J. Lei, Y. Sun, miR-PREFeR: an accurate, fast and easy-to-use plant miRNA prediction tool using small RNA-seq data, Bioinformatics. 30 (2014) 2837–2839.

[78] M. Evers, M. Huttner, A. Dueck, G. Meister, J.C. Engelmann, miRA: adaptable novel miRNA identification in plants using small RNA sequencing data, BMC Bioinforma. 16 (2015) 370.

[79] C. Paicu, I. Mohorianu, M. Stocks, P. Xu, A. Coince, M. Billmeier, T. Dalmay, V. Moulton, S. Moxon, miRCat2: accurate prediction of plant and animal microRNAs from next-generation sequencing data sets, Bioinformatics. 33 (2017) 2446–2454.

[80] M. Sheikh Hassani, J.R. Green, Multi-view co-training for microRNA prediction, Sci. Rep. 9 (2019) 10931.

[81] G.C. Tan, E. Chan, A. Molnar, R. Sarkar, D. Alexieva, I.M. Isa, S. Robinson, S. Zhang, P. Ellis, C.F. Langford, P.V. Guillot, A. Chandrashekran, N.M. Fisk, L. Castellano, G. Meister, R.M. Winston, W. Cui, D. Baulcombe, N.J. Dibb, 5′ isomiR variation is of functional and evolutionary importance, Nucleic Acids Res. 42 (2014) 9424–9435.

[82] A.G. Telonis, P. Loher, Y. Jing, E. Londin, I. Rigoutsos, Beyond the one-locus-one-miRNA paradigm: microRNA isoforms enable deeper insights into breast cancer heterogeneity, Nucleic Acids Res. 43 (2015) 9158–9175.

[83] F. Yu, K.A. Pillman, C.T. Neilsen, J. Toubia, D.M. Lawrence, A. Tsykin, M.P. Gantier, D.F. Callen, G.J. Goodall, C.P. Bracken, Naturally existing isoforms of miR-222 have distinct functions, Nucleic Acids Res. 45 (2017) 11371–11385.

[84] P. Sheng, C. Fields, K. Aadland, T. Wei, O. Kolaczkowski, T. Gu, B. Kolaczkowski, M. Xie, Dicer cleaves 5′-extended microRNA precursors originating from RNA polymerase II transcription start sites, Nucleic Acids Res. 46 (2018) 5737–5752.

[85] B. Kim, K. Jeong, V.N. Kim, Genome-wide mapping of DROSHA cleavage sites on primary microRNAs and noncanonical substrates, Mol. Cell 66 (2017) 258–269.e5.

[86] C.T. Neilsen, G.J. Goodall, C.P. Bracken, IsomiRs – the overlooked repertoire in the dynamic microRNAome, Trends Genet. 28 (2012) 544–549.

[87] X. Bofill-De Ros, A. Yang, S. Gu, IsomiRs: expanding the miRNA repression toolbox beyond the seed, Biochim. Biophys. Acta - Gene Regul. Mech. (2019) 194373.

[88] G. Urgese, G. Paciello, A. Acquaviva, E. Ficarra, isomiR-SEA: an RNA-seq analysis tool for miRNAs/isomiRs expression level profiling and miRNA-mRNA interaction sites evaluation, BMC Bioinforma. 17 (2016) 148.

[89] Y. Zhang, Q. Zang, B. Xu, W. Zheng, R. Ban, H. Zhang, Y. Yang, Q. Hao, F. Iqbal, A. Li, Q. Shi, IsomiR Bank: a research resource for tracking IsomiRs, Bioinformatics. 32 (2016) 2069–2071.

[90] X. Bofill-De Ros, K. Chen, S. Chen, N. Tesic, D. Randjelovic, N. Skundric, S. Nesic, V. Varjacic, E.H. Williams, R. Malhotra, M. Jiang, S. Gu, QuagmiR: a cloud-based application for isomiR big data analytics, Bioinformatics. 35 (2019) 1576–1578.

[91] X. Bofill-De Ros, W.K. Kasprzak, Y. Bhandari, L. Fan, Q. Cavanaugh, M. Jiang, L. Dai, A. Yang, T.-J. Shao, B.A. Shapiro, Y.-X. Wang, S. Gu, Structural differences between pri-miRNA paralogs promote alternative Drosha cleavage and expand target repertoires, Cell Rep. 26 (2019) 447–459.e4.

[92] H. Kim, J. Kim, K. Kim, H. Chang, K. You, V.N. Kim, Bias-minimized quantification of microRNA reveals widespread alternative processing and 3′ end modification, Nucleic Acids Res. 47 (2019) 2630–2640.

[93] A. Schamberger, T.I. Orbán, 3’ IsomiR species and DNA contamination influence reliable quantification of microRNAs by stem-loop quantitative PCR, PLoS One 9 (2014) e106315.

[94] T.H. Vu, N.V. Chuyen, T. Li, A.R. Hoffman, M. Blick, F. Fornari, N. Zanesi, H. Alder, G. D'Elia, L. Gramantieri, L. Bolondi, G. Lanza, P. Querzoli, A. Angioni, C.M. Croce, M. Negrini, Loss of imprinting of IGF2 sense and antisense transcripts in Wilms' tumor, Cancer Res. 63 (2003) 1900–1905.

[95] D. Ferland-McCollough, D.S. Fernandez-Twinn, I.G. Cannell, H. David, M. Warner, A.A. Vaag, J. Bork-Jensen, C. Brøns, T.W. Gant, A.E. Willis, K. Siddle, M. Bushell, S.E. Ozanne, Programming of adipose tissue miR-483-3p and GDF-3 expression by maternal diet in type 2 diabetes, Cell Death Differ. 19 (2012) 1003–1012.

[96] F. Pepe, S. Pagotto, S. Soliman, C. Rossi, P. Lanuti, C. Braconi, R. Mariani-Costantini, R. Visone, A. Veronese, Regulation of miR-483-3p by the O-linked N-acetylglucosamine

transferase links chemosensitivity to glucose metabolism in liver cancer cells, Oncogenesis. 6 (2017) e328.

[97] M. Liu, A. Roth, M. Yu, R. Morris, F. Bersani, M.N. Rivera, J. Lu, T. Shioda, S. Vasudevan, S. Ramaswamy, S. Maheswaran, S. Diederichs, D.A. Haber, The IGF2 intronic miR-483 selectively enhances transcription from IGF2 fetal promoters and enhances tumorigenesis, Genes Dev. 27 (2013) 2543–2548.

[98] A. Veronese, L. Lupini, J. Consiglio, R. Visone, M. Ferracin, F. Fornari, N. Zanesi, H. Alder, G. D'Elia, L. Gramantieri, L. Bolondi, G. Lanza, P. Querzoli, A. Angioni, C.M. Croce, M. Negrini, Oncogenic role of miR-483-3p at the IGF2/483 locus, Cancer Res. 70 (2010) 3140–3149.

[99] H. Zhang, J. Hu, L. Liu, MiR-200a modulates TGF- β 1-induced endothelial-to-mesenchymal shift via suppression of GRB2 in HAECs, Biomed. Pharmacother. 95 (2017) 215–222.

[100] K. Wang, B. Long, J.-Q. Jiao, J.-X. Wang, J.-P. Liu, Q. Li, P.-F. Li, miR-484 regulates mitochondrial network through targeting Fis1, Nat. Commun. 3 (2012) 781.

Journal of Animal Science
and Biotechnology
BMC

Co-expression network analysis predicts a key role of microRNAs in the adaptation of the porcine skeletal muscle to nutrient supply

Mármol-Sánchez, E.[1], Ramayo-Caldas, Y.[2], Quintanilla, R.[2], Cardoso, T. F.[1,3], Tibau, J.[2] and Amills, M.[1,4*]

[1]Department of Animal Genetics, Centre for Research in Agricultural Genomics (CRAG), CSIC-IRTA-UAB-UB, Campus de la Universitat Autònoma de Barcelona, Bellaterra, Spain. [2]Animal Breeding and Genetics Programme, Institute for Research and Technology in Food and Agriculture (IRTA), Torre Marimon, Caldes de Montbui, Spain. [3]Embrapa Pecuária Sudeste, Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA), São Carlos, SP 13560-970 Brazil. [4]Departament de Ciència Animal i dels Aliments, Universitat Autònoma de Barcelona, Bellaterra, Spain.

[*]Corresponding author: Marcel Amills. marcel.amills@uab.cat

J. Anim. Sci. Biotechnol. (2020) 11, 10. doi: 10.1186/s40104-019-0412-z.

Abstract

Background

The role of non-coding RNAs in the porcine muscle metabolism is poorly understood, with few studies investigating their expression patterns in response to nutrient supply. Therefore, we aimed to investigate the changes in microRNAs (miRNAs), long intergenic non-coding RNAs (lincRNAs) and mRNAs muscle expression before and after food intake.

Results

We measured the miRNA, lincRNA and mRNA expression levels in the *gluteus medius* muscle of 12 gilts in a fasting condition (AL-T0) and 24 gilts fed *ad libitum* during either 5 h. (AL-T1, N = 12) or 7 h. (AL-T2, N = 12) prior to slaughter. The small RNA fraction was extracted from muscle samples retrieved from the 36 gilts and sequenced, whereas lincRNA and mRNA expression data were already available. In terms of mean and variance, the expression profiles of miRNAs and lincRNAs in the porcine muscle were quite different than those of mRNAs. Food intake induced the differential expression of 149 (AL-T0/AL-T1) and 435 (AL-T0/AL-T2) mRNAs, 6 (AL-T0/AL-T1) and 28 (AL-T0/AL-T2) miRNAs and none lincRNAs, while the number of differentially dispersed genes was much lower. Among the set of differentially expressed miRNAs, we identified ssc-miR-148a-3p, ssc-miR-22-3p and ssc-miR-1, which play key roles in the regulation of glucose and lipid metabolism. Besides, co-expression network analyses revealed several miRNAs that putatively interact with mRNAs playing key metabolic roles and that also showed differential expression before and after feeding. One case example was represented by seven miRNAs (ssc-miR-148a-3p, ssc-miR-151-3p, ssc-miR-30a-3p, ssc-miR-30e-3p, ssc-miR-421-5p, ssc-miR-493-5p and ssc-miR-503) which putatively interact with the *PDK4* mRNA, one of the master regulators of glucose utilization and fatty acid oxidation.

Conclusions

As a whole, our results evidence that microRNAs are likely to play an important role in the porcine skeletal muscle metabolic adaptation to nutrient availability.

Keywords: Co-expression analysis, lincRNAs, microRNAs, Pig, Regulatory impact factor, Skeletal muscle

Background

The majority of nutrigenomic studies in domestic animals have investigated the effects of dietary factors on the mean expression of messenger RNAs (mRNAs) [1], whereas the potential consequences of nutrition on the expression profiles of microRNAs (miRNAs) and long intergenic non-coding RNAs (lincRNAs) have not been explored in depth. Although changes in the expression of porcine genes in response to dietary and genetic factors have been reported in previous studies [2–6], the regulatory co-expression networks underlying such changes have not been fully elucidated yet [3, 7, 8]. Moreover, gene expression variance (GEV), also referred as gene dispersion, has been often overlooked, being considered just as experimental noise without any biological significance [9]. Few methods have been explicitly designed for modeling GEV across samples in RNA-seq experiments [10, 11], despite the fact that changes in gene expression in response to a specific stimulus might have a biologically meaningful individual component that should not be confounded with experimental noise. Indeed, metabolic responses to nutritional factors are often driven by complex signaling pathways and gene-to-gene interactions that are not necessarily identical across the whole cohort of analyzed biological replicates, adding an intrinsic source of variation in gene expression patterns that is often ignored or modeled as a constant variable [11]. A widely accepted estimator of GEV is the biological coefficient of variation (BCV) [12]. In contrast with the canonical coefficient of variation (CV), the BCV effectively integrates both technical and biological variability, thus avoiding the dependence on count size that CV commonly shows.

When the expression patterns of two experimental groups are compared, differences in the magnitudes of average gene expression (differential gene expression) and GEV (differential gene dispersion) can be observed. Differential dispersion might be particularly useful to identify regulatory changes induced by the experimental factor under study. For instance, it is assumed that genes with low GEV are central members of signal transduction pathways while those with high GEV tend to occupy more peripheral positions in gene networks [13]. However, the central or peripheral position of a given gene in a network is not necessarily stable across time and it could also be altered by the experimental factor being analyzed. Differential dispersion could be a useful parameter to detect such source of biological variation as well as to infer its potential consequences.

In a previous study, we investigated how the patterns of mRNA expression change in response to food intake by comparing the muscle transcriptomes of fasting vs fed gilts [5]. Herewith, we wanted to determine how the expression profiles of miRNAs and lincRNAs vary in response to nutrient supply by using mRNA profiles as a reference [5]. This analysis took into consideration both changes in the mean (differential expression) and the variance (differential dispersion) of gene expression. Moreover, we have used a co-expression network approach to elucidate potential regulatory interactions between expressed miRNAs and differentially expressed (DE) mRNA genes as well as to investigate the relationship between gene co-expression modules and meat quality and fatty acid composition traits recorded in the *gluteus medius* skeletal muscle of Duroc pigs.

Materials and methods

Animal material and phenotypic recording

The Duroc pig population used in the current work has been previously described [5]. Thirty-six female Duroc piglets were transported to the IRTA-Pig Experimental Farm at Monells (Girona, Spain) after weaning (age = 3–4 weeks). Gilts were kept in transition devices and fed *ad libitum* with a standard transition diet until they reached approximately 2 months of age (around 18 kg of live weight). Subsequently, all gilts were transferred to fattening pens, where they were housed individually and fed *ad libitum* until reaching approximately 155 d of age. Nutritional details about the feed provided to gilts between 60 and 155 d have been previously reported in [6]. During fattening (60 to 125 d), gilts received feed *ad libitum* with 14.6% crude protein, 4.25% crude fat, 4.8% crude fiber, 4.9% ashes, 0.92% lysine, 0.58% methionine + cysteine and 3190 kcal/kg. During the finishing period (126 to 155 d), gilts were also fed *ad libitum* with a diet containing 14.4% crude protein, 5.53% crude fat, 5.1% crude fiber, 4.9% ashes, 0.86% lysine, 0.53% methionine + cysteine and 3238 kcal/kg. Gilts were slaughtered in the IRTA Experimental Slaughterhouse in Monells (Girona, Spain) in accordance with relevant Spanish welfare regulations. Before slaughter, the 36 gilts were fasted for 12 h. Subsequently, 12 gilts were slaughtered in a fasting condition (AL-T0, N = 12), and the remaining ones were slaughtered 5 h. (AL-T1, N = 12) and 7 h. (AL-T2, N = 12) after receiving food. High concentrations of CO_2 were used to stun the gilts before bleeding. After slaughter, samples of the *gluteus medius* skeletal muscle were taken from the 36 gilts, submerged in RNAlater (Thermo Fisher Scientific, Barcelona, Spain) and stored at − 80 °C. The whole experimental design used in the current work is depicted in Figure 1.

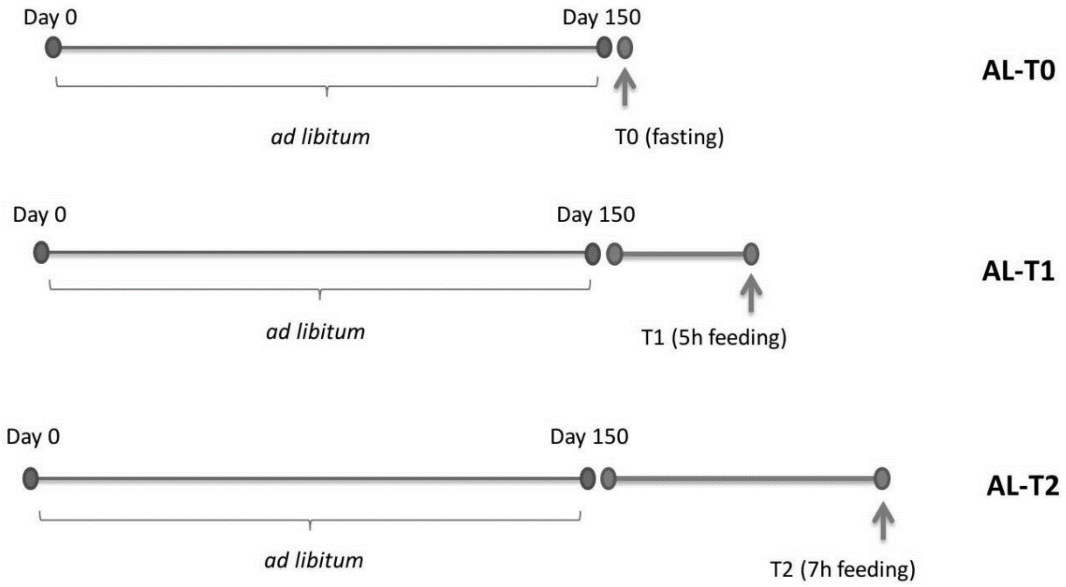


Figure 1: Depiction of the experimental design used in our study. Gilts were fed *ad libitum* (N = 36, N = 12 per group) with a commercial feeding diet during the whole growth period. Prior to slaughter, the 36 gilts were fasted for 12 h. The day of slaughter, 12 gilts (AL-T0) were killed under fasting conditions. The remaining 24 gilts were fed during 5 h. (AL-T1) and 7 h. (AL-T2) and they were subsequently slaughtered.

Phenotypes listed in Additional file 1: Table S1 were recorded in the 36 Duroc gilts. Meat quality traits were measured as described in [14, 15]. Total muscle cholesterol content was determined following Cayuela et al. [16], whereas intramuscular fatty acids content and composition were determined in accordance with previous reports [17].

RNA isolation, library preparation and sequencing of small RNAs

The *gluteus medius* skeletal muscle RNA-seq data set employed in the analysis of lincRNA and mRNA expression comprised a total of 36 individuals (12 AL-T0, 12 AL-T1 and 12 AL-T2 gilts). Details about the RNA extraction and sequencing protocols can be found in [5]. Briefly, *gluteus medius* skeletal muscle samples were pulverized and subsequently homogenized in 1 mL of TRI Reagent (Thermo Fisher Scientific, Barcelona, Spain). The RiboPure kit (Ambion, Austin, TX) was used to isolate the total RNA fraction, and its concentration and purity were determined with a Nanodrop ND-1000 spectrophotometer (Thermo Fisher Scientific, Barcelona, Spain). RNA integrity was assessed with a Bioanalyzer-2100 equipment (Agilent Technologies Inc., Santa Clara, CA) by using the Agilent RNA 6000 Nano Kit (Agilent Technologies, Inc., Santa Clara, CA). Libraries were prepared with the TruSeq SBS Kit v3-HS (Illumina Inc. CA) and paired-end sequenced (2 × 75 bp) in a HiSeq 2000 platform (Illumina Inc., CA) at the Centro Nacional de Análisis Genómico (https://www.cnag.crg.eu).

In the present study, we have generated an additional *gluteus medius* skeletal muscle RNA-seq data set specifically targeting small RNAs and comprising the same 36 individuals cited above. Total RNA was purified as reported above. The percentage of small-RNA over total RNA was determined with the Agilent Small RNA Kit (Agilent Technologies Inc., Santa Clara, CA). All 36 samples met the quality threshold (i.e. 0.2–2 μg total RNA with RIN > 7 and miRNA percentage over total RNA > 0.5%) to be sequenced in Sistemas Genómicos S.L. (https://www.sistemasgenomicos.com). Individual libraries for each sample (N = 36) were prepared with the TruSeq Small RNA Sample Preparation Kit (Illumina Inc., CA) according to the protocols of the manufacturer. Small RNA libraries were then subjected to single-end (1 × 50 bp) sequencing in a HiSeq 2500 platform (Illumina Inc., CA).

Quality assessment, mapping and count estimation

Quality control of paired-end reads was performed with the FASTQC software (Babraham Bioinformatics, http://www.bioinformatics.babraham.ac.uk./projects/fastqc/) and filtered reads were trimmed for any remaining sequencing adapters with the Trimmomatic v.0.22 tool [18], as described in [5, 6]. In the case of single-end sequenced reads derived from small RNA molecules, sequencing adapters were trimmed and filtered with the Cutadapt software [19], and reads outside a window of 15–25 nucleotides were discarded. Paired-end trimmed raw reads from RNA-seq sequences were mapped to the porcine Sscrofa.11.1 reference assembly by using the HISAT2 aligner [20] with default parameters. The StringTie software [21] was subsequently employed to estimate mRNA and lincRNA abundances. Single-end trimmed raw reads derived from small RNAs were also mapped to the Ssscrofa.11.1 assembly with the Bowtie Alignment v.1.2.1.1 software [22], and the following specifications for aligning short miRNA reads were taken into consideration: 1) allowing no mismatches in the alignment, 2) removing reads with more than 20 putative mapping sites and 3) reporting first single best stratum alignment (*bowtie -n 0 -l 25 -m 20 -k 1 --best --strata*). The featureCounts software tool [23] was then used to summarize counts of unambiguously mapped reads from miRNA-seq sequences.

Differential expression and differential dispersion estimates

Raw expression matrices generated on the basis of count estimates obtained with StringTie (mRNAs and lincRNAs) or featureCounts (miRNAs) [21, 23] were normalized with the trimmed mean of M-values normalization method [24]. Sequencing depth and read count per gene were calculated for each sequenced sample (Additional file 15: Figure S1). On the basis of this analysis, the AL-T0 7197 sample was removed from RNA-seq and miRNA-seq count matrices due to the low read coverage observed in the RNA-seq sequencing data set. The presence of influential outliers for each estimate of gene expression was corrected by capping expression values laying outside the boundaries of 1.5 times inter-quartile range per gene and fitting them within the 10th and 90th percentiles. For estimating GEV, the BCV was computed for each detected annotated gene as described in the *edgeR* protocol [25], and further discussed in [12]. The BCV encapsulates all sources of inter-library variation between replicates, including the contribution of library preparation biases [12].

Differentially expressed (DE) and dispersed (DD) genes were determined by comparing the means and variances of gene expression in the two AL-T0/AL-T1 and AL-T0/AL-T2 contrasts. Only mRNAs and miRNAs showing an average expression value above 1 count-per-million (CPM) in at least 50% (N = 12) of the samples (considering all AL-T0, AL-T1 and AL-T2 samples) were retained for further analyses. Because lincRNAs are much less expressed than mRNAs and miRNAs, all lincRNAs (N = 352) annotated in the Sscrofa11.1 reference assembly (v.97) were considered for differential expression and dispersion analyses (a filtering step imposing an expression threshold above 1 CPM would have implied the removal of as much as 80% of annotated lincRNA loci). The *edgeR* [25] and *MDSeq* [10] packages with default parameters were used for performing differential expression and dispersion analyses, respectively. The *edgeR* protocol uses the quantile-adjusted conditional maximum likelihood method for detecting differences in gene expression between two groups. Once negative binomial models are fitted to the input counts and dispersion estimates are obtained, differential expression is determined by using an exact test of significance. Correction for multiple hypothesis testing is implemented by using the Benjamini-Hochberg false discovery rate approach [26]. The *MDSeq* method implements a re-parametrization of the real-valued negative binomial distribution to allow the modelling of gene expression variability [10]. Correction for multiple hypothesis testing across genes is implemented with the Benjamini-Yekutieli procedure [27]. The DE and DD genes obtained with *MDSeq* and *edgeR* were considered to be significant at a fold change > |1.5| and q-value < 0.05.

Gene Ontology and pathway enrichment analysis

The lists of mRNA genes detected as DE in the AL-T0/AL-T1 and AL-T0/AL-T2 contrasts were used as inputs for Gene Ontology (GO) and pathway enrichment analyses. The ClueGO v2.5.0 plug-in application [28] embedded in the Cytoscape 3.5.1 software [29] was used for determining enriched Reactome and KEGG pathways, as well as Biological Process enriched GO terms. A two-sided hypergeometric test of significance was applied for determining enriched terms and multiple testing correction for pathway enrichment analyses was implemented with a false discovery rate approach [26], whereas a Bonferroni-based multiple testing correction was used in the GO enrichment analysis.

Building of co-expression networks

Significant connections between predicted interacting gene pairs were identified with the partial correlation with information theory (PCIT) network inference algorithm [30]. By using first-order partial correlation coefficients estimated for each trio of genes along with an information theory approach, this tool identifies meaningful gene-to-gene putative interactions. The PCIT approach has been widely used to reconstruct co-expression regulatory networks from expression data with good performance [31]. The main aim of this analysis was to determine truly informative correlations between node pairs (genes in our context), once the influence of other nodes in the network has been considered.

Pearson's pairwise correlation coefficients (*r*) were calculated for each expressed miRNA and DE mRNAs in each of the two contrasts (AL-T0/AL-T1 and AL-T0/AL-T2). The *pcit* function from the PCIT R package [30, 32] was then used for detecting meaningful co-expressed gene pairs. To further identify the putative miRNA-to-mRNA interaction pairs with biological interest, a repressor effect of miRNAs on mRNA expression was assumed [33] and, in consequence, only miRNA-to-mRNA co-expressed pairs showing $r < -0.5$ were retained. Furthermore, we only considered miRNA-to-mRNA interactions with perfect 7mer-m8 pairing between the miRNA-seed and the 3′-UTR of the putative mRNA targets, hence removing spurious miRNA-to-mRNA significant correlations with no robust biological meaning. To this end, we downloaded the full set of annotated 3′-UTR sequences in the porcine Sscrofa11.1 assembly available at BioMart Ensembl repositories (http://www.ensembl.org/biomart/martview/). Seed portions (2nd to 8th 5′ nucleotides in the mature miRNA) of the annotated set of porcine miRNAs were reverse-complemented and interrogated along the 3′-UTR sequence regions of mRNA genes by making use of the SeqKit toolkit [34]. Additionally, we selected four highly expressed and DE miRNAs (ssc-miR-148a-3p, ssc-miR-1, ssc-miR-493-5p and ssc-let-7/ssc-miR-98) and used the TargetScan webserver to evaluate the evolutionary conservation of their binding sites in the 3'-UTR of predicted mRNA targets [35]. Only conserved target mRNAs with TargetScan context++ scores above the 75% percentile were considered as confidently cross-validated. The context++ score described by Agarwal et al. [35] incorporates the information of 14 estimated features in order to rank the probability of all the predicted target sites to be biologically functional.

For those mRNAs predicted to interact with miRNAs, we also investigated if they also interact with other mRNA-encoding genes. In order to focus on relevant putative mRNA-to-mRNA gene interactions, we only retained those meaningful mRNA co-expressed pairs showing $|r| > 0.7$, as assessed with the PCIT algorithm. We applied this threshold, which is more stringent than the one used for miRNA-to-mRNA interactions, because correlations between expressed mRNAs tend to be higher than those between mRNAs and miRNAs [36]. Hub genes within selected mRNA-to-mRNA gene interactions (i.e. those mRNAs showing a higher degree of meaningful connectivity according to the PCIT algorithm), were also identified by calculating a hub score per gene (K_i), defined as:

$$K_i = \frac{x_i}{\overline{X}}$$

where x_i is the number of selected significant connections ($|r| > 0.7$) reported by the PCIT algorithm and X is the average connectivity within the mRNA-to-mRNA co-expression network among DE mRNA genes. Gene co-expression networks were visualized with the Cytoscape 3.5.1 software [29].

Besides, for each selected miRNA-to-mRNA predicted interactions, we calculated the regulatory impact factor (RIF) of the corresponding miRNAs [37]. The RIF algorithm aims to identify regulator genes contributing to the observed differential expression in the analyzed contrasts. Its implementation results in two different and inter-connected RIF scores: while RIF1 score represents those transcriptional regulators that are most differentially co-expressed with the most highly abundant and highly DE genes, the RIF2 score highlights those regulators that show the most altered ability to act as predictors of the changes in the expression levels of DE genes [37]. Both RIF values capture different regulatory impact features and hence, they can be considered as two independent measurements of the putative relevance of miRNAs as gene expression regulators. The RIF1 values for each i^{th} regulatory factor were calculated as follows:

$$RIF1_i = \frac{1}{n_{de}} \sum_{j=1}^{j=n_{de}} PIF_j \times DW_{ij}^2$$

where n_{de} is the number of DE genes and phenotype impact factor (PIF) and differential wiring (DW) are denoted by:

$$PIF_j = \frac{1}{2}\left(e1_j^2 - e2_j^2\right)$$

$$DW_{ij} = r1_{ij} - r2_{ij}$$

being $e1_j$ and $e2_j$ the expression of the j^{th} differentially expressed gene in both conditions 1 and 2, respectively, whereas $r1_{ij}$ and $r2_{ij}$ represent the co-expression correlation between the i^{th} regulatory factor (miRNAs in our case) and the j^{th} DE mRNA gene in conditions 1 and 2, respectively.

The RIF2 values for each i^{th} regulatory factors were defined as:

$$RIF2_i = \frac{1}{n_{de}} \sum_{j=1}^{j=n_{de}} \left[\left(e1_j \times r1_{ij}\right)^2 - \left(e2_j \times r2_{ij}\right)^2\right]$$

The positive or negative sign of the RIF1 score is mainly determined by the magnitude of the PIF estimates, and hence is dependent on the directionality of the defined contrast (i.e. the AL-T0/AL-T2 vs. AL-T2/AL-T0 contrasts would generate RIF1 scores with opposite signs). In contrast, the sign of the RIF2 score reflects the altered ability of the regulators to act as predictors of the abundance of DE genes [37].

Association between muscle phenotypes and weighted gene co-expression networks

Significant associations between key co-expressed genes and meat quality and fatty acids composition traits measured in the *gluteus medius* skeletal muscle samples (Additional file 1: Table S1) were determined with the weighted gene correlation network analysis (WGCNA) approach [38]. We used the WGCNA R package [38] for building signed weighted gene co-expression modules based on mRNA and miRNA genes present in the AL-T0/AL-T1 and AL-T0/AL-T2 count matrices and displaying a minimum expression of 1 CPM in at least 50% of samples. Weighted adjacency matrices were built for each expression data set by using a power soft threshold (β) = 16, as recommended by Langfelder and Horvath [38] for estimating signed correlations based on the number of replicates used in our experimental design. The obtained weighted adjacency matrices were subsequently transformed into topological overlapping matrices (TOM) and corresponding dissimilarities were calculated to minimize the effect of noise and spurious co-expression patterns. Hierarchical clustering was then applied to the dissimilarity matrices (1-TOM) and co-expressed genes were merged into modules through dynamic tree branch cutting. Highly inter-connected modules were finally merged by calculating their eigengenes and setting a minimum height cut of 0.25 and a minimum module size of 30 genes for each identified gene co-expression module.

To further elucidate whether the inferred gene co-expression modules were significantly associated with the variation of meat quality and fatty acids composition traits (Additional file 1: Table S1), module eigengenes (MEs) were defined as the first principal component calculated with the principal component analysis (PCA) algorithm. In this way, MEs summarize the co-expression patterns of all genes within each module into a single variable. Measured phenotypes were then correlated with each defined ME. Correlated phenotype-module pairs were considered to be significant when P-value < 0.05. Co-expressed miRNA-only modules were discarded for further analyses. A Student asymptotic P-value approach was finally used for determining the significance of the contribution of each gene within the co-expression modules to the correlation coefficient between MEs and each one of the recorded phenotypes. Relevant genes within significant modules were selected based on the estimates of gene significance (GS, P-value < 0.05) obtained for each phenotype-module significant association.

Additionally, hub genes within each detected gene co-expression module showing significant correlations with phenotypic traits were assessed. WGCNA inferred networks were converted to edge graphs by using the RNAseqDE wrapper R package (https://github.com/jtlovell/RNAseqDE). Subsequently, hub scores for each gene in the selected co-expression modules were calculated by computing the scaled Kleinberg's hub centrality score as described in the igraph tool (https://igraph.org) [39].

Results

Comparing the expression patterns of coding and non-coding RNAs expressed in the porcine skeletal muscle

The RNA-seq data set employed for mRNA and lincRNA quantification encompassed an average of 48.6 million paired-end reads per sample, and approximately 93% of them mapped successfully to the Sscrofa11.1 assembly. Roughly, 76% of unambiguously mapped reads were assigned to annotated features (genes) after quantification. With regard to the miRNA-seq experiment, an average of 8.2 million single-end reads per sample were generated, which were reduced to approximately 6.8 million reads per sample after quality-check and adapter trimming. From these, approximately 77% mapped to the porcine assembly, and an average of 42% single-end mapped reads were successfully assigned to annotated microRNAs in the Sscrofa11.1 assembly. The accuracy of the RNA-seq procedures employed in the current work were previously validated by Cardoso et al. [40], analyzing the differential expression of eight genes based on RNA-seq results and real-time quantitative PCR measurements of gene expression. Such comparison showed a high concordance between the results obtained with these two independent methods [40].

We have characterized and compared the muscle expression profiles of lincRNAs, miRNAs and mRNAs in three groups of pigs (Figure 1): AL-T0 (fasted), AL-T1 (5 h after feeding) and AL-T2 (7 h after feeding). The computed BCVs measuring the range of variability in gene expression across biological replicates within the same group were markedly elevated for lincRNAs, moderate for mRNAs and low for miRNAs, which ultimately showed a very stable and homogeneous expression profile across samples (Figure 2a). Moreover, as expressed by the regularized $\log_2$ (Rlog) transformation of gene counts according to Love et al. [41], the average expression of lincRNAs was much lower than that of mRNAs, while miRNAs occupied an intermediate position between these two extremes (Figure 2b).

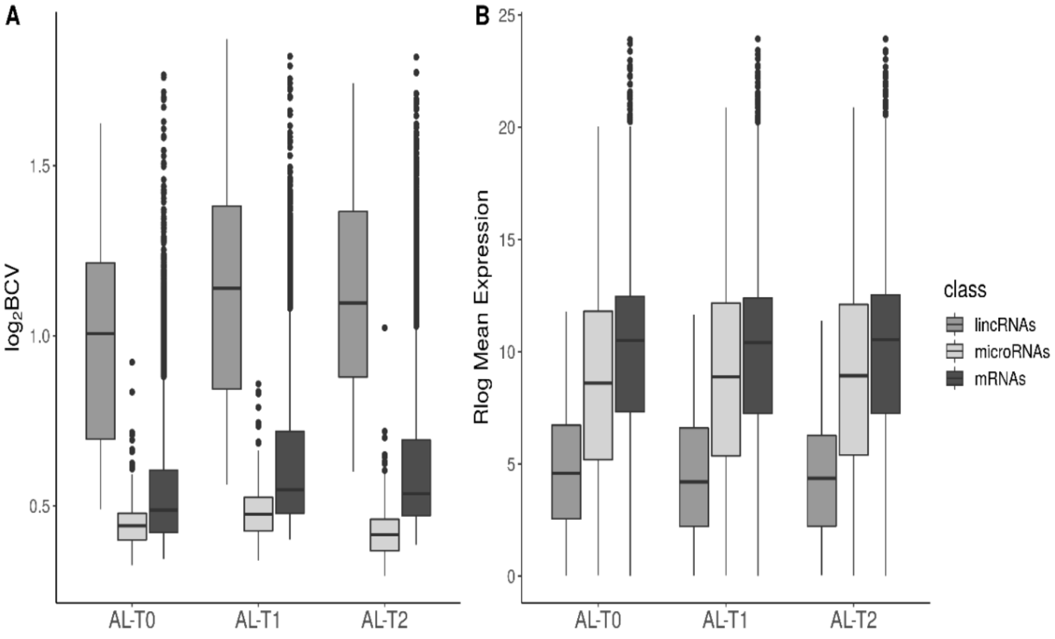


Figure 2: Expression variability and quantification of expression levels of mRNAs, microRNAs and lincRNAs. (**A**) Biological coefficient of variation (BCV) distribution across transcript types within each analyzed group. (**B**) *DESeq2* regularized $\log_2$ mean expression (Rlog) values across transcript types within each analyzed group.

In general, lowly expressed genes displayed higher BCVs than genes with high levels of expression (Figure 3). This pattern was especially relevant for mRNAs (Figure 3a), with an average estimated background BCV of 0.53 (i.e. 53% of mean variability in gene expression across biological replicates expected for mRNA genes), and lincRNAs (mean BCV = 115%, Figure 3c). In strong contrast, miRNAs showed a narrow range of gene expression variability (mean BCV = 37%). Indeed, we did not detect miRNA genes with extremely high BCV values even when we considered miRNAs expressed at marginal levels below 1 CPM

(Figure 3b). With *MDSeq* tool [10], we computed fold changes (FC) for dispersion estimates. For each contrast, $\log_2$FC dispersion values were then plotted against $\log_2$CPM gene expression values (Figure 4). In general, protein-coding genes with medium to low expression levels (Figure 4a) showed higher dispersion FC values than those that were highly expressed. This antagonistic relationship was much less obvious for miRNAs or lincRNAs than for mRNAs (Figure 4b, c).

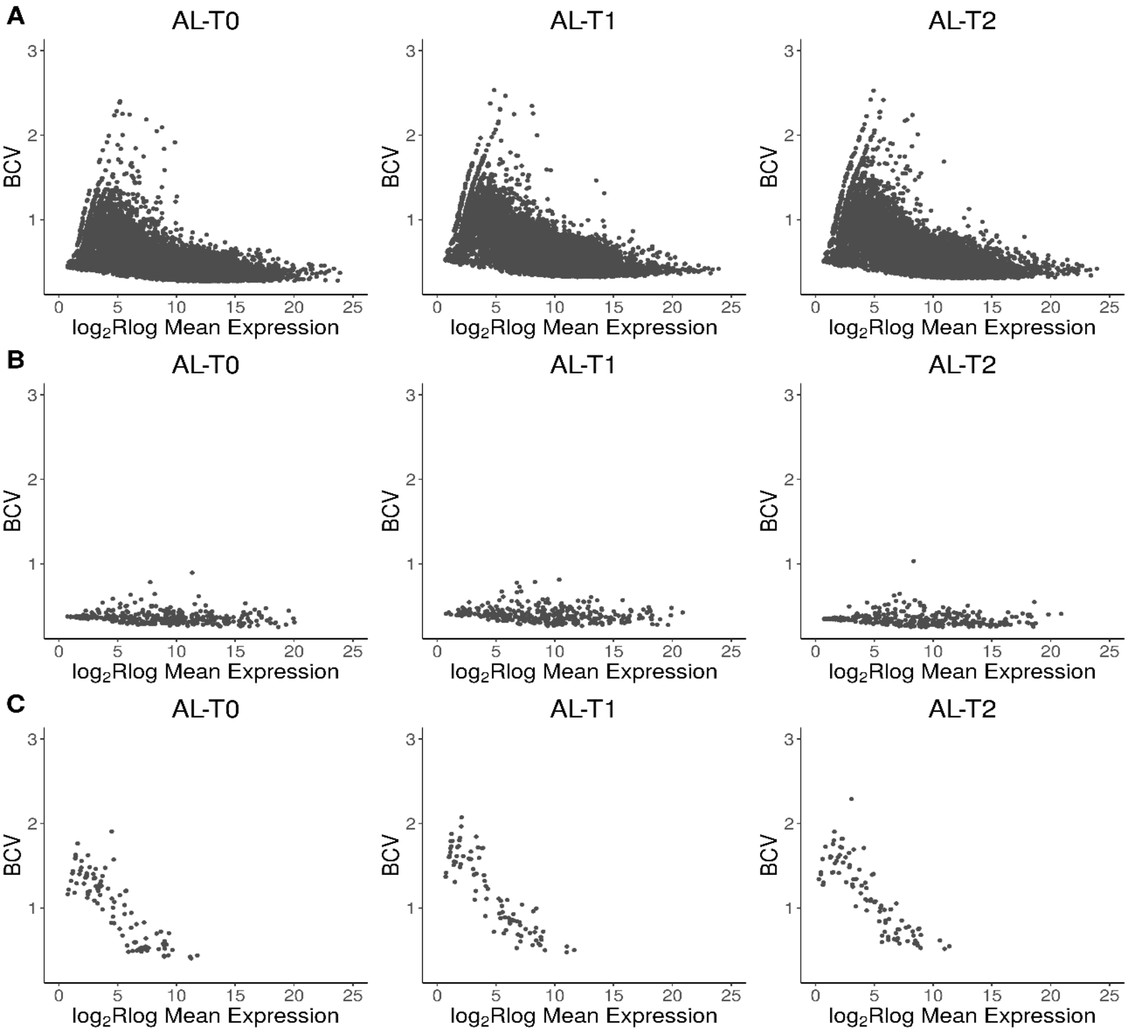


Figure 3: Biological coefficient of variation (BCV) vs. *DESeq2* regularized $\log_2$ mean expression (Rlog) of (**A**) mRNAs, (**B**) microRNAs and (**C**) lincRNAs in each of the analyzed groups (AL-T0, AL-T1 and AL-T2).

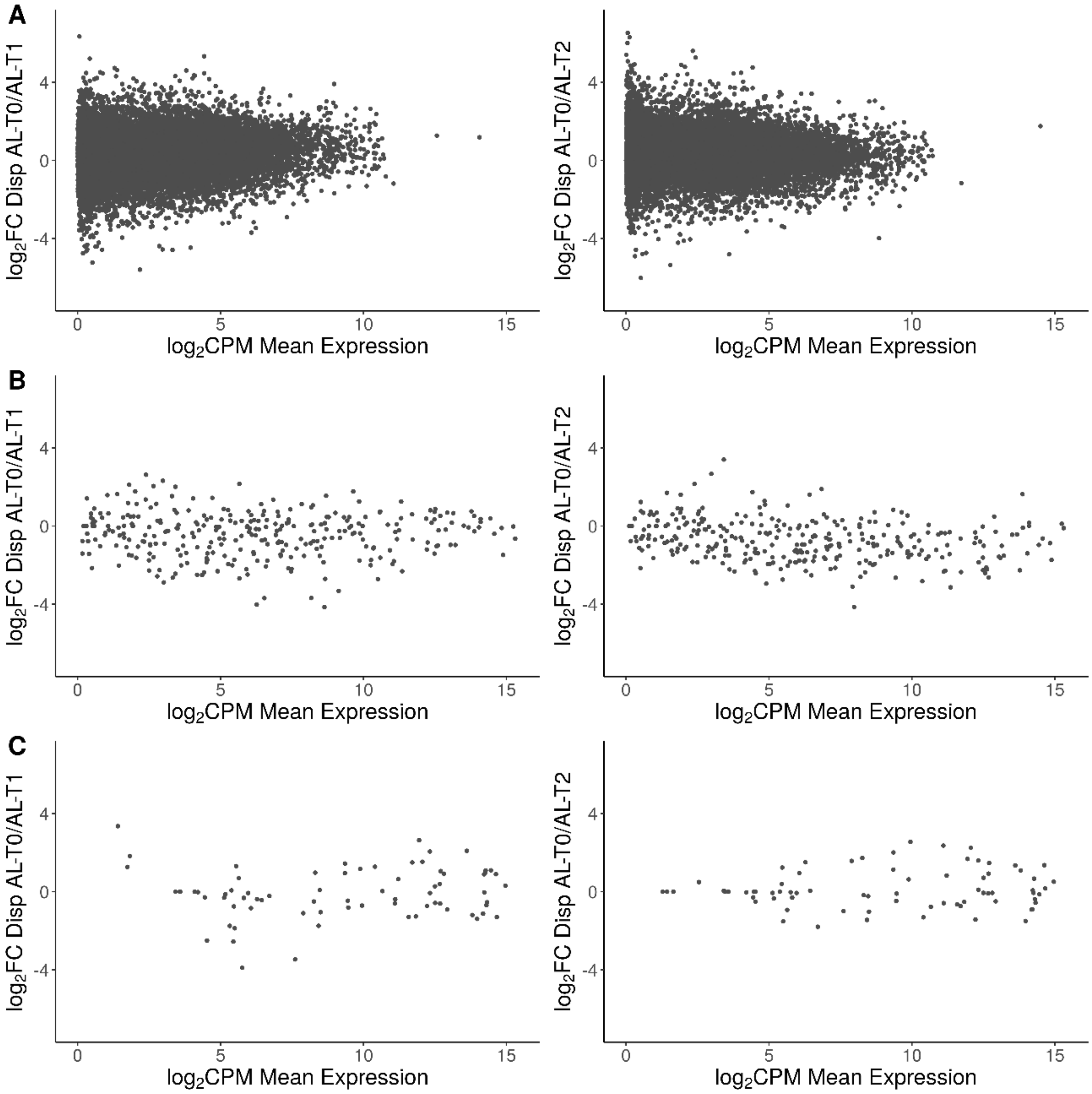


Figure 4: Log$_2$ fold change (FC) of the dispersion values estimated with *MDSeq* tools vs. log$_2$ mean expression (counts-per-million, CPM) of (**A**) mRNAs, (**B**) microRNAs and (**C**) lincRNAs expression patterns in the AL-T0/AL-T1 (left column) and AL-T0/AL-T2 contrasts (right column).

Identification of differentially expressed and dispersed genes

Principal component analysis showed a clear clustering of samples according to their group of origin (AL-T0, AL-T1 and AL-T2) when we considered mRNA expression patterns (Figure 5a), and this was particularly true in the AL-T0/AL-T2 contrast. This outcome agrees well with previously reported results using the same experimental data [5]. In contrast, the clustering of samples based on their miRNA expression patterns was more diffuse (Figure 5b), and in the case of lincRNAs, no evident pattern of clustering was observed (Figure 5c). This lack of sample clustering could be due, at least in part, to the low and very low numbers of annotated pig miRNAs and lincRNAs, respectively. Moreover, the highly variable expression of lincRNAs across samples could also contribute to this lack of clustering. Joint PCA clustering considering all three contrast groups is depicted in Additional file 15: Figure S2.

As previously said, statistical analyses for DE and DD miRNA and mRNA genes were restricted to loci with expression levels above 1 CPM in each contrast and in at least 50% (N = 12) of the samples (each contrast includes 23 samples), whereas all annotated lincRNAs, irrespective of their expression levels, were considered. These filtering criteria reduced approximately by half the number of analyzed mRNAs, i.e. 10,648 (AL-T0/AL-T1) and 10,714 (AL-T0/AL-T2,) expressed mRNAs from a total of 22,342 annotated protein-coding genes were selected for further analyses. Regarding miRNAs, 35% of annotated miRNAs did not reach the expression threshold of 1 CPM (286 expressed miRNAs out of 442 annotated miRNA genes in both AL-T0/AL-T1 and AL-T0/AL-T2).

Differential expression and/or dispersion results generated with *MDSeq* and *edgeR* approaches reflected evident changes in the skeletal muscle transcriptomic profile of pigs after feed intake. These changes were particularly intense in the case of mRNA genes, with 149 and 435 DE mRNAs in AL-T0/AL-T1 and AL-T0/AL-T2, respectively (Additional file 2: Table S2). Moreover, 6 and 28 miRNAs (*q*-value < 0.05; |FC| > 1.5) were classified by *edgeR* as DE in AL-T0/AL-T1 and AL-T0/AL-T2 respectively (Table 1), whereas no lincRNAs showed significant DE in any of the two contrasts. When we considered a less stringent FC threshold for miRNAs and lincRNAs (|FC| > 1.2), we were able to recover 5 additional DE miRNAs in the AL-T0/AL-T2 contrast (Table 1). With regard to differential dispersion, 27 and 30 DD mRNAs were detected with *MDSeq* in the AL-T0/AL-T1 and AL-T0/AL-T2 contrasts, respectively (Additional file 3: Table S3), and several of these mRNAs were also differentially expressed (Additional file 2: Table S2). Few DD miRNAs (i.e. 5 in AL-T0/AL-T1 and 1 in AL-T0/AL-T2) and only two DD lincRNAs (in AL-T0/AL-T1) were detected (Table 2).

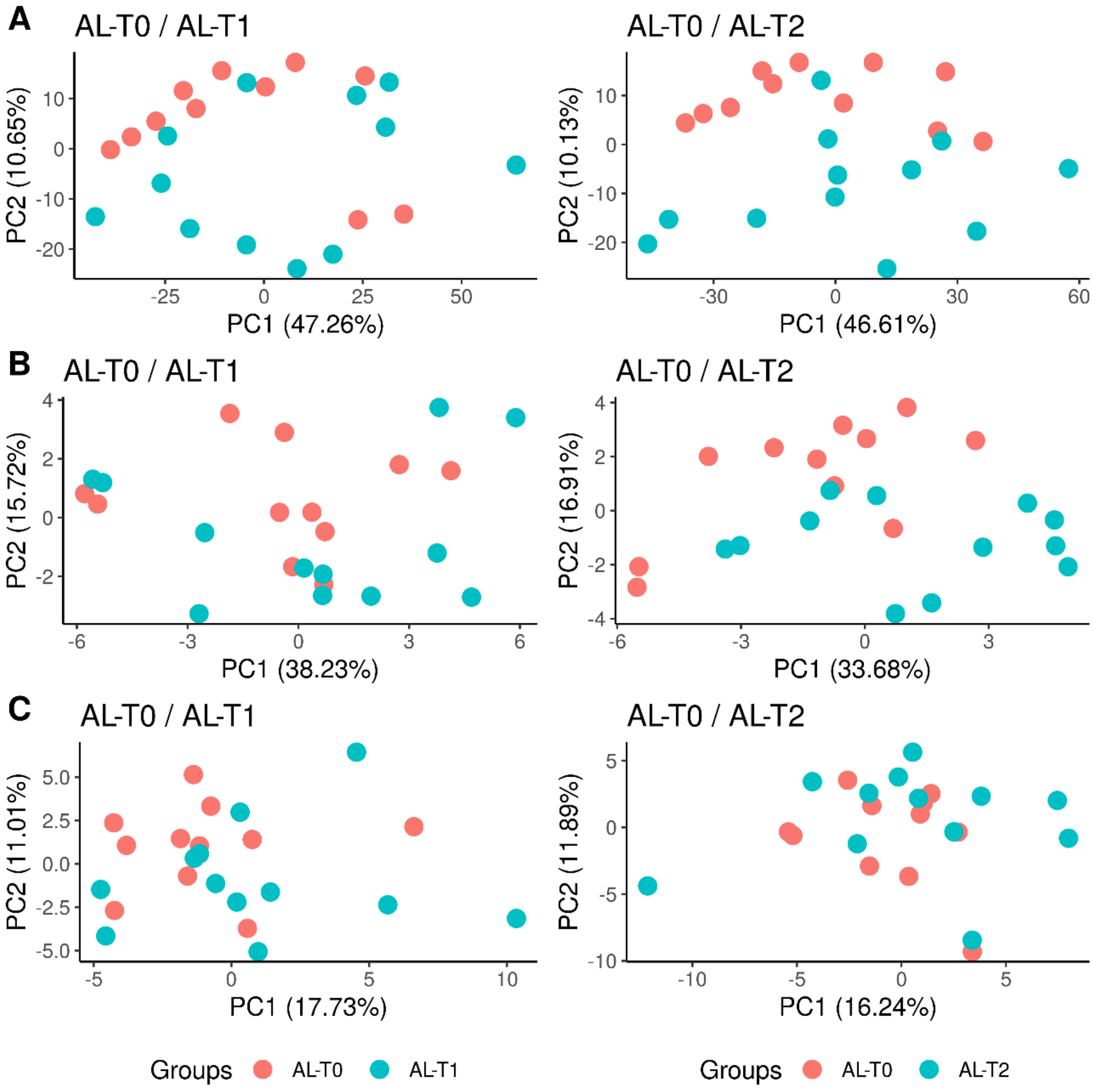


Figure 5: Principal component analysis (PCA) clustering of *gluteus medius* skeletal muscle samples (11 AL-T0, 12 AL-T1 and 12 AL-T2 gilts) according to the expression profiles of (**A**) mRNAs, (**B**) microRNAs and (**C**) lincRNAs.

Table 1: MicroRNAs detected by *edgeR* as differentially expressed when comparing AL-T0 (fasted) gilts with their AL-T1 (5 h after eating) and AL-T2 (7 h after eating) counterparts.

Contrast	miRNA	log$_2$FC^b	*P*-value	*q*-value^c	log$_2$CPM AL-T0^d	log$_2$CPM AL-T1^d	log$_2$CPM AL-T2^d
AL-T0/AL-T1[a]	ssc-miR-7-5p	0.9978	7.56E-05	2.16E-02	6.8416	7.5115	-
	ssc-miR-374a-3p	0.8568	4.73E-04	3.81E-02	6.9201	7.5034	-
	ssc-miR-7	0.9229	5.26E-04	3.81E-02	6.4206	7.0178	-
	ssc-miR-148a-3p	0.8989	5.97E-04	3.81E-02	13.8692	14.5105	-
	ssc-miR-1	0.7686	6.66E-04	3.81E-02	16.8124	17.3183	-
	ssc-miR-32	1.2420	9.92E-04	4.73E-02	2.9824	3.6098	-
AL-T0/AL-T2[a]	ssc-miR-1285	-2.9830	3.47E-09	9.92E-07	7.9799	-	5.5849
	ssc-miR-148a-3p	1.3831	2.39E-06	2.83E-04	13.8692	-	14.9315
	ssc-miR-7-5p	1.1592	2.97E-06	2.83E-04	6.8416	-	7.6948
	ssc-miR-493-5p	0.7464	3.84E-05	2.37E-03	6.4846	-	7.1191
	ssc-miR-7	1.0724	4.14E-05	2.37E-03	6.4206	-	7.1910
	ssc-miR-22-3p	-0.9814	1.01E-04	4.20E-03	12.6857	-	11.7583
	ssc-miR-421-5p	1.2893	1.03E-04	4.20E-03	2.6775	-	3.7359
	ssc-miR-758	-0.7536	1.24E-04	4.43E-03	5.4106	-	4.5480
	ssc-miR-339	-0.876	1.68E-04	5.34E-03	2.8919	-	2.0274
	ssc-let-7f-1	0.7735	2.36E-04	6.43E-03	14.0981	-	147031
	ssc-let-7f-5p	0.7641	2.74E-04	6.43E-03	12.4788	-	13.0761
	ssc-miR-374a-3p	0.9025	2.75E-04	6.43E-03	6.9201	-	7.5867
	ssc-miR-30a-3p	0.66	3.36E-04	6.43E-03	9.8397	-	10.3833
	ssc-miR-151-3p	0.694	3.37E-04	6.43E-03	12.3832	-	12.9732
	ssc-miR-129a-3p	-1.3858	4.09E-04	7.05E-03	4.7123	-	3.0830
	ssc-miR-296-5p	-0.9342	4.79E-04	7.61E-03	5.1094	-	3.9239
	ssc-miR-30e-3p	0.643	7.45E-04	1.12E-02	10.8497	-	11.3840
	ssc-miR-98	0.7127	1.24E-03	1.69E-02	9.6818	-	10.2075
	ssc-let-7a-1	0.466	1.13E-03	2.53E-02	13.7761	-	14.1002
	ssc-let-7a-2	0.459	1.43E-03	2.53E-02	12.4875	-	12.8046
	ssc-miR-503	0.4912	1.12E-03	2.53E-02	7.8380	-	8.1776
	ssc-miR-181c	-0.6665	2.02E-03	2.56E-02	3.0770	-	2.3722
	ssc-miR-32	1.1189	2.11E-03	2.56E-02	2.9824	-	3.5525
	ssc-miR-1	0.6586	2.15E-03	2.56E-02	16.8124	-	17.2479
	ssc-miR-450b-3p	0.9689	2.78E-03	2.95E-02	1.3512	-	2.0993
	ssc-miR-136-5p	0.9319	2.78E-03	2.95E-02	3.4211	-	3.8968
	ssc-miR-7857-3p	-1.1003	3.03E-03	3.09E-02	2.3255	-	1.5283
	ssc-miR-125b	-0.5858	1.88E-03	3.20E-02	13.0432	-	12.3566
	ssc-miR-361-5p	-0.5109	3.02E-03	4.45E-02	7.4597	-	6.8061
	ssc-miR-362	-0.5567	3.45E-03	4.61E-02	6.5327	-	5.8194
	ssc-miR-218b	0.7746	4.75E-03	4.62E-02	48429	-	5.2796
	ssc-miR-532-3p	-0.6865	4.84E-03	4.62E-02	7.4422	-	6.5930
	ssc-miR-365-3p	-0.7367	5.36E-03	4.79E-02	9.9681	-	8.9921

[a]AL-T0: Duroc gilts in a fasting condition (N = 11); AL-T1: Duroc gilts slaughtered after 5 h of food intake (N = 12); AL-T2: Duroc gilts slaughtered after 7 h of food intake (N = 12). [b]Log$_2$FC: estimated log$_2$ fold change mean expression levels. [c]*q*-value: *P*-value corrected for multiple testing with the Benjamini-Hochberg procedure. [d]Log$_2$CPM: estimated log$_2$ counts-per-million (CPM) mean expression levels in AL-T0, AL-T1 and AL-T2 groups.

Table 2: MicroRNAs and lincRNAs detected by *MDSeq* as differentially expressed when comparing AL-T0 (fasted) gilts with their AL-T1 (5 h after eating) and AL-T2 (7 h after eating) counterparts.

AL-T0/AL-T1[a]	**log$_2$FC[b]**	***P*-value**	***q*-value[c]**	**Log$_2$CPM AL-T0[d]**	**Log$_2$CPM AL-T1[d]**	**Log$_2$CPM AL-T2[d]**
miRNA						
ssc-miR-17-5p	-4.0190	3.20E-06	2.81E-03	6.2654	5.7652	-
ssc-miR-186-5p	-4.1486	1.66E-06	2.81E-03	8.6343	8.2410	-
ssc-miR-362	-3.6875	1.64E-05	9.48E-03	6.5327	5.8885	-
ssc-miR-451	-3.6825	2.16E-05	9.48E-03	8.1730	8.1217	-
ssc-miR-29a-3p	-3.3204	1.16E-04	4.07E-02	9.1335	8.7859	-
lincRNA						
ENSSSCG00000032301	3.3076	3.60E-05	1.32E-02	1.4722	5.8072	-
ENSSSCG00000031192	-3.8178	1.59E-04	2.93E-02	5.8864	1.7548	-
AL-T0/AL-T2[a]						
miRNA						
ssc-miR-1285	-4.1428	4.60E-06	8.04E-03	7.9799	-	5.5849

[a]AL-T0: Duroc gilts in a fasting condition (N = 11); AL-T1: Duroc gilts slaughtered after 5 h of food intake (N = 12); AL-T2: Duroc gilts slaughtered after 7 h of food intake (N = 12). [b]Log$_2$FC: estimated log$_2$ fold change mean dispersion levels. [c]*q*-value: *P*-value corrected with the Benjamini-Yekutieli procedure. [d]Log$_2$CPM: estimated log$_2$ counts-per-million (CPM) mean expression levels in AL-T0, AL-T1 and AL-T2 groups.

Functional annotation and pathway enrichment of differentially expressed genes

A total of 26 Reactome and 8 KEGG significantly enriched pathways were detected in the AL-T0/AL-T1 contrast, whereas 16 Reactome and 14 KEGG enriched pathways were identified for the AL-T0/AL-T2 contrast (*q*-value < 0.05). Gene ontology biological process enrichment analyses resulted in 65 and 107 significant GO terms for AL-T0/AL-T1 and AL-T0/AL-T2, respectively. A complete list of enriched pathways and GO terms is shown in

Additional files 4: Table S4 (AL-T0/AL-T1) and 5: Table S5 (AL-T0/AL-T2). Among the most highly enriched pathways, those related with circadian clock regulation appeared in both contrasts, as well as other pathways associated with myogenesis, nuclear receptor transcription or NOTCH1, and interleukin 4 and 13 signaling. Regarding the GO enriched terms, many biological processes triggered by nutrient availability after food intake were activated, such as skeletal muscle differentiation (GO:0035914), carbohydrate biosynthetic process (GO:0016051), regulation of gluconeogenesis (GO:0035947), glycogen biosynthetic process (GO:0005978), gluconeogenesis (GO:0006094), energy reserve metabolic process (GO:0006112), activation of transcription from RNA polymerase II promoter (GO:0006366), response to lipids (GO:0033993), adipose tissue development (GO:006012), regulation of fat cell differentiation (GO:0045598), circadian regulation of gene expression (GO:0032922), cellular response to external stimulus (GO:0071496), response to starvation (GO:0042594) or regulation of energy homeostasis (GO:2000505), to mention a few (Additional files 4 and 5: Table S4 and S5).

Construction of co-expression networks and measurement of regulatory impact factors

We also aimed to determine whether the expression of miRNAs is associated with that of mRNAs in each one of the experimental contrasts. With the PCIT algorithm, we detected 24 (AL-T0/AL-T1) and 55 (AL-T0/AL-T2) miRNAs co-expressed ($r < -0.50$) with sets of differentially expressed putative mRNA targets (Additional file 6: Table S6). For mRNA-to-mRNA connections, only meaningful co-expression relationships with $|r| > 0.7$ were considered (Additional file 7: Table S7). Hub genes showing a high degree of connectivity were prioritized by means of their estimated hub score values (K). A list of selected mRNA genes and their K values is available in Additional file 8: Table S8. Among the genes with the top (5%) hub scores, it is worth mentioning the following ones: (1) AL-T0/AL-T1: Rev-Erb-β (*NR1D2*), BTB domain and CNC homolog 1 (*BACH1*), ETS proto-oncogene 1 (*ETS1*) and the cAMP responsive element binding protein 1 (*CREB1*), and (2) AL-T0/AL-T2: secretory carrier membrane protein 2 (*SCAMP2*), neuraminidase 3 (*NEU3*), pyruvate dehydrogenase kinase 4 (*PDK4*), fatty acid transport protein 4 (*SLC27A4*), thiamine transporter 1 (*SLC19A2*), NAD kinase (*NADK*), BTB domain and CNC homolog 2 (*BACH2*) and ARID domain-containing protein 5B (*ARID5B*). We have also compared the results based on K estimates with the sets of hub genes forming part of the co-expression modules generated with the WGCNA algorithm [38]. By doing so, we found several genes that in both approaches were identified as top central players in the metabolic response to food intake. For instance, *BACH1* and *CREB1* genes were among the top hubs in the Blue co-expression module corresponding to the AL-T0/AL-T1 contrast (Additional file 9: Table S9). With respect to AL-T0/ALT2, *SCAMP2*, *NEU3* and *PDK4* genes within the Green co-expression module were also among the top hub transcripts, whereas *BACH2* and *ARID5B* occupied intermediate positions in the ranking of hub genes (Additional file 9: Table S9).

Additionally, we used the TargetScan algorithm to evaluate the accuracy of the miRNA-to-mRNA interactions predicted with PCIT and 3′-UTR seed matching. Four highly expressed DE miRNA*s (*ssc-miR-148a-3p, ssc-miR-1, ssc-miR-493-5p and ssc-let-7/ssc-miR-98) were selected for this task. From a total of 30 different mRNA genes predicted to be targets of the selected miRNAs (Additional file 6: Table S6), 14 showed conserved and putatively valid interactions (context++ score > 75% percentile) according to predictions made with the TargetScan algorithm (Additional file 10: Table S10).

Particularly interesting was the case of the miRNAs predicted to bind the 3′-UTR sequence of the *PDK4* mRNA (Additional file 11: Table S11), which happened to be the most highly downregulated gene in the AL-T0/AL-T2 contrast (Additional file 2: Table S2). Among the 7 predicted miRNAs with putative 7mer-m8 binding sites in the *PDK4* 3′-UTR, only two sites appeared to be consistently conserved when compared against the corresponding orthologous regions in other phylogenetically related species (Additional file 15: Figure S3, Additional file 10: Table S10). Noteworthy, the two conserved sites are predicted to bind to ssc-miR-148a-3p and ssc-miR-493-5p, which were two of the most highly DE miRNAs in the AL-T0/AL-T2 contrast (Table 1).

Besides, after estimating the RIF score for each co-expressed miRNA, results were ranked according to their regulatory relevance. A complete list of all RIF values for miRNAs is presented in Additional file 12: Table S12. Moreover, a list of the top 5 ranking positive and negative regulatory miRNAs according to their RIF1 and RIF2 scores is presented in Tables 3 and 4, respectively. Interestingly, we observed a high correspondence between miRNAs classified as DE with the *edgeR* tool and miRNAs categorized by the PCIT and RIF algorithms as meaningful regulators (Tables 1, 3 and 4, Additional files 6 and 12: Table S6 and S12). For instance, ssc-miR-32, which was DE in the two considered contrasts, ranked as the second (AL-T0/AL-T1) and third (AL-T0/AL-T2) most relevant miRNA in terms of RIF1 (Table 3**,** Additional file 12: Table S12). The DE miRNAs (AL-T0/AL-T2) ssc-miR-339 and ssc-miR-1 were also detected as relevant in terms of RIF1 score (Table 3). When considering RIF2 and AL-T0/AL-T2, the ssc-miR-1285, ssc-miR-129a-3p, ssc-miR-296-5p, ssc-miR-374a-3p and ssc-miR-7-5p DE miRNAs happened to be among the top predicted regulators (Table 4). In the AL-T0/AL-T2 contrast, several additional DE miRNAs also belonged to the group of the top 10 most relevant regulators according to their RIF scores, e.g. ssc-miR-22-3p for RIF1 and ssc-miR-148a-3p or ssc-miR-493-5p for RIF2 (Additional file 12: Table S12).

Table 3: Top five positive and negative regulatory microRNAs according to their regulatory impact factor 1 (RIF1).

miRNA	RIF1
AL-T0/AL-T1[a]	
ssc-miR-450b-5p	1.7939
ssc-miR-32	1.7041
ssc-miR-136-5p	1.3928
ssc-miR-542-3p	1.2969
ssc-miR-19a	1.2620
ssc-miR-339-3p	-0.9864
ssc-miR-421-5p	-0.9871
ssc-miR-503	-1.1680
ssc-miR-326	-1.2569
ssc-miR-128	-1.2830
AL-T0/AL-T2[a]	
ssc-miR-9858-5p	2.7536
ssc-miR-148b-5p	2.4587
ssc-miR-32	2.3825
ssc-miR-129a-5p	1.9010
ssc-miR-7139-5p	1.3797
ssc-let-7g	-1.0629
ssc-miR-130b-5p	-1.1300
ssc-miR-339	-1.2069
ssc-miR-1	-1.2630
ssc-miR-326	-1.3955

[a] AL-T0: Duroc gilts in a fasting condition (N = 11); AL-T1: Duroc gilts slaughtered after 5 h of food intake (N = 12); AL-T2: Duroc gilts slaughtered after 7 h of food intake (N = 12).

Table 4: Top five positive and negative regulatory microRNAs according to their regulatory impact factor 2 (RIF2).

miRNA	RIF2
AL-T0/AL-T1[a]	
ssc-miR-129a-3p	1.8373
ssc-miR-219a	1.4996
ssc-miR-128	1.4256
ssc-miR-503	1.2053
ssc-miR-450b-3p	1.0913
ssc-miR-455-5p	-0.9408
ssc-miR-296-5p	-1.0613
ssc-miR-143-3p	-1.3923
ssc-miR-542-3p	-1.4893
ssc-miR-450b-5p	-1.5585
AL-T0/AL-T2[a]	
ssc-miR-1285	2.2089
ssc-miR-206	1.7993
ssc-let-7d-5p	1.7537
ssc-miR-129a-3p	1.5109
ssc-miR-129a-5p	1.3630
ssc-miR-296-5p	-1.6368
ssc-miR-374a-3p	-1.6758
ssc-miR-148b-5p	-1.8280
ssc-miR-7-5p	-2.0613
ssc-miR-7139-5p	-2.6767

[a]AL-T0: Duroc gilts in a fasting condition (N = 11); AL-T1: Duroc gilts slaughtered after 5 h of food intake (N = 12); AL-T2: Duroc gilts slaughtered after 7 h of food intake (N = 12).

Relationship between weighted gene co-expression modules and meat quality and muscle fatty acids composition traits

The WGCNA algorithm applied to mRNA and miRNA expression estimates in the AL-T0/AL-T1 and AL-T0/AL-T2 matrices made possible the identification of 5 and 10 gene co-expression modules, respectively (Additional file 15: Figure S4 and S5), excluding miRNA-only co-expression modules. Among these, the identified modules for the AL-T0/AL-T1 contrast were significantly associated with the following meat quality and fatty acids composition phenotypes measured in the *gluteus medius* muscle: meat lightness (L*), intramuscular pH (PHGM), intramuscular fat content (GMIMF), palmitic acid content (C16:0), linoleic acid content (C18:2-ω6), arachidonic acid content (C20:4), omega-6 fatty acids content (ω6), omega-6/omega-3 ratio (ω6/ω3), polyunsaturated fatty acids content (PUFA) and polyunsaturated/saturated fatty acids ratio (PUFA/SFA), as shown in Additional file 13: Table S13. Regarding the AL-T0/AL-T2 contrast, *gluteus medius* phenotypes showing significant associations with co-expression modules were: meat redness (a*), pH measured 45 min post-mortem (PH45GM), linoleic acid content (C18:2-ω6), arachidonic acid content (C20:4), omega-3 (ω3), omega-6/omega-3 ratio (ω6/ω3), unsaturated fatty acids content (UFA) and polyunsaturated/saturated fatty acids ratio (PUFA/SFA) and saturated/unsaturated fatty acids ratio (SFA/UFA) (Additional file 14: Table S14). A detailed list of all analyzed phenotypes is shown in Additional file 1: Table S1. *P*-values measuring the significance of the contribution of each gene within co-expression modules to significantly correlated phenotypic traits can be found in Additional files 13: Table S13 (AL-T0/AL-T1) and 14: Table S14 (AL-T0/AL-T2).

Discussion

Coding and non-coding RNAs show highly divergent patterns of expression in the porcine muscle

By comparing mRNAs, miRNAs and lincRNAs expression patterns, we have observed that the expression of mRNAs in the porcine skeletal muscle is, on average, substantially higher than that of miRNAs and lincRNAs (Figure 2). This finding was expected because previous studies in humans have reflected the same trend for lincRNAs [42, 43] and miRNAs [44]. On the other hand, we have also observed an inverse relationship between the expression means of mRNA and lincRNA genes and the magnitude of BCVs (Figure 3a, c), whereas such trend was not obvious for miRNAs (Figure 3b).

With regard to differential dispersion, the number of DD mRNA and miRNA genes was much lower than that of DE mRNA and miRNA genes, indicating that nutrient supply has a stronger impact on the mean expression of genes rather than on their BCV. Of course, these two parameters are closely related, so decreases in the mean expression of genes are usually accompanied by increases in the variance of expression (and vice versa), being such trend particularly true for mRNAs and lincRNAs. In contrast, miRNAs showed a very resilient and stable pattern of expression across replicates (Figs. 3b and 4b).

While nutrient supply induced substantial changes in the expression of mRNAs (Additional file 2: Table S2), the absolute number of DE miRNAs was much lower (Table 1), whereas no DE lincRNAs were detected. This result is probably not due to a limited accuracy of RNA-seq in detecting differential gene expression, because previous experiments [40] showed a high consistency between differential gene expression results obtained with RNA-seq and real time quantitative PCR data in the same experimental system. However, it should be taken into account that the absolute numbers of annotated porcine miRNAs and lincRNAs are much smaller than those of mRNAs. Indeed, when the number of DE genes is expressed as a proportion (i.e. number of DE genes/number of total analyzed expressed genes), the total amount of DE mRNAs happened to be 1.39% (AL-T0/AL-T1) and 4.06% (AL-T0/AL-T2). In the case of miRNAs, such proportions were 2.09% (AL-T0/AL-T1) and 9.79% (AL-T0/AL-T2). Moreover, the average |FC| of DE mRNAs was 2.12-fold and 2.02-fold in AL-T0/AL-T1 and AL-T0/AL-T2 respectively, while for miRNAs, changes of 1.9-fold (AL-T0/AL-T1) and 1.85-fold (AL-T0/AL-T2) were detected. In the light of these results, it should be concluded that both mRNAs and miRNAs show consistent patterns of differential expression in response to food intake, while no conclusive evidence has been obtained for lincRNAs. This latter observation could be due to the poor annotation of lincRNAs as well as to their low expression levels and elevated within group expression variability (Figs. 2 and 3c), which ultimately would make the differential expression analysis much less powerful to detect significant differences.

Nevertheless, the high variance in the expression of lincRNAs contrasted strongly with the stable patterns of expression across contrasts displayed by miRNAs (Figs. 2 and 3b, c). This high stability might be due to the fact that the expression and silencing activity of miRNAs are decoupled to some extent [36]. There are several factors that explain such circumstance. For instance, miRNAs can be sequestered by pseudogene, mRNA, lincRNA or circular RNA transcripts with repeated miRNA antisense sequences (the so-called miRNA sponges), thus limiting their availability to regulate the expression of target RNAs [45,46,47]. Moreover, compelling evidence has been accumulating during past years highlighting the exceptional stability of certain miRNAs, which show half-lives of days [48, 49]. This long half-life might be explained by the protective effect of the Argonaute protein in isolating naked single-stranded small miRNA molecules from exonucleases within the cell environment [50]. Besides, miRNAs might localize to cell compartments other than the cytosol, where they exert functions unrelated with the modulation of mRNA levels [51]. Last but not least, the expression levels of miRNAs do not necessarily correlate with their functional availability as a part of the RNA-induced silencing complex [36].

Differentially expressed and dispersed miRNAs are related with the regulation of key metabolic processes in the skeletal muscle

As shown in Tables 1 and 2, several miRNAs were detected as either being DE and/or DD in the AL-T0/AL-T1 and AL-T0/AL-T2 contrasts. Among the DE miRNAs, we found that ssc-miR-1 and ssc-miR-148a were two of the most expressed and DE miRNAs in AL-T0/AL-T1 and AL-T0/AL-T2 contrasts (Table 1), whereas ssc-miR-7-5p was the most highly differentially upregulated miRNA in AL-T1 gilts. Both miR-7 and miR-1 regulate the mTOR-related cell response to nutrient availability. For instance, miR-1 was found to be directly upregulated by the myogenic differentiation 1 (*MYOD1*) gene [52], which is a transcription factor essential for skeletal muscle development and myocyte fusion [53] and also functions as a circadian modulator in the peripheral muscle clock [54]. Noteworthy, *MYOD1* was also significantly upregulated in the AL-T0/AL-T2 contrast (Additional file 2: Table S2), a finding that agrees well with the observed upregulation of ssc-miR-1 (Table 1). Additionally, miR-7 has been also associated with the Akt-mTOR and PI3K/Akt signaling by targeting the insulin receptor substrate 2 (*IRS2*) and the phosphoinositide 3-kinase catalytic subunit δ (*PIK3CD*) [55, 56], two genes that are integrated in the coordinated signaling cascade in response to nutrient supply to promote skeletal muscle growth and differentiation.

Regarding the miR-148 family, it has been reported that these miRNAs play a key role in cholesterol metabolism [57,58,59] and insulin homeostasis [60]. In a fasting/feeding study resembling ours, Goedeke et al. [59] reported that miR-148a binds the 3′-UTR of the low density lipoprotein receptor (*LDLR*) mRNA leading to the accumulation of low-density lipoprotein (LDL) cholesterol in blood plasma. Similar results were reported by Rotllan et al. [61]. Furthermore, Goedeke et al. [59] suggested that the sterol regulatory element-binding transcription factor 1 (*SREBF1*) may activate the expression of miR-148a by targeting conserved E-box motifs in the miRNA promoter. In the same study, the role of the ATP-binding cassette 1 (*ABCA1*) gene in the regulation of high-density lipoprotein (HDL) cholesterol levels was explored, and a binding site for miR-148a in the 3′-UTR of *ABCA1* transcripts was predicted, thus providing a functional explanation for the inhibitory effect of miR-148a on plasma HDL cholesterol levels [59]. Other studies have also linked miRNAs belonging to the miR-148 family with angiogenesis and glucose metabolism through insulin like growth factor 1 receptor (*IGF1R*) target inhibition [62].

With respect to other relevant DE miRNAs detected in our study, the miR-30 family and miR-503 have been described to be involved in skeletal muscle differentiation and fiber-type composition [63, 64]. Moreover, they also regulate adipogenesis [65], a role that has also been reported for miR-148a [66] and miR-22 [67]. Furthermore, the observed downregulation of miR-22 after food ingestion (Table 1) could be the consequence of the active influx of glucose within muscle cells after nutrient supply. Indeed, the glucose transporter 1 (*GLUT1*) mRNA is targeted by miR-22 [68]. A similar reasoning could be extended to miR-17-5p, which binds to the glucose transporter 4 (*GLUT4*) mRNA [69] and that was DD but not DE after feed intake (Table 2).

Relevant miRNA-to-mRNA regulatory interactions in response to nutrient supply

Co-expression network analyses highlighted that the majority of DE miRNAs were also potentially meaningful regulatory factors (Tables 1, 3 and 4, Additional file 12: Table S12). Other miRNAs also emerged as potential regulators (Tables 3 and 4, Additional file 12: Table S12) despite not being detected as significantly DE, a finding that would be in agreement with the very stable and low expression levels detected for most miRNAs (Figs. 2 and 3b). These results evidence the interest of reconstructing regulatory networks in order to gain new biological insights that canonical differential expression analysis cannot yield [70]. Several critical downregulated transcription factors in AL-T1 animals were identified as potential co-expressed targets of ssc-miR-1 and ssc-miR-148a-3p DE miRNAs (Additional files 2 and 6: Table S2 and Table S6), e.g. the myogenic factor 6 (*MYF6*), FOS-related antigen 2 (*FOSL2*) and arrestin domain-containing protein 3 (*ARRDC3*) for ssc-miR-1, and thioredoxin interacting protein (*TXNIP*) and fasting-induced gene protein (*DEPP1*) for ssc-miR-148a.

The *MYF6* gene has been previously associated with the regulation of myogenesis and skeletal muscle cell differentiation [8, 71]. A proliferation modulating function has also been described for *TXNIP* [72] as well as for *FOSL2* [73], which is also involved in leptin expression regulation [74], whereas *DEPP1* downregulation has been associated with autophagy inhibition [75]. Moreover, ssc-miR-32 and ssc-miR-7-5p, two miRNAs that were differentially upregulated in AL-T1 gilts (Table 1), were predicted to target several relevant genes (Additional file 6: Table S6) such as the activating transcription factor 3 (*ATF3*), a key regulator of glucose and energy metabolism [76, 77] which was significantly downregulated in both AL-T0/AL-T1 and AL-T0/AL-T2 contrasts (Additional file 2: Table S2). Other relevant additional transcripts that formed part of the miRNA-to-mRNA interconnected networks were, to mention a few, the Kruppel-like factor 15 (*KLF15*), early growth factor 1 (*EGR1*) and ARID domain-containing protein 5B (*ARID5B*), all of which play key roles in muscle lipid metabolism [8, 78, 79], or myogenin (*MYOG*), a gene that is crucial for muscle development and differentiation [80].

With regard to AL-T2 gilts, it is worth mentioning the *PDK4* gene, which happened to be the most extremely downregulated mRNA transcript (Additional file 2: Table S2) and was also detected as DD in the AL-T0/AL-T2 contrast (Additional file 3: Table S3). After reconstructing meaningful miRNA-to-mRNA interactions, seven miRNAs (ssc-miR-148a-3p, ssc-miR-151-3p, ssc-miR-30a-3p, ssc-miR-30e-3p, ssc-miR-421-5p, ssc-miR-493-5p and ssc-miR-503) were predicted to have putative binding sites in the *PDK4* 3′-UTR (Additional files 6 and 11: Table S6 and Table S11). Noteworthy, all these miRNAs were significantly upregulated in the skeletal muscle of AL-T2 gilts (Table 1), with the only exception of ssc-miR-503, (Table 1). Our findings agree well with a cooperative and synergistic interaction between the aforementioned miRNAs and the *PDK4* mRNA, that would result in its strong downregulation observed in AL-T2 pigs (Additional file 2: Table S2). Interestingly, among the set of miRNAs significantly co-expressed with *PDK4* mRNAs, and also predicted to interact with its 3′-UTR, ssc-miR-148a-3p and ssc-miR-493-5p were two of the most significantly upregulated miRNAs in AL-T2 gilts (Table 1). Moreover, the TargetScan analysis [35] showed that both miRNAs have evolutionarily conserved binding sites in the 3′-UTR of the *PDK4* gene (Additional file 15: Figure S3, Additional file 10: Table S10). We may hypothesize that ssc-miR-148a-3p and ssc-miR-493-5p play a key role in the downregulation of the *PDK4* mRNA after food intake, but such hypothesis still needs experimental verification.

Co-expression network analysis also indicated that the *PDK4* gene might interact with a broad array of mRNA transcripts (Figure 6). Among these, several have been already mentioned (*MYF6*, *FOSL2*, *KLF15*, *ARID5B*, *DEPP1*, *MYOG* or *TXNIP*) while others have not, e.g. aryl hydrocarbon receptor nuclear translocator like (*ARNTL*), forkhead box O1 (*FOXO1*), neuronal PAS domain protein 2 (*NPAS2*), BTB domain and CNC homolog 2 (*BACH2*) or the period circadian regulator 2 (*PER2*). The *PDK4* gene is one of the master regulators of glucose and lipid metabolism [81]. Moreover, the PDK4 protein is located in the matrix of the mitochondria and inhibits the pyruvate dehydrogenase complex, which catalyzes the conversion of pyruvate to acetyl-CoA, and hence it is responsible of the decrease in glucose utilization and the upregulation of fatty acid oxidation in energy-deprived cells under fasting conditions [82, 83].

The observed coordinated downregulation of both *PDK4* and *FOXO1* mRNAs in the AL-T0/AL-T2 contrast (Additional file 2: Table S2) is consistent with the active energy production and fatty acid synthesis of muscle cells in response to nutrient supply, as already reported by Cardoso et al. [5]. In fact, the activation of *FOXO1* is known to enhance *PDK4* transcription by binding to its promoter region [84, 85]. Besides, the *BACH2* transcription factor was also predicted to be regulated by ssc-miR-148a-3p (Additional files 6 and 10: Table S6 and Table S10) as well as to interact with both *FOXO1* and *PDK4* mRNAs **(**Figure 6). These findings agree well with the previously described role of *BACH2* as a transcriptional activator of *FOXO1* by binding to its promoter region [86,87,88]. The presence of genes involved in the maintenance of circadian rhythms (*NPAS2*, *ARNTL* and *PER2*) was also relevant, as the expression of the *PDK4* mRNA is subjected to circadian fluctuations in response to light shifting and insulin and fatty acids availability [89,90,91]. Noteworthy, the potential implications of nutrition in the regulation of the porcine peripheral clocks was already discussed in two previous studies using the very same animal material and experimental design reported herewith [5, 40], a result that would be in agreement with the reconstructed *PDK4* miRNA-to-mRNA interaction network reported in this study.

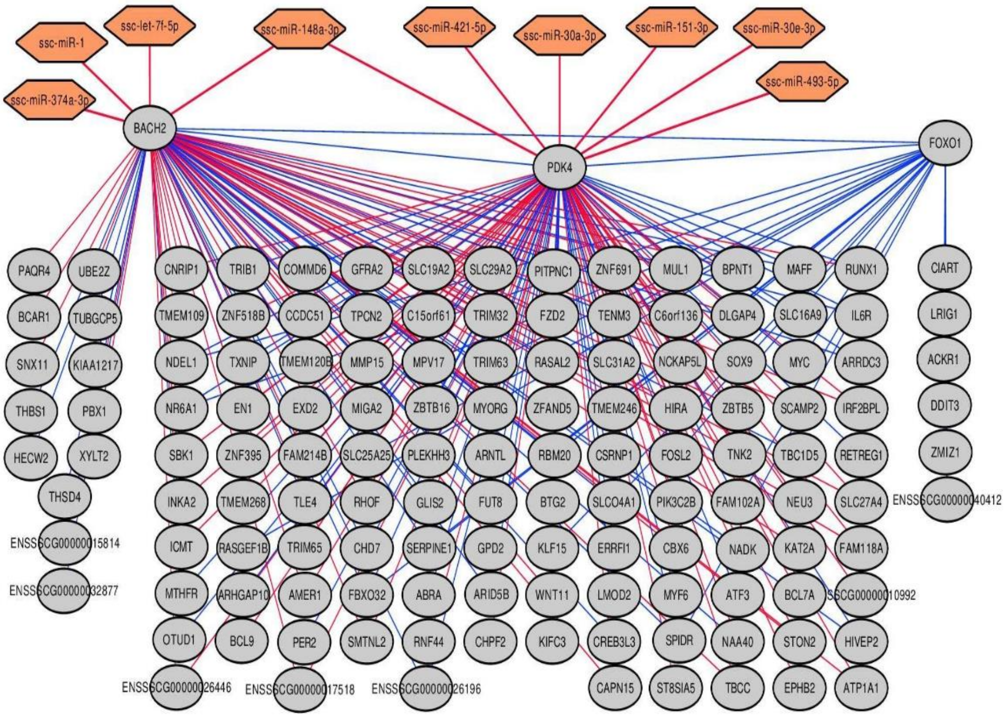


Figure 6: Selected miRNA-to-mRNA and mRNA-to-mRNA co-expression network according to the PCIT algorithm in the AL-T0/AL-T2 contrast. Differentially expressed miRNAs and mRNAs were considered. Only significant correlations below − 0.5 for miRNA-to-mRNA and above |0.7| for mRNA-to-mRNA interactions where selected. Red and blue edges indicate negative and positive correlations in the co-expression network, respectively.

mRNA-to-mRNA hub genes reveal glucose and lipid metabolism changes induced by food intake

Hub scoring of meaningful mRNA genes from selected co-expression interaction networks also allowed the identification of several relevant transcripts involved in organizing the cell response to nutrient availability (Additional file 8: Table S8), and several of these were also detected as hub genes in WGCNA analyses (Additional file 9: Table S9). With respect to AL-T0/AL-T1, the *NR1D2* gene was the most prominent hub gene among all other transcripts, despite the fact that it was not detected as DE. This transcription factor and its paralog Rev-Erbα (*NR1D1*) contribute to establish links between circadian rhythms and cell metabolism regulation [92]. Remarkably, other relevant top hub genes were not DE, e.g. the *BACH1* transcription factor, whose inhibition has been associated with an increased protection against oxidative stress [93], *ETS1*, which mediates *FOXO1* acetylation and regulates gluconeogenesis in fasting-feeding cycles [94] or *CREB1*, an important cofactor for the peroxisome proliferator-activated receptor γ coactivator 1-α (*PPARGC1A*), a gene that plays a key role in insulin-mediated glucose uptake [95].

Regarding hub genes detected in the AL-T0/AL-T2 contrast (Additional file 8: Table S8), *SCAMP2* has been related to glucose transporters trafficking during insulin stimulation [96], whereas *NEU3*, which was also highly upregulated in fed gilts (Additional file 2: Table S2), stimulates insulin sensitivity and glucose tolerance [97]. Other relevant examples are: *SLC27A4*, responsible for long chain fatty acids metabolism and trafficking [98], *SLC19A2*, also highly downregulated in fed gilts (Additional file 2: Table S2) and reported as being negatively regulated by glucose uptake [99], and NADK, a protein that phosphorylates NAD$^+$ to generate NADP$^+$, a metabolite tightly linked with the regulation of circadian rhythms [100].

These findings agree well with data previously reported by Cardoso et al. [5], as well as with enrichment analyses described in this study (Additional files 4 and 5: Table S4 and Table S5), where many DE genes associated with diverse glucose and lipid metabolism pathways and GO terms were highlighted. Other biological processes like muscle proliferation associated to nutrient availability and circadian regulation provided compelling evidence about the complex machinery triggered in the skeletal muscle to respond to nutrient supply after food ingestion.

Weighted co-expression analyses revealed hub genes related with lipids metabolism regulation

Among the gene co-expression modules detected with the WGCNA approach [38], the so-called Red and Purple clusters (Additional file 14: Table S14), corresponding to the AL-T0/AL-T2 contrast, contained several relevant lipid metabolism-related genes such as the fatty acid binding protein 4 (*FABP4*), carbohydrate-responsive element-binding protein (*MLXIPL*), fatty acid synthase (*FASN*), thyroid hormone responsive protein (*THRSP*), stearoyl-CoA desaturase (*SCD*), acetyl-CoA carboxylase α (*ACACA*) or the secreted frizzled-related proteins 1 and 5 (*SFRP1* and *SFRP5*), as well as other loci such as the cholinergic receptor nicotinic δ subunit (*CHRND*). From these, the *MLXIPL*, *FASN*, *SCD*, *SFRP1*, *SFRP5* and *THRSP* genes were also significantly upregulated in AL-T2 gilts after feeding (Additional file 2: Table S2).

Interestingly, the active/non-active conformation of the muscle acetylcholine receptor function regulating motor nerve-muscle communication and muscle contraction is tightly associated with the concentration of certain surrounding fatty acid components, contributing to stabilize or destabilize its functionality [101], a phenomenon that could explain the observed association between its δ subunit (*CHRND*) and the content of ω-3 fatty acids and ω6/ω3 content ratio in the *gluteus medius*, as shown in Additional file 14: Table S14.

Other genes that are key regulators of lipid metabolism such as *SCD*, *ACACA*, *FABP4*, *SFRP1*, *THRSP* or the hub genes *SFRP5* and *FASN* (Additional file 9: Table S9), also clustered in a tight co-expression module and they were significantly associated with linoleic and arachidonic fatty acids content in the *gluteus medius* muscle (Additional file 14: Table S14). The SFRP5 protein has been thoroughly studied as a central regulator of lipid accumulation and adipocytes differentiation, which are a result of an increased mitochondrial respiration promoted by SFRP5 blocking of Wnt signaling, hence repressing Wnt-induced oxidative metabolism [102]. The other identified SFRP element (*SFRP1*) has also been reported to be located in a genomic region overlapping a QTL for meat marbling [103, 104]. Moreover, the *THRSP*, *MLXIPL* and *FASN* upregulation detected in our analyses (Additional file 2: Table S2), as well as their contribution to intramuscular lipid content (Additional files 9 and 14: Table S9 and Table S14) could be a reflection of the intramuscular adipocyte proliferation triggered by the nutrient supply provided to AL-T2 fed gilts [105]. Indeed, the *MLXIPL* is a key carbohydrate-signaling transcription factor whose activity is enhanced by glucose metabolites, thus binding to carbohydrate response elements (ChoREs) present in the promoters of several key lipid genes such as *FASN* [106].

Conclusions

In conclusion, we have demonstrated that the profiles of expression of lincRNAs and miRNAs in the *gluteus medius* muscle of pigs are very different than those observed for mRNAs. For instance, the mean and the variance of gene expression are closely interdependent parameters in the case of mRNAs, while miRNAs do not show such trend. We have also demonstrated that feeding induces changes mainly in the mean expression of genes rather than on their expression variance, a parameter which remains relatively unaffected by nutrient supply. Finally, co-expression network analyses predict that miRNAs and hub mRNA genes may play an essential role in the regulation of mRNAs showing differential expression upon feeding. Such regulatory interactions predicted with *in silico* tools should be validated experimentally in order to verify their occurrence as well as to infer their biological significance in the context of porcine muscle metabolism and nutrition.

Supplementary Information

Additional file 1: Table S1. Meat quality and *gluteus medius* fatty acids composition traits recorded in AL-T0, AL-T1 and AL-T2 Duroc gilts.

Additional file 2: Table S2. Differentially expressed mRNAs in the ALT0/AL-T1 and AL-T0/AL-T2 contrasts.

Additional file 3: Table S3. Differentially dispersed mRNAs in the ALT0/AL-T1 and AL-T0/AL-T2 contrasts.

Additional file 4: Table S4. Pathway Enrichment and GO Enrichment analyses in the AL-T0/AL-T1 contrast.

Additional file 5: Table S5. Pathway Enrichment and GO Enrichment analyses in the AL-T0/AL-T2 contrast.

Additional file 6: Table S6. miRNA-to-mRNA significant interactions detected with the PCIT algorithm in the AL-T0/AL-T1 and AL-T0/AL-T2 contrasts.

Additional file 7: Table S7. mRNA-to-mRNA significant interactions detected with the PCIT algorithm in the AL-T0/AL-T1 and AL-T0/AL-T2 contrasts.

Additional file 8: Table S8. Estimated hub scores (K) values for genes differentially expressed in the AL-T0/AL-T1 and AL-T0/AL-T2 contrasts.

Additional file 9: Table S9. Scaled Kleinberg's hub scores per gene for each WGCNA module significantly correlated with phenotypic traits in the AL-T0/AL-T1 and AL-T0/AL-T2 contrasts.

Additional file 10: Table S10. Identification of conserved mRNA targets for selected highly expressed and differentially expressed miRNAs based on the TargetScan Context++ score.

Additional file 11: Table S11. miRNAs predicted to bind the 3′-UTR of the porcine *PDK4* gene.

Additional file 12: Table S12. Regulatory impact factor scores (RIF1 and RIF2) for miRNAs classified by the PCIT algorithm as meaningful regulators in the AL-T0/AL-T1 and AL-T0/AL-T2 contrasts.

Additional file 13: Table S13. Gene co-expression modules significantly associated with meat quality and fatty acids composition traits according to the WGCNA algorithm (AL-T0/AL-T1 contrast).

Additional file 14: Table S14. Gene co-expression modules significantly associated with meat quality and fatty acids composition traits according to the WGCNA algorithm (AL-T0/AL-T2 contrast).

Additional file 15: Figure S1. Sequencing depth obtained for samples analyzed in each one of the two contrasts (AL-T0/AL-T1 and AL-T0/AL-T2). **Figure S2.** Joint principal component analysis (PCA) clustering of *gluteus medius* skeletal muscle samples (11 AL-T0, 12 AL-T1 and 12 AL-T2 samples) according to the expression profiles of (**A**) mRNAs, (**B**) microRNAs and (**C**) lincRNAs. **Figure S3.** Phylogenetically conserved 7mer-8 m predicted binding sites in the 3′- UTR of the pig *PDK4* gene for (**A**) ssc-miR-148a-3p and (**B**) ssc-miR-493-5p porcine miRNAs. The TargetScan software was used for generating conservation graphs across the investigated mammalian species. Red nucleotides show complementary matching base-pairs between the seed of the mature miRNA and the 3'-UTR of the pig *PDK4* gene. **Figure S4.** Gene co-expression module association with meat quality and fatty acids composition traits in the AL-T0/ AL-T1 contrast as determined with the WGCNA tool.

Figure S5. Gene co-expression module association with meat quality and fatty acids composition traits in the AL-T0/AL-T2 contrast as determined with the WGCNA tool.

Abbreviations

BCV: Biological coefficient of variation; CPM: Counts-per-million; CV: Coefficient of variation; DD: Differentially dispersed; DE: Differentially expressed; FC: Fold change; GEV: Gene expression variability; GO: Gene ontology; GS: Gene significance; K: Hub score; lincRNA: Long intergenic noncoding RNA; MEs: Module eigengenes; miRNA: MicroRNA; mRNA: Messenger RNA; PCA: Principal component analysis; PCIT: Partial correlation with information theory; PIF: Phenotype impact factor; r: Pearson's correlation coefficient; RIF: Regulatory impact factor; Rlog: Regularized $\log_2$; TOM: Topological overlapping matrix; WGCNA: Weighted gene correlation network analysis.

Acknowledgements

The authors are indebted to Selección Batallé S.A. for providing the animal material. We gratefully acknowledge to J. Reixach (Selecció Batallé), J. Soler (IRTA), C. Millan (IRTA), A. Quintana (IRTA) and A. Rossell (IRTA) for their collaboration in the experimental protocols and pig management. Thanks also to the CERCA Programme of the Generalitat de Catalunya. We would also like to thank dr. Alex Clop for critical comments that greatly improved the content of the manuscript.

Author contributions

The authors' responsibilities were as follows: MA and RQ designed the research. MA, RQ, JT, TFC, RGP and EMS conducted the research. EMS analyzed the data. YRC and RGP contributed to the integrative analyses. MA and RQ secured funding for the study. EMS and MA drafted the manuscript. All authors read and approved the final manuscript.

Funding

The research presented in this publication was funded by grants AGL2013– 48742-C2–1-R and AGL2013–48742-C2–2-R awarded by the Spanish Ministry of Economy and Competitivity. We also acknowledge the support of the Spanish Ministry of Economy and Competitivity for the Center of Excellence Severo Ochoa 2016–2019 (SEV-2015-0533) grant awarded to the Centre for Research in Agricultural Genomics (CRAG). E. Mármol-Sánchez was funded with a Ph.D. fellowship FPU15/01733 awarded by the Spanish Ministry of Education and Culture (MECD). Y. Ramayo-Caldas is financially supported by the European Union H2020 Research and Innovation Programme under Marie Skłodowska-Curie grant (P-Sphere) agreement N° 6655919. T. F. Cardoso was funded with a fellowship from the CAPES Foundation-Coordination of Improvement of Higher Education, Ministry of Education of the Federal Government of Brazil. Thanks also to the CERCA Programme of the Generalitat de Catalunya.

Data Availability

The small RNA-seq data set used and/or analyzed in the current study is available at the sequence read archive (SRA) database (BioProject: PRJNA595998). The previously published RNA-seq data set was also submitted to the SRA database (BioProject: PRJNA386796).

Ethics approval

Animal care and management procedures followed national guidelines for the Good Experimental Practices and were approved by the Ethical Committee of the Institut de Recerca i Tecnologia Agroalimentàries (IRTA).

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Benítez R, Núñez Y, Óvilo C. Nutrigenomics in farm animals. J Invest Genomics. 2017; 4:1.

2. Puig-Oliveras A, Ramayo-Caldas Y, Corominas J, Estellé J, Pérez-Montarelo D, Hudson NJ, et al. Differences in muscle transcriptome among pigs phenotypically extreme for fatty acid composition. PLoS One. 2014; 9:e99720.

3. Ayuso M, Fernández A, Núñez Y, Benítez R, Isabel B, Barragán C, et al. Comparative analysis of muscle transcriptome between pig genotypes identifies genes and regulatory mechanisms associated to growth, fatness and metabolism. PLoS One. 2015; 10:e0145162.

4. Cardoso TF, Cánovas A, Canela-Xandri O, González-Prendes R, Amills M, Quintanilla R. RNA-seq based detection of differentially expressed genes in the skeletal muscle of Duroc pigs with distinct lipid profiles. Sci Rep. 2017; 7:40005.

5. Cardoso TF, Quintanilla R, Tibau J, Gil M, Mármol-Sánchez E, González-Rodríguez O, et al. Nutrient supply affects the mRNA expression profile of the porcine skeletal muscle. BMC Genomics. 2017; 18:603.

6. Ballester M, Amills M, González-Rodríguez O, Cardoso TF, Pascual M, González-Prendes R, et al. Role of AMPK signaling pathway during compensatory growth in pigs. BMC Genomics. 2018; 19:682.

7. Jia Cunling, Kong Xiaoyan, Koltes James E., Gou Xiao, Yang Shuli, Yan Dawei, Lu Shaoxiong, Wei Zehui. Gene Co-Expression Network Analysis Unraveling Transcriptional Regulation of High-Altitude Adaptation of Tibetan Pig. PLoS One. 2016; 11:e0168161.

8. Muñoz M, García-Casco JM, Caraballo C, Fernández-Barroso MA, Sánchez-Esquiliche F, Gómez F, et al. Identification of candidate genes and regulatory factors underlying intramuscular fat content through *longissimus dorsi* transcriptome analyses in heavy Iberian pigs. Front Genet. 2018; 9:608.

9. Komurov K, Ram PT. Patterns of human gene expression variance show strong associations with signaling network hierarchy. BMC Syst Biol. 2010; 4:154.

10. Ran D, Daye ZJ. Gene expression variability and the analysis of large-scale RNA-seq studies with the MDSeq. Nucleic Acids Res. 2017; 45:e127.

11. Ma C, Ji T. Detecting differentially expressed genes for syndromes by considering change in mean and dispersion simultaneously. BMC Bioinformatics. 2018; 19:330.

12. McCarthy DJ, Chen Y, Smyth GK. Differential expression analysis of multifactor RNA-seq experiments with respect to biological variation. Nucleic Acids Res. 2012; 40:4288–4297.

13. Chalancon Guilhem, Ravarani Charles N.J., Balaji S., Martinez-Arias Alfonso, Aravind L., Jothi Raja, Babu M. Madan. Interplay between gene expression noise and regulatory network architecture. Trends in Genetics. 2012; 28(5):221–232.

14. Eusebi PG, González-Prendes R, Quintanilla R, Tibau J, Cardoso TF, Clop A, et al. A genome-wide association analysis for carcass traits in a commercial Duroc pig population. Anim Genet. 2017; 48:466–469.

15. Mármol-Sánchez E., Quintanilla R., Jordana J., Amills M. An association analysis for 14 candidate genes mapping to meat quality quantitative trait loci in a Duroc pig population reveals that the ATP 1A2 genotype is highly associated with muscle electric conductivity. Animal Genetics. 2019; 51(1):95–100.

16. Cayuela JM, Garrido MD, Bañón SJ, Ros JM. Simultaneous HPLC analysis of α-tocopherol and cholesterol in fresh pig meat. J Agric Food Chem. 2003; 51:1120–1124.

17. Mach N, Devant M, Díaz I, Font-Furnols M, Oliver MA, García JA, et al. Increasing the amount of n-3 fatty acid in meat from young Holstein bulls through nutrition. J Anim Sci. 2006; 84:3039–3048.

18. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014; 30:2114–2120.

19. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal. 2011; 17:10.

20. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nat Methods. 2015; 12:357–360.

21. Pertea M, Pertea GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol. 2015; 33:290–295.

22. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. 2009; 10:R25.

23. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014; 30:923–930.

24. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. 2010; 11:R25.

25. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010; 26:139–140.

26. Benjamini Yoav, Hochberg Yosef. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Stat Soc B. 1995; 57(1):289–300.

27. Benjamini Y, Yekutieli D. The control of the false discovery rate in multiple testing under dependency. Ann Stat. 2001; 29:1165–1188.

28. Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, et al. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics. 2009; 25:1091–1093.

29. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003; 13:2498–2504.

30. Reverter A, Chan EKF. Combining partial correlation and an information theory approach to the reversed engineering of gene co-expression networks. Bioinformatics. 2008; 24:2491–2497.

31. Bellot P, Olsen C, Salembier P, Oliveras-Vergés A, Meyer PE. NetBenchmark: a bioconductor package for reproducible benchmarks of gene regulatory network inference. BMC Bioinformatics. 2015; 16:312.

32. Watson-Haigh NS, Kadarmideen HN, Reverter A. PCIT: an R package for weighted gene co-expression networks based on partial correlation and information theory approaches. Bioinformatics. 2010; 26:411–413.

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

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

35. Agarwal V, Bell GW, Nam J-W, Bartel DP. Predicting effective microRNA target sites in mammalian mRNAs. eLife. 2015; 4:e05005.

36. Mayya VK, Duchaine TF. On the availability of microRNA-induced silencing complexes, saturation of microRNA-binding sites and stoichiometry. Nucleic Acids Res. 2015; 43:7556–7565.

37. Reverter A, Hudson NJ, Nagaraj SH, Pérez-Enciso M, Dalrymple BP. Regulatory impact factors: unraveling the transcriptional regulation of complex traits from expression data. Bioinformatics. 2010; 26:896–904.

38. Langfelder P, Horvath S. WGCNA: and R package for weighted correlation network analysis. BMC Bioinformatics. 2008; 9:559.

39. Csardi G, Nepusz T. The igraph software package for complex network research. InterJournal Complex Syst. 2006; 1695.

40. Cardoso TF, Quintanilla R, Castelló A, Mármol-Sánchez E, Ballester M, Jordana J, et al. Analysing the expression of eight clock genes in five tissues from fasting and fed sows. Front Genet. 2018; 9:475.

41. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014; 15:550.

42. Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, et al. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev. 2011; 25:1915–1927.

43. Iyer MK, Niknafs YS, Malik R, Singhal U, Sahu A, Hosono Y, et al. The landscape of long noncoding RNAs in the human transcriptome. Nat Genet. 2015; 47:199–208.

44. de Rie D, Abugessaisa I, Alam T, Arner E, Arner P, Ashoor H, et al. An integrated expression atlas of miRNAs and their promoters in human and mouse. Nat Biotechnol. 2017; 35:872–878.

45. Ebert MS, Sharp PA. Emerging roles for natural microRNA sponges. Curr Biol. 2010; 20:R858–R861.

46. Migault M, Donnou-Fournet E, Galibert M-D, Gilot D. Definition and identification of small RNA sponges: focus on miRNA sequestration. Methods. 2017; 117:35–47.

47. Pan X, Wenzel A, Jensen LJ, Gorodkin J. Genome-wide identification of clusters of predicted microRNA binding sites as microRNA sponge candidates. PLoS One. 2018; 13:e0202369.

48. Bail S, Swerdel M, Liu H, Jiao X, Goff LA, Hart RP, et al. Differential regulation of microRNA stability. RNA. 2010; 16:1032–1039.

49. Guo Y, Liu J, Elfenbein SJ, Ma Y, Zhong M, Qiu C, et al. Characterization of the mammalian miRNA turnover landscape. Nucleic Acids Res. 2015; 43:2326–2341.

50. Lv Z, Wei Y, Wang D, Zhang C-Y, Zen K, Li L. Argonaute 2 in cell-secreted microvesicles guides the function of secreted miRNAs in recipient cells. PLoS One. 2014; 9:e103599.

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

52. Sun Y, Ge Y, Drnevich J, Zhao Y, Band M, Chen J. Mammalian target of rapamycin regulates miRNA-1 and follistatin in skeletal myogenesis. J Cell Biol. 2010; 189:1157–1169.

53. Zhang Y, Yu B, He J, Chen D. From nutrient to microRNA: a novel insight into cell signaling involved in skeletal muscle development and disease. Int J Biol Sci. 2016; 12:1247.

54. Hodge BA, Zhang X, Gutierrez-Monreal MA, Cao Y, Hammers DW, Yao Z, et al. *MYOD1* functions as a clock amplifier as well as a critical co-factor for downstream circadian gene expression in muscle. eLife. 2019; 8:e43017.

55. Kefas B, Godlewski J, Comeau L, Li Y, Abounader R, Hawkinson M, et al. microRNA-7 inhibits the epidermal growth factor receptor and the Akt pathway and is down-regulated in glioblastoma. Cancer Res. 2008; 68:3566–3572.

56. Fang Y, Xue J-L, Shen Q, Chen J, Tian L. MicroRNA-7 inhibits tumor growth and metastasis by targeting the phosphoinositide 3-kinase/Akt pathway in hepatocellular carcinoma. Hepatology. 2012; 55:1852–1862.

57. Goedeke L, Aranda JF, Fernández-Hernando C. microRNA regulation of lipoprotein metabolism. Curr Opin Lipidol. 2014; 25:282–288.

58. Wagschal A, Najafi-Shoushtari SH, Wang L, Goedeke L, Sinha S, deLemos AS, et al. Genome-wide identification of microRNAs regulating cholesterol and triglyceride homeostasis. Nat Med. 2015; 21:1290–1297.

59. Goedeke Leigh, Rotllan Noemi, Canfrán-Duque Alberto, Aranda Juan F, Ramírez Cristina M, Araldi Elisa, Lin Chin-Sheng, Anderson Norma N, Wagschal Alexandre, de Cabo Rafael, Horton Jay D, Lasunción Miguel A, Näär Anders M, Suárez Yajaira, Fernández-Hernando Carlos. MicroRNA-148a regulates LDL receptor and ABCA1 expression to control circulating lipoprotein levels. Nature Medicine. 2015; 21(11):1280–1289.

60. Gastebois C, Chanon S, Rome S, Durand C, Pelascini E, Jalabert A, et al. Transition from physical activity to inactivity increases skeletal muscle miR-148b content and triggers insulin resistance. Phys Rep. 2016; 4:e12902.

61. Rotllan N, Price N, Pati P, Goedeke L, Fernández-Hernando C. microRNAs in lipoprotein metabolism and cardiometabolic disorders. Atherosclerosis. 2016; 246:352–360.

62. Xu Q, Jiang Y, Yin Y, Li Q, He J, Jing Y, et al. A regulatory circuit of miR-148a/152 and DNMT1 in modulating cell transformation and tumor angiogenesis through IGF-IR and IRS1. J Mol Cell Biol. 2013; 5:3–13.

63. Sarkar S, Dey BK, Dutta A. MiR-322/424 and −503 are induced during muscle differentiation and promote cell cycle quiescence and differentiation by down-regulation of Cdc25A. Mol Biol Cell. 2010; 21:2138–2149.

64. Jia H, Zhao Y, Li T, Zhang Y, Zhu D. miR-30e is negatively regulated by myostatin in skeletal muscle and is functionally related to fiber-type composition. Acta Biochim Biophys Sin. 2017; 49:392–399.

65. Zaragosi L-E, Wdziekonski B, Le Brigand K, Villageois P, Mari B, Waldmann R, et al. Small RNA sequencing reveals miR-642a-3p as a novel adipocyte-specific microRNA and miR-30 as a key regulator of human adipogenesis. Genome Biol. 2011; 12:R64.

66. Shi C, Zhang M, Tong M, Yang L, Pang L, Chen L, et al. miR-148a is associated with obesity and modulates adipocyte differentiation of mesenchymal stem cells through Wnt signaling. Sci Rep. 2015; 5:9930.

67. Schweisgut J, Schutt C, Wüst S, Wietelmann A, Ghesquière B, Carmeliet P, et al. Sex-specific, reciprocal regulation of ERα and miR-22 controls muscle lipid metabolism in male mice. EMBO J. 2017; 36:1199–1214.

68. Chen B, Tang H, Liu X, Liu P, Yang L, Xie X, et al. miR-22 as a prognostic factor targets glucose transporter protein type 1 in breast cancer. Cancer Lett. 2015; 356:410–417.

69. Xiao D, Zhou T, Fu Y, Wang R, Zhang H, Li M, et al. MicroRNA-17 impairs glucose metabolism in insulin-resistant skeletal muscle via repressing glucose transporter 4 expression. Eur J Pharmacol. 2018; 838:170–176.

70. Hudson NJ, Dalrymple BP, Reverter A. Beyond differential expression: the quest for causal mutations and effector molecules. BMC Genomics. 2012; 13:356.

71. Óvilo C, Benítez R, Fernández A, Núñez Y, Ayuso M, Fernández AI, et al. *Longissimus dorsi* transcriptome analysis of purebred and crossbred Iberian pigs differing in muscle characteristics. BMC Genomics. 2014; 15:413.

72. Li J, Yue Z, Xiong W, Sun P, You K, Wang J. *TXNIP* overexpression suppresses proliferation and induces apoptosis in SMMC7221 cells through ROS generation and MAPK pathway activation. Oncol Rep. 2017; 37:3369–3376.

73. Ling L, Zhang S-H, Zhi L-D, Li H, Wen Q-K, Li G, et al. MicroRNA-30e promotes hepatocyte proliferation and inhibits apoptosis in cecal ligation and puncture-induced sepsis through the JAK/STAT signaling pathway by binding to *FOSL2*. Biomed Pharmacother. 2018; 104:411–419.

74. Wrann CD, Eguchi J, Bozec A, Xu Z, Mikkelsen T, Gimble J, et al. *FOSL2* promotes leptin gene expression in human and mouse adipocytes. J Clin Invest. 2012; 122:1010–1021.

75. Salcher S, Hermann M, Kiechl-Kohlendorfer U, Ausserlechner MJ, Obexer P. C10ORF10/DEPP-mediated ROS accumulation is a critical modulator of *FOXO3*-induced autophagy. Mol Cancer. 2017; 16:95.

76. Lee Y-S, Sasaki T, Kobayashi M, Kikuchi O, Kim H-J, Yokota-Hashimoto H, et al. Hypothalamic ATF3 is involved in regulating glucose and energy metabolism in mice. Diabetologia. 2013; 56:1383–1393.

77. Allison MB, Pan W, MacKenzie A, Patterson C, Shah K, Barnes T, et al. Defining the transcriptional targets of leptin reveals a role for *Atf3* in leptin action. Diabetes. 2018; 67:1093–1104.

78. Boyle KB, Hadaschik D, Virtue S, Cawthorn WP, Ridley SH, O'Rahilly S, et al. The transcription factors Egr1 and Egr2 have opposing influences on adipocyte differentiation. Cell Death Differ. 2009; 16:782–789.

79. Haldar SM, Jeyaraj D, Anand P, Zhu H, Lu Y, Prosdocimo DA, et al. Kruppel-like factor 15 regulates skeletal muscle lipid flux and exercise adaptation. Proc Natl Acad Sci U S A. 2012; 109:6739–6744.

80. Ganassi M, Badodi S, Ortuste Quiroga HP, Zammit PS, Hinits Y, Hughes SM. Myogenin promotes myocyte fusion to balance fibre number and size. Nat Commun. 2018; 9:4232.

81. Jeong JY, Jeoung NH, Park K-G, Lee I-K. Transcriptional regulation of pyruvate dehydrogenase kinase. Diabetes Metab J. 2012; 36:328–335.

82. Holness MJ, Sugden MC. Regulation of pyruvate dehydrogenase complex activity by reversible phosphorylation. Biochem Soc Trans. 2003; 31:1143–1151.

83. Zhang S, Hulver MW, McMillan RP, Cline MA, Gilbert ER. The pivotal role of pyruvate dehydrogenase kinases in metabolic flexibility. Nutr Metab. 2014; 11:10.

84. Piao L, Sidhu VK, Fang Y-H, Ryan JJ, Parikh KS, Hong Z, et al. FOXO1-mediated upregulation of pyruvate dehydrogenase kinase-4 (PDK4) decreases glucose oxidation and impairs right ventricular function in pulmonary hypertension: therapeutic benefits of dichloroacetate. J Mol Med. 2013; 91:333–346.

85. Gopal K, Saleme B, Al Batran R, Aburasayn H, Eshreif A, Ho KL, et al. FoxO1 regulates myocardial glucose oxidation rates via transcriptional control of pyruvate dehydrogenase kinase 4 expression. Am J Physiol Circ Physiol. 2017; 313:H479–H490.

86. Ouyang Weiming, Liao Will, Luo Chong T., Yin Na, Huse Morgan, Kim Myoungjoo V., Peng Min, Chan Pamela, Ma Qian, Mo Yifan, Meijer Dies, Zhao Keji, Rudensky Alexander

Y., Atwal Gurinder, Zhang Michael Q., Li Ming O. Novel Foxo1-dependent transcriptional programs control Treg cell function. Nature. 2012; 491(7425):554–559.

87. Kim Eui Ho, Gasper David J., Lee Song Hee, Plisch Erin Hemmila, Svaren John, Suresh M. Bach2 Regulates Homeostasis of Foxp3+ Regulatory T Cells and Protects against Fatal Lung Disease in Mice. The Journal of Immunology. 2013; 192(3):985–995.

88. Itoh-Nakadai A, Matsumoto M, Kato H, Sasaki J, Uehara Y, Sato Y, et al. A Bach2-Cebp gene regulatory network for the commitment of multipotent hematopoietic progenitors. Cell Rep. 2017; 18:2401–2414.

89. Reznick J, Preston E, Wilks DL, Beale SM, Turner N, Cooney GJ. Altered feeding differentially regulates circadian rhythms and energy metabolism in liver and muscle of rats. Biochim Biophys Acta Mol basis Dis. 1832; 2013:228–238.

90. Dyar KA, Ciciliot S, Wright LE, Biensø RS, Tagliazucchi GM, Patel VR, et al. Muscle insulin sensitivity and glucose metabolism are controlled by the intrinsic muscle clock. Mol Metab. 2014; 3:29–41.

91. Yamaguchi S, Moseley AC, Almeda-Valdes P, Stromsdorfer KL, Franczyk MP, Okunade AL, et al. Diurnal variation in PDK4 expression is associated with plasma free fatty acid availability in people. J Clin Endocrinol Metab. 2018; 103:1068–1076.

92. Everett Logan J., Lazar Mitchell A. Nuclear receptor Rev-erbα: up, down, and all around. Trends in Endocrinology & Metabolism. 2014; 25(11):586–592.

93. Zhang Xinyue, Guo Jieyu, Wei Xiangxiang, Niu Cong, Jia Mengping, Li Qinhan, Meng Dan. Bach1: Function, Regulation, and Involvement in Disease. Oxidative Medicine and Cellular Longevity. 2018; 2018:1–8.

94. Li K, Qiu C, Sun P, Liu D, Wu T, Wang K, et al. Ets1-mediated acetylation of FoxO1 is critical for gluconeogenesis regulation during feed-fast cycles. Cell Rep. 2019; 26:2998–3010.e5.

95. Besse-Patin A, Jeromson S, Levesque-Damphousse P, Secco B, Laplante M, Estall JL. *PGC1A* regulates the IRS1:IRS2 ratio during fasting to influence hepatic metabolism downstream of insulin. Proc Natl Acad Sci U S A. 2019; 116:4285–4290.

96. Laurie SM, Cain CC, Lienhard GE, Castle JD. The glucose transporter GluT4 and secretory carrier membrane proteins (SCAMPs) colocalize in rat adipocytes and partially segregate during insulin stimulation. J Biol Chem. 1993; 268:19110–19117.

97. Yoshizumi S, Suzuki S, Hirai M, Hinokio Y, Yamada T, Yamada T, et al. Increased hepatic expression of ganglioside-specific sialidase, *NEU3*, improves insulin sensitivity and glucose tolerance in mice. Metabolism. 2007; 56:420–429.

98. Jia Z, Moulson CL, Pei Z, Miner JH, Watkins PA. Fatty acid transport protein 4 is the principal very long chain fatty acyl-CoA synthetase in skin fibroblasts. J Biol Chem. 2007; 282:20573–20583.

99. Larkin James R., Zhang Fang, Godfrey Lisa, Molostvov Guerman, Zehnder Daniel, Rabbani Naila, Thornalley Paul J. Glucose-Induced Down Regulation of Thiamine Transporters in the Kidney Proximal Tubular Epithelium Produces Thiamine Insufficiency in Diabetes. PLoS One. 2012; 7(12):e53175.

100. Rey G, Valekunja UK, Feeney KA, Wulund L, Milev NB, Stangherlin A, et al. The pentose phosphate pathway regulates the circadian clock. Cell Metab. 2016; 24:462–473.

101. Baenziger JE, Hénault CM, Therien JPD, Sun J. Nicotinic acetylcholine receptor–lipid interactions: mechanistic insight and biological function. Biochim Biophys Acta Biomembr. 1848; 2015:1806–1817.

102. Liu LB, Chen XD, Zhou XY, Zhu Q. The Wnt antagonist and secreted frizzled-related protein 5: Implications on lipid metabolism, inflammation, and type 2 diabetes mellitus. Biosci Rep. 2018; 38:BSR20180011.

103. Casas E, Shackelford SD, Keele JW, Stone RT, Kappes SM, Koohmaraie M. Quantitative trait loci affecting growth and carcass composition of cattle segregating alternate forms of myostatin. J Anim Sci. 2000; 78:560.

104. Nalaila SM, Stothard P, Moore SS, Li C, Wang Z. Whole-genome QTL scan for ultrasound and carcass merit traits in beef cattle using Bayesian shrinkage method. J Anim Breed Genet. 2012; 129:107–119.

105. Schering L, Albrecht E, Komolka K, Kühn C, Maak S. Increased expression of thyroid hormone responsive protein (THRSP) is the result but not the cause of higher intramuscular fat content in cattle. Int J Biol Sci. 2017; 13:532–544.

106. Ortega-Prieto P, Postic C. Carbohydrate sensing through the transcription factor ChREBP. Front Genet. 2019; 10:47.

CHAPTER IV. DISCUSSION

The main goals of the present discussion are to explain the rationale of the experiments made in the current thesis as well as to discuss issues and additional topics, thematically linked to our results, that were not treated with enough detail in the six papers that form part of the Ph.D. thesis.

4.1. Identifying causal mutations in regions containing QTL for meat quality traits: limitations and prospects

In previous studies, González-Prendes et al. (2017, 2019b) carried out GWAS for meat color, pH, muscle electric conductivity (CE) and IMF content and composition traits recorded in the Lipgen population formed by 350 Duroc pigs, thus accomplishing one of the main goals of the AGL2010-22208-C02-02 project. González-Prendes et al. (2017) found 17 QTL for color traits such as redness (a*), and lightness (L*), as well as for pH and CE of the *longissimus dorsi* (LD) and *gluteus medius* (GM) skeletal muscles. In an additional work, González-Prendes et al. (2019b) further determined QTL regions associated with IMF content and composition traits in the GM and LD muscles. One of the goals of the present Ph.D. thesis was to identify potential genes of interest mapping to meat quality QTL reported by González Prendes et al. (2017, 2019b) and to investigate whether their variability might be associated with such traits. The final goal of this research would be to identify mutations which might be good candidates to have causal effects, thus deserving further validation. To achieve this goal, we performed WGS of the five Duroc founders of the Lipgen population. This provided a comprehensive view of all SNP variants mapping to genomic regions of the five founders that, in the Lipgen population, have been identified as containing meat quality QTL.

An interesting approach for detecting putative causal mutations in candidate genes within QTL of interest would have been the fine-mapping of these causal polymorphisms by using the WGS data from the founders to impute genotypes in the Lipgen population (Marchini and Howie, 2010). Through this approach, the number of markers within QTL would have increased substantially and, in principle, the causal mutations would be comprised in the set of imputed variants. This would lead to an increased accuracy of GWAS predictions and a higher number of reported QTL within refined intervals (Druet et al., 2014).

Unfortunately, in our case this approach was unfeasible due to the very low number of sequenced individuals (N = 5), a feature that made imputations highly unreliable. Given this limitation, we used a much less powerful approach consisting of identifying high impact mutations in functional candidate genes mapping to meat quality QTL. This approach has certain limitations because the biological factors that regulate meat quality traits are mostly unknown and the annotation of the pig genome is still quite limited, making it difficult to select the set of candidate genes and polymorphisms to be genotyped. In the end, we selected a set of 14 candidate genes, and a total of 19 SNPs that were located in QTL regions associated with meat quality and IMF composition traits. In this regard, we were able to detect 8 SNPs significantly associated with a*, CE and C18:1-n7 fatty acid content in the LD muscle at the nominal level, and seven of them conserved their significance after multiple testing correction, as described in Paper I. The significance at the nominal level of these seven SNPs was also maintained when analyzing their association at the chromosome-wide level, but only 1 SNP in the *ATP1A2* gene remained significant at the chromosome-wide level after multiple testing correction.

Why the majority of the SNPs mapping to QTL were not associated with the trait regulated by the QTL? The most obvious reason would be the lack of sufficient linkage disequilibrium with the causal mutation. The analysis of the Manhattan plot shown in Figure 1 of paper I clearly indicates that SNPs located at a close distance behave differently, i.e. they show highly divergent associations with electrical conductivity in the LD muscle, going from the complete lack of association to highly significant associations. The amount of LD between nearby mutations depends on many factors such as local recombination rate, selection and demography (Pritchard and Przeworski, 2001; Tenesa et al., 2007; Amaral et al., 2008). In a recent study, the association of 4 SNPs mapping to the *SLC45A2* gene with the red and blond color of Mangalitza pigs was investigated, and two SNPs displayed a very significant association while the remaining two did not show any evidence of association despite being located in the same gene (our unpublished data).

The only SNP that showed chromosome-wide significance, after multiple testing correction, when merged with 3,899 SNPs mapping to SSC4 chromosome, was the splice region variant rs344748241 (c.1653G>A) located in the ATPase Na^+/K^+ transporting α_2 subunit (*ATP1A2*) gene. As outlined in Paper I, this finding was relevant because the *ATP1A2* gene has been previously reported in a number of studies as being associated with several porcine meat quality traits such as fat cut percentage, backfat thickness, carcass length or muscle mass (Cepica et al., 2003; Davoli et al., 2006; Fontanesi et al., 2012). More importantly, this gene encodes the catalytic subunit of the ATPase Na^+/K^+ enzyme, which is responsible for the hydrolysis of ATP coupled with the exchange of Na^+ and K^+ ions across the plasma membrane. Such mechanism contributes to maintain the electrochemical balance within the cell, providing the energy needed for the active transport of several nutrients and the electrical excitability of nerves and muscles (Suhail, 2010).

A high linkage disequilibrium between the rs344748241 *ATP1A2* SNP and the reported QTL lead variant with the highest significance in the study of González-Prendes et al. (2017) was revealed (Paper I). Despite this result, it is important to note that the site with the highest significance within the QTL interval not always contains the causal mutation (Schaid et al., 2018). As previously stated, SNP panels are designed to capture linkage disequilibrium blocks throughout the genome, rather than pre-defined causal SNPs. Indeed, even if the underlying causal mutation is genotyped, the statistical significance obtained for the association with the phenotypic measure highly depends on allele frequency, so we cannot assume that always causal mutations will be located at the peak of QTL because if their frequencies are very low the estimates of genotypic means will be inaccurate unless very large populations are used in the GWAS.

In summary, as discussed in Paper I, the polymorphism of the *ATP1A2* gene might modulate the electrochemical gradient across the cellular membrane of skeletal muscle cells (Suhail, 2010), a parameter intimately ligated with electrical conductivity in the meat. Nevertheless, the confirmation of such hypothesis was not part of our work. Further experiments should be conducted in order to determine the functional implications of the *ATP1A2* gene and variants within its sequence. For instance, building a comprehensive catalogue of *ATP1A2* variants and genotyping them in the Lipgen population, as well as in other pig populations with electrical conductivity data, could facilitate the identification of a potential causal variant. In the Ensembl database, as much as 2,945 SNP variants have been detected in the porcine *ATP1A2* gene and only 54 of them might have functional effects in the coding region (i.e., missense or splice site variants). On the other hand, in our set of mutations to be genotyped, we did not include indels because calling for such type of variants is prone to error, generating many false positives because most dedicated tools lack accurate methods for identifying sequencing errors before indel calling (Hasan et al., 2015). Moreover, the observed concordance among tools when annotating indels is reported to be quite low (Fang et al., 2014). However, indels can have important consequences on gene expression and protein structure, so any gene-centric study should take them into account. Once one or several candidate mutations are detected, the next step would be the implementation of functional assays. In the case of a candidate regulatory mutation, the creation of fusion plasmid constructs to be co-transfected to cell lines jointly with luciferase reporters and subsequent measurement of the transcriptional activity of each allele through a luciferase reporter assays system could be a good strategy. This method has proved to be a successful approach to unravel the functional consequences of variants with uncertain significance (Woods et al., 2016). In the case of a missense mutation, enzyme activity could be measured in cell cultures transfected with different constructs (one for each allele). In this way, structural and physicochemical alterations in the Na^{+}/K^{+} ATPase enzymes could be investigated, as several pathological conditions have been previously reported for dysfunctions in this enzymatic complex (Suhail, 2010; Sampedro Castañeda et al., 2018).

4.2. Loss-of-function variant detection and the relevance of accurate gene annotation

The segregation of deleterious variants has a considerable impact on the fitness and fertility of pig populations, as lethal or sublethal mutations can cause reproductive dysfunctions as well as pre and postnatal losses in the form of mummification, abortions or stillborn events (Dron et al., 2014; Verardo et al., 2016; Derks et al., 2017, 2019a). The WGS of the five Duroc boars that founded the Lipgen population allowed us to identify putative loss-of-function (LoF) mutations and examine their segregation in the offspring formed by 350 Duroc barrows. From a functional point of view, stop gained polymorphisms are particularly relevant because they might cause the inactivation of gene function through nonsense mediated decay or due to the generation of truncated non-functional or impaired proteins (Miller and Pearce, 2014), thus leading to observable phenotypic or genotypic (e.g. lack or depletion of one of the homozygous genotypes) consequences.

First, we would like to discuss the experimental workflow that led to the identification of a putative stop gained mutation in the *ASS1* gene. This workflow is no presented in the Animal Genetics paper II, due to space constraints, but it is essential to understand the motivation of the experiments that led to its publication. An initial screening of the whole-genome sequences of the five founders of the Lipgen population using the Sscrofa10.2 assembly annotation revealed the existence of 432 predicted stop gained mutations, from which a total of 225 were found in heterozygous states (our unpublished data). Among these, we selected a total of 60 variants for genotyping, based on their location at genes with relevant predicted functions. After genotyping the 60 stop gained variants, only 27 of them segregated in the Lipgen population. Such results revealed the extent of false positives included in the predicted variant effects using the Sscrofa10.2 annotation, a finding that could be explained by the fact that the initial porcine assembly was the least complete among all sequenced domestic species (Seemann et al., 2015). A good example of this is the number of genomic regions with low quality and low coverage in the Sscrofa10.2 assembly. Warr et al. (2015) investigated this issue and found that regions with low quality and low coverage encompassed as much as 33.07% of the pig genome, and that half of annotated SNPs according to dbSNP database (https://www.ncbi.nlm.nih.gov/snp/) were located in these regions. This could provide an explanation to the fact that when we carefully assessed the predicted effects of the selected 27 segregating variants using the current Sscrofa11.1 annotation, only 7 of them conserved a stop gained predicted effect.

The level of discrepancy among different porcine annotations should be considered as clearly indicative of the limited accuracy of functional predictions based on mere sequence assessment, particularly in poorly annotated species. It is therefore crucial that a careful manual curation of the observed LoF mutations is performed, in order to exclude as many false positives as possible before considering any further analyses. Indeed, a substantial number of false positives for LoF polymorphism are expected to arise when performing variant-calling from WGS data. For instance, MacArthur et al. (2012) identified a total of 2,951 putative LoF variants from a cohort of 185 human whole-genome sequences. From these, 2,809 corresponded to SNPs or short indels, whereas 142 belonged to large deletions. By applying diverse filtering steps, MacArthur and collaborators were able to retain 1,285 (43,5%) out of the initial 2,951 polymorphisms. Among the removed variants, there was a tendency to have increased alternative allele frequencies and a higher distribution at the beginning or the end of the annotated transcripts, highlighting their probable non-deleteriousness or at least partial harmful effects.

The are many factors which can cause the erroneous annotation of LoF variants, like, for instance, the presence of adjacent indels or compensatory mutations, the presence of pseudogenes, which can be an important source of error as they tend to accumulate deleterious mutations (Sen and Ghosh, 2013), insufficient coverage of the region containing the SNP, splice sites located in non-canonical sites or deficient knowledge of gene structure, which ultimately requires a thorough curation (Warr et al., 2015). Several of these factors were very difficult to control in our experimental design due to the poor quality of the 10.2 version of the pig genome assembly (Seemann et al., 2015; Warr et al., 2015), leading to genotype many putative stop gained mutations that in reality did not have such effect.

Even in very well characterized species such as humans, obtaining a set of high-confidence LoF mutations is very challenging. This topic is not discussed in the six papers that form part of the current thesis, but we think it needs to be commented in this global discussion. As previously said, we consider that these difficulties were produced not only by the intrinsic complexity of evaluating the functional consequences of LoF mutations (it is even hard in well characterized species such as humans) but also by the poor annotation of the 10.2 assembly of the pig genome. In a couple of cases, however, we found mutations that are, in principle, *bona fide* stop gained mutations with potential functional effects. One of them was the rs339148947 (c.1087C>T) variant within the glycerol-3 phosphate acyltransferase 2 (*GPAT2*) gene (manuscript in preparation). We observed in the Lipgen population a complete lack of TT pigs, and a significant deviation of the Hardy-Weinberg equilibrium (*P*-value = 0.0044). It is well known that one of the hallmarks of lethal or sublethal mutations is the lack or depletion of homozygous genotypes (Casellas et al., 2012; Derks et al., 2017, 2019a; Pausch et al., 2015; VanRaden et al., 2011). Knockout mice for *GPAT2* display male infertility associated with reduced testis weight, impaired spermatogenesis, azoospermia, abnormal DNA methylation during gametogenesis, and increased male germ cell apoptosis (http://www.informatics.jax.org/marker/MGI:2684962). Indeed, GPAT2 protein has key roles in regulating spermatogenesis and biogenesis of piwi-interacting RNAs (piRNAs), as described by Shiromoto et al. (2019).

The second mutation of interest was rs81212146 (c.944T>A), located in the *ASS1* gene, which was analyzed in depth in Paper II. Genotyping of this polymorphism confirmed, to our surprise, the existence of seven AA homozygous animals in the Lipgen population, and a relatively high proportion of heterozygous individuals (MAF = 0.1433). Such findings were counterintuitive because highly harmful variants usually have very low frequencies and, moreover, individuals with homozygous genotypes are usually not viable. Interestingly, when we surveyed the available literature, we became aware that Groenen et al. (2012) already reported the existence of pigs homozygous for the same stop gained mutation detected by us.

These authors attributed such finding to the fact that homozygous AA pigs might suffer a mild form of citrullinemia, not harmful enough to determine the death or any reproductive impairment of the affected animals (Häberle et al., 2003). This assumption did not take into account that homozygous individuals for LoF mutations that abolish partially ASS1 function usually present a severely hampered detoxification of ammonia metabolites, thus leading to type I citrullinemia. Typical symptoms of the partial inactivation of ASS1 are hyperammonemia and hepatic encephalopathy that usually manifests with lethargy, reduced feed intake and growth rate, seizures, comma and sometimes death (Burton, 2000). In the case of individuals with a complete inactivation of ASS1, the urea cycle is not functional and they die soon after birth. The contradiction between the severe predicted harmful effects of ASS1 dysfunction and the apparent healthy status of homozygous recessive pigs in our population (they completed their production cycle and showed a normal weight at slaughter), motivated us to further investigate the possible underlying causes.

In our mind, we were aware that individuals with homozygous genotypes for lethal mutations that are alive and perfectly healthy have been reported in the literature (Chen et al., 2016b). Such paradoxical findings could be explained by the existence of compensatory mechanisms like the read-through of the nonsense codon in the ribosome, skipping of the exon carrying the lethal mutation during mRNA maturation, or the co-segregation of variants abolishing the effect of the nonsense mutation by modifying the open reading frame (ORF) of the transcript (MacArthur et al., 2012; Rausell et al., 2014). We suspected that one of these mechanisms was the true cause of the non-lethality of the c.944T>A mutation in the *ASS1* gene.

We carried out the sequencing of the genomic region containing the stop gained mutation both at the genomic DNA and cDNA levels, in order to detect any possible structural compensatory mechanism, such as alternative splicing events, that could avoid the nonsense mutation to be present in the mRNA sequence. A thorough analysis of the obtained sequences allowed us to discover a compensatory mutation (rs81212145, c.943T>C) located immediately before the rs81212146 variant, being both in complete linkage disequilibrium. This result was confirmed by analyzing the segregation of both mutations in 120 whole-genome sequenced pigs obtained from NCBI SRA public repositories, which evidenced that both SNPs co-segregate perfectly and reach high frequencies in Asian pigs. In the light of these results, rather than considering the existence of two SNPs, each one with its particular annotation and functional prediction, we should consider the existence of a single dinucleotide missense polymorphism causing a benign amino acid change of leucine by glutamine at position 315 of the mature protein.

From a practical point of view, one of the applications of identifying harmful mutations would be the design of genotyping techniques aiming to screen boars with the final aim of planning mating schemes resulting in the total or at least partial removal of such variation. This approach would result in the improvement of numerical productivity, which is one of the main parameters which determines the economic profitability of pig farms.

4.3. RNA-seq analyses from porcine skeletal muscle transcriptome revealed candidate polymorphisms regulating gene expression profiles

In paper I, we reported the genotyping of high impact SNPs located in genes with potential roles in meat quality and mapping to meat quality QTL. These meat quality QTL were previously identified by González-Prendes et al (2017) by using a GWAS approach. As previously stated, the moderate sample size of the Lipgen population (N = 350) and thus the limited statistical power, made it highly likely that many genomic regions with true but small additive effects on meat quality traits were missed in the GWAS carried out by González-Prendes et al. (2017). On the other hand, Cardoso et al. (2017b) investigated the effect of food intake on the GM skeletal muscle transcriptome in 36 Duroc gilts subjected to fasting-feeding conditions. This study led to the identification of a wide range of transcripts showing differential expression profiles when comparing fasting and fed gilts. Among the set of detected DE genes, those associated with the regulation of peripheral circadian clock, as well as with glucose metabolism and energy homeostasis were among the ones which showed a prominent differential expression. In this regard, in paper III we aimed to investigate whether the variability of several of the differentially expressed genes detected by Cardoso et al. (2017b) might display associations with meat quality traits. The main emphasis of this experiment was put on circadian genes *ARNTL2*, *CIART*, *CRY2*, *NPAS2*, *PER1* and *PER2*. The main reason of this is that there is a tight link between circadian genes and lipid metabolism. For instance, genes involved in lipid absorption show dysregulated expression patterns and are irresponsive to feed restriction in knockout mice lacking the clock circadian regulator (*CLOCK*) gene (Pan et al., 2010). In peripheral tissues, the role of *CLOCK* is accomplished by *NPAS2*, in conjunction with *ARNTL2* (Landgraf et al., 2016). Additionally, *CLOCK* impairment also produces altered rhythmic expression patterns of lipolytic and lipogenic genes (Shostak et al., 2013). The lack of another important circadian zeitgeber, *ARNTL*, a paralog of *ARNTL2* in the central nervous system, severely impacts the expression of key lipogenic factors such as *PPARG*, CCAAT/enhancer-binding protein α (*CEBPA*), *SREBF1* or *FASN* (Shimba et al., 2005). Triglycerides serum concentrations are also affected by altered expression of *PER* and *CRY* elements, as *PER2* inhibits *PPARG*-mediated activation of lipogenic genes (Grimaldi et al., 2010), while *CRY1/2* inhibition leads to increased susceptibility to diet-induced obesity (Barclay et al., 2013). We also investigated two genes related with energy homeostasis (*MIGA2*) and glucose metabolism (*PCK1*), as both processes are also tightly linked with circadian rhythmicity (Lamia et al., 2011; Zhao et al., 2012; Barclay et al., 2013; Zani et al., 2013).

The polymorphisms to be genotyped (N = 20) in the Lipgen population were retrieved from 52 Duroc pigs with RNA-seq transcriptomic data for the GM muscle (Cardoso et al., 2017a), as well as from the WGS of the five Duroc boars that founded the Lipgen population. Performance of association analyses revealed that 10 out of 20 genotyped SNPs showed nominal associations with at least one lipid trait, but the majority of these associations lost significance after correction for multiple testing (Paper III). It should be noticed that here correction for multiple testing is much less stringent that in the case of a GWAS, hence we conclude that the majority of SNPs mapping to circadian genes did not have any impact on the lipid traits under study. One interpretation of this result is that circadian genes are master regulators of gene expression and a disruption of their function could have dramatic consequences on viability, so purifying selection is probably intense at removing most polymorphisms with functional effects. For instance, Layeghifard et al. (2008) explored the evolutionary mechanisms underlying gene duplication and functional divergence in circadian regulator genes, and found that the rate of nonsynonymous substitutions per nonsynonymous site rate was well below the rate of synonymous substitutions per synonymous site rate, while purifying selection was the evolutionary dominant process guiding sequence change in several of the analyzed genes such as *CLOCK* and *NPAS2* or the casein kinases δ and ε (*CSNK1D* and *CSNK1E*). Other key circadian regulators like timeless (*TIM*), which downregulates the activation of *PER*1 by *CLOCK/ARNTL*, are also under strong purifying selection (Gu et al., 2014). Indeed, this evolutionary constraint would have removed strong deleterious mutations from such regulatory circadian genes, given their paramount importance in maintaining a correct shifting between day-night cycles and the metabolic balance in cells. So, we conclude that the majority of the variants that we genotyped were benign and did not have noticeable effects on lipid metabolism, probably because purifying selection has a strong effect on the variability of circadian genes.

Apart from circadian genes, we were particularly interested in exploring the effects of *PCK1* gene variations. In this regard, it is worth mentioning the rs343196765 missense mutation (Met139Leu) located in the 4th exon of *PCK1* gene, which was among the ones we genotyped in the Lipgen population. Latorre et al. (2016) investigated this variant in several porcine populations, and reported its strong association with reduced IMF content and glucose metabolism due to a reduced enzymatic activity in the glyceroneogenic direction. However, we did not find any significant association of this variant on IMF content or on other lipid traits, with the only exception of C17:0 in the LD muscle, but only at the nominal level. Additionally, other polymorphisms located at *CIART, PER1* and *PCK1* genes were associated with backfat thickness, α-linolenic (C18:3) and margaric (C17:0) fatty acids, respectively, although only at the nominal level.

Moreover, as discussed in paper III, increased C18:0 fatty acid content in LD and LDL serum concentration were revealed for animals carrying the mutated alleles for rs320439526 (c.-6C>T) and rs330779504 (c.1455G>A) polymorphisms at *CRY2* and *MIGA2* genes, respectively. Only these two SNPs showed associations with meat quality traits that remained significant after correction for multiple testing. For these variants, we decided to investigate their association with the mRNA expression of the corresponding genes. Through these analyses, we were able to detect an association between rs320439526 and rs330779504 polymorphisms and the expression of *CRY2* and *MIGA2* transcripts, respectively. In this way, pigs with homozygous genotypes for the alternative alleles showed a reduced mRNA expression of the corresponding genes when compared with pigs heterozygous or homozygous for the reference alleles using microarray expression data (Paper III, Figure 1).

A complementary analysis that is not published in paper III but that was carried out during the present thesis was to confirm whether these observed differences in microarray probes expression profiles could be reproduced in the RNA-seq data employed for detecting candidate SNPs (Cardoso et al., 2017a), as microarray profiles have limited resolution to measure with confidence the mRNA levels of lowly expressed genes (Black et al., 2013). This is an important issue because regulatory transcription factors like those encoded by circadian genes usually are expressed at low levels (Vaquerizas et al., 2009). Although this confirmatory analysis has not been published, we believe that it is worth briefly reporting it in this global discussion because it provides an additional and valuable perspective about the results presented in paper III. Indeed, contrary to our expectation, no significant association were detected between the two aforementioned *CRY2* and *MIGA2* SNPs and the mRNA levels of the corresponding genes measured by RNA-seq technique (Figure 1). This result is probably due to the scarce number of available animals with GM muscle RNA-Seq data (N = 52). Nevertheless, we observed a statistical tendency towards reduced expression levels in pigs homozygous for the alternative alleles for rs320439526 (TT) and rs330779504 (AA) polymorphisms. So, there is a certain consistency between the association analyses performed with microarray and RNA-seq data, although statistical significances are not the same.

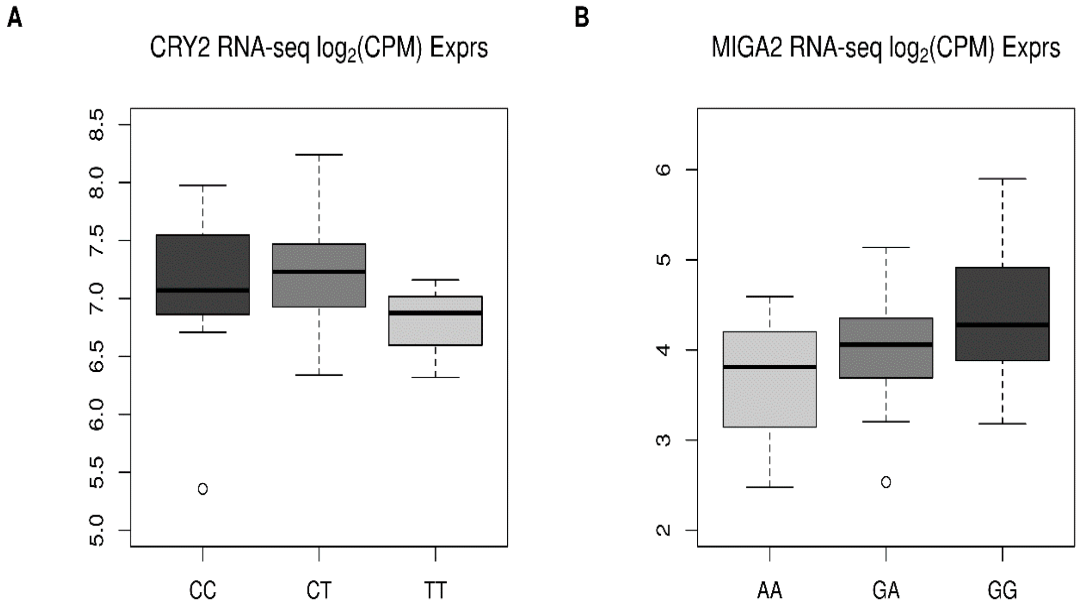


Figure 1: (**A**) Boxplots depicting the median and the distribution of *CRY2* mRNA $\log_2$ counts-per-million (CPM) expression levels in the *gluteus medius* skeletal muscle for each one of the three rs320439526 genotypes: CC (N = 11), CT (N = 30) and TT (N = 3). Homozygous TT animals for the alternative allele showed a reduced expression of *CRY2* compared with their homozygous CC and heterozygous CT counterparts, although not at a significant level (*P*-value = 0.392). (**B**) Boxplots depicting the median and the distribution of *MIGA2* mRNA $\log_2$ counts-per-million (CPM) expression levels in the *gluteus medius* skeletal muscle for each one of the three rs330779504 genotypes: GG (N = 24), GA (N = 18) and AA (N = 3). Homozygous AA animals for the alternative allele showed a reduced expression of *MIGA2* compared with their homozygous GA and heterozygous GG counterparts, although not at a significant level (*P*-value = 0.0953).

The inclusion of chromosome-wide genotyping information for the two polymorphic sites rs320439526 and rs330779504 SNPs at *CRY2* and *MIGA2*, respectively, obtained by González-Prendes et al. (2017), did not yield significant results. In other words, there were several SNPs in the chip that displayed associations with C18:0 fatty acid content in LD and LDL serum concentration that were more significant than the ones obtained for the rs320439526 and rs330779504 SNPs, respectively (Paper III, Figure 2). This outcome inclines us to think that these SNPs do not have causal effects. However, as discussed in Paper I, causal mutations not always display the most significant association with the trait under study (Schaid et al., 2018) and, moreover, QTL are not always produced by a solitary causal mutation but by the combined effects of multiple SNPs in close linkage disequilibrium.

Overall, it should not be concluded from our results that the variability of the circadian genes does not have effects on the lipid metabolism of pigs, but probably alleles with strong effects are very rare and can only be detected by sequencing or genotyping large populations. Indeed, some studies in humans have identified SNPs from core circadian regulating genes including *CLOCK*, *ARNTL*, and *CRY2*, being associated with lipid levels and metabolic syndrome (Scott et al., 2008; Englund et al., 2009; Garaulet et al., 2009; Sookoian et al., 2010; Tsuzaki et al., 2010; Garcia-Rios et al., 2012; Kovanen et al., 2015; Lin et al., 2017), so we consider that the potential consequences of the polymorphism of circadian genes on traits of genomic interest in pigs should be further explored. On the other hand, our results indicate that differential expression analysis can be used as a source of information to select candidate genes but only in combination with other criteria, being one of the most important ones the positional information generated in GWAS experiments.

4.4. About the amount and distribution of variation in porcine microRNA genes

One of the main goals of paper IV was to analyze the distribution of polymorphic sites within miRNA loci in four selected porcine populations of wild boars (WB) and domestic breeds (DM) from Asian (A) and European (E) origins. For this purpose, we made use of publicly available WGS data from a set of 120 pigs covering different breeds and geographical origins, as detailed in papers II and IV. We selected a set of *bona fide* miRNA loci annotated in the porcine genome (N = 370) to assess the presence of SNPs within their sequences, and a total of 285 variants segregating in at least one of the four defined porcine populations (i.e. ADM, AWB, EDM or EWB) were identified. A thorough analysis of the distribution of these SNPs within these populations evidenced a clear differentiation between pigs from Asian and European origins. This result was expectable because the analysis of genetic markers and of porcine WGS has evidenced that *Sus scrofa* emerged in Southeast Asia in the early Pliocene, 5.3–3.5 Myr ago (Larson et al., 2007, 2011; Frantz et al., 2013), and subsequently spread westwards and reached Europe 0.8 Mya (Frantz et al., 2013). Following this initial dispersal of wild boars across Eurasia, there was a long period of geographic isolation between the European and Asian gene pools, probably due to a colder climate that promoted the high genetic differentiation that we have been able to corroborate in our miRNA data set. Initial estimates obtained with mitochondrial data suggested that these two gene pools diverged 500,000 YBP, while more recent estimates inferred from WGS indicate a much older time of divergence, i.e. 1.6-0.8 Myr (Groenen et al., 2012). We have also detected (Paper IV, Figure 1) that Asian pigs and wild boars are much more diverse than their European counterparts, a finding probably motivated by the strong founder effect that the first wild boars colonizing Europe underwent as a consequence of a dispersal process initiated in Southeast Asia, a very distant location (Groenen et al., 2012). It should be also mentioned that, in the last centuries, the long-lasting geographic isolation of the European and Asian gene pools was disrupted by the transfer of pigs in both directions (Frantz et al., 2015). It is well known, for instance, that Chinese sows were massively imported into England during the 18th-19th centuries, and that as a result of this, many European breeds, and particularly Large White, carry Asian alleles at relatively high frequencies (Giuffra et al., 2000). The opposite is also true, so European breeds have been imported into China to improve the genetics of the national pig industry. In any case, the PCAs depicted in Figure 1 of paper IV agrees well with the general idea, backed up by the results of many studies (Ramos-Onsins et al., 2014), that the magnitude of these migrations was not strong enough to dilute the genetic differences produced by millennia of geographic isolation.

These genetic differences, although detectable, where less evident when comparing wild boars and domestic breeds, especially in animals from the European lineage. It is worth noting that, despite the limited number of analyzed miRNA loci (N = 370), the population differentiation between ADM and AWB pigs was still discernible, although less evident compared with other PCAs based on a much higher number of SNPs. These differences were also evidenced when analyzing SNPs at a whole-genome scale, as well as SNPs located in 3’-UTRs and in putative predicted 7mer-m8 and 8mer miRNA target sites. The current view is that the relatively weak differentiation between pigs and wild boars is mainly explained by the occurrence of recurrent hybridization events soon after domestication. Indeed, Frantz et al. (2015) demonstrated the existence of a post-domestication gene flow between pigs and wild boars, and they also evidenced that this gene flow was strongly asymmetrical, mainly following the direction that goes from wild boars to pigs. The most probable cause of this outcome is that, at least in Europe, pigs were not kept in sties, but instead they could roam freely, scavenging food, until the 18th century (White, 2011), thus providing a broad window of opportunity for the occurrence of hybridization events with their wild counterparts.

Moreover, when we analyzed the distribution of SNPs within miRNA loci, we detected that the majority of miRNA genes showed 1 or 2 polymorphic sites at most, and almost half of the variants presented reduced alternative allele frequencies (MAF < 0.1). Besides, the probability of finding a SNP in the seed region of miRNAs (the miRNA portion that ultimately determines its targeting affinity and regulatory effects) was reduced by half compared with the remaining of the mature miRNA sequence, and more than twice compared with miRNA precursor regions. Indeed, the average SNP density observed in the seeds (~0.42 SNPs per 100 bp) strongly contrasted with that observed in other non-miRNA regions of the genome (~2.5 SNPs per 100 bp), thus evidencing that mutations within the seed of miRNA genes have been strongly selected against. This is the likely signature of the ongoing purifying selection in regions that are crucial to ensure the binding affinity of the miRNA. These results were in accordance with other similar previous reports analyzing miRNA variability in diverse vertebrate species (Saunders et al., 2007; Sun et al., 2009; Gong et al., 2012; Zorc et al., 2012; Omariba et al., 2020). Accordingly, provided that microRNA loci seem to evolve under selective constraints, population differentiation observed at these sites should be relatively low. Such hypothesis was reinforced by the scarce F_{ST} significance observed for SNPs located at miRNA loci among the four porcine populations. In fact, only two SNPs within ssc-miR-4335 and ssc-miR-9835 showed significant F_{ST} estimates when contrasting allele frequencies between ADM and AWB pigs, and between EWB and AWB pigs, respectively (Paper IV, Additional Table S4).

The existence of a strong purifying selection removing mutations, mostly in the seed region (2nd to 8th 5’ nucleotides in the mature miRNA), is somewhat expected given the importance of preserving the seed integrity to ensure a correct targeting of the mRNAs regulated by a given miRNA. Gains and losses of target sites by means of seed modifications have the potential to completely disrupt entire gene regulatory networks, a reasoning supported by the involvement of miRNAs as central players in dysregulated pathways and aberrant gene expression patterns observed in cancer (Ryan et al., 2010; Wilk and Braun, 2018). It is worth mentioning that polymorphisms in the miRNA binding sites of 3'-UTRs, rather than in the sequence of the miRNA itself, seem to account for the majority of observed gains and losses of miRNA targets (Marco, 2015). That is, the emergence of novel regulatory networks is biased towards the variability observed in 3'-UTRs, compared with the highly conserved miRNA loci. Indeed, while many miRNA families seem to have remained almost unaltered during millions of years across species, it is at their target binding sites that evolutionary differences seem to be more pervasive (Hui et al., 2013). Nevertheless, strong purifying selection has also been reported for miRNA target gains, as the appearance of novel regulatory effects of miRNAs should be mainly deleterious (Chen and Rajewsky, 2006; Saunders et al., 2007; Hatlen and Marco, 2020). In contrast, population differentiation via positive selection has also been described for target sites (Gardner and Vinther, 2008; Li et al., 2012), but this phenomenon seems to be linked to the loss of miRNA binding sites where the functional target site is the ancestral allele, and losses in a given miRNA-mRNA might be neutral or even beneficial in some cases (Helmy et al., 2019; Hatlen and Marco, 2020).

To the best of our knowledge, these effects have not been thoroughly studied in the porcine species. Although partially envisaged in paper IV, in which we demonstrate the low variability of porcine miRNA genes, further studies analyzing miRNA target binding sites could provide novel hints about the genetic divergence across porcine populations in miRNA-related regulatory pathways.

Of particular interest is our observation that, even outside the seed, the rate of polymorphic sites across the mature miRNA sequence is not uniform. In fact, we found a high heterogeneity in the SNP density across the mature miRNA, with well-defined intervals that matched regions with potential functional properties in the miRNA binding affinity. In the first 5' nucleotide of the miRNA, we found a SNP density similar to that of the seed (~0.49 SNPs per 100 bp). This nucleotide is important as it provides an anchor to the miRNA for correctly attaching to the Argonaute protein in the miRISC complex. The four nucleotides (9th to 12th) immediately following the seed gathered an increased SNP density (~0.98 SNPs per 100 bp), which was in accordance with their probable lack of involvement in the miRNA binding process (Bartel, 2018). Following nucleotides 13th to 18th experienced, once again, a decrease in their observed SNP density, more prominent towards nucleotides 16th and 17th (Paper IV, Figure 3). This finding is consistent with the existence of a supplementary pairing in the miRNA that stabilizes the binding to the 3'-UTR of target mRNAs (Bartel, 2018). The remaining nucleotides until the end of the mature miRNA displayed elevated SNP densities, similar to those observed for the 9th to 12th nt interval. These results were in accordance to those previously reported by Gong et al. (2012), who described a non-uniform distribution of SNPs along miRNA sequences. In summary the observed pattern of SNP density variation in pig miRNA genes is consistent with the functional significance of the regions under analysis (Bartel, 2018).

4.5. The variability of miRNA genes is associated with the mRNA expression of their predicted targets

A total of 15 SNPs located in miRNA genes were genotyped in the Lipgen population (N = 350), and association analyses were carried out using microarray expression profiles in GM and liver tissues from a set of animals belonging to the Lipgen population. Our results highlighted several significant associations between SNPs in the seed (rs322514450, n.16G>A), and also outside the seed (rs333787816, n.65T>C; rs319154814, n.46G>A; rs335924546, n.72C>T), with the expression of their putative mRNA targets. From these, the rs322514450 SNP located in the seed of ssc-miR-9792-5p was significantly associated with the mRNA expression of the *NUDT6* and *RLBP1* genes in GM and liver tissues, respectively. The targeting of the 3'-UTR of both genes took place when the A allele was present in the seed of ssc-miR-9792-5p, hence their expression might be lowered in pigs homozygous for the mutated seed of ssc-miR-9792-5p. Indeed, individuals with AA genotypes for this SNP showed reduced expression profiles of *NUDT6* and *RLBP1* transcripts with respect to their GG and GA counterparts, as depicted in Figure 6 of paper IV. The significance of such expression variation was better exemplified when we contrasted the mean mRNA expression of each genotype measured by microarray hybridization (our unpublished data, Table 1). However, the fact that only two putative gained mRNA targets for the mutated seed of ssc-miR-9792-5p showed significant, yet not extremely dramatic changes in their expression, evidences the probable existence of compensatory mechanisms buffering the effects of potentially deleterious SNPs located in the seed region, such as functional redundancy of genes and miRNAs (Ventura et al., 2008; Alvarez-Saavedra and Horvitz, 2010; Park et al., 2010; Concepcion et al., 2012) or compensatory pairing mechanisms, and/or imperfect seed matching (Chipman and Pasquinelli, 2019). Indeed, the knockdown of miRNAs not always implies neither substantial changes in the expression of its target mRNAs nor readily observable phenotypic effects (Park et al., 2010; Gong et al., 2012).

Interestingly, the SNP showing the most consistent associations with gene expression, i.e. rs319154814 (n.46G>A) in the apical loop of ssc-miR-326, was located outside the seed (Table 1). These changes in gene expression involving miRNAs not polymorphic in their seed, but in other parts of their sequences, shed light on the existence of potential mechanisms related with the maturation process, that could affect the function of porcine miRNAs without altering their binding properties. In this way, these SNPs would impact on the expression profiles of the miRNAs, thus leading to changes in the expression levels of their mRNA targets. Polymorphisms located at the apical loop of miRNA hairpins like the one reported by us in paper IV or elsewhere (Fernandez et al., 2017), and also variants in the basal junction (Li et al., 2020; Nguyen et al., 2020) or in the middle of the hairpin stem (Omariba et al., 2020), have the potential to modify the pairing process during the folding of the hairpin, thus stabilizing or destabilizing the miRNA hairpin (Omariba et al., 2020). Moreover, these SNPs are able to affect crucial determinants for the processing machinery, hence hampering or directly abrogating the expression of the miRNA (Li et al., 2020; Nguyen et al., 2020).

Table 1: Results of significance tests for differences in normalized microarray expression profiles of mRNAs putatively regulated by ssc-miR-9792 and ssc-miR-326.[a]

SNP	Type	Gene	ANOVA *P*-value	Tukey's HSD		
				GA vs AA	GG vs AA	GG vs GA
rs322514450 (8:110922752)	ssc-miR-9792 seed region (G/A)	*NUDT6*	3.670E-05	9.918E-01	4.544E-02	6.990E-05
		RLBP1	3.200E-04	9.760E-01	7.301E-02	5.885E-04
rs319154814 (9:9581989)	ssc-miR-326 apical loop (G/A)	*CFLAR*	3.120E-03	1.850E-02	6.389E-03	6.177E-01
		PPP1CC	8.060E-03	1.512E-01	6.465E-03	6.177E-01
		FSTL1	9.610E-03	3.161E-01	6.687E-03	1.167E-01
		SF3A3	1.930E-03	8.079E-03	6.475E-03	7.689E-01
		ELAVL1	1.340E-02	3.993E-02	2.824E-02	8.003E-01
		NAA50	1.600E-02	1.294E-01	1.580E-02	4.052E-01

[a] Expression values for each genotype of rs322514450 (GG = 56 in GM and 58 in liver, GA = 25, AA = 5) and rs319154814 (GG = 17, GA = 37, AA = 32) were contrasted applying an ANOVA test and further stratified between groups with a Tukey's range test for honestly significant differences (HSD).

As we did for data reported in paper III, in this global discussion we have investigated whether the associations between miRNA SNPs and mRNA levels measured with microarrays could be reproduced using RNA-seq data generated by Cardoso et al. (2017a) in the GM skeletal muscle of 52 Duroc pigs from the Lipgen population. It is important to emphasize that only the associations between rs333787816 (n.65T>C) in ssc-miR-23a and *NUP50, PAFAH1B2, CSNK1G3, UBE2R2, AGO1* and *AGO2* transcript levels, as well as that between rs322514450 (n.16G>A) in ssc-miR-9792 and *NUDT6*, were detected in the GM muscle. In contrast, the remaining significant associations were only valid for transcripts expressed in the liver tissue. When using GM RNA-seq expression phenotypes, no significant associations were observed for any miRNA genotype (Table 2). For those genes that were detected as significantly associated with miRNA SNPs in the liver tissue, these results were expected, as miRNA expression signatures are independent for each tissue (Ludwig et al., 2016). Indeed, they should be fairly different between the GM skeletal muscle and the liver, two tissues that greatly differ in their metabolic functions and gene expression profiles (González-Prendes et al., 2019a). In this way, co-expression associations for miRNA-mRNA interactions should also reproduce any tissue and/or developmental stage specificity observed for the expression patterns of either miRNA or mRNA transcripts (Nowakowski et al., 2018).

As previously commented in the section 4.3 of the present general discussion, the limited number of available samples with GM RNA-seq profiles in the Lipgen population (N = 52) could be hampering the statistical power to detect any meaningful differences. Moreover, contrary to what was shown for *CRY2* and *MIGA2* RNA-seq profiles, no tendency towards a reduced expression for homozygous animals carrying the mutated alleles was observed when comparing ssc-miR-23a and ssc-miR-9792 miRNA genotypes (data not shown). To what extent this would be attributable to the reduced number of available samples or to inherent differences in the inference of transcript abundances between microarrays and RNA-seq approaches, remains unclear. Nevertheless, it should be noticed that, contrary to *cis*-expression effects observed for *CRY2* and *MIGA2* polymorphisms investigated in paper III, expression differences in mRNAs described in paper IV are dependent on *trans*- effects elicited by SNPs mapping to miRNAs that putatively target these transcripts. Contrary to the direct *cis*- effects reported in paper III, confounding variables such as the expression level of the miRNAs, their subcellular location or their loading into active miRISC complexes to exert functional interactions with their mRNA targets, can cause important distortions in the magnitude and significance of the associations detected between miRNA genotypes and the mRNA levels of the corresponding miRNA targets. Moreover, our inference of putative mRNA targets for the analyzed miRNAs relied on *in silico* predictions based on the detection of 7mer-m8 binding sites between the miRNA seeds and short sequences in the 3'-UTRs of target mRNAs, as well as on experimental validations reported in humans and available in public databases (Karagkouni et al., 2018). Such approach did not take into account neither the existence of alternative seed matching types for miRNA-mRNA interactions nor the possibility that experimental results validated in humans might not be directly reproducible in pigs.

Table 2: Results of significance tests for differences in $\log_2$ counts-per-million (CPM) RNA-seq expression profiles of mRNA targets putatively regulated by ssc-miR-23a, ssc-miR-9792, ssc-miR-326 and ssc-miR-1224.[a]

SNP	Type	Tissue	Gene	***P*-value**
rs333787816 (2:65308181)	ssc-miR-23a precursor stem (T/C)	GM	*NUP50*	6.40E-01
			PAFAH1B2	7.78E-01
			CSNK1G3	9.83E-01
			UBE2R2	5.91E-01
			AGO1	5.68E-01
			AGO2	5.04E-01
rs322514450 (8:110922752)	ssc-miR-9792 seed region (G/A)	GM	*NUDT6*	1.23E-01
rs319154814 (9:9581989)	ssc-miR-326 apical loop (G/A)	LIVER	*CFLAR*	9.40E-01
			PPP1CC	8.96E-01
			SF3A3	3.94E-01
			FSTL1	7.31E-01
			CELF1	3.32E-01
			NAA50	7.08E-01
			ELAVL1	2.86E-01
rs335924546 (13:122141078)	ssc-miR-1224 precursor stem (C/T)	LIVER	*MKRN1*	8.25E-01

[a]Expression values for rs333787816 (TT = 10, TC = 28, CC = 8), rs322514450 (GG = 31, GA = 11, AA = 3), rs319154814 (GG = 8, GA = 17, AA = 20) and rs335924546 (CC = 31, CT = 14, TT = 0) genotypes were contrasted applying an ANOVA test.

4.6. Polymorphisms within porcine microRNA genes are associated with lipid-related traits

Another goal that we aimed to achieve in paper IV was to analyze the association of the genotyped miRNA SNPs with the phenotypic variation of lipid-related traits measured in the Lipgen population. In this regard, most of the associations that remained significant after multiple testing correction involved the rs319154814 SNP (n.46G>A) in the apical loop of ssc-miR-326. Indeed, this polymorphism was associated with myristic acid content in both LD and GM muscles, as well as with the gadoleic acid content and the ratio between PUFA and MUFA in the LD muscle. Other several significant associations at the nominal level for this and other miRNA SNPs were also detected (Paper IV, Table 3). Although no previous surveys have been conducted investigating if ssc-miR-326 is involved in the regulation of lipid metabolism, it is well known that miRNAs can act as key regulators of lipogenesis, lipolysis, cholesterol metabolism, and many other lipid-related pathways. For instance, miR-122 is one of the most expressed miRNAs in the liver, and its downregulation has been associated with a marked decrease in total serum cholesterol and triglyceride levels (Esau et al., 2006; Elmén et al., 2007). Besides, this miRNA is also transcribed in a circadian fashion (Gatfield et al., 2009), being involved in the circadian regulation of its mRNA targets (Kojima et al., 2010). The miRNA-33 family can directly target *ABCA1*, a key regulator of cholesterol metabolism, resulting in a decreased cholesterol efflux and lowered nascent HDL levels (Marquart et al., 2010; Najafi-Shoushtari et al., 2010; Rayner et al., 2010). These two outcomes are anticipated to have protective effects against atherosclerosis progression (Horie et al., 2012). Taniguchi et al. (2014) also reported ssc-miR-33b as a key player in lipogenesis in the porcine adipose tissue. Mice lacking miR-33b and fed with a high-fat diet during a long period of time, developed obesity and liver steatosis, and they also displayed an increased expression of two miR-33b targets, i.e. *FASN* and *ACACA*, two essential players in lipogenesis (Horie et al., 2013; Goedeke et al., 2014). The miR-148a, which was reported as DE in paper VI, can target the LDL receptor transcript (*LDLR*), hence contributing to increase the plasma concentration of LDL (Goedeke et al., 2015; Rotllan et al., 2016). Moreover, miR-21 overexpression blocks intracellular lipid accumulation by targeting the fatty acid-binding protein 7 (*FABP7*) in high-fat diet-fed mice (Ahn et al., 2012) and miR-378 is directly regulated by the lipogenic factors CCAAT/enhancer-binding protein α and β (*CEBPA* and *CEBPB*) (Gerin et al., 2010; John et al., 2012).

The number of associations that we have found between miRNA SNPs and lipid phenotypes are relatively low. The most probable reason is the functional redundancy that exists for many miRNAs, in which even the consequences of a complete knockout are not readily evident because loss of function can be compensated by other highly related miRNAs with similar binding properties. For instance, functional overlaps have been described for the miR-34a/b/c and miR-449a/b/c loci (Concepcion et al., 2012), as well as for miR-17~92 family and miR-106a/b (Ventura et al., 2008), where partial losses of any but not all of the miRNAs of the family are compensated by the expression of other existing miRNA paralogs (Alvarez-Saavedra and Horvitz, 2010). Another important reason for this lack of association could be purifying selection, that tends to remove potentially deleterious mutations altering miRNA function. In this context, such harmful mutations would be extremely rare and unlikely to be detected in a population such as the Lipgen, which has a modest sample size (N = 350). Finally, the genetic determinism of lipid-related traits is very complex and depends on many genetic determinants with variable and sometimes opposed effects on the trait under study, thus providing an additional mechanism of genetic compensation that obscures the consequences of miRNA variability. For instance, a mutation in the seed of a miRNA could be counteracted by another suppressor mutation in the binding site of the 3’-UTR of one of its targets, thus restoring the affinity between both molecules. Our impression is that the most reliable mechanism to understand the functional consequences of miRNAs on complex phenotypes is exploring their relationship with intermediate and simpler expression phenotypes, which depend on many fewer variables and can be dissected more easily. Of course, this involves not only the performance of association analyses but of functional tests to confidently determine the sets of mRNAs targeted by specific miRNAs. We foresee that the exploration of the functions of non-coding RNAs will be one of the most active fields of research aiming to elucidate the biological basis and molecular physiology of production traits in livestock.

4.7. Development of the eMIRNA pipeline to annotate porcine microRNA genes

As previously outlined, the miRNA annotation in pigs is still incomplete if compared with that of cows and chicken, or with the annotations of well characterized organisms such as humans and mice (Introduction, Figure 10).

Such circumstance motivated us to further investigate the existence of miRNA genes not yet annotated in the current porcine assembly. In principle, a plethora of tools for the homology-based search of miRNAs and their identification from small transcriptome data are available (Bortolomeazzi et al., 2017). Many different approaches have been implemented, ranging from rule-based algorithms using a series of heuristics for calculating the odds of a sequence to be a true miRNA (Friedländer et al., 2008, 2012; Mathelier and Carbone, 2010; Barturen et al., 2014; Aparicio-Puerta et al., 2019), to more advanced classifier algorithms relying on machine learning techniques, and further discussed in Stegmayer et al. (2018). Nevertheless, the majority of them do not implement end-to-end pipelines for the discovery and functional annotation of miRNAs, but are mostly focused on maximizing their predictive performance compared with other tools. In this way, hairpin reconstruction from mature miRNA sequences (one of the most critical steps in identifying novel miRNA genes) is often overlooked. Moreover, pre-miRNA boundaries are commonly defined by flanking sequence motifs around the miRNA gene (Auyeung et al., 2013; Fang and Bartel, 2015), a source of information that could be used for better delineating the hairpin sequences to be used as candidates for prediction.

In paper V, we aimed to tackle this issue by using a motif-informed ranking approach based on the randfold algorithm (Bonnet et al., 2004). Briefly, we analyzed the positional occurrence of miRNA flanking motifs in a set of annotated *bona fide* porcine miRNAs (N = 370), determining their most common location (e.g. -13/-12 for basal UG upstream motif, and +18/+21 for downstream CNNC motif), which coincided with previous surveys characterizing such motifs in human miRNAs (Auyeung et al., 2013; Fang and Bartel, 2015). Candidate sequence reconstruction from aligned mature miRNAs were generated using several elongation patterns (i.e. 15/60, 30/60, 15/70, 30/70, 15/80 and 30/80 nucleotides were added upstream and downstream, respectively). Motif positional information was then incorporated to generate motif-corrected candidate sequences. In this way, we were able to generate 12 putative candidate sequences per each aligned mature miRNA. Their folding thermodynamic stability was assessed with the randfold algorithm (Bonnet et al., 2004), and those showing a more stable secondary structure of the hairpin were selected to be interrogated by the classifier algorithm.

As detailed in Figure 1 from paper V, and further described in the github repository for the eMIRNA pipeline (https://github.com/emarmolsanchez/eMIRNA), the *eMIRNA.Hunter* module was explicitly designed for aligning mature miRNAs to any given genome assembly and, subsequently, performing the reconstruction of the sequence candidate by (i) taking advantage of user-defined elongation patterns, and (ii) simultaneously adjusting the boundaries according to flanking motifs, if present. Once candidate elongated sequences are generated, the *eMIRNA.Structural.Pscore* module can be optionally used for assessing the stability of the folding, thus filtering those sequences that do not surpass a given threshold (i.e. $p < 0.05$). This implementation relies on a user-guided decision on which elongation patterns need to be applied to the data, as well as on the optional use of p scores calculation for removing structurally unstable candidate sequences. Users could implement a similar procedure as reported in paper V, where we tested several elongation patterns and selected those showing a more stable folding. Another option would be to apply a fixed elongation pattern (i.e. 15-30/60-70) and further remove all the candidates that do not show a p score < 0.05.

However, such approach assumes the informed decision of the users to choose which elongation patterns to test, as well as whether to implement candidate sequence filtering based on structural stability or bypassing such step and continue with further analyses.

A better approach to this task would be the development of an automated dynamic search for proper elongation and motif search correction, which would directly generate the better possible candidate sequence for each mature miRNA. This could be achieved in the same line to what has been reported by Evers et al. (2015) and Paicu et al. (2017). In this regard, a sufficiently long window around the mature miRNA under study should be defined (i.e. 80-100 nts). Following steps would include the generation of multiple secondary folding hairpin candidates within the defined window, by using, for instance, the RNALfold algorithm (Lorenz et al., 2011). Subsequently, the most stable hairpin candidate would be prioritized by applying any given thresholding rule such as that provided by the randfold algorithm, or by using the minimum free energy of the folding or both. With this approach, an easier-to-use implementation for generating candidate sequences to be tested could be achieved.

For miRNA prediction, we made use of a machine learning semi-supervised transductive graph-based approach, as reported by Yones et al. (2018). Semi-supervised schemes have the advantage to allow the inclusion of unlabeled data to the training process, hence increasing the amount of data used for training the model, which, in the case of miRNAs, is a very limiting factor, given the scarce number of annotated miRNAs that are often available to be used as a reference. To the best of our knowledge, only one pipeline, apart from the one published by Yones and collaborators, is currently available to perform such task (Sheikh Hassani and Green, 2019). This case exemplifies the still limited exploration of semi-supervised frameworks for miRNA prediction. Furthermore, it is worth noting that, despite the wide range of ML algorithms that have been employed for miRNA prediction throughout the years, with SVM and RF being the most used ones (Introduction, Table 3), good performances are often reported when evaluating the ability of the classifiers to correctly discern between miRNAs and other sequences. Indeed, benchmarking of several other ML algorithms showed acceptable performances for correctly identifying annotated porcine miRNAs (Paper V, Table 2), with the lGBM algorithm yielding performance metrics equivalent to those achieved by our semi-supervised approach. In the light of these results, we are inclined to think that, other than the training algorithm, the correct selection and diversity of training sequences and their defined features are the key factors affecting the real performance of miRNA prediction tools. It is therefore of paramount importance that researchers focus their efforts on correctly determining sequence candidates with adjusted boundaries, jointly with a careful estimation of their defining features. At the same time, it is highly advisable to incorporate as much sequences as possible to the training process.

Moreover, as an extension of the discussion of paper V, we would like to emphasize that the annotation nomenclature of novel detected miRNAs and their isoforms (isomiRs) should be carefully considered, taking into account recent unified systems for miRNA classification as the ones proposed by Fromm et al. (2015) or Desvignes et al. (2019).

The functional annotation of novel and annotated miRNAs is also a relevant task, which, once again, is not commonly covered by many of the reported tools for miRNA prediction. On the contrary, this task is usually carried out by integrated tools for miRNA bioinformatic analyses such as the UEA sRNA workbench (Beckers et al., 2017) or the sRNAtoolbox (Aparicio-Puerta et al., 2019). More importantly, the inference of miRNA-mRNA interaction networks by making use of expression data and miRNA targeting properties is also a very useful approach in order to discern functional interactions between miRNAs and their mRNA targets. To do so, several methods have been proposed, which commonly start from a set of expression data for miRNA and mRNA genes that are used to infer co-expression patterns among miRNAs and their putative mRNA targets. In order to exclude spurious correlations, prior miRNA targeting rules are often embedded, as well as experimental information about true miRNA-mRNA interactions. One of the most reported approaches aiming at integrating miRNA and mRNA expression data sets through gene regulatory networks is the MAGIA tool (Sales et al., 2010), and its updated version MAGIA^2 (Bisognin et al., 2012). This web-based software incorporates *in silico* and experimental miRNA target predictions based on MicroCosm (Griffiths-Jones et al., 2008), miRanda (John et al., 2004), DIANA-microT (Maragkakis et al., 2009; Paraskevopoulou et al., 2013), miRDB (Wang, 2008; Liu and Wang, 2019), PicTar (Krek et al., 2005), PITA (Kertesz et al., 2007), rna22 (Miranda et al., 2006), and TargetScan (Agarwal et al., 2015, 2018) as prior knowledge. Then, co-expression modules between miRNA and mRNA expression data are inferred based on correlation metrics in order to generate informed miRNA-mRNA regulatory interaction networks. Other more recent tools are also available, such as micrographite (Calura et al., 2014), ToppMiR (Wu et al., 2014a), DIANA-mirEXTra v2.0 (Vlachos et al., 2016), spidermiR (Cava et al., 2017) or miRmapper (da Silveira et al., 2018).

We consider that integrating co-expression miRNA-mRNA networks with miRNA prediction and target mRNAs identification would allow researchers to infer putative functional relationships emerging from their predicted novel miRNA candidates, hence highlighting hidden regulatory relationships that would have been overlooked otherwise.

In order to implement such approach, we have developed three additional modules for the eMIRNA pipeline, which can be found in the dedicated github repository (https://github.com/emarmolsanchez/eMIRNA). A detailed scheme is also shown in Figure 2. First, similarly to what was implemented in papers IV and VI for miRNA target prediction, we made use of the SeqKit Toolkit (Shen et al., 2016) for searching miRNA binding sites in 3'-UTR sequences of putative target mRNAs. The *eMIRNA.Target* module was developed for such purpose. It accepts two FASTA files, one with mature miRNA sequences, and the other one encompassing 3’-UTR sequences of mRNAs to screen for binding sites. This module allows the user to choose between four different types of miRNA target sites (i.e. 6mer, 7mer-A1, 7mer-m8 or 8mer binding types). After running the module, the identified putative miRNA-mRNA interactions can be used to feed the following module, *eMIRNA.Network*. In order to detect meaningful co-expression regulatory networks, this module makes use of miRNA and mRNA expression datasets, jointly with the prior knowledge of miRNA-mRNA interactions inferred with the *eMIRNA.Target*, as well as first-order partial correlation coefficients deduced with the PCIT algorithm (Reverter and Chan, 2008; Watson-Haigh et al., 2010).

Finally, the *eMIRNA.RIF* module is able to identify key regulatory miRNAs from the set of expressed miRNAs by means of the RIF algorithm (Reverter et al., 2010). The miRNA and mRNA expression data sets, along with meaningful miRNA-mRNA interactions from *eMIRNA.Network* and, optionally, a list of differentially expressed mRNAs, are needed for running the RIF algorithm.

Overall, the collection of bioinformatic modules described in paper V and discussed herewith, allowed us to identify a total of 47 putative novel porcine miRNAs. Besides, 20 of them were detected as expressed in a small RNA-seq data set from the GM skeletal muscle of Duroc gilts. The remaining ones were inferred by an homology-based approach using orthologous annotated human mature miRNAs. Furthermore, three of them were successfully profiled as expressed in the LD skeletal muscle and liver tissues from an independent population of Göttingen minipigs. Our pipeline demonstrated good performance compared with other state-of-the-art algorithms, and showed an improved ability to recover the miRNA boundaries with regard to miRDeep2, another widely used tool for miRNA prediction (Friedländer et al., 2012). In the light of these results, we believe that the eMIRNA pipeline reported in paper V, jointly with further ongoing developments such as those mentioned above, constitutes a useful tool for the discovery and functional annotation of novel miRNAs in any given species with an available genome assembly. We also think that the development of this and other tools will be essential to improve the annotation of miRNAs in domestic species, a step that is crucial to understand their role in the determinism of traits of economic importance.

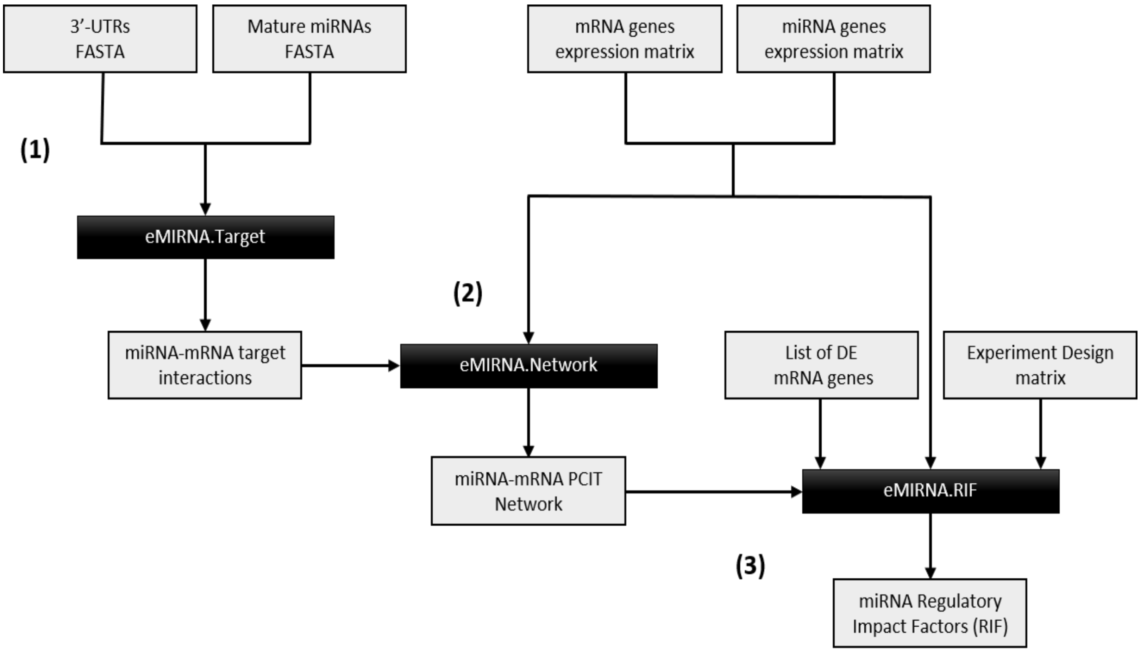


Figure 2: eMIRNA pipeline scheme for assessing meaningful miRNA-mRNA interactions and key regulatory miRNAs from expression data sets. **(1)** mature miRNA and 3'-UTR sequences are used for predicting miRNA-mRNA target interactions. **(2)** Expression data from miRNA and mRNA genes belonging to the same experimental conditions are used to predict meaningful miRNA-mRNA interactions based on a partial correlations and information theory (PCIT) approach. **(3)** The regulatory impact factor (RIF) of each considered miRNA is calculated based on mRNA and miRNA expression data, as well as with meaningful PCIT interactions.

4.8. Analyzing the changes in the expression mean and variance of genes transcribed in the skeletal muscle of fasting and fed gilts

In paper VI, we first compared the expression profiles of protein coding mRNAs, miRNAs and lincRNAs, making use of RNA-seq data previously reported by Cardoso et al. (2017b) and small RNA-seq generated within the framework of the present Ph.D. thesis, using the same Duroc pig population of 36 gilts subjected to fasting-feeding experimental conditions.

We modeled the amount of expression of each of these three types of transcripts by means of their regularized $\log_2$ normalized expression, which showed that, overall, mRNA transcripts were more expressed than miRNAs and lincRNAs. With few exceptions, lincRNAs had low or very low expression levels in the pig skeletal muscle (Paper IV, Figure 2). Such results coincided with the fact that lincRNA median expression is only about a tenth than that of protein coding mRNAs (Cabili et al., 2011; Ulitsky and Bartel, 2013), although both transcript types have similar half-life distributions (Clark et al., 2012; Tani et al., 2012).

In addition, we aimed to analyze the inherent intragroup expression variability observed for each gene, i.e. we computed the biological coefficient of variation (BCV) for each annotated mRNA, miRNA and lincRNA in each of the defined dietary groups (T0, T1 and T2), in order to obtain an estimate of how variable each transcript would be across biological replicates. In this way, we observed that, on average, lowly expressed mRNAs and lincRNAs showed increased BCV values, meaning that lowly expressed genes presented increased variance in their expression profiles when compared to loci expressed at higher rates (Paper VI, Figure 3), and this trend was particularly true for mRNAs and lincRNAs. In contrast, miRNAs showed very stable and resilient expression profiles across samples, irrespective of their expression values, with few exceptions showing mildly increased variances. Moreover, we computed the fold changes not only for the mean expression of genes but also for the variances associated with these means in each experimental condition. By doing so, we observed that mRNA genes with low expression profiles tend to display higher fold changes of their expression variances that genes which are highly expressed. This pattern was not observed in the case of miRNA or lincRNA genes (Paper VI, Figure 4).

These results were in accordance with previous surveys characterizing the rate of production, accumulation and decay of mRNAs (Carninci et al., 2005; Söllner et al., 2017), lincRNAs (Cabili et al., 2011; Iyer et al., 2015) and miRNAs (de Rie et al., 2017). Indeed, the overall stable expression profile observed for miRNA transcripts, probably reflects the particular decoupling between synthesis and functional activity that small RNAs often experience (Mayya and Duchaine, 2015; Reichholf et al., 2019).

The existence of intrinsic variation in gene expression and the transcriptional noise produced by such phenomenon has been reported since the beginning of the characterization of transcriptional profiles (Tu et al., 2002), but their biological meaning remained obscure until high-throughput sequencing techniques allowed researchers to investigate the transcriptional landscape of multiple tissues, developmental stages and biological conditions (Suntsova et al., 2019). In this regard, it is important to discern between true gene expression variation and other possible sources of variation that might be introduced throughout the generation of the data, either at collecting samples, or in subsequent steps while performing library preparation, sequencing, and further processing of the data. Biological variation, in essence, reflects not only the noisy nature of cell dynamics across tissues, but also to which extent the individuals comprised in a group respond differently to the same stimulus. Indeed, when we observe a change in the mean mRNA levels of a given gene in response to an experimental condition, very often this fold change of the mean does not necessarily have an identical magnitude (or even direction) in all the members of the experimental group. Thus, changes in the variance of the expression of genes capture a source of information that cannot be inferred from a simple comparison of means, paving the way to understand why individuals subjected to the same experimental condition respond, sometimes, in vastly different manners. Of course, the main challenge is to model and treat the many sources of technical variation to minimize their influence in the data set and to correctly estimate the amount of biological variation. The so-called BCV metric was reported to successfully account for this purpose (McCarthy et al., 2012), and was hence used in paper VI for estimating the extent of gene variation across samples, while controlling for technical sources of background noise.

Several authors have intended to explore the biological meaning of the variance of gene expression (also termed as gene dispersion). In fact, the stochasticity of gene expression and its inverse correlation with transcription and translation rates is well known to have a relevant impact in the cell metabolism (Raj and van Oudenaarden, 2008). Moreover, how intrinsic and extrinsic gene stochasticity might affect the gene-to-gene interactome rewiring has been also subjected to detailed analysis (Chalancon et al., 2012). For instance, Komurov and Ram (2010) described how the extent of transcriptional activation and expression levels was associated to the hierarchy of signaling cascades and the centrality or peripheral positioning of genes in regulatory networks. Highly expressed and lowly variable genes with active transcription tend to be central regulators of cellular metabolism such as energy homeostasis, transcription and processing rates or protein synthesis and degradation. In contrast, lowly expressed genes with high variability often correspond to extracellular effectors and regulatory proteins with peripheral and sparser locations within the regulatory networks

(Komurov and Ram, 2010). Kaneko et al. (2011) extended the notion of gene expression variability, due to intrinsic noise in transcription rate, to dispersion effects caused by mutations altering the expression of genes and their regulatory networks. Both types of noise (due to gene expression dynamics and to mutations altering transcriptional regulation) had significant effects on the observed phenotypic variability and the plasticity of gene-to-gene interplay, thus evolving in response to environmental forces. Age-dependent patterns of gene expression noise have been also reported, affecting phenotypes such as redox homeostasis or fatty acids metabolism. These patterns are sometimes produced by age-related diseases caused by epigenetic modifications under environmental aggressions rather than by a genetically-driven decline in regulatory functions after genome damage caused by age (Viñuela et al., 2017).

Nevertheless, it is worth commenting that, given the fact that high dispersion estimates are commonly found in lowly expressed genes, such is the case of lincRNAs and protein-coding regulatory elements like transcription factors (Vaquerizas et al., 2009), the modelling of gene dispersion needs to be adjusted for the zeroes inflation that arises from such expression patterns. This issue makes canonical models for RNA-seq data based on negative binomial distributions particularly prone to underestimate the true signals of differential expression between conditions, and, unfortunately, few methods have been proposed to tackle the zero-inflated problem (Ran and Daye, 2017; Li et al., 2019b).

In paper VI, we applied the *MDseq* method proposed by Ran and Daye (2017) in order to infer differential dispersion signals between fasting (T0) and fed (T1 and T2) pigs at the mRNA, miRNA and lincRNA levels. Few results compared to those obtained with canonical differential expression analyses based on the mean were obtained (Paper VI, Table 2 and Additional Table S3), leading to the conclusion that feeding did not produce dramatic changes on the variance of the expression of genes. Among the differentially dispersed mRNAs, to mention a few, *NEU3* was detected as overdispersed and also overexpressed after food intake in both T0/T1 and T0/T2 contrasts. This protein is located in the plasma membrane and stimulates insulin sensitivity and glucose tolerance (Yoshizumi et al., 2007). In contrast, the fold change of the variance of the *PDK4* gene was significantly reduced in fed gilts (T2) compared with their fasting counterparts (T0), in agreement with the marked downregulation of its expression in T2 animals (Paper VI, Additional Table S2). About miRNAs, ssc-miR-17-5p and ssc-miR-451 showed reduced dispersion values in fed pigs (T1) with respect to fasting individuals (T1), hence indicating stabilized expression profiles upon feeding. However, no significant downregulation was observed for their mean expression values in the same contrast, but a trend to reduced expression was indeed detected (Paper VI, Table 3). The ssc-miR-17-5p molecule is able to target several key transcripts regulating lipid metabolism such as *FABP4* or *PPARG* (Han et al., 2017), whereas ssc-miR-451 is downregulated in pigs with increased fatness (Xing et al., 2019). Although two lincRNAs (ENSSSCG00000032301 and ENSSSCG00000031192) showed slight significant differences in their dispersion profiles, no putative functional meaning could be deduced due to the lack of biological information.

Taking into account these results, several points should be further discussed here:

First, for miRNAs and particularly for lincRNAs, the reduced number of investigated transcripts (286 expressed miRNAs and 352 annotated lincRNAs) would definitely limit our ability to find significant dispersion signals. In recent genome annotations for the pig assembly, the number of annotated lincRNAs increased substantially, so further re-analyzing our muscle expression data might provide a better representation of the long non-coding RNA fraction of the transcriptome than what was reported in paper VI. Despite this, the limited annotation of such non-coding regions might still hamper the detection of significant differentially dispersed transcripts. Besides, the possibility that the applied *MDseq* method might still be too conservative for detecting the whole landscape of gene dispersion differences should not be discarded, as well as technical distortions produced throughout the pre-processing, normalization and outlier correction procedures that were applied to the expression data sets.

On the other hand, the effects of gene dispersion in sequencing experiments has gained momentum with recent advances in profiling gene expression at the single-cell level. In single-cell sequencing techniques (scRNA-seq), researchers have the ability of tracking gene expression heterogeneity and determine cell subpopulations within a given tissue at single-cell scale resolution (Jaitin et al., 2014). In this way, gene dispersion can be seen as a manifestation of cell-to-cell differences in a single tissue sample, and different approaches have been reported to model this phenomenon (Yip et al., 2018). Such refined analysis is unfeasible with bulk RNA-seq methods. Therefore, it is relevant to remark that RNA-seq experiments rely on the bulk sequencing of an homogenate tissue sample, and that expression profiles obtained using this approach are formed by a complex mixture of single-cell transcriptomic contributions. Although sample collection for RNA-seq analyses is commonly implemented in such a way that tissue integrity and purity are prioritized, it is obvious that tissues such as muscle, liver, fat etc, are composed by diverse cell populations in different developmental stages, a feature which is known to significantly contribute to the observed gene variability across and within biological replicates (Osorio et al., 2019).

In the light of this, RNA-seq-based experiments might not be powerful enough to fully capture the intrinsic nature of gene dispersion within treatment groups. To what extent the apparent absence of gene variability in our experimental design is due to true biological homogeneity or to other sources of variation remains unclear. For instance, sequence variants located at promoter regions can affect the accessibility of transcription factors, hence changing the expression dynamics of genes in a *cis*- action manner. This phenomenon, known as allelic differential expression (ADE), can be produced by any regulatory mutation and probably explains in part the existence of differential dispersion in RNA-seq data sets (Serre et al., 2008; Reddy et al., 2012). In addition, other more in depth analyses might also be applied in order to determine the biological implications of gene dispersion, using, for instance, scRNA-seq techniques, which are currently surpassing RNA-seq as a method of choice in the molecular biology research field (Chen et al., 2019).

4.9. Analyzing gene co-expression modules associated with energy homeostasis and lipid metabolism in response to nutrient supply

The second goal we aimed to achieve in paper VI was to identify mRNA co-expression modules based on skeletal muscle transcriptomic data obtained from fasting (T0) and fed (T1 = 5 hours after feeding, and T2 = 7 hours after feeding) Duroc gilts. We were also interested in determining to which extent these modules are associated with meat quality and lipid-related phenotypic traits recorded in the analyzed gilts (Paper VI, Additional Table S1).

For this purpose, we made use of the WGCNA software (Langfelder and Horvath, 2008), a widely used and cited tool particularly designed for finding gene co-expression modules and hub genes in gene expression data sets and for relating modules (clusters of genes with correlated expression) with phenotype measurements.

Apart from the WGCNA approach, we computed an additional hub score metric (K) describing the degree of connectivity of each analyzed gene according to meaningful co-expression interactions defined with the PCIT algorithm (Reverter and Chan, 2008; Watson-Haigh et al., 2010). Our goal was to detect hub genes and compare the reproducibility of the hub score K metric with Kleinberg's hub centrality score metric by means of the WGCNA algorithm.

Our analyses evidenced the presence of highly co-expressed genes in both T0/T1 and T0/T2 contrasts, although the muscle metabolic response to nutrient supply was more intense after 7 hours of feeding (T2) than after 5 hours (T1). Indeed, the number of DE mRNA genes increased by nearly 3-fold in T0/T2 (N = 435 genes) when compared to T0/T1 (N = 149 genes). As previously discussed in paper VI, it is worth mentioning the case of *BACH1, ETS1* and *CREB1* genes in the T0/T1 contrast. These genes formed part of a co-expression module significantly associated with the C16:0 content of the GM muscle (*r* = 0.45; *P*-value = 0.03). Although none of these genes was DE in T0/T1, they were identified as hub genes in the WGCNA and K metric analyses (Table 3). In addition, the *NR1D2* gene was also classified as a hub gene according to the analysis based on K values, but such result was not confirmed with the WGCNA tool. More importantly, these genes are regulatory cofactors of key metabolic processes such as the maintenance of circadian rhythms (Everett and Lazar, 2014), protection against oxidative stress (Zhang et al., 2018), gluconeogenesis in fasting-feeding transitions (Li et al., 2019a) and glucose uptake (Besse-Patin et al., 2019).

With regard to the T0/T2 contrast, the *SCAMP2*, *NEU3*, *PDK4*, *BACH2* and *ARID5B* genes showed significant differential expression, as well as high interconnectivity in both the WGCNA analysis and PCIT-based determination of hub genes (Table 3). Moreover, they formed a co-expression module which showed expression patterns negatively correlated with the pH of the GM skeletal muscle after slaughtering. Nevertheless, the meat pH measured for T0 and T2 gilts after slaughter did not show significant differences in the T0/T2 contrast (*P*-value = 0.467), i.e. we only observed a slight pH reduction in T2 gilts (average pH = 6.46, SD = 0.16) when compared with T0 gilts (average pH = 6.58, SD = 0.11). Despite the fact that *PDK4* was the most highly downregulated gene in T2 gilts, and that it was also highly interconnected within its co-expression module according to the WGCNA analysis, its contribution to the slight pH reduction observed after nutrient supply (T2) was not significant. The *SCAMP2, NEU3, PDK4* and *ARID5B* genes are involved in regulating glucose and lipid metabolism (Laurie et al., 1993; Yoshizumi et al., 2007; Jeong et al., 2012; Muñoz et al., 2018), and their expression patterns might reflect the activated glycolysis profile that is established in response to food intake in order to produce ATP for energy storage.

Besides, it is worth mentioning the observed Red co-expression module (Paper VI, Additional Table S9 and S14) composed by several lipid-related mRNA genes, among which the *MLXIPL*, *FASN* and *SCD* genes showed high to moderate hub scores by using the PCIT-based K metric (Table 3). These genes were also detected as significantly upregulated in T2 gilts, and were among the most relevant contributors to the association of this co-expression module with the content of palmitic and arachidonic fatty acids in the GM muscle, according to the WGCNA algorithm. Once again, these observed significant associations with fatty acids content did not imply relevant changes of their contents in the T0/T2 contrast (C16:0 *P*-value = 0.699; C20:4 *P*-value = 0.501). Despite this, the presence of *FASN* as a highly interconnected and overexpressed hub gene in fed gilts is suggestive of a coordinated response between the activated glycolysis in response to nutrient supply and the *de novo* fatty acids synthesis of palmitate, which is used as a precursor for generating more complex fatty acids chains through elongation and desaturation processes (Figure 3). Indeed, the final products obtained after glycolysis are two molecules of pyruvate plus 2 ATPs per each catabolized glucose molecule. Pyruvate molecules accumulated in the cytosol after active glycolysis can then be imported into the mitochondrial matrix and used as a substrate to produce acetyl-CoA by means of the pyruvate dehydrogenase (PDH) complex (Ameer et al., 2014). The PDK4 kinase (highly downregulated in fed gilts, Table 3) is in fact responsible of phosphorylating two sites of the PDH enzyme, hence inhibiting its functionality (Holness and Sugden, 2003; Zhang et al., 2014). Subsequently, the produced acetyl-CoA molecules are imported into the tricarboxylic acids (TCA) cycle to generate citric acid that can be exported back to the cytosol to generate acetyl-CoA by the ATP citrate lyase (ACLY) enzyme. These cytosolic acetyl-CoA molecules accumulated after active glycolysis are transformed by the acetyl-CoA carboxylase α enzyme (encoded by the *ACACA* gene) into malonyl-CoA, which is subsequently used by the fatty acid synthase (FASN) as a substrate to generate palmitic (C16:0) fatty acid molecules (Ameer et al., 2014). In this regard, *ACACA* mRNA expression was also included in the Red co-expression module and was among the top contributors to the association with C16:0 fatty acid content in GM (*P*-value = 9.82E-04).

Table 3: Summary of integrated metrics for DE analyses, WGCNA co-expression modules and hub scores for relevant mRNA genes in the T0/T1 and T0/T2 contrasts. Genes that commonly showed significant DE, hub scores and associations with GM phenotypes are depicted in bold.

Genes	log$_2$FC[b]	DE *q*-value[c]	WGCNA Hub score[d]	K Hub score[e]	Module[f]	*P*-value[g]	Phenotype	*r*[h]	*P*-value*[i]
T0/T1[a]									
NR1D2	-0.4212	1.00E+00	-	3.0387	Blue	-	C16:0[j]	0.45	0.03
BACH1	-0.7830	7.54E-01	0.9583	3.0118		1.17E-01			
ETS1	-0.4667	7.55E-01	0.7114	2.9042		1.90E-01			
CREB1	-0.0802	1.00E+00	0.9604	2.9042		5.66E-02			
T0/T2[a]									
SCAMP2	**1.0292**	**5.50E-07**	**0.7646**	**3.2596**	**Green**	**3.25E-02**	**PH45GM[k]**	**-0.46**	**0.03**
NEU3	**1.9862**	**3.48E-14**	**0.9445**	**3.1914**		**5.68E-03**			
PDK4	-4.9403	2.00E-18	0.7089	3.0548		1.14E-01			
BACH2	**-2.0980**	**3.28E-10**	**0.4938**	**2.7988**		**3.06E-02**			
ARID5B	**-2.5951**	**5.33E-14**	**0.5397**	**2.6282**		**2.81E-02**			
MLXIPL	0.9659	3.89E-02	-	0.2731	**Red**	6.18E-04	C18:2[l]	-0.42	0.04
						1.41E-02	C20:4[m]	-0.47	0.03
FASN	**1.4127**	**2.17E-02**	**0.8601**	**0.4096**		**8.73E-03**	**C18:2[l]**	**-0.42**	**0.04**
						1.79E-02	**C20:4[m]**	**-0.47**	**0.03**
SCD	1.9187	6.32E-03	-	0.3755		4.00E-05	C18:2[l]	-0.42	0.04
						7.82E-04	C20:4[m]	-0.47	0.03

[a]T0, T1, T2: Duroc gilts slaughtered in a fasting condition (T0, N = 11) and after 5 h (T1, N = 12) and 7 h (T2, N = 12) of food intake. [b]Log$_2$FC: estimated log$_2$ fold change mean expression levels. [c]DE *q*-value: *P*-value corrected for multiple testing with the Benjamini-Hochberg procedure, indicating the significance of differential expression between T0 and T1 gilts, as well as between T0 and T2 gilts. [d]WGCNA Hub score: Scaled Kleinberg's hub centrality score for co-expression modules according to the WGCNA algorithm. [e]K Hub score: Hub score per gene according to mRNA-mRNA meaningful interactions with the PCIT algorithm. [f]Module: WGCNA co-expression module color identification. [g]*P*-value: Gene significance *P*-value within the co-expression module according to the WGCNA algorithm. [h]*r*: Pearson correlation coefficient. [i]*P*-value*: P-value of the Pearson correlation between co-expression WGCNA modules and measured phenotypes. [j]C16:0: Palmitic acid content in the GM muscle. [k]PH45GM: intramuscular pH measured 45 minutes post mortem in the GM muscle. [l]C18:2: Linoleic acid content in the GM muscle. [m]C20:4: Arachidonic acid content in the GM muscle.

On the contrary, *ACACA* was neither detected as DE in the T0/T2 contrast, nor highly interconnected by WGCNA or PCIT algorithms.

Overall, these results evidenced the existence of an active metabolic interconnection and common regulation between the glycolytic process for energy storage and the *de novo* lipogenesis in the cytoplasm (Figure 3), which would be in accordance with the observed differential expression patterns of the aforementioned genes in response to nutrient supply.

4.10. Analyzing miRNA-mRNA interactions associated with energy homeostasis and lipid metabolism in response to nutrient supply

In paper VI, we also aimed to analyze the expression profiles of miRNAs in the GM skeletal muscle after food intake, and how these profiles would correlate with the observed downregulation of mRNA transcripts putatively targeted by the analyzed miRNAs. We sequenced the small RNA fraction of the GM muscle samples that were previously analyzed by Cardoso et al. (2017b), and further explored in paper VI, to assess gene co-expression modules. It is important to remark that the use of bulk RNA-seq sequencing data previously generated by Cardoso et al. (2017b) would have not allowed us to characterize the expression profiles of small RNAs, including miRNAs. The reason is that this bulk RNA-seq data set was generated by Cardoso and collaborators by using the TruSeq Stranded mRNA Library Preparation Kit (Illumina Inc., CA) before paired-end sequencing (2 × 75 bp) in a HiSeq 2000 platform (Illumina Inc., CA). This library preparation protocol is specifically designed to target mRNA transcripts via poly(A) affinity selection to enrich for polyadenylated RNA sequences. This enriched fraction would include the pri-miRNA precursor forms of mature miRNAs, which are transcribed by Pol-II in the form of capped and polyadenylated transcripts (Cai et al., 2004). Such transcripts would have the sufficient size (~80-100 nts long) to be captured, at least partially, by mRNA sequencing protocols. However, the most abundant miRNA molecules present in the cytoplasm and also the functional effectors responsible for downregulating target mRNAs are the processed mature miRNAs. Indeed, mature miRNAs almost double the concentration of pri and pre-miRNA molecules (Gan and Denecke, 2013), they are much smaller in size (~18-22 nts) and they do not have poly(A) tails themselves, so they would not be enriched by library preparation kits that perform poly(A) selection as it is the case of the TruSeq protocol.

With this caveat in mind, we decided to perform a single-end small RNA-seq sequencing of the total RNA fraction extracted from GM muscle samples in the same T0, T1 and T2 gilts for which RNA-seq data were available. A HiSeq 2500 platform (Illumina Inc., CA) was used for sequencing. Through this approach, we were able to fully capture the small RNA molecules present in our RNA homogenates, which would have been otherwise missed or misrepresented if making use of the original RNA-seq sequencing data reported by Cardoso et al. (2017b).

Once pre-processing was completed and specific alignment for small RNA sequences and quantification of annotated porcine miRNAs were performed, we explored the differences in miRNA expression between fasting (T0) and fed (T1 and T2) gilts. As described in paper VI, a total of 6 and 28 DE miRNAs were detected in the T0/T1 and T0/T2 contrasts, respectively. Differences in the numbers of DE miRNAs in the T0/T1 and T0/T2 contrasts were consistent with the numbers of DE mRNAs obtained for the same contrasts (149 for T0/T1 and 435 for T0/T2). However, the overall detected absolute FC measures were slightly lower for miRNAs (~1.9) than for mRNAs (~2.1).

Additionally, we aimed to integrate both mRNA and miRNA expression profiles by inferring putative miRNA-mRNA interactions that might partially explain the muscle metabolic responses to nutrient supply. In this way, we predicted *in silico* DE mRNAs putatively targeted by the observed DE miRNAs by means of sequence searching of 7mer-m8 binding sites in the 3’-UTRs of these putative mRNA targets. To do so, we made use of the *eMIRNA.Target* tool previously described in this general discussion (Section 4.7), as a further development of our eMIRNA pipeline. Meaningful negative correlations (as a reflection of the repressor activity of miRNAs over targeted mRNAs) between miRNAs and mRNAs were also assessed using the *eMIRNA.Network* module. The regulatory impact of the DE miRNAs was also analyzed by applying the RIF algorithm (Reverter et al., 2010) with the *eMIRNA.RIF* module. Our results highlighted several interesting meaningful miRNA-mRNA interactions that merited further discussion in paper VI.

Regarding the T0/T1 contrast, ssc-miR-32 was among the upregulated miRNAs (FC = 1.242, *q*-value = 4.73E-02) detected as highly influential according to the RIF1 metric (Paper VI, Table 3). Moreover, it was significantly associated with the downregulation of the *EGR1* and *ARID5B* differentially expressed mRNAs. These two genes are involved in the regulation of lipid metabolism (Boyle et al., 2009; Muñoz et al., 2018). Other relevant miRNA-mRNA interactions were, to mention a few, those observed between ssc-miR-1, scc-miR-148a-3p and ssc-miR-7-5p and several mRNAs such as *MYF6*, *FOSL2*, *ARRDC3*, *TXNIP*, *ATF3* or *MYOG*, all of which play key roles in regulating muscle growth and differentiation (Óvilo et al., 2014; Li et al., 2017; Muñoz et al., 2018), as well as in modulating glucose and lipid metabolism (Wrann et al., 2012; Lee et al., 2013; Allison et al., 2018; Ling et al., 2018).

Besides, in the T0/T2 contrast one of the most relevant miRNA-mRNA interactions was that between several upregulated miRNAs (ssc-miR-148a-3p, ssc-miR-151-3p, ssc-miR-30a-3p, ssc-miR-30e-3p, ssc-miR-421-5p and ssc-miR-493-5p) and the *PDK4* mRNA transcript, which was strongly downregulated in fed gilts (T2) compared with their fasting counterparts (T0). Among these, the ssc-miR-148a-3p and ssc-miR-493-5p showed conserved 7mer-m8 binding sites in the *PDK4* 3'-UTR across closely related species according to TargetScan analyses (Paper VI, Additional Table S10). These findings give support to the significant inverse relationship between the levels of these two miRNAs and *PDK4* expression as inferred with the PCIT-based co-expression analyses. Moreover, ssc-miR-148a-3p and ssc-miR-493-5p were among the top 10 impactful regulators according to RIF2 metric (Paper VI, Additional Table S12).

In summary, we were able to detect several meaningful putative miRNA-mRNA interactions in the experiment comparing gilts in fasting-feeding conditions, as further discussed in paper VI. More importantly, a suggestive regulatory interconnection between active glycolysis and energy storage, upon nutrient supply in the GM skeletal muscle, and the synthesis and elongation of fatty acids, was envisaged (Figure 3), thus shedding light on the induction and coordination of carbohydrate and lipid metabolism in the porcine skeletal muscle as a consequence of nutrient availability.

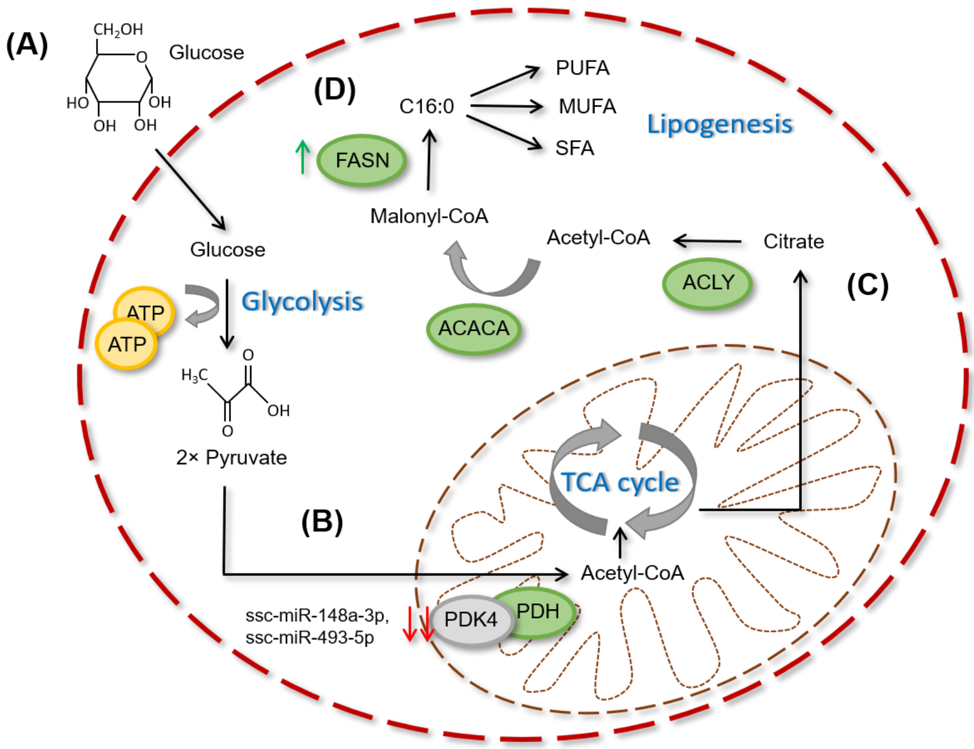


Figure 3: Scheme of the coordination of carbohydrate and lipid metabolism in the porcine skeletal muscle. **(A)** Carbohydrate supply to GM skeletal muscle cells in the form of glucose is converted into pyruvate and ATP through glycolysis. **(B)** The pyruvate dehydrogenase (PDH) enzyme can convert pyruvate to acetyl-CoA in the mitochondria, which is imported into the TCA cycle to generate citrate. Pyruvate kinase 4 (PDK4) enzyme can inhibit the PDH enzyme by phosphorylation, but its expression was strongly downregulated in fed gilts (T2) by the putative predicted interaction of several miRNAs such as ssc-miR-148a-3p or ssc-miR-493-5p, which were differentially upregulated in T2 compared with fasting (T0) pigs. **(C)** The citrate molecules are transformed to cytosolic acetyl-CoA by the ATP citrate lyase (ACLY) enzyme, so that acetyl-CoA can then be converted into malonyl-CoA by the acetyl CoA carboxylase α enzyme (ACACA). **(D)** Malonyl-CoA is the precursor used to synthetize palmitic fatty acid (C16:0) by the action of the fatty acid synthase (FASN) enzyme, the expression of which was significantly upregulated after food intake (T2). Palmitate molecules are finally used as precursors for the synthesis of a broad array of elongated saturated fatty acids (SFA), as well as of monounsaturated and polyunsaturated fatty acids (MUFA and PUFA).

CHAPTER V. CONCLUSIONS

The main conclusions that can be extracted from the present Ph.D. thesis are as follows:

1. We have genotyped 19 SNPs located in 14 genes with well characterized metabolic functions and mapping to meat quality QTL in 345 Duroc pigs (Lipgen population). The majority of these SNPs did not show significant associations with the traits under investigation after correcting for multiple testing, with the only exception of one SNP located in the *ATP1A2* gene, which was highly associated with electric conductivity in the *longissimus dorsi* muscle. Given the important function of the ATP1A2 protein in the induction of an electrochemical gradient across the plasma membrane of cells, this is an interesting candidate gene for modulating electric conductivity that should be further investigated.

2. We have identified one putative stop gained mutation (rs81212146, c.944T>A) in the pig *ASS1* gene segregating in a Duroc commercial population (Lipgen). Individuals homozygous for these mutations were perfectly healthy despite the fact that ASS1 inactivation has lethal consequences due to the disruption of the urea cycle. Sequencing of this region revealed the existence of an additional compensatory mutation (rs81212145, c.943T>A) immediately adjacent to c.944T>A, which suppresses the emergence of a premature stop codon.

3. We have investigated the association between the phenotypic variation of lipid-related traits recorded in the Duroc Lipgen population and polymorphic sites of six circadian genes (*CRY2*, *NPAS2*, *CIART*, *ARNTL2*, *PER1* and *PER2*) and two loci (*PCK1* and *MIGA2*) with important metabolic functions. Out of 20 genotyped SNPs, only two SNPs in the *CRY2* (rs320439526, c.-6C>T) and *MIGA2* (rs330779504, c.1455G>A) genes showed significant associations, after correction for multiple testing, with stearic acid content in the *longissimus dorsi* muscle and with LDL serum concentration at 190 days, respectively. However, chromosome-wide level association analyses did not yield significant results, indicating that these SNPs are unlikely to have causal effects on the phenotypic variance of the aforementioned traits.

4. The analysis of the variability of porcine microRNA genes has shown divergent patterns in Asian vs European pigs and wild boars, and a dramatic reduction of polymorphic sites in the seed region, likely due to the presence of purifying selection removing mutations that alter the binding ability of the microRNA. Genotyping of 15 SNPs in Duroc pigs with microarray expression data from *gluteus medius* muscle (N = 89) and liver (N = 87) tissues revealed several associations. The most interesting one was featured by one SNP in the apical loop of ssc-miR-326 (rs319154814, n.46G>A), which displayed significant associations with several of its potential mRNA targets (e.g. *PPP1CC, CFLAR, SF3A3* or *FSTL1*). This polymorphism might exert its effect by influencing the efficiency of the maturation of the microRNA.

5. The development of a machine learning-based transductive approach for the discovery and annotation of miRNA genes allowed us to identify 20 unreported porcine miRNAs expressed in the *gluteus medius* muscle of 48 Duroc pigs, as well as 27 additional miRNAs with orthologous sequences in humans. We verified that the use of a dynamic sequence motif search for the reconstruction of candidate miRNA sequences improved the predictive accuracy of the miRNA classifier, allowing a better determination of the boundaries of miRNA genes. Comparison with the miRDeep2 software demonstrated that our approach makes possible to identify an increased number of miRNAs as well as to determine more accurately the boundaries of miRNA genes.

6. Comparison of the expression patterns of mRNAs, miRNAs and lincRNAs expressed in the *gluteus medius* muscle of Duroc pigs showed that protein-coding genes were generally the most expressed transcripts, followed by miRNAs and lincRNAs. Moreover, lincRNAs displayed the highest variance in expression, while microRNAs showed the lowest, indicating that they have a narrow range of expression probably due to their key regulatory role.

7. The *gluteus medius* miRNA expression profiles of fasted pigs (T0) and pigs sampled 5 h (T1) and 7h (T2) after feeding were determined in the current thesis. A total of 149 (T0 vs T1) and 435 (T0 vs T2) mRNAs, 6 (T0 vs T1) and 28 (T0 vs T2) miRNAs and none lincRNAs were detected as differentially expressed in fasted vs fed pigs. Such results allowed us to infer that the expression of ssc-miR-148a-3p, ssc-miR-151-3p, ssc-miR-30a-3p, ssc-miR-30e-3p, ssc-miR-421-5p, ssc-miR-493-5p and ssc-miR-503 is significantly associated with the observed downregulation of *PDK4*, a gene involved in regulating glucose metabolism and fatty acids oxidation. Additionally, co-expression modules were identified including relevant differentially expressed genes related with lipid metabolism such as *MLXIPL*, *FASN*, *SCD*, *SFRP1*, *SFRP5* or *THRSP*.

CHAPTER VI. REFERENCES

Abrams, Z. B., Johnson, T. S., Huang, K., Payne, P. R. O., and Coombes, K. (2019). A protocol to evaluate RNA sequencing normalization methods. *BMC Bioinformatics* 20, 679.

Agarwal, V., Bell, G. W., Nam, J. W., and Bartel, D. P. (2015). Predicting effective microRNA target sites in mammalian mRNAs. *Elife* 4, e05005.

Agarwal, V., Subtelny, A. O., Thiru, P., Ulitsky, I., and Bartel, D. P. (2018). Predicting microRNA targeting efficacy in Drosophila. *Genome Biol.* 19, 152.

Aguet, F., Brown, A. A., Castel, S. E., Davis, J. R., He, Y., Jo, B., et al. (2017). Genetic effects on gene expression across human tissues. *Nature* 550, 204–213.

Ahn, J., Lee, H., Jung, C. H., and Ha, T. (2012). Lycopene inhibits hepatic steatosis via microRNA-21-induced downregulation of fatty acid-binding protein 7 in mice fed a high-fat diet. *Mol. Nutr. Food Res.* 56, 1665–1674.

Allison, M. B., Pan, W., MacKenzie, A., Patterson, C., Shah, K., Barnes, T., et al. (2018). Defining the transcriptional targets of leptin reveals a role for *Atf3* in leptin action. *Diabetes* 67, 1093–1104.

Altay, G., and Emmert-Streib, F. (2010). Inferring the conservative causal core of gene regulatory networks. *BMC Syst. Biol.* 4, 132.

Alvarez-Saavedra, E., and Horvitz, H. R. (2010). Many Families of *C. elegans* MicroRNAs Are Not Essential for Development or Viability. *Curr. Biol.* 20, 367–373.

Amaral, A. J., Megens, H. J., Crooijmans, R. P. M. A., Heuven, H. C. M., and Groenen, M. A. M. (2008). Linkage disequilibrium decay and haplotype block structure in the pig. *Genetics* 179, 569–579.

Amarasinghe, S. L., Su, S., Dong, X., Zappia, L., Ritchie, M. E., and Gouil, Q. (2020). Opportunities and challenges in long-read sequencing data analysis. *Genome Biol.* 21, 1–16.

Ambros, V., Bartel, B., Bartel, D. P., Burge, C. B., Carrington, J. C., Chen, X., et al. (2003). A uniform system for microRNA annotation. *RNA* 9, 277–9.

Ameer, F., Scandiuzzi, L., Hasnain, S., Kalbacher, H., and Zaidi, N. (2014). De novo lipogenesis in health and disease. *Metabolism.* 63, 895–902.

Ameres, S. L., and Zamore, P. D. (2013). Diversifying microRNA sequence and function. *Nat. Rev. Mol. Cell Biol.* 14, 475–488.

An, J., Lai, J., Lehman, M. L., and Nelson, C. C. (2013). miRDeep*: an integrated application tool for miRNA identification from RNA sequencing data. *Nucleic Acids Res.* 41, 727–737.

Anders, S., Pyl, P. T., and Huber, W. (2015). HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics* 31, 166–169.

Andersson, L., Haley, C. S., Ellegren, H., Knott, S. A., Johansson, M., Andersson, K., et al. (1994). Genetic mapping of quantitative trait loci for growth and fatness in pigs. *Science* 263, 1771–1774.

Andrés-León, E., Núñez-Torres, R., and Rojas, A. M. (2016). miARma-Seq: a comprehensive tool for miRNA, mRNA and circRNA analysis. *Sci. Rep.* 6, 25749.

Aparicio-Puerta, E., Lebrón, R., Rueda, A., Gómez-Martín, C., Giannoukakos, S., Jaspez, D., et al. (2019). sRNAbench and sRNAtoolbox 2019: intuitive fast small RNA profiling and differential expression. *Nucleic Acids Res.* 47, W530–W535.

Auyeung, V. C., Ulitsky, I., McGeary, S. E., and Bartel, D. P. (2013). Beyond secondary structure: primary-sequence determinants license pri-miRNA hairpins for processing. *Cell* 152, 844–58.

Ayuso, M., Fernández, A., Núñez, Y., Benítez, R., Isabel, B., Fernández, A. I., et al. (2016). Developmental stage, muscle and genetic type modify muscle transcriptome in pigs: Effects on gene expression and regulatory factors involved in growth and metabolism. *PLoS One* 11, e0167858.

Babiarz, J. E., Ruby, J. G., Wang, Y., Bartel, D. P., and Blelloch, R. (2008). Mouse ES cells express endogenous shRNAs, siRNAs, and other microprocessor-independent, dicer-dependent small RNAs. *Genes Dev.* 22, 2773–2785.

Backes, C., Fehlmann, T., Kern, F., Kehl, T., Lenhof, H.-P., Meese, E., et al. (2018). miRCarta: a central repository for collecting miRNA candidates. *Nucleic Acids Res.* 46, D160–D167.

Baek, D., Villén, J., Shin, C., Camargo, F. D., Gygi, S. P., and Bartel, D. P. (2008). The impact of microRNAs on protein output. *Nature* 455, 64–71.

Balov, N. (2013). A categorical network approach for discovering differentially expressed regulations in cancer. *BMC Med. Genomics* 6, S1.

Bar, M., Wyman, S. K., Fritz, B. R., Qi, J., Garg, K. S., Parkin, R. K., et al. (2008). MicroRNA Discovery and Profiling in Human Embryonic Stem Cells by Deep Sequencing of Small RNA Libraries. *Stem Cells* 26, 2496–2505.

Barbosa, S., Niebel, B., Wolf, S., Mauch, K., and Takors, R. (2018). A guide to gene regulatory network inference for obtaining predictive solutions: Underlying assumptions and fundamental biological and data constraints. *BioSystems* 174, 37–48.

Barclay, J. L., Shostak, A., Leliavski, A., Tsang, A. H., Jöhren, O., Müller-Fielitz, H., et al. (2013). High-fat diet-induced hyperinsulinemia and tissue-specific insulin resistance in Cry-deficient mice. *Am. J. Physiol. - Endocrinol. Metab.* 304, E1053-63.

Bartel, D. P. (2018). Metazoan MicroRNAs. *Cell* 173, 20–51.

Barturen, G., Rueda, A., Hamberg, M., Alganza, A., Lebron, R., Kotsyfakis, M., et al. (2014). sRNAbench: profiling of small RNAs and its sequence variants in single or multi-species high-throughput experiments. *Methods Next Gener. Seq.* 1, 21-31.

Batuwita, R., and Palade, V. (2009). microPred: effective classification of pre-miRNAs for human miRNA gene prediction. *Bioinformatics* 25, 989–995.

Bazzoni, F., Rossato, M., Fabbri, M., Gaudiosi, D., Mirolo, M., Mori, L., et al. (2009). Induction and regulatory function of miR-9 in human monocytes and neutrophils exposed to proinflammatory signals. *Proc. Natl. Acad. Sci. U. S. A.* 106, 5282–5287.

Beckers, M., Mohorianu, I., Stocks, M., Applegate, C., Dalmay, T., and Moulton, V. (2017). Comprehensive processing of high-throughput small RNA sequencing data including quality checking, normalization, and differential expression analysis using the UEA sRNA Workbench. *RNA* 23, 823–835.

Bellot, P., Olsen, C., Salembier, P., Oliveras-Vergés, A., and Meyer, P. E. (2015). NetBenchmark: a bioconductor package for reproducible benchmarks of gene regulatory network inference. *BMC Bioinformatics* 16, 312.

Benítez, R., Trakooljul, N., Núñez, Y., Isabel, B., Murani, E., De Mercado, E., et al. (2019).

Breed, diet, and interaction effects on adipose tissue transcriptome in iberian and duroc pigs fed different energy sources. *Genes.* 10, 589.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* 57, 289–300.

Benjamini, Y., and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Stat.* 29, 1165–1188.

Bentwich, I., Avniel, A., Karov, Y., Aharonov, R., Gilad, S., Barad, O., et al. (2005). Identification of hundreds of conserved and nonconserved human microRNAs. *Nat. Genet.* 37, 766–770.

Besse-Patin, A., Jeromson, S., Levesque-Damphousse, P., Secco, B., Laplante, M., and Estall, J. L. (2019). PGC1A regulates the IRS1:IRS2 ratio during fasting to influence hepatic metabolism downstream of insulin. *Proc. Natl. Acad. Sci.* 116, 4285–4290.

Bhaskaran, M., and Mohan, M. (2014). MicroRNAs: history, biogenesis, and their evolving role in animal development and disease. *Vet. Pathol.* 51, 759–774.

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

Bisognin, A., Sales, G., Coppe, A., Bortoluzzi, S., and Romualdi, C. (2012). MAGIA^2: From miRNA and genes expression data integrative analysis to microRNA-transcription factor mixed regulatory circuits (2012 update). *Nucleic Acids Res.* 40, W13-W21.

Black, M. B., Parks, B. B., Pluta, L., Chu, T.-M., Allen, B. C., Wolfinger, R. D., et al. (2013). Comparison of microarrays and RNA-seq for gene expression analyses of dose-response experiments. *Toxicol. Sci.* 137, 385–403.

Blüher, M., Williams, C. J., Klöting, N., Hsi, A., Ruschke, K., Oberbach, A., et al. (2007). Gene expression of adiponectin receptors in human visceral and subcutaneous adipose tissue is related to insulin resistance and metabolic parameters and is altered in response to physical training. *Diabetes Care* 30, 3110–3115.

Bogdanova, N., Horst, J., Chlystun, M., Croucher, P. J. P., Nebel, A., Bohring, A., et al. (2007). A common haplotype of the annexin A5 (*ANXA5*) gene promoter is associated with recurrent pregnancy loss. *Hum. Mol. Genet.* 16, 573–578.

Bolger, A. M., Lohse, M., and Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114–2120.

Bonnet, E., Wuyts, J., Rouze, P., and Van de Peer, Y. (2004). Evidence that microRNA precursors, unlike other non-coding RNAs, have lower folding free energies than random sequences. *Bioinformatics* 20, 2911–2917.

Borodina, T., Adjaye, J., and Sultan, M. (2011). A strand-specific library preparation protocol for RNA sequencing. *Methods Enzymol.* 500, 79–98.

Bortolomeazzi, M., Gaffo, E., and Bortoluzzi, S. (2017). A survey of software tools for microRNA discovery and characterization using RNA-seq. *Brief. Bioinform.* 20, 918-930.

Bosse, M., Megens, H. J., Derks, M. F. L., de Cara, Á. M. R., and Groenen, M. A. M. (2019). Deleterious alleles in the context of domestication, inbreeding, and selection. *Evol. Appl.* 12, 6–17.

Bosse, M., Megens, H. J., Madsen, O., Frantz, L. A. F., Paudel, Y., Crooijmans, R. P. M. A., et al. (2014). Untangling the hybrid nature of modern pig genomes: A mosaic derived from biogeographically distinct and highly divergent Sus scrofa populations. *Mol. Ecol.* 23, 4089–4102.

Boyle, K. B., Hadaschik, D., Virtue, S., Cawthorn, W. P., Ridley, S. H., O'Rahilly, S., et al. (2009). The transcription factors Egr1 and Egr2 have opposing influences on adipocyte differentiation. *Cell Death Differ.* 16, 782–9.

Bray, N. L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* 34, 525–527.

Brown, R. A. M., Epis, M. R., Horsham, J. L., Kabir, T. D., Richardson, K. L., and Leedman, P. J. (2018). Total RNA extraction from tissues for microRNA and target gene expression analysis: not all kits are created equal. *BMC Biotechnol.* 18, 16.

Bruun, C. S., Jorgensen, C. B., Nielsen, V. H., Andersson, L., and Fredholm, M. (2006). Evaluation of the porcine melanocortin 4 receptor (*MC4R*) gene as a positional candidate for a fatness QTL in a cross between Landrace and Hampshire. *Anim. Genet.* 37, 359–

362.

Budak, H., Bulut, R., Kantar, M., and Alptekin, B. (2015). MicroRNA nomenclature and the need for a revised naming prescription. *Brief. Funct. Genomics* 15, elv026.

Burton, B. K. (2000). Urea cycle disorders. *Clin. Liver Dis.* 4, 815–830.

Butte, A. J., and Kohane, I. S. (2000). Mutual information relevance networks: functional genomic clustering using pairwise entropy measurements. *Pac. Symp. Biocomput.*, 418–429.

Cabili, M. N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., et al. (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* 25, 1915–27.

Cai, X., Hagedorn, C. H., and Cullen, B. R. (2004). Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *RNA* 10, 1957–1966.

Calura, E., Martini, P., Sales, G., Beltrame, L., Chiorino, G., D'Incalci, M., et al. (2014). Wiring miRNAs to pathways: a topological approach to integrate miRNA and mRNA expression profiles. *Nucleic Acids Res.* 42, e96.

Cánovas, A., Pena, R. N., Gallardo, D., Ramírez, O., Amills, M., and Quintanilla, R. (2012). Segregation of regulatory polymorphisms with effects on the *gluteus medius* transcriptome in a purebred pig population. *PLoS One* 7, e35583.

Cao, C., Zhang, Y., Jia, Q., Wang, X., Zheng, Q., Zhang, H., et al. (2019). An exonic splicing enhancer mutation in *DUOX2* causes aberrant alternative splicing and severe congenital hypothyroidism in Bama pigs. *Dis. Model. Mech.* 12, dmm036616.

Cardoso, T. F., Cánovas, A., Canela-Xandri, O., González-Prendes, R., Amills, M., and Quintanilla, R. (2017a). RNA-seq based detection of differentially expressed genes in the skeletal muscle of Duroc pigs with distinct lipid profiles. *Sci. Rep.* 7, 40005.

Cardoso, T. F., Quintanilla, R., Tibau, J., Gil, M., Mármol-Sánchez, E., González-Rodríguez, O., et al. (2017b). Nutrient supply affects the mRNA expression profile of the porcine skeletal muscle. *BMC Genomics* 18, 603.

Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M. C., Maeda, N., et al. (2005).

The transcriptional landscape of the mammalian genome. *Science* 309, 1559–63.

Casellas, J., Gularte, R. J., Farber, C. R., Varona, L., Mehrabian, M., Schadt, E. E., et al. (2012). Genome scans for transmission ratio distortion regions in mice. *Genetics* 191, 247–259.

Casellas, J., Noguera, J. L., Reixach, J., Díaz, I., Amills, M., and Quintanilla, R. (2010). Bayes factor analyses of heritability for serum and muscle lipid traits in Duroc pigs1. *J. Anim. Sci.* 88, 2246–2254.

Casellas, J., Vidal, O., Pena, R. N., Gallardo, D., Manunza, A., Quintanilla, R., et al. (2013). Genetics of serum and muscle lipids in pigs. *Anim. Genet.* 44, 609–619.

Cava, C., Colaprico, A., Bertoli, G., Graudenzi, A., Silva, T. C., Olsen, C., et al. (2017). SpidermiR: An R/bioconductor package for integrative analysis with miRNA data. *Int. J. Mol. Sci.* 18, 274.

Cepica, S., Stratil, A., Kopecny, M., Blazkova, P., Schroffel, J., Davoli, R., et al. (2003). Linkage and QTL mapping for *Sus scrofa* chromosome 4. *J. Anim. Breed. Genet.* 120, 28–37.

Chalancon, G., Ravarani, C. N. J., Balaji, S., Martinez-Arias, A., Aravind, L., Jothi, R., et al. (2012). Interplay between gene expression noise and regulatory network architecture. *Trends Genet.* 28, 221–232.

Charlesworth, D., and Willis, J. H. (2009). The genetics of inbreeding depression. *Nat. Rev. Genet.* 10, 783–796.

Chen, C. Y. A., and Shyu, A. Bin (2011). Mechanisms of deadenylation-dependent decay. *Wiley Interdiscip. Rev. RNA* 2, 167–183.

Chen, G., Ning, B., and Shi, T. (2019). Single-cell RNA-seq technologies and related computational data analysis. *Front. Genet.* 10, 317.

Chen, J., Wang, X., and Liu, B. (2016a). iMiRNA-SSF: Improving the identification of microRNA precursors by combining negative sets with different distributions. *Sci. Rep.* 6, 19062.

Chen, J.-F., Mandel, E. M., Thomson, J. M., Wu, Q., Callis, T. E., Hammond, S. M., et al. (2006). The role of microRNA-1 and microRNA-133 in skeletal muscle proliferation and differentiation. *Nat. Genet.* 38, 228–33.

Chen, K., and Rajewsky, N. (2006). Natural selection on human microRNA binding sites inferred from SNP data. *Nat. Genet.* 38, 1452–1456.

Chen, P. Y., Manninga, H., Slanchev, K., Chien, M., Russo, J. J., Ju, J., et al. (2005). The developmental miRNA profiles of zebrafish as determined by small RNA cloning. *Genes Dev.* 19, 1288–93.

Chen, R., Shi, L., Hakenberg, J., Naughton, B., Sklar, P., Zhang, J., et al. (2016b). Analysis of 589,306 genomes identifies individuals resilient to severe Mendelian childhood diseases. *Nat. Biotechnol.* 34, 531–538.

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34, i884–i890.

Cheng, H. Y. M., Papp, J. W., Varlamova, O., Dziema, H., Russell, B., Curfman, J. P., et al. (2007). microRNA modulation of circadian-clock period and entrainment. *Neuron* 54, 813–829.

Chipman, L. B., and Pasquinelli, A. E. (2019). miRNA Targeting: Growing beyond the Seed. *Trends Genet.* 35, 215–222.

Cho, I. C., Park, H. B., Ahn, J. S., Han, S. H., Lee, J. B., Lim, H. T., et al. (2019). A functional regulatory variant of *MYH3* influences muscle fiber-type composition and intramuscular fat content in pigs. *PLoS Genet.* 15, e1008279.

Cho, I. C., Yoo, C. K., Lee, J. B., Jung, E. J., Han, S. H., Lee, S. S., et al. (2015). Genome-wide QTL analysis of meat quality-related traits in a large F2 intercross between Landrace and Korean native pigs. *Genet. Sel. Evol.* 47, 7.

Cho, I. S., Kim, J., Seo, H. Y., Lim, D. H., Hong, J. S., Park, Y. H., et al. (2010). Cloning and characterization of microRNAs from porcine skeletal muscle and adipose tissue. *Mol. Biol. Rep.* 37, 3567–3574.

Choi, I., Steibel, J. P., Bates, R. O., Raney, N. E., Rumph, J. M., and Ernst, C. W. (2011). Identification of carcass and meat quality QTL in an F2 Duroc × Pietrain pig resource population using different least-squares analysis models. *Front. Genet.* 2, 18.

Chrusciel, M., Rekawiecki, R., Ziecik, A. J., and Andronowska, A. (2010). mRNA and protein expression of FGF-1, FGF-2 and their receptors in the porcine umbilical cord during pregnancy. *Folia Histochem. Cytobiol.* 48, 572–580.

Cirera, S., Birck, M., Busk, P. K., and Fredholm, M. (2010). Expression profiles of miRNA-122 and its target *CAT1* in minipigs (*Sus scrofa*) fed a high-cholesterol diet. *Comp. Med.* 60, 136–41.

Clark, M. B., Johnston, R. L., Inostroza-Ponta, M., Fox, A. H., Fortini, E., Moscato, P., et al. (2012). Genome-wide analysis of long noncoding RNA stability. *Genome Res.* 22, 885–898. doi:10.1101/gr.131037.111.

Concepcion, C. P., Han, Y. C., Mu, P., Bonetti, C., Yao, E., D'Andrea, A., et al. (2012). Intact p53-dependent responses in miR-34-deficient mice. *PLoS Genet.* 8, e1002797.

Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., et al. (2016). A survey of best practices for RNA-seq data analysis. *Genome Biol.* 17, 13.

Costa, E. V., Diniz, D. B., Veroneze, R., Resende, M. D. V., Azevedo, C. F., Guimaraes, S. E. F., et al. (2015). Estimating additive and dominance variances for complex traits in pigs combining genomic and pedigree information. *Genet. Mol. Res.* 14, 6303–6311.

Cover, T. M., and Thomas, J. A. (2005). Elements of information theory. *Second edition*. *John Wiley & Sons*.

da Silveira, W. A., Renaud, L., Simpson, J., Glen, W. B., Hazard, E. S., Chung, D., et al. (2018). miRmapper: A tool for interpretation of miRNA–mRNA interaction networks. *Genes* 9, 458.

Davoli, R., Fontanesi, L., Braglia, S., Nisi, I., Scotti, E., Buttazzoni, L., et al. (2006). Investigation of SNPs in the *ATP1A2*, *CA3* and *DECR1* genes mapped to porcine chromosome 4: Analysis in groups of pigs divergent for meat production and quality traits. *Ital. J. Anim. Sci.* 5, 249–263.

Davoli, R., Zappaterra, M., and Zambonelli, P. (2019). Genome-wide association study identifies markers associated with meat ultimate pH in Duroc pigs. *Anim. Genet.* 50, 154–156.

de Matos Simoes, R., and Emmert-Streib, F. (2012). Bagging statistical network inference from large-scale gene expression data. *PLoS One* 7, e33624.

de Rie, D., Abugessaisa, I., Alam, T., Arner, E., Arner, P., Ashoor, H., et al. (2017). An integrated expression atlas of miRNAs and their promoters in human and mouse. *Nat. Biotechnol.* 35, 872–878.

Dekkers, J. C. M., Mathur, P. K., and Knol, E. F. (2011). Genetic improvement of the pig. *The Genetics of the Pig: Second Edition*. *CABI*, 390–425.

Derks, M. F. L., Gjuvsland, A. B., Bosse, M., Lopes, M. S., van Son, M., Harlizius, B., et al. (2019a). Loss of function mutations in essential genes cause embryonic lethality in pigs. *PLoS Genet.* 15, e1008055.

Derks, M. F. L., Harlizius, B., Lopes, M. S., Greijdanus-van der Putten, S. W. M., Dibbits, B., Laport, K., et al. (2019b). Detection of a frameshift deletion in the *SPTBN4* gene leads to prevention of severe myopathy and postnatal mortality in pigs. *Front. Genet.* 10, 1226.

Derks, M. F. L., Lopes, M. S., Bosse, M., Madsen, O., Dibbits, B., Harlizius, B., et al. (2018). Balancing selection on a recessive lethal deletion with pleiotropic effects on two neighboring genes in the porcine genome. *PLoS Genet.* 14, e1007661.

Derks, M. F. L., Megens, H. J., Bosse, M., Lopes, M. S., Harlizius, B., and Groenen, M. A. M. (2017). A systematic survey to identify lethal recessive variation in highly managed pig populations. *BMC Genomics* 18, 858.

Desvignes, T., Loher, P., Eilbeck, K., Ma, J., Urgese, G., Fromm, B., et al. (2019). Unification of miRNA and isomiR research: the mirGFF3 format and the mirtop API. *Bioinformatics* 36, 698-703.

Ding, J., Zhou, S., and Guan, J. (2010). MiRenSVM: towards better prediction of microRNA precursors using an ensemble SVM classifier with multi-loop features. *BMC Bioinformatics* 11 Suppl 11, S11.

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., et al. (2012). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15–21.

Dron, N., Hernández-Jover, M., Doyle, R. E., and Holyoake, P. K. (2014). Investigating risk factors and possible infectious aetiologies of mummified fetuses on a large piggery in Australia. *Aust. Vet. J.* 92, 472–478.

Druet, T., Macleod, I. M., and Hayes, B. J. (2014). Toward genomic prediction from whole-genome sequence data: Impact of sequencing design on genotype imputation and accuracy of predictions. *Heredity* 112, 39–47.

Duijvesteijn, N., Knol, E. F., Merks, J. W. M., Crooijmans, R. P. M. A., Groenen, M. A. M., Bovenhuis, H., et al. (2010). A genome-wide association study on androstenone levels in pigs reveals a cluster of candidate genes on chromosome 6. *BMC Genet.* 11, 42.

Dumortier, O., Hinault, C., and Van Obberghen, E. (2013). MicroRNAs and metabolism crosstalk in energy homeostasis. *Cell Metab.* 18, 312–324.

Eden, E., Navon, R., Steinfeld, I., Lipson, D., and Yakhini, Z. (2009). GOrilla: A tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinformatics* 10, 48.

Eiring, A. M., Harb, J. G., Neviani, P., Garton, C., Oaks, J. J., Spizzo, R., et al. (2010). miR-328 functions as an RNA decoy to modulate hnRNP E2 regulation of mRNA translation in leukemic blasts. *Cell* 140, 652–665.

Elmén, J., Lindow, M., Silahtaroglu, A., Bak, M., Christensen, M., Lind-Thomsen, A., et al. (2007). Antagonism of microRNA-122 in mice by systemically administered LNA-antimiR leads to up-regulation of a large set of predicted target mRNAs in the liver. *Nucleic Acids Res.* 36, 1153–1162.

Emmert-Streib, F., Dehmer, M., and Haibe-Kains, B. (2014). Gene regulatory networks and their applications: understanding biological and medical problems in terms of networks. *Front. Cell Dev. Biol.* 2, 38.

Ender, C., Krek, A., Friedländer, M. R., Beitzinger, M., Weinmann, L., Chen, W., et al. (2008). A Human snoRNA with microRNA-like functions. *Mol. Cell* 32, 519–528.

Englund, A., Kovanen, L., Saarikoski, S. T., Haukka, J., Reunanen, A., Aromaa, A., et al. (2009). *NPAS2* and *PER2* are linked to risk factors of the metabolic syndrome. *J. Circadian Rhythms* 7, 5.

Ernst, C. W., and Steibel, J. P. (2013). Molecular advances in QTL discovery and application in pig breeding. *Trends Genet.* 29, 215–224.

Esau, C., Davis, S., Murray, S. F., Yu, X. X., Pandey, S. K., Pear, M., et al. (2006). miR-122 regulation of lipid metabolism revealed by in vivo antisense targeting. *Cell Metab.* 3, 87–

98.

Esfandyari, H., Sørensen, A. C., and Bijma, P. (2015). Maximizing crossbred performance through purebred genomic selection. *Genet. Sel. Evol.* 47, 16.

Estany, J., Ros-Freixedes, R., Tor, M., and Pena, R. N. (2014). A functional variant in the stearoyl-CoA desaturase gene promoter enhances fatty acid desaturation in pork. *PLoS One* 9, e86177.

Evans, G. J., Giuffra, E., Sanchez, A., Kerje, S., Davalos, G., Vidal, O., et al. (2003). Identification of quantitative trait loci for production traits in commercial pig populations. *Genetics* 164, 621–627.

Everett, L. J., and Lazar, M. A. (2014). Nuclear receptor Rev-erbα: up, down, and all around. *Trends Endocrinol. Metab.* 25, 586–592.

Evers, M., Huttner, M., Dueck, A., Meister, G., and Engelmann, J. C. (2015). miRA: adaptable novel miRNA identification in plants using small RNA sequencing data. *BMC Bioinformatics* 16, 370.

Faith, J. J., Hayete, B., Thaden, J. T., Mogno, I., Wierzbowski, J., Cottarel, G., et al. (2007). Large-scale mapping and validation of *Escherichia coli* transcriptional regulation from a compendium of expression profiles. *PLoS Biol.* 5, e8.

Fang, H., Wu, Y., Narzisi, G., O'Rawe, J. A., Barrón, L. T. J., Rosenbaum, J., et al. (2014). Reducing INDEL calling errors in whole genome and exome sequencing data. *Genome Med.* 6, 89.

Fang, W., and Bartel, D. P. (2015). The menu of features that define primary microRNAs and enable *de novo* design of microRNA genes. *Mol. Cell* 60, 131–45.

Feng, J., Xing, W., and Xie, L. (2016). Regulatory roles of microRNAs in diabetes. *Int. J. Mol. Sci.* 17, 1729.

Fernandez, N., Cordiner, R. A., Young, R. S., Hug, N., MacIas, S., and Cáceres, J. F. (2017). Genetic variation and RNA structure regulate microRNA biogenesis. *Nat. Commun.* 8, 15114.

Fiannaca, A., La Rosa, M., La Paglia, L., Rizzo, R., and Urso, A. (2016). MiRNATIP: a SOM-based miRNA-target interactions predictor. *BMC Bioinformatics*. 17, 321.

Fontanesi, L., Galimberti, G., Calò, D. G., Fronza, R., Martelli, P. L., Scotti, E., et al. (2012). Identification and association analysis of several hundred single nucleotide polymorphisms within candidate genes for back fat thickness in Italian Large White pigs using a selective genotyping approach1. *J. Anim. Sci.* 90, 2450–2464.

Frank, F., Sonenberg, N., and Nagar, B. (2010). Structural basis for 5′-nucleotide base-specific recognition of guide RNA by human AGO2. *Nature* 465, 818–822.

Frantz, L. A. F., Schraiber, J. G., Madsen, O., Megens, H. J., Bosse, M., Paudel, Y., et al. (2013). Genome sequencing reveals fine scale diversification and reticulation history during speciation in Sus. *Genome Biol.* 14, R107.

Frantz, L. A. F., Schraiber, J. G., Madsen, O., Megens, H. J., Cagan, A., Bosse, M., et al. (2015). Evidence of long-term gene flow and selection during domestication from analyses of Eurasian wild and domestic pig genomes. *Nat. Genet.* 47, 1141–1148.

Friedländer, M. R., Chen, W., Adamidi, C., Maaskola, J., Einspanier, R., Knespel, S., et al. (2008). Discovering microRNAs from deep sequencing data using miRDeep. *Nat. Biotechnol.* 26, 407–15.

Friedländer, M. R., MacKowiak, S. D., Li, N., Chen, W., and Rajewsky, N. (2012). MiRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res.* 40, 37–52.

Friedman, N., Linial, M., Nachman, I., and Pe’er, D. (2000). Using Bayesian networks to analyze expression data. *J. Comput. Biol.* 601–620.

Friedman, R. C., Farh, K. K. H., Burge, C. B., and Bartel, D. P. (2009). Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res.* 19, 92–105.

Fromm, B., Billipp, T., Peck, L. E., Johansen, M., Tarver, J. E., King, B. L., et al. (2015). A uniform system for the annotation of vertebrate microRNA genes and the evolution of the human microRNAome. *Annu. Rev. Genet.* 49, 213–242.

Fromm, B., Domanska, D., Høye, E., Ovchinnikov, V., Kang, W., Aparicio-Puerta, E., et al. (2019). MirGeneDB 2.0: the metazoan microRNA complement. *Nucleic Acids Res.* 48, D132-D141.

Fujii, J., Otsu, K., Zorzato, F., De Leon, S., Khanna, V. K., Weiler, J. E., et al. (1991).

Identification of a mutation in porcine ryanodine receptor associated with malignant hyperthermia. *Science.* 253, 448–451.

Gan, L., and Denecke, B. (2013). Profiling pre-microRNA and mature microRNA expressions using a single microarray and avoiding separate sample preparation. *Microarrays* 2, 24–33.

Garaulet, M., Lee, Y. C., Shen, J., Parnell, L. D., Arnett, D. K., Tsai, M. Y., et al. (2009). *CLOCK* genetic variation and metabolic syndrome risk: Modulation by monounsaturated fatty acids. *Am. J. Clin. Nutr.* 90, 1466–1475.

Garcia-Rios, A., Perez-Martinez, P., Delgado-Lista, J., Phillips, C. M., Gjelstad, I. M. F., Wright, J. W., et al. (2012). A Period 2 genetic variant interacts with plasma SFA to modify plasma lipid concentrations in adults with metabolic syndrome. *J. Nutr.* 142, 1213–1218.

Gardner, P. P., and Vinther, J. (2008). Mutation of miRNA target sequences during human evolution. *Trends Genet.* 24, 262–265.

Gatfield, D., Le Martelot, G., Vejnar, C. E., Gerlach, D., Schaad, O., Fleury-Olela, F., et al. (2009). Integration of microRNA miR-122 in hepatic circadian gene expression. *Genes Dev.* 23, 1313–1326.

Gerin, I., Bommer, G. T., McCoin, C. S., Sousa, K. M., Krishnan, V., and MacDougald, O. A. (2010). Roles for miRNA-378/378* in adipocyte gene expression and lipogenesis. *Am. J. Physiol. Endocrinol. Metab.* 299, E198-206.

Gerlach, D., Kriventseva, E. V, Rahman, N., Vejnar, C. E., and Zdobnov, E. M. (2009). miROrtho: computational survey of microRNA genes. *Nucleic Acids Res.* 37, D111-D117.

Gianola, D., De Los Campos, G., Hill, W. G., Manfredi, E., and Fernando, R. (2009). Additive genetic variability and the Bayesian alphabet. *Genetics* 183, 347–363.

Giuffra, E., Kijas, J. M. H., Amarger, V., Carlborg, Ö., Jeon, J. T., and Andersson, L. (2000). The origin of the domestic pig: Independent domestication and subsequent introgression. *Genetics* 154, 1785–1791.

Giuffra, E., Tuggle, C. K., and the FAANG Consortium (2019). Functional annotation of animal genomes (FAANG): Current achievements and roadmap. *Annu. Rev. Anim. Biosci.* 7, 65–88.

Giurato, G., De Filippo, M., Rinaldi, A., Hashim, A., Nassa, G., Ravo, M., et al. (2013). iMir: an integrated pipeline for high-throughput analysis of small non-coding RNA data obtained by smallRNA-Seq. *BMC Bioinformatics* 14, 362.

Gjerlaug-Enger, E., Aass, L., Ødegård, J., and Vangen, O. (2010). Genetic parameters of meat quality traits in two pig breeds measured by rapid methods. *Animal* 4, 1832–1843.

Gkirtzou, K., Tsamardinos, I., Tsakalides, P., and Poirazi, P. (2010). MatureBayes: A probabilistic algorithm for identifying the mature miRNA within novel precursors. *PLoS One* 5, e11843.

Goddard, M. (2009). Genomic selection: Prediction of accuracy and maximisation of long term response. *Genetica* 136, 245–257.

Goedeke, L., Rotllan, N., Canfrán-Duque, A., Aranda, J. F., Ramírez, C. M., Araldi, E., et al. (2015). MicroRNA-148a regulates LDL receptor and *ABCA1* expression to control circulating lipoprotein levels. *Nat. Med.* 21, 1280–1289.

Goedeke, L., Salerno, A., Ramírez, C. M., Guo, L., Allen, R. M., Yin, X., et al. (2014). Long-term therapeutic silencing of miR-33 increases circulating triglyceride levels and hepatic lipid accumulation in mice. *EMBO Mol. Med.* 6, 1133–1141.

Gong, J., Tong, Y., Zhang, H. M., Wang, K., Hu, T., Shan, G., et al. (2012). Genome-wide identification of SNPs in MicroRNA genes and the SNP effects on MicroRNA target binding and biogenesis. *Hum. Mutat.* 33, 254–263.

González-Peña, D., Knox, R. V., Macneil, M. D., and Rodriguez-Zas, S. L. (2015). Genetic gain and economic values of selection strategies including semen traits in three- and four-way crossbreeding systems for swine production. *J. Anim. Sci.* 93, 879–891.

González-Prendes, R., Mármol-Sánchez, E., Quintanilla, R., Castelló, A., Zidi, A., Ramayo-Caldas, Y., et al. (2019a). About the existence of common determinants of gene expression in the porcine liver and skeletal muscle. *BMC Genomics* 20, 518.

González-Prendes, R., Quintanilla, R., Cánovas, A., Manunza, A., Figueiredo Cardoso, T., Jordana, J., et al. (2017). Joint QTL mapping and gene expression analysis identify positional candidate genes influencing pork quality traits. *Sci. Rep.* 7, 39830.

González-Prendes, R., Quintanilla, R., Mármol-Sánchez, E., Pena, R. N., Ballester, M., Cardoso, T. F., et al. (2019b). Comparing the mRNA expression profile and the genetic determinism of intramuscular fat traits in the porcine *gluteus medius* and *longissimus dorsi* muscles. *BMC Genomics* 20, 170.

Grandjean, V., Gounon, P., Wagner, N., Martin, L., Wagner, K. D., Bernex, F., et al. (2009). The miR-124-Sox9 paramutation: RNA-mediated epigenetic control of embryonic and adult growth. *Development* 136, 3647–3655.

Griffiths-Jones, S., Hui, J. H. L., Marco, A., and Ronshaugen, M. (2011). MicroRNA evolution by arm switching. *EMBO Rep.* 12, 172–177.

Griffiths-Jones, S., Saini, H. K., Van Dongen, S., and Enright, A. J. (2008). miRBase: Tools for microRNA genomics. *Nucleic Acids Res.* 36, 154–158.

Grimaldi, B., Bellet, M. M., Katada, S., Astarita, G., Hirayama, J., Amin, R. H., et al. (2010). *PER2* controls lipid metabolism by direct regulation of *PPARγ. Cell Metab.* 12, 509–520.

Groenen, M. A. M., Archibald, A. L., Uenishi, H., Tuggle, C. K., Takeuchi, Y., Rothschild, M. F., et al. (2012). Analyses of pig genomes provide insight into porcine demography and evolution. *Nature* 491, 393–398.

Große-Brinkhaus, C., Jonas, E., Buschbell, H., Phatsara, C., Tesfaye, D., Jüngst, H., et al. (2010). Epistatic QTL pairs associated with meat quality and carcass composition traits in a porcine Duroc × Pietrain population. *Genet. Sel. Evol.* 42, 39.

Gu, H. F., Xiao, J. H., Niu, L. M., Wang, B., Ma, G. C., Dunn, D. W., et al. (2014). Adaptive evolution of the circadian gene timeout in insects. *Sci. Rep.* 4, 4212.

Gudyś, A., Szcześniak, M. W., Sikora, M., and Makałowska, I. (2013). HuntMi: an efficient and taxon-specific approach in pre-miRNA identification. *BMC Bioinformatics* 14, 83.

Guo, T., Gao, J., Yang, B., Yan, G., Xiao, S., Zhang, Z., et al. (2019). A whole genome sequence association study of muscle fiber traits in a White Duroc × Erhualian F_2 resource population. *Asian-Australasian J. Anim. Sci.* 33, 704-711.

Ha, M., and Kim, V. N. (2014). Regulation of microRNA biogenesis. *Nat. Rev. Mol. Cell Biol.* 15, 509–524.

Haas, B. J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P. D., Bowden, J., et al. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* 8, 1494–1512.

Häberle, J., Pauli, S., Schmidt, E., Schulze-Eilfing, B., Berning, C., and Koch, H. G. (2003). Mild citrullinemia in Caucasians is an allelic variant of argininosuccinate synthetase deficiency (citrullinemia type 1). *Mol. Genet. Metab.* 80, 302–306.

Habier, D., Fernando, R. L., Kizilkaya, K., and Garrick, D. J. (2011). Extension of the bayesian alphabet for genomic selection. *BMC Bioinformatics* 12, 186.

Hackenberg, M., Sturm, M., Langenberger, D., Falcón-Pérez, J. M., and Aransay, A. M. (2009). miRanalyzer: a microRNA detection and analysis tool for next-generation sequencing experiments. *Nucleic Acids Res.* 37, W68–W76.

Häggman, J., and Uimari, P. (2017). Novel harmful recessive haplotypes for reproductive traits in pigs. *J. Anim. Breed. Genet.* 134, 129–135.

Han, H. S., Kang, G., Kim, J. S., Choi, B. H., and Koo, S. H. (2016). Regulation of glucose metabolism from a liver-centric perspective. *Exp. Mol. Med.* 48, e218.

Han, H., Gu, S., Chu, W., Sun, W., Wei, W., Dang, X., et al. (2017). miR-17-5p regulates differential expression of *NCOA3* in pig intramuscular and subcutaneous adipose tissue. *Lipids* 52, 939–949.

Hansen, T. B., Venø, M. T., Kjems, J., and Damgaard, C. K. (2014). miRdentify: high stringency miRNA predictor identifies several novel animal miRNAs. *Nucleic Acids Res.* 42, e124–e124.

Hasan, M. S., Wu, X., and Zhang, L. (2015). Performance evaluation of indel calling tools using real short-read data. *Hum. Genomics* 9, 20.

Hatlen, A., and Marco, A. (2020). Pervasive selection against microRNA target sites in human populations. *bioRxiv*, 420646.

Haury, A. C., Mordelet, F., Vera-Licona, P., and Vert, J. P. (2012). TIGRESS: Trustful inference of gene regulation using stability selection. *BMC Syst. Biol.* 6, 145.

Hayes, B., and Goddard, M. E. (2001). The distribution of the effects of genes affecting quantitative traits in livestock. *Genet. Sel. Evol.* 33, 209.

Helmy, M., Hatlen, A., and Marco, A. (2019). The impact of population variation in the analysis of microRNA target sites. *Non-coding RNA* 5, 42.

Helvik, S. A., Snove, O., and Saetrom, P. (2007). Reliable prediction of Drosha processing sites improves microRNA gene prediction. *Bioinformatics* 23, 142–149.

Hertel, J., and Stadler, P. F. (2006). Hairpins in a Haystack: recognizing microRNA precursors in comparative genomics data. *Bioinformatics* 22, e197–e202.

Holness, M. J., and Sugden, M. C. (2003). Regulation of pyruvate dehydrogenase complex activity by reversible phosphorylation. *Biochem. Soc. Trans.* 31, 1143–51.

Homer, H. A., McDougall, A., Levasseur, M., Yallop, K., Murdoch, A. P., and Herbert, M. (2005). Mad2 prevents aneuploidy and premature proteolysis of cyclin B and securin during meiosis I in mouse oocytes. *Genes Dev.* 19, 202–207.

Horie, T., Baba, O., Kuwabara, Y., Chujo, Y., Watanabe, S., Kinoshita, M., et al. (2012). MicroRNA-33 deficiency reduces the progression of atherosclerotic plaque in ApoE$^{-/-}$ mice. *J. Am. Heart Assoc.* 1, e003376.

Horie, T., Nishino, T., Baba, O., Kuwabara, Y., Nakao, T., Nishiga, M., et al. (2013). MicroRNA-33 regulates sterol regulatory element-binding protein 1 expression in mice. *Nat. Commun.* 4, 1–12.

Horodyska, J., Wimmers, K., Reyer, H., Trakooljul, N., Mullen, A. M., Lawlor, P. G., et al. (2018). RNA-seq of muscle from pigs divergent in feed efficiency and product quality identifies differences in immune response, growth, and macronutrient and connective tissue metabolism. *BMC Genomics* 19, 791.

Houston, R. D., Cameron, N. D., and Rance, K. A. (2004). A melanocortin-4 receptor (*MC4R*) polymorphism is associated with performance traits in divergently selected large white pig populations. *Anim. Genet.* 35, 386–390.

Hu, H. Y., Yan, Z., Xu, Y., Hu, H., Menzel, C., Zhou, Y. H., et al. (2009). Sequence features associated with microRNA strand selection in humans and flies. *BMC Genomics* 10, 413.

Hu, Z. L., Dracheva, S., Jang, W., Maglott, D., Bastiaansen, J., Rothschild, M. F., et al. (2005). A QTL resource and comparison tool for pigs: PigQTLDB. *Mamm. Genome* 16, 792–800.

Huang, T. H., Zhu, M. J., Li, X. Y., and Zhao, S. H. (2008). Discovery of porcine microRNAs and profiling from skeletal muscle tissues during development. *PLoS One* 3, e3225.

Huang, T.-H., Fan, B., Rothschild, M. F., Hu, Z.-L., Li, K., and Zhao, S.-H. (2007). MiRFinder: An improved approach and software implementation for genome-wide fast microRNA precursor scans. *BMC Bioinformatics* 8, 341.

Hui, J. H. L., Marco, A., Hunt, S., Melling, J., Griffiths-Jones, S., and Ronshaugen, M. (2013). Structure, evolution and function of the bi-directionally transcribed iab-4/iab-8 microRNA locus in arthropods. *Nucleic Acids Res.* 41, 3352–3361.

Hutvágner, G., McLachlan, J., Pasquinelli, A. E., Bálint, É., Tuschl, T., and Zamore, P. D. (2001). A cellular function for the RNA-interference enzyme dicer in the maturation of the let-7 small temporal RNA. *Science* 293, 834–838.

Huynh-Thu, V. A., Irrthum, A., Wehenkel, L., and Geurts, P. (2010). Inferring Regulatory Networks from Expression Data Using Tree-Based Methods. *PLoS One* 5, e12776.

Iacomino, G., and Siani, A. (2017). Role of microRNAs in obesity and obesity-related diseases. *Genes Nutr.* 12, 1–16.

Iwasaki, S., Kobayashi, M., Yoda, M., Sakaguchi, Y., Katsuma, S., Suzuki, T., et al. (2010). Hsc70/Hsp90 chaperone machinery mediates ATP-dependent RISC loading of small RNA duplexes. *Mol. Cell* 39, 292–299.

Iyer, M. K., Niknafs, Y. S., Malik, R., Singhal, U., Sahu, A., Hosono, Y., et al. (2015). The landscape of long noncoding RNAs in the human transcriptome. *Nat. Genet.* 47, 199–208.

Jaitin, D. A., Kenigsberg, E., Keren-Shaul, H., Elefant, N., Paul, F., Zaretsky, I., et al. (2014). Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science* 343, 776–779.

Jeon, J. T., Carlborg, O., Tornsten, A., Giuffra, E., Amarger, V., Chardon, P., et al. (1999). A paternally expressed QTL affecting skeletal and cardiac muscle mass in pigs maps to the *IGF2* locus. *Nat. Genet.* 21, 157–158.

Jeong, J. Y., Jeoung, N. H., Park, K.-G., and Lee, I.-K. (2012). Transcriptional regulation of pyruvate dehydrogenase kinase. *Diabetes Metab. J.* 36, 328–35.

Jha, A., and Shankar, R. (2013). miReader: Discovering novel miRNAs in species without sequenced genome. *PLoS One* 8, e66857.

Jiang, L., Zhang, J., Xuan, P., and Zou, Q. (2016). BP neural network could help improve pre-miRNA identification in various species. *Biomed Res. Int.* 9565689.

Jiang, P., Wu, H., Wang, W., Ma, W., Sun, X., and Lu, Z. (2007). MiPred: classification of real and pseudo microRNA precursors using random forest prediction model with combined features. *Nucleic Acids Res.* 35, W339-44.

Jiao, X., Sherman, B. T., Huang, D. W., Stephens, R., Baseler, M. W., Lane, H. C., et al. (2012). DAVID-WS: a stateful web service to facilitate gene/protein list analysis. *Bioinformatics* 28, 1805–1806.

John, B., Enright, A. J., Aravin, A., Tuschl, T., Sander, C., and Marks, D. S. (2004). Human microRNA targets. *PLoS Biol.* 2, e363.

John, E., Wienecke-Baldacchino, A., Liivrand, M., Heinäniemi, M., Carlberg, C., and Sinkkonen, L. (2012). Dataset integration identifies transcriptional regulation of microRNA genes by PPARγ in differentiating mouse 3T3-L1 adipocytes. *Nucleic Acids Res.* 40, 4446–4460.

Jonas, S., and Izaurralde, E. (2015). Towards a molecular understanding of microRNA-mediated gene silencing. *Nat. Rev. Genet.* 16, 421–433.

Kadri, S., Hinman, V., and Benos, P. V (2009). HHMMiR: efficient de novo prediction of microRNAs using hierarchical hidden Markov models. *BMC Bioinformatics* 10 Suppl 1, S35.

Kalsotra, A., Singh, R. K., Gurha, P., Ward, A. J., Creighton, C. J., and Cooper, T. A. (2014). The Mef2 transcription network is disrupted in myotonic dystrophy heart tissue, dramatically altering miRNA and mRNA expression. *Cell Rep.* 6, 336–345.

Kaneko, K. (2011). Proportionality between variances in gene expression induced by noise and mutation: Consequence of evolutionary robustness. *BMC Evol. Biol.* 11, 27.

Karagkouni, D., Paraskevopoulou, M. D., Chatzopoulos, S., Vlachos, I. S., Tastsoglou, S., Kanellos, I., et al. (2018). DIANA-TarBase v8: A decade-long collection of experimentally supported miRNA-gene interactions. *Nucleic Acids Res.* 46, D239–D245.

Kawamata, T., and Tomari, Y. (2010). Making RISC. *Trends Biochem. Sci.* 35, 368–376.

Kenny, N. J., Sin, Y. W., Hayward, A., Paps, J., Chu, K. H., and Hui, J. H. L. (2015). The phylogenetic utility and functional constraint of microRNA flanking sequences. *Proc. R. Soc. B Biol. Sci.* 282, 20142983.

Kertesz, M., Iovino, N., Unnerstall, U., Gaul, U., and Segal, E. (2007). The role of site accessibility in microRNA target recognition. *Nat. Genet.* 39, 1278–1284.

Khvorova, A., Reynolds, A., and Jayasena, S. D. (2003). Functional siRNAs and miRNAs exhibit strand bias. *Cell* 115, 209–216.

Kim, D. H., Saetrom, P., Snove, O., and Rossi, J. J. (2008a). MicroRNA-directed transcriptional gene silencing in mammalian cells. *Proc. Natl. Acad. Sci. U. S. A.* 105, 16230–16235.

Kim, D., Paggi, J. M., Park, C., Bennett, C., and Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* 37, 907–915.

Kim, H. J., Cui, X. S., Kim, E. J., Kim, W. J., and Kim, N. H. (2006). New porcine microRNA genes found by homology search. *Genome* 49, 1283–1286.

Kim, J., Cho, I. S., Hong, J. S., Choi, Y. K., Kim, H., and Lee, Y. S. (2008b). Identification and characterization of new microRNAs from pig. *Mamm. Genome* 19, 570–580.

Kim, K., Duc Nguyen, T., Li, S., and Anh Nguyen, T. (2018). SRSF3 recruits DROSHA to the basal junction of primary microRNAs. *RNA* 24, 892–898.

Kim, Y. K., Kim, B., and Kim, V. N. (2016). Re-evaluation of the roles of DROSHA, Exportin 5, and DICER in microRNA biogenesis. *Proc. Natl. Acad. Sci. U. S. A.* 113, E1881–E1889.

Kleftogiannis, D., Theofilatos, K., Likothanassis, S., and Mavroudi, S. (2015). YamiPred: A novel evolutionary method for predicting pre-miRNAs and selecting relevant features. *IEEE/ACM Trans. Comput. Biol. Bioinforma.* 12, 1183–1192.

Kogelman, L. J. A., Cirera, S., Zhernakova, D. V., Fredholm, M., Franke, L., and Kadarmideen, H. N. (2014). Identification of co-expression gene networks, regulatory genes and pathways for obesity based on adipose tissue RNA Sequencing in a porcine model. *BMC Med. Genomics* 7, 57.

Kojima, S., Gatfield, D., Esau, C. C., and Green, C. B. (2010). MicroRNA-122 modulates the rhythmic expression profile of the circadian deadenylase nocturnin in mouse liver. *PLoS One* 5, e11264.

Komurov, K., and Ram, P. T. (2010). Patterns of human gene expression variance show strong associations with signaling network hierarchy. *BMC Syst. Biol.* 4, 154.

Kovaka, S., Zimin, A. V., Pertea, G. M., Razaghi, R., Salzberg, S. L., and Pertea, M. (2019). Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biol.* 20, 278.

Kovanen, L., Donner, K., Kaunisto, M., and Partonen, T. (2015). *CRY1*, *CRY2* and *PRKCDBP* genetic variants in metabolic syndrome. *Hypertens. Res.* 38, 186–192.

Kozomara, A., Birgaoanu, M., and Griffiths-Jones, S. (2019). MiRBase: From microRNA sequences to function. *Nucleic Acids Res.* 47, D155–D162.

Krek, A., Grün, D., Poy, M. N., Wolf, R., Rosenberg, L., Epstein, E. J., et al. (2005). Combinatorial microRNA target predictions. *Nat. Genet.* 37, 495–500.

Krüger, J., and Rehmsmeier, M. (2006). RNAhybrid: MicroRNA target prediction easy, fast and flexible. *Nucleic Acids Res.* 34, W451-W454.

Kuleshov, M. V, Jones, M. R., Rouillard, A. D., Fernandez, N. F., Duan, Q., Wang, Z., et al. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* 44, W90–W97.

Kume, H., Hino, K., Galipon, J., and Ui-Tei, K. (2014). A-to-I editing in the miRNA seed region regulates target mRNA selection and silencing efficiency. *Nucleic Acids Res.* 42, 10050–60.

Kwon, S. C., Baek, S. C., Choi, Y.-G., Yang, J., Lee, Y., Woo, J.-S., et al. (2019). Molecular basis for the single-nucleotide precision of primary microRNA processing. *Mol. Cell* 73, 505-518.e5.

Lagos-Quintana, M., Rauhut, R., Lendeckel, W., and Tuschl, T. (2001). Identification of novel genes coding for small expressed RNAs. *Science* 294, 853–8.

Lai, E. C., Tomancak, P., Williams, R. W., and Rubin, G. M. (2003). Computational identification of *Drosophila* microRNA genes. *Genome Biol.* 4, R42.

Lamia, K. A., Papp, S. J., Yu, R. T., Barish, G. D., Uhlenhaut, N. H., Jonker, J. W., et al. (2011). Cryptochromes mediate rhythmic repression of the glucocorticoid receptor. *Nature* 480, 552–6.

Landgraf, D., Wang, L. L., Diemer, T., and Welsh, D. K. (2016). NPAS2 Compensates for loss of *CLOCK* in peripheral circadian oscillators. *PLoS Genet.* 12, e1005882.

Langfelder, P., and Horvath, S. (2008). WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics* 9, 559.

Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10, R25.

Larson, G., Cucchi, T., and Dobney, K. (2011). Genetic aspects of pig domestication. *The Genetics of the Pig: Second Edition*. *CABI*, 14–37.

Larson, G., Cucchi, T., Fujita, M., Matisoo-Smith, E., Robins, J., Anderson, A., et al. (2007). Phylogeny and ancient DNA of Sus provides insights into neolithic expansion in Island Southeast Asia and Oceania. *Proc. Natl. Acad. Sci. U. S. A.* 104, 4834–4839.

Latorre, P., Burgos, C., Hidalgo, J., Varona, L., Carrodeguas, J. A., and López-Buesa, P. (2016). C.A2456C-substitution in Pck1 changes the enzyme kinetic and functional properties modifying fat distribution in pigs. *Sci. Rep.* 6, 19617.

Lau, N. C., Lim, L. P., Weinstein, E. G., and Bartel, D. P. (2001). An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science* 294, 858–862.

Laurie, S. M., Cain, C. C., Lienhard, G. E., and Castle, J. D. (1993). The glucose transporter GluT4 and secretory carrier membrane proteins (SCAMPs) colocalize in rat adipocytes and partially segregate during insulin stimulation. *J. Biol. Chem.* 268, 19110–7.

Layeghifard, M., Rabani, R., Pirhaji, L., and Yakhchali, B. (2008). Evolutionary mechanisms underlying the functional divergence of duplicate genes involved in vertebrates' circadian rhythm pathway. *Gene* 426, 65–71.

Lee, R. C., and Ambros, V. (2001). An extensive class of small RNAs in *Caenorhabditis* *elegans*. *Science* 294, 862–864.

Lee, R. C., Feinbaum, R. L., and Ambros, V. (1993). The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* 75, 843–854.

Lee, Y., Kim, M., Han, J., Yeom, K.-H., Lee, S., Baek, S. H., et al. (2004). MicroRNA genes are transcribed by RNA polymerase II. *EMBO J.* 23, 4051–4060.

Lee, Y.-S., Sasaki, T., Kobayashi, M., Kikuchi, O., Kim, H.-J., Yokota-Hashimoto, H., et al. (2013). Hypothalamic ATF3 is involved in regulating glucose and energy metabolism in mice. *Diabetologia* 56, 1383–1393.

Legarra, A., Aguilar, I., and Misztal, I. (2009). A relationship matrix including full pedigree and genomic information. *J. Dairy Sci.* 92, 4656–4663.

Levy, S. E., and Myers, R. M. (2016). Advancements in next-generation sequencing. *Annu. Rev. Genomics Hum. Genet.* 17, 95–115.

Lewis, B. P., Burge, C. B., and Bartel, D. P. (2005). Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* 120, 15–20.

Li, G., Li, Y., Li, X., Ning, X., Li, M., and Yang, G. (2011). MicroRNA identity and abundance in developing swine adipose tissue as determined by solexa sequencing. *J. Cell. Biochem.* 112, 1318–1328.

Li, J., Liu, Y., Xin, X., Kim, T. S., Cabeza, E. A., Ren, J., et al. (2012). Evidence for positive selection on a number of microRNA regulatory interactions during recent human evolution. *PLoS Genet.* 8, e1002578.

Li, J., Yue, Z., Xiong, W., Sun, P., You, K., and Wang, J. (2017). *TXNIP* overexpression suppresses proliferation and induces apoptosis in SMMC7221 cells through ROS generation and MAPK pathway activation. *Oncol. Rep.* 37, 3369–3376.

Li, K., Qiu, C., Sun, P., Liu, D., Wu, T., Wang, K., et al. (2019a). Ets1-mediated acetylation of FoxO1 is critical for gluconeogenesis regulation during feed-fast cycles. *Cell Rep.* 26, 2998-3010.e5.

Li, Q., Yu, X., Chaudhary, R., Slebos, R. J. C., Chung, C. H., and Wang, X. (2019b). lncDIFF: a novel quasi-likelihood method for differential expression analysis of non-coding RNA. *BMC Genomics* 20, 539.

Li, S., Nguyen, T. D., Nguyen, T. L., and Nguyen, T. A. (2020). Mismatched and wobble base pairs govern primary microRNA processing by human Microprocessor. *Nat. Commun.* 11, 1926.

Li, X., Kim, S. W., Choi, J. S., Lee, Y. M., Lee, C. K., Choi, B. H., et al. (2010). Investigation of porcine *FABP3* and *LEPR* gene polymorphisms and mRNA expression for variation in intramuscular fat content. *Mol. Biol. Rep.* 37, 3931–3939.

Liao, Y., Smyth, G. K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923–930.

Lim, L. P., Lau, N. C., Weinstein, E. G., Abdelhakim, A., Yekta, S., Rhoades, M. W., et al. (2003). The microRNAs of *Caenorhabditis elegans*. *Genes Dev.* 17, 991–1008.

Lin, E., Kuo, P. H., Liu, Y. L., Yang, A. C., Kao, C. F., and Tsai, S. J. (2017). Effects of circadian clock genes and healthrelated behavior on metabolic syndrome in a Taiwanese population: Evidence from association and interaction analysis. *PLoS One* 12, e0173861.

Ling, L., Zhang, S.-H., Zhi, L.-D., Li, H., Wen, Q.-K., Li, G., et al. (2018). MicroRNA-30e promotes hepatocyte proliferation and inhibits apoptosis in cecal ligation and puncture-induced sepsis through the JAK/STAT signaling pathway by binding to FOSL2. *Biomed. Pharmacother.* 104, 411–419.

Liu, B., Fang, L., Chen, J., Liu, F., and Wang, X. (2015). miRNA-dis: microRNA precursor identification based on distance structure status pairs. *Mol. Biosyst.* 11, 1194–204.

Liu, G., Jennen, D. G. J., Tholen, E., Juengst, H., Kleinwächter, T., Hölker, M., et al. (2007). A genome scan reveals QTL for growth, fatness, leanness and meat quality in a Duroc-Pietrain resource population. *Anim. Genet.* 38, 241–52.

Liu, J., Carmell, M. A., Rivas, F. V., Marsden, C. G., Thomson, J. M., Song, J. J., et al. (2004). Argonaute2 is the catalytic engine of mammalian RNAi. *Science* 305, 1437–1441.

Liu, W., and Wang, X. (2019). Prediction of functional microRNA targets by integrative modeling of microRNA binding and target expression data. *Genome Biol.* 20, 18.

Liu, X., Li, Y. I., and Pritchard, J. K. (2019a). Trans effects on gene expression can drive omnigenic inheritance. *Cell* 177, 1022-1034.e6.

Liu, X., Zhou, L., Xie, X., Wu, Z., Xiong, X., Zhang, Z., et al. (2019b). Muscle glycogen level and occurrence of acid meat in commercial hybrid pigs are regulated by two low-frequency causal variants with large effects and multiple common variants with small effects. *Genet. Sel. Evol.* 51, 46.

Lopes, I. de, Schliep, A., and de L. F. de Carvalho, A. P. (2016). Automatic learning of pre-miRNAs from different species. *BMC Bioinformatics* 17, 224.

Lorenz, R., Bernhart, S. H., Höner zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P. F., et al. (2011). ViennaRNA Package 2.0. *Algorithms Mol. Biol.* 6, 26.

Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550.

Lu, Y., Baras, A. S., and Halushka, M. K. (2018). miRge 2.0 for comprehensive analysis of microRNA sequencing data. *BMC Bioinformatics* 19, 275.

Ludwig, N., Leidinger, P., Becker, K., Backes, C., Fehlmann, T., Pallasch, C., et al. (2016). Distribution of miRNA expression across human tissues. *Nucleic Acids Res.* 44, 3865–3877.

Lund, E., Güttinger, S., Calado, A., Dahlberg, J. E., and Kutay, U. (2004). Nuclear export of microRNA precursors. *Science* 303, 95–98.

Lynn, F. C. (2009). Meta-regulation: microRNA regulation of glucose and lipid metabolism. *Trends Endocrinol. Metab.* 20, 452–459.

MacArthur, D. G., Balasubramanian, S., Frankish, A., Huang, N., Morris, J., Walter, K., et al. (2012). A systematic survey of loss-of-function variants in human protein-coding genes. *Science* 335, 823–8.

MacLennan, D. H., Duff, C., Zorzato, F., Fujii, J., Phillips, M., Korneluk, R. G., et al. (1990). Ryanodine receptor gene is a candidate for predisposition to malignant hyperthermia. *Nature* 343, 559–561.

Makino, T., Rubin, C.-J., Carneiro, M., Axelsson, E., Andersson, L., and Webster, M. T. (2018). Elevated proportions of deleterious genetic variation in domestic animals and plants. *Genome Biol. Evol.* 10, 276–290.

Malek, M., Dekkers, J. C. M., Lee, H. K., Baas, T. J., and Rothschild, M. F. (2001). A molecular genome scan analysis to identify chromosomal regions influencing economic traits in the pig. I. Growth and body composition. *Mamm. Genome* 12, 630–636.

Maragkakis, M., Reczko, M., Simossis, V. A., Alexiou, P., Papadopoulos, G. L., Dalamagas, T., et al. (2009). DIANA-microT web server: elucidating microRNA functions through target prediction. *Nucleic Acids Res.* 37, W273–W276.

Marchini, J., and Howie, B. (2010). Genotype imputation for genome-wide association studies. *Nat. Rev. Genet.* 11, 499–511.

Marco, A. (2015). Selection against maternal microRNA target sites in maternal transcripts. *G3 Genes, Genomes, Genet.* 5, 2199–2207.

Margolin, A. A., Nemenman, I., Basso, K., Wiggins, C., Stolovitzky, G., Favera, R. D., et al. (2006). ARACNE: An algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinformatics* 7, S7.

Markham, N. R., and Zuker, M. (2008). UNAFold: software for nucleic acid folding and hybridization. *Methods Mol. Biol.* 453, 3–31.

Marquart, T. J., Allen, R. M., Ory, D. S., and Baldán, Á. (2010). miR-33 links SREBP-2 induction to repression of sterol transporters. *Proc. Natl. Acad. Sci. U. S. A.* 107, 12228–12232.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17, 10.

Mathelier, A., and Carbone, A. (2010). MIReNA: finding microRNAs with high accuracy and no learning at genome scale and from deep sequencing data. *Bioinformatics* 26, 2226–2234.

Mayya, V. K., and Duchaine, T. F. (2015). On the availability of microRNA-induced silencing complexes, saturation of microRNA-binding sites and stoichiometry. *Nucleic Acids Res.* 43, 7556–65.

McCarthy, D. J., Chen, Y., and Smyth, G. K. (2012). Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res.* 40, 4288–4297.

McDaneld, T. G., Smith, T. P. L., Doumit, M. E., Miles, J. R., Coutinho, L. L., Sonstegard, T. S., et al. (2009). MicroRNA transcriptome profiles during swine skeletal muscle development. *BMC Genomics* 10, 77.

McDaneld, T. G., Smith, T. P. L., Harhay, G. P., and Wiedmann, R. T. (2012). Next-generation sequencing of the porcine skeletal muscle transcriptome for computational prediction of microRNA gene targets. *PLoS One* 7, e42039.

Mentzel, C. M. J., Alkan, F., Keinicke, H., Jacobsen, M. J., Gorodkin, J., Fredholm, M., et al. (2016). Joint profiling of miRNAs and mRNAs reveals miRNA mediated gene regulation in the Göttingen minipig obesity model. *PLoS One* 11, e0167285.

Mentzel, C. M. J., Anthon, C., Jacobsen, M. J., Karlskov-Mortensen, P., Bruun, C. S., Jørgensen, C. B., et al. (2015). Gender and obesity specific microRNA expression in adipose tissue from lean and obese pigs. *PLoS One* 10, e0131650.

Meuwissen, T., Hayes, B., and Goddard, M. (2013). Accelerating improvement of livestock with genomic selection. *Annu. Rev. Anim. Biosci.* 1, 221–237.

Meyer, P. E., Lafitte, F., and Bontempi, G. (2008). Minet: A r/bioconductor package for inferring large transcriptional networks using mutual information. *BMC Bioinformatics* 9, 461.

Mi, H., Muruganujan, A., Huang, X., Ebert, D., Mills, C., Guo, X., et al. (2019). Protocol update for large-scale genome and gene function analysis with the PANTHER classification system (v.14.0). *Nat. Protoc.* 14, 703–721.

Milan, D., Jeon, J. T., Looft, C., Amarger, V., Robic, A., Thelander, M., et al. (2000). A mutation in *PRKAG3* associated with excess glycogen content in pig skeletal muscle. *Science* 288, 1248–1251.

Miller, J. N., and Pearce, D. A. (2014). Nonsense-mediated decay in genetic disease: Friend or foe? *Mutat. Res. Rev. Mutat. Res.* 762, 52–64.

Miranda, K. C., Huynh, T., Tay, Y., Ang, Y. S., Tam, W. L., Thomson, A. M., et al. (2006). A pattern-based method for the identification of microRNA binding sites and their corresponding heteroduplexes. *Cell* 126, 1203–1217.

Müller, S., Rycak, L., Winter, P., Kahl, G., Koch, I., and Rotter, B. (2013). omiRas: A web server for differential expression analysis of miRNAs derived from small RNA-Seq data. *Bioinformatics* 29, 2651–2652.

Muñoz, M., García-Casco, J. M., Caraballo, C., Fernández-Barroso, M. Á., Sánchez-Esquiliche, F., Gómez, F., et al. (2018). Identification of candidate genes and regulatory factors underlying intramuscular fat content through *longissimus dorsi* transcriptome analyses in heavy Iberian pigs. *Front. Genet.* 9, 608.

Muñoz, M., Rodríguez, M. C., Alves, E., Folch, J. M., Ibañez-Escriche, N., Silió, L., et al. (2013). Genome-wide analysis of porcine backfat and intramuscular fat fatty acid composition using high-density genotyping and expression data. *BMC Genomics* 14, 845.

Nagamine, Y., Haley, C. S., Sewalem, A., and Visschert, P. M. (2003). Quantitative trait loci variation for growth and obesity between and within lines of pigs (*Sus scrofa*). *Genetics* 164, 629–635.

Najafi-Shoushtari, S. H., Kristo, F., Li, Y., Shioda, T., Cohen, D. E., Gerszten, R. E., et al. (2010). MicroRNA-33 and the SREBP host genes cooperate to control cholesterol homeostasis. *Science* 328, 1566–1569.

Nam, J.-W., Kim, J., Kim, S.-K., and Zhang, B.-T. (2006). ProMiR II: A web server for the probabilistic prediction of clustered, nonclustered, conserved and nonconserved microRNAs. *Nucleic Acids Res.* 34, W455-8.

Nam, J.-W., Shin, K.-R., Han, J., Lee, Y., Kim, V. N., and Zhang, B.-T. (2005). Human microRNA prediction through a probabilistic co-learning model of sequence and structure. *Nucleic Acids Res.* 33, 3570–81.

Neeteson-van Nieuwenhoven, A.-M., Knap, P., and Avendaño, S. (2013). The role of sustainable commercial pig and poultry breeding for food security. *Anim. Front.* 3, 52–57.

Neilsen, C. T., Goodall, G. J., and Bracken, C. P. (2012). IsomiRs – the overlooked repertoire in the dynamic microRNAome. *Trends Genet.* 28, 544–549.

Nelson, P. T., Baldwin, D. A., Kloosterman, W. P., Kauppinen, S., Plasterk, R. H. A., and Mourelatos, Z. (2006). RAKE and LNA-ISH reveal microRNA expression and localization in archival human brain. *RNA* 12, 187–91.

Nguyen, T. A., Jo, M. H., Choi, Y. G., Park, J., Kwon, S. C., Hohng, S., et al. (2015). Functional anatomy of the human microprocessor. *Cell* 161, 1374–1387.

Nguyen, T. L., Nguyen, T. D., Bao, S., Li, S., and Nguyen, T. A. (2020). The internal loops in the lower stem of primary microRNA transcripts facilitate single cleavage of human Microprocessor. *Nucleic Acids Res.* 48, 2579-2593.

Nica, A. C., and Dermitzakis, E. T. (2013). Expression quantitative trait loci: Present and future. *Philos. Trans. R. Soc. B Biol. Sci.* 368, 20120362.

Nii, M., Hayashi, T., Mikawa, S., Tani, F., Niki, A., Mori, N., et al. (2005). Quantitative trait loci mapping for meat quality and muscle fiber traits in a Japanese wild boar x Large White intercross. *J. Anim. Sci.* 83, 308–315.

Nowakowski, T. J., Rani, N., Golkaram, M., Zhou, H. R., Alvarado, B., Huch, K., et al. (2018). Regulation of cell-type-specific transcriptomes by microRNA networks during human brain development. *Nat. Neurosci.* 21, 1784–1792.

Obayashi, T., and Kinoshita, K. (2009). Rank of correlation coefficient as a comparable measure for biological significance of gene coexpression. *DNA Res.* 16, 249–260.

Omariba, G., Xu, F., Wang, M., Li, K., Zhou, Y., and Xiao, J. (2020). Genome-wide analysis of microRNA-related single nucleotide polymorphisms (SNPs) in mouse genome. *Sci. Rep.* 10, 1–9.

Opgen-Rhein, R., and Strimmer, K. (2007). From correlation to causation networks: A simple approximate learning algorithm and its application to high-dimensional plant gene expression data. *BMC Syst. Biol.* 1, 37.

Osorio, D., Yu, X., Zhong, Y., Li, G., Serpedin, E., Huang, J. Z., et al. (2019). Single-cell expression variability implies cell function. *Cells* 9, 14.

Oulas, A., Boutla, A., Gkirtzou, K., Reczko, M., Kalantidis, K., and Poirazi, P. (2009). Prediction of novel microRNA genes in cancer-associated genomic regions--a combined computational and experimental approach. *Nucleic Acids Res.* 37, 3276–87.

Óvilo, C., Benítez, R., Fernández, A., Núñez, Y., Ayuso, M., Fernández, A. I., et al. (2014). *Longissimus dorsi* transcriptome analysis of purebred and crossbred Iberian pigs differing in muscle characteristics. *BMC Genomics* 15, 413.

Óvilo, C., Fernández, A., Fernández, A. I., Folch, J. M., Varona, L., Benítez, R., et al. (2010). Hypothalamic expression of porcine leptin receptor (*LEPR*), neuropeptide y (*NPY*), and cocaine- and amphetamine-regulated transcript (*CART*) genes is influenced by *LEPR* genotype. *Mamm. Genome* 21, 583–591.

Ozgur, S., Basquin, J., Kamenska, A., Filipowicz, W., Standart, N., and Conti, E. (2015). Structure of a human 4E-T/DDX6/CNOT1 complex reveals the different interplay of DDX6-binding proteins with the CCR4-NOT complex. *Cell Rep.* 13, 703–711.

Paicu, C., Mohorianu, I., Stocks, M., Xu, P., Coince, A., Billmeier, M., et al. (2017). miRCat2: accurate prediction of plant and animal microRNAs from next-generation sequencing datasets. *Bioinformatics* 33, 2446–2454.

Pan, X., Zhang, Y., Wang, L., and Mahmood Hussain, M. (2010). Diurnal regulation of MTP and plasma triglyceride by *CLOCK* is mediated by *SHP*. *Cell Metab.* 12, 174–186.

Pant, S. D., Karlskov-Mortensen, P., Jacobsen, M. J., Cirera, S., Kogelman, L. J. A., Bruun, C. S., et al. (2015). Comparative analyses of QTLs influencing obesity and metabolic phenotypes in pigs and humans. *PLoS One* 10, e0137356.

Paraskevopoulou, M. D., Georgakilas, G., Kostoulas, N., Vlachos, I. S., Vergoulis, T., Reczko, M., et al. (2013). DIANA-microT web server v5.0: Service integration into miRNA functional analysis workflows. *Nucleic Acids Res.* 41, W169–W173.

Park, C. Y., Choi, Y. S., and McManus, M. T. (2010). Analysis of microRNA knockouts in mice. *Hum. Mol. Genet.* 19, R169–R175.

Partin, A. C., Ngo, T. D., Herrell, E., Jeong, B. C., Hon, G., and Nam, Y. (2017). Heme enables proper positioning of Drosha and DGCR8 on primary microRNAs. *Nat. Commun.* 8, 1737.

Pasquinelli, A. E., Reinhart, B. J., Slack, F., Martindale, M. Q., Kuroda, M. I., Maller, B., et al. (2000). Conservation of the sequence and temporal expression of let-7 heterochronic regulatory RNA. *Nature* 408, 86–89.

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., and Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* 14, 417–419.

Pausch, H., Schwarzenbacher, H., Burgstaller, J., Flisikowski, K., Wurmser, C., Jansen, S., et al. (2015). Homozygous haplotype deficiency reveals deleterious mutations compromising reproductive and rearing success in cattle. *BMC Genomics* 16, 312.

Peng, Y., and Croce, C. M. (2016). The role of microRNAs in human cancer. *Signal Transduct. Target. Ther.* 1, 1–9.

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

Peñagaricano, F., Valente, B. D., Steibel, J. P., Bates, R. O., Ernst, C. W., Khatib, H., et al. (2015). Exploring causal networks underlying fat deposition and muscularity in pigs through the integration of phenotypic, genotypic and transcriptomic data. *BMC Syst. Biol.* 9, 58.

Pérez-Enciso, M., Clop, A., Noguera, J. L., Óvilo, C., Coll, A., Folch, J. M., et al. (2000). A QTL on pig chromosome 4 affects fatty acid metabolism: Evidence from an Iberian by Landrace intercross. *J. Anim. Sci.* 78, 2525–2531.

Pérez-Montarelo, D., Fernández, A., Barragán, C., Noguera, J. L., Folch, J. M., Rodríguez, M. C., et al. (2013). Transcriptional characterization of porcine leptin and leptin receptor genes. *PLoS One* 8, e66398.

Pérez-Montarelo, D., Madsen, O., Alves, E., Rodríguez, M. C., Folch, J. M., Noguera, J. L., et al. (2014). Identification of genes regulating growth and fatness traits in pig through hypothalamic transcriptome analysis. *Physiol. Genomics* 46, 195–206.

Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T.-C., Mendell, J. T., and Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* 33, 290–295.

Peterson, K. J., Dietrich, M. R., and McPeek, M. A. (2009). MicroRNAs and metazoan macroevolution: insights into canalization, complexity, and the Cambrian explosion. *BioEssays* 31, 736–747.

Pilcher, C. M., Jones, C. K., Schroyen, M., Severin, A. J., Patience, J. F., Tuggle, C. K., et al. (2015). Transcript profiles in *longissimus dorsi* muscle and subcutaneous adipose tissue: A comparison of pigs with different postweaning growth rates. *J. Anim. Sci.* 93, 2134–2143.

Pimentel, H., Bray, N. L., Puente, S., Melsted, P., and Pachter, L. (2017). Differential analysis of RNA-seq incorporating quantification uncertainty. *Nat. Methods* 14, 687–690.

Pla, A., Zhong, X., and Rayner, S. (2018). miRAW: A deep learning-based approach to predict microRNA targets by analyzing whole microRNA transcripts. *PLoS Comput. Biol.* 14, e1006185.

Podolska, A., Kaczkowski, B., Busk, P. K., Søkilde, R., Litman, T., Fredholm, M., et al. (2011). Microrna expression profiling of the porcine developing. *PLoS One* 6, e14494.

Ponsuksili, S., Du, Y., Murani, E., Schwerin, M., and Wimmers, K. (2012). Elucidating molecular networks that either affect or respond to plasma cortisol concentration in target tissues of liver and muscle. *Genetics* 192, 1109–1122.

Ponsuksili, S., Jonas, E., Murani, E., Phatsara, C., Srikanchai, T., Walz, C., et al. (2008). Trait correlated expression combined with expression QTL analysis reveals biological pathways and candidate genes affecting water holding capacity of muscle. *BMC Genomics* 9, 367.

Ponsuksili, S., Murani, E., Brand, B., Schwerin, M., and Wimmers, K. (2011). Integrating expression profiling and whole-genome association for dissection of fat traits in a porcine model. *J. Lipid Res.* 52, 668–678.

Ponsuksili, S., Murani, E., Schwerin, M., Schellander, K., and Wimmers, K. (2010). Identification of expression QTL (eQTL) of genes expressed in porcine M. *longissimus dorsi* and associated with meat quality traits. *BMC Genomics* 11, 572.

Ponsuksili, S., Murani, E., Trakooljul, N., Schwerin, M., and Wimmers, K. (2014). Discovery of candidate genes for muscle traits based on GWAS supported by eQTL-analysis. *Int. J. Biol. Sci.* 10, 327–337.

Pritchard, J. K., and Przeworski, M. (2001). Linkage disequilibrium in humans: Models and data. *Am. J. Hum. Genet.* 69, 1–14.

Puig-Oliveras, A., Ramayo-Caldas, Y., Corominas, J., Estellé, J., Pérez-Montarelo, D., Hudson, N. J., et al. (2014). Differences in muscle transcriptome among pigs phenotypically extreme for fatty acid composition. *PLoS One* 9, e99720.

Puig-Oliveras, A., Revilla, M., Castelló, A., Fernández, A. I., Folch, J. M., and Ballester, M.

(2016). Expression-based GWAS identifies variants, gene interactions and key regulators affecting intramuscular fatty acid content and composition in porcine meat. *Sci. Rep.* 6, 31803.

Qian, K., Auvinen, E., Greco, D., and Auvinen, P. (2012). miRSeqNovel: An R based workflow for analyzing miRNA sequencing data. *Mol. Cell. Probes* 26, 208–211.

Rahman, M. E., Islam, R., Islam, S., Mondal, S. I., and Amin, M. R. (2012). MiRANN: A reliable approach for improved classification of precursor microRNA using Artificial Neural Network model. *Genomics* 99, 189–194.

Raj, A., and van Oudenaarden, A. (2008). Nature, nurture, or chance: Stochastic gene expression and its consequences. *Cell* 135, 216–226.

Ramayo-Caldas, Y., Ballester, M., Sánchez, J. P., González-Rodríguez, O., Revilla, M., Reyer, H., et al. (2018). Integrative approach using liver and duodenum RNA-Seq data identifies candidate genes and pathways associated with feed efficiency in pigs. *Sci. Rep.* 8, 558.

Ramayo-Caldas, Y., Mercadé, A., Castelló, A., Yang, B., Rodríguez, C., Alves, E., et al. (2012). Genome-wide association study for intramuscular fatty acid composition in an Iberian × Landrace cross. *J. Anim. Sci.* 90, 2883–2893.

Ramos, A. M., Crooijmans, R. P. M. A., Affara, N. A., Amaral, A. J., Archibald, A. L., Beever, J. E., et al. (2009). Design of a high density SNP genotyping assay in the pig using SNPs identified and characterized by next generation sequencing technology. *PLoS One* 4, e6524.

Ramos-Onsins, S. E., Burgos-Paz, W., Manunza, A., and Amills, M. (2014). Mining the pig genome to investigate the domestication process. *Heredity* 113, 471–484.

Ran, D., and Daye, Z. J. (2017). Gene expression variability and the analysis of large-scale RNA-seq studies with the MDSeq. *Nucleic Acids Res.* 45, e127–e127.

Rausell, A., Mohammadi, P., McLaren, P. J., Bartha, I., Xenarios, I., Fellay, J., et al. (2014). Analysis of stop-gain and frameshift variants in human innate immunity genes. *PLoS Comput. Biol.* 10, e1003757.

Rayner, K. J., Suárez, Y., Dávalos, A., Parathath, S., Fitzgerald, M. L., Tamehiro, N., et al.

(2010). MiR-33 contributes to the regulation of cholesterol homeostasis. *Science* 328, 1570–1573.

Reddy, T. E., Gertz, J., Pauli, F., Kucera, K. S., Varley, K. E., Newberry, K. M., et al. (2012). Effects of sequence variation on differential allelic transcription factor occupancy and gene expression. *Genome Res.* 22, 860–869.

Reichholf, B., Herzog, V. A., Fasching, N., Manzenreither, R. A., Sowemimo, I., and Ameres, S. L. (2019). Time-resolved small RNA sequencing unravels the molecular principles of microRNA homeostasis. *Mol. Cell* 75, 756-768.e7.

Reimand, J., Arak, T., Adler, P., Kolberg, L., Reisberg, S., Peterson, H., et al. (2016). g:Profiler—a web server for functional interpretation of gene lists (2016 update). *Nucleic Acids Res.* 44, W83–W89.

Reinhart, B. J., Slack, F. J., Basson, M., Pasquinelli, A. E., Bettinger, J. C., Rougvie, A. E., et al. (2000). The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. *Nature* 403, 901–906.

Reverter, A., and Chan, E. K. F. (2008). Combining partial correlation and an information theory approach to the reversed engineering of gene co-expression networks. *Bioinformatics* 24, 2491–2497.

Reverter, A., Hudson, N. J., Nagaraj, S. H., Pérez-Enciso, M., and Dalrymple, B. P. (2010). Regulatory impact factors: unraveling the transcriptional regulation of complex traits from expression data. *Bioinformatics* 26, 896–904.

Revilla, M., Ramayo-Caldas, Y., Castelló, A., Corominas, J., Puig-Oliveras, A., Ibáñez-Escriche, N., et al. (2014). New insight into the SSC8 genetic determination of fatty acid composition in pigs. *Genet. Sel. Evol.* 46, 28.

Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–40.

Roden, C., Gaillard, J., Kanoria, S., Rennie, W., Barish, S., Cheng, J., et al. (2017). Novel determinants of mammalian primary microRNA processing revealed by systematic evaluation of hairpin-containing transcripts and human genetic variation. *Genome Res.* 27, 374–384.

Rohrer, G. A., Thallman, R. M., Shackelford, S., Wheeler, T., and Koohmaraie, M. (2006). A genome scan for loci affecting pork quality in a Duroc-Landrace F2 population. *Anim. Genet.* 37, 17–27.

Ros-Freixedes, R., Gol, S., Pena, R. N., Tor, M., Ibáñez-Escriche, N., Dekkers, J. C. M., et al. (2016). Genome-wide association study singles out *SCD* and *LEPR* as the two main loci influencing intramuscular fat content and fatty acid composition in duroc pigs. *PLoS One* 11, e0152496.

Rothschild, M. F., Ruvinsky, A., and C.A.B. International. (2011). *The genetics of the pig*. *Second edition*. *CABI*.

Rotllan, N., Price, N., Pati, P., Goedeke, L., and Fernández-Hernando, C. (2016). microRNAs in lipoprotein metabolism and cardiometabolic disorders. *Atherosclerosis* 246, 352–360.

Ruby, J. G., Jan, C. H., and Bartel, D. P. (2007). Intronic microRNA precursors that bypass Drosha processing. *Nature* 448, 83–86.

Rückert, C., Stratz, P., Preuss, S., and Bennewitz, J. (2012). Mapping quantitative trait loci for metabolic and cytological fatness traits of connected F2 crosses in pigs. *J. Anim. Sci.* 90, 399–409.

Rueda, A., Barturen, G., Lebrón, R., Gómez-Martín, C., Alganza, Á., Oliver, J. L., et al. (2015). sRNAtoolbox: an integrated collection of small RNA research tools. *Nucleic Acids Res.* 43, W467–W473.

Ryan, B. M., Robles, A. I., and Harris, C. C. (2010). Genetic variation in microRNA networks: The implications for cancer research. *Nat. Rev. Cancer* 10, 389–402.

Saçar, D. M. D. (2019). MicroRNA prediction based on 3D graphical representation of RNA secondary structures. *Turkish J. Biol.* 43, 274–280.

Saçar, D. M. D., Baumbach, J., and Allmer, J. (2017). On the performance of pre-microRNA detection algorithms. *Nat. Commun.* 8, 330.

Saçar, D. M. D., Hamzeiy, H., and Allmer, J. (2013). Can miRBase provide positive data for machine learning for the detection of miRNA hairpins? *J. Integr. Bioinform.* 10, 1–11.

Sadeghi, B., Ahmadi, H., Azimzadeh-Jamalkandi, S., Nassiri, M. R., and Masoudi-Nejad, A. (2014). BosFinder: a novell pre-microRNA gene prediction algorithm in *Bos taurus*.

Anim. Genet. 45, 479–484.

Sales, G., Coppe, A., Bisognin, A., Biasiolo, M., Bortoluzzi, S., and Romualdi, C. (2010). MAGIA, a web-based tool for miRNA and genes integrated analysis. *Nucleic Acids Res.* 38, W352–W359.

Samorè, A. B., and Fontanesi, L. (2016). Genomic selection in pigs: state of the art and perspectives. *Ital. J. Anim. Sci.* 15, 211–232.

Sampedro Castañeda, M., Zanoteli, E., Scalco, R. S., Scaramuzzi, V., Marques Caldas, V., Conti Reed, U., et al. (2018). A novel *ATP1A2* mutation in a patient with hypokalaemic periodic paralysis and CNS symptoms. *Brain* 141, 3308–3318.

Sato, S., Uemoto, Y., Kikuchi, T., Egawa, S., Kohira, K., Saito, T., et al. (2017). Genome-wide association studies reveal additional related loci for fatty acid composition in a Duroc pig multigenerational population. *Anim. Sci. J.* 88, 1482–1490.

Saunders, M. A., Liang, H., and Li, W. H. (2007). Human polymorphism at microRNAs and microRNA target sites. *Proc. Natl. Acad. Sci. U. S. A.* 104, 3300–3305.

Sawera, M., Gorodkin, J., Cirera, S., and Fredholm, M. (2005). Mapping and expression studies of the mir17-92 cluster on pig Chromosome 11. *Mamm. Genome* 16, 594–598.

Schäfer, J., and Strimmer, K. (2005). An empirical Bayes approach to inferring large-scale gene association networks. *Bioinformatics* 21, 754–764.

Schaid, D. J., Chen, W., and Larson, N. B. (2018). From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nat. Rev. Genet.* 19, 491–504.

Schirle, N. T., Sheu-Gruttadauria, J., Chandradoss, S. D., Joo, C., and MacRae, I. J. (2015). Water-mediated recognition of t1-adenosine anchors Argonaute2 to microRNA targets. *Elife* 4, e07646.

Schulman, B. R. M., Esquela-Kerscher, A., and Slack, F. J. (2005). Reciprocal expression of *lin - 41* and the microRNAs *let - 7* and *mir - 125* during mouse embryogenesis. *Dev. Dyn.* 234, 1046–1054.

Scott, E. M., Carter, A. M., and Grant, P. J. (2008). Association between polymorphisms in the *Clock* gene, obesity and the metabolic syndrome in man. *Int. J. Obes.* 32, 658–662.

Seemann, S. E., Anthon, C., Palasca, O., and Gorodkin, J. (2015). Quality assessment of domesticated animal genome assemblies. *Bioinform. Biol. Insights* 9, 49–58.

Sempere, L. F., Sokol, N. S., Dubrovsky, E. B., Berger, E. M., and Ambros, V. (2003). Temporal regulation of microRNA expression in *Drosophila melanogaster* mediated by hormonal signals and broad-complex gene activity. *Dev. Biol.* 259, 9–18.

Sen, K., and Ghosh, T. C. (2013). Pseudogenes and their composers: delving in the ‘debris’ of human genome. *Brief. Funct. Genomics* 12, 536–547.

Serre, D., Gurd, S., Ge, B., Sladek, R., Sinnett, D., Harmsen, E., et al. (2008). Differential allelic expression in the human genome: A robust approach to identify genetic and epigenetic Cis-acting mechanisms regulating gene expression. *PLoS Genet.* 4, e1000006.

Sewer, A., Paul, N., Landgraf, P., Aravin, A., Pfeffer, S., Brownstein, M. J., et al. (2005). Identification of clustered microRNAs using an ab initio prediction method. *BMC Bioinformatics* 6, 267.

Sheikh Hassani, M., and Green, J. R. (2019). Multi-view co-training for microRNA prediction. *Sci. Rep.* 9, 10931.

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

Sheng, Y., Engström, P. G., and Lenhard, B. (2007). Mammalian microRNA prediction through a Support Vector Machine model of sequence and structure. *PLoS One* 2, e946.

Shimba, S., Ishii, N., Ohta, Y., Ohno, T., Watabe, Y., Hayashi, M., et al. (2005). Brain and muscle Arnt-like protein-1 (*BMAL1*), a component of the molecular clock, regulates adipogenesis. *Proc. Natl. Acad. Sci. U. S. A.* 102, 12071–12076.

Shiromoto, Y., Kuramochi-Miyagawa, S., Nagamori, I., Chuma, S., Arakawa, T., Nishimura, T., et al. (2019). GPAT2 is required for piRNA biogenesis, transposon silencing, and maintenance of spermatogonia in mice. *Biol. Reprod.* 101, 248–256.

Shostak, A., Meyer-Kovac, J., and Oster, H. (2013). Circadian regulation of lipid mobilization in white adipose tissues. *Diabetes* 62, 2195–2203.

Silva, K. M., Bastiaansen, J. W. M., Knol, E. F., Merks, J. W. M., Lopes, P. S., Guimarães, S. E. F., et al. (2011). Meta-analysis of results from quantitative trait loci mapping studies on pig chromosome 4. *Anim. Genet.* 42, 280–292.

Simkin, A., Geissler, R., McIntyre, A. B. R., and Grimson, A. (2020). Evolutionary dynamics of microRNA target sites across vertebrate evolution. *PLoS Genet.* 16, e1008285.

Söllner, J. F., Leparc, G., Hildebrandt, T., Klein, H., Thomas, L., Stupka, E., et al. (2017). An RNA-Seq atlas of gene expression in mouse and rat normal tissues. *Sci. Data* 4, 170185.

Sookoian, S., Gianotti, T. F., Burgueno, A., and Pirola, C. J. (2010). Gene-gene interaction between serotonin transporter (*SLC6A4*) and clock modulates the risk of metabolic syndrome in rotating shiftworkers. *Chronobiol. Int.* 27, 1202–1218.

Stegmayer, G., Di Persia, L. E., Rubiolo, M., Gerard, M., Pividori, M., Yones, C., et al. (2018). Predicting novel microRNA: a comprehensive comparison of machine learning approaches. *Brief. Bioinform.* 20, 1607-1620.

Stegmayer, G., Yones, C., Kamenetzky, L., and Milone, D. H. (2017). High class-imbalance in pre-miRNA prediction: A novel approach based on deepSOM. *IEEE/ACM Trans. Comput. Biol. Bioinforma.* 14, 1316–1326.

Su, G., Christensen, O. F., Ostersen, T., Henryon, M., and Lund, M. S. (2012). Estimating additive and non-additive genetic variances and predicting genetic merits using genome-wide dense single nucleotide polymorphism markers. *PLoS One* 7, e45293.

Suhail, M. (2010). Na+, K+-ATPase: Ubiquitous multifunctional transmembrane protein and its relevance to various pathophysiological conditions. *J. Clin. Med. Res.* 2, 1-17.

Sun, G., Yan, J., Noltner, K., Feng, J., Li, H., Sarkis, D. A., et al. (2009). SNPs in human miRNA genes affect biogenesis and function. *RNA* 15, 1640–1651.

Sun, Z., Evans, J., Bhagwate, A., Middha, S., Bockol, M., Yan, H., et al. (2014). CAP-miRSeq: A comprehensive analysis pipeline for microRNA sequencing data. *BMC Genomics* 15, 423.

Suntsova, M., Gaifullin, N., Allina, D., Reshetun, A., Li, X., Mendeleeva, L., et al. (2019). Atlas of RNA sequencing profiles for normal human tissues. *Sci. Data* 6, 36.

Suzuki, K., Ishida, M., Kadowaki, H., Shibata, T., Uchida, H., and Nishida, A. (2006). Genetic correlations among fatty acid compositions in different sites of fat tissues, meat production, and meat quality traits in Duroc pigs. *J. Anim. Sci.* 84, 2026–2034.

Tang, X., and Sun, Y. (2019). Fast and accurate microRNA search using CNN. *BMC*

Bioinformatics 20, 646.

Tani, H., Mizutani, R., Salam, K. A., Tano, K., Ijiri, K., Wakamatsu, A., et al. (2012). Genome-wide determination of RNA stability reveals hundreds of short-lived noncoding transcripts in mammals. *Genome Res.* 22, 947–956.

Taniguchi, M., Nakajima, I., Chikuni, K., Kojima, M., Awata, T., and Mikawa, S. (2014). MicroRNA-33b downregulates the differentiation and development of porcine preadipocytes. *Mol. Biol. Rep.* 41, 1081–1090.

Tanzer, A., and Stadler, P. F. (2004). Molecular evolution of a microRNA cluster. *J. Mol. Biol.* 339, 327–335.

Tarazona, S., Furió-Tarí, P., Turrà, D., Pietro, A. Di, Nueda, M. J., Ferrer, A., et al. (2015). Data quality aware analysis of differential expression in RNA-seq with NOISeq R/Bioc package. *Nucleic Acids Res.* 43, e140–e140.

Tav, C., Tempel, S., Poligny, L., and Tahi, F. (2016). miRNAFold: A web server for fast miRNA precursor prediction in genomes. *Nucleic Acids Res.* 44, W181–W184.

Tenesa, A., Navarro, P., Hayes, B. J., Duffy, D. L., Clarke, G. M., Goddard, M. E., et al. (2007). Recent human effective population size estimated from linkage disequilibrium. *Genome Res.* 17, 520–526.

Terai, G., Komori, T., Asai, K., and Kin, T. (2007). miRRim: A novel system to find conserved miRNAs with high sensitivity and specificity. *RNA* 13, 2081–90.

Thomas, J., Thomas, S., and Sael, L. (2017). DP-miRNA: An improved prediction of precursor microRNA using deep learning model. *IEEE International Conference on Big Data and Smart Computing, BigComp 2017*.

Thompson, D., Regev, A., and Roy, S. (2015). Comparative analysis of gene regulatory networks: From network reconstruction to evolution. *Annu. Rev. Cell Dev. Biol.* 31, 399–428.

Thorslund, C. A. H., Aaslyng, M. D., and Lassen, J. (2017). Perceived importance and responsibility for market-driven pig welfare: Literature review. *Meat Sci.* 125, 37–45.

Torres-Rovira, L., Astiz, S., Caro, A., Lopez-Bote, C., Ovilo, C., Pallares, P., et al. (2012). Diet-induced swine model with obesity/leptin resistance for the study of metabolic syndrome and type 2 diabetes. *Sci. World J.* 2012, 1–8.

Tortereau, F., Gilbert, H., Heuven, H. C., Bidanel, J. P., Groenen, M. A. M., and Riquet, J. (2010). Combining two Meishan F2 crosses improves the detection of QTL on pig chromosomes 2, 4 and 6. *Genet. Sel. Evol.* 42, 42.

Tran, V. D. T., Tempel, S., Zerath, B., Zehraoui, F., and Tahi, F. (2015). miRBoost: boosting support vector machines for microRNA precursor classification. *RNA* 21, 775–85.

Truesdell, S. S., Mortensen, R. D., Seo, M., Schroeder, J. C., Lee, J. H., Letonqueze, O., et al. (2012). MicroRNA-mediated mRNA translation activation in quiescent cells and oocytes involves recruitment of a nuclear microRNP. *Sci. Rep.* 2, 842.

Tsuzaki, K., Kotani, K., Sano, Y., Fujiwara, S., Takahashi, K., and Sakane, N. (2010). The association of the *Clock* 3111 T/C SNP with lipids and lipoproteins including small dense low-density lipoprotein: Results from the Mima study. *BMC Med. Genet.* 11,150.

Tu, Y., Stolovitzky, G., and Klein, U. (2002). Quantitative noise analysis for gene expression microarray experiments. *Proc. Natl. Acad. Sci. U. S. A.* 99, 14031–14036.

Ulitsky, I., and Bartel, D. P. (2013). lincRNAs: genomics, evolution, and mechanisms. *Cell* 154, 26–46.

Van Eenennaam, A. L., Weigel, K. A., Young, A. E., Cleveland, M. A., and Dekkers, J. C. M. (2014). Applied animal genomics: Results from the field. *Annu. Rev. Anim. Biosci.* 2, 105–139.

VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91, 4414–4423.

VanRaden, P. M., Olson, K. M., Null, D. J., and Hutchison, J. L. (2011). Harmful recessive effects on fertility detected by absence of homozygous haplotypes. *J. Dairy Sci.* 94, 6153–6161.

Vaquerizas, J. M., Kummerfeld, S. K., Teichmann, S. A., and Luscombe, N. M. (2009). A census of human transcription factors: Function, expression and evolution. *Nat. Rev. Genet.* 10, 252–263.

Vasudevan, S., Tong, Y., and Steitz, J. A. (2007). Switching from repression to activation: MicroRNAs can up-regulate translation. *Science* 318, 1931–1934.

Ventura, A., Young, A. G., Winslow, M. M., Lintault, L., Meissner, A., Erkeland, S. J., et al. (2008). Targeted deletion reveals essential and overlapping functions of the miR-17∼92 family of miRNA clusters. *Cell* 132, 875–886.

Verardo, L. L., Silva, F. F., Lopes, M. S., Madsen, O., Bastiaansen, J. W. M., Knol, E. F., et al. (2016). Revealing new candidate genes for reproductive traits in pigs: Combining Bayesian GWAS and functional pathways. *Genet. Sel. Evol.* 48, 9.

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

Villaverde, A. F., Becker, K., and Banga, J. R. (2018). PREMER: A Tool to Infer Biological Networks. *IEEE/ACM Trans. Comput. Biol. Bioinforma.* 15, 1193–1202.

Villaverde, A. F., Ross, J., Morán, F., and Banga, J. R. (2014). MIDER: Network inference with mutual information distance and entropy reduction. *PLoS One* 9, e96732.

Viñuela, A., Brown, A. A., Buil, A., Tsai, P.-C., Davies, M. N., Bell, J. T., et al. (2017). Age-dependent changes in mean and variance of gene expression across tissues in a twin cohort. *Hum. Mol. Genet.* 27, 732–741.

Vitsios, D. M., Kentepozidou, E., Quintais, L., Benito-Gutiérrez, E., van Dongen, S., Davis, M. P., et al. (2017). Mirnovo: Genome-free prediction of microRNAs from small RNA sequencing data and single-cells using decision forests. *Nucleic Acids Res.* 45, e177–e177.

Vlachos, I. S., Vergoulis, T., Paraskevopoulou, M. D., Lykokanellos, F., Georgakilas, G., Georgiou, P., et al. (2016). DIANA-mirExTra v2.0: Uncovering microRNAs and transcription factors with crucial roles in NGS expression data. *Nucleic Acids Res.* 44, W128–W134.

Võsa, U., Claringbould, A., Westra, H.-J., Bonder, M. J., Deelen, P., Zeng, B., et al. (2018). Unraveling the polygenic architecture of complex traits using blood eQTL metaanalysis. *bioRxiv*, 447367.

Waller, A. P., Kohler, K., Burns, T. A., Mudge, M. C., Belknap, J. K., and Lacombe, V. A. (2011). Naturally occurring compensated insulin resistance selectively alters glucose transporters in visceral and subcutaneous adipose tissues without change in AS160 activation. *Biochim. Biophys. Acta - Mol. Basis Dis.* 1812, 1098–1103.

Wang, X. (2008). miRDB: A microRNA target prediction and functional annotation database with a wiki interface. *RNA* 14, 1012–1017.

Warr, A., Robert, C., Hume, D., Archibald, A. L., Deeb, N., and Watson, M. (2015). Identification of low-confidence regions in the pig reference genome (Sscrofa10.2). *Front. Genet.* 6, 338–338.

Watanabe, T., Takeda, A., Mise, K., Okuno, T., Suzuki, T., Minami, N., et al. (2005). Stage-specific expression of microRNAs during *Xenopus* development. *FEBS Lett.* 579, 318–324.

Watson-Haigh, N. S., Kadarmideen, H. N., and Reverter, A. (2010). PCIT: An R package for weighted gene co-expression networks based on partial correlation and information theory approaches. *Bioinformatics* 26, 411–413.

Wei, W., Li, B., Liu, K., Jiang, A., Dong, C., Jia, C., et al. (2018). Identification of key microRNAs affecting drip loss in porcine *longissimus dorsi* by RNA-Seq. *Gene* 647, 276–282.

Wernersson, R., Schierup, M. H., Jørgensen, F. G., Gorodkin, J., Panitz, F., Stærfeldt, H. H., et al. (2005). Pigs in sequence space: A 0.66× coverage pig genome survey based on shotgun sequencing. *BMC Genomics* 6, 70.

Westra, H. J., and Franke, L. (2014). From genome to function by studying eQTLs. *Biochim. Biophys. Acta - Mol. Basis Dis.* 1842, 1896–1902.

Wheeler, B. M., Heimberg, A. M., Moy, V. N., Sperling, E. A., Holstein, T. W., Heber, S., et al. (2009). The deep evolution of metazoan microRNAs. *Evol. Dev.* 11, 50–68.

White, S. (2011). From globalized pig breeds to capitalist pigs: A study in animal cultures and evolutionary history. *Environ. Hist.* 16, 94–120.

Wilk, G., and Braun, R. (2018). regQTLs: Single nucleotide polymorphisms that modulate microRNA regulation of gene expression in tumors. *PLoS Genet.* 14, e1007837.

Wolf, J. B. W. (2013). Principles of transcriptome analysis and gene expression quantification: An RNA-seq tutorial. *Mol. Ecol. Resour.* 13, 559–572.

Wood, J. D., Enser, M., Fisher, A. V., Nute, G. R., Sheard, P. R., Richardson, R. I., et al.

(2008). Fat deposition, fatty acid composition and meat quality: A review. *Meat Sci.* 78, 343–358.

Woods, N. T., Baskin, R., Golubeva, V., Jhuraney, A., De-Gregoriis, G., Vaclova, T., et al. (2016). Functional assays provide a robust tool for the clinical annotation of genetic variants of uncertain significance. *npj Genomic Med.* 1, 1–9.

Wrann, C. D., Eguchi, J., Bozec, A., Xu, Z., Mikkelsen, T., Gimble, J., et al. (2012). FOSL2 promotes leptin gene expression in human and mouse adipocytes. *J. Clin. Invest.* 122, 1010–1021.

Wu, C., Bardes, E. E., Jegga, A. G., and Aronow, B. J. (2014a). ToppMiR: ranking microRNAs and their mRNA targets based on biological functions and context. *Nucleic Acids Res.* 42, W107–W113.

Wu, J., Liu, Q., Wang, X., Zheng, J., Wang, T., You, M., et al. (2013). mirTools 2.0 for non-coding RNA discovery, profiling, and functional annotation based on high-throughput sequencing. *RNA Biol* 10, 1087–1092.

Wu, J., Xiao, J., Zhang, Z., Wang, X., Hu, S., and Yu, J. (2014b). Ribogenomics: The Science and Knowledge of RNA. *Genomics, Proteomics Bioinforma.* 12, 57–63.

Wu, Y., Wei, B., Liu, H., Li, T., and Rayner, S. (2011). MiRPara: A SVM-based software tool for prediction of most probable microRNA coding regions in genome scale sequences. *BMC Bioinformatics* 12, 107.

Xi, Y., Liu, H., Zhao, Y., Li, J., Li, W., Liu, G., et al. (2018). Comparative analyses of longissimus muscle miRNAomes reveal microRNAs associated with differential regulation of muscle fiber development between Tongcheng and Yorkshire pigs. *PLoS One* 13, e0200445.

Xie, S. S., Li, X. Y., Liu, T., Cao, J. H., Zhong, Q., and Zhao, S. H. (2011). Discovery of porcine microRNAs in multiple tissues by a solexa deep sequencing approach. *PLoS One* 6, e16235.

Xie, S., Li, X., Qian, L., Cai, C., Xiao, G., Jiang, S., et al. (2019). An integrated analysis of mRNA and miRNA in skeletal muscle from myostatin-edited Meishan pigs. *Genome* 62, 305–315.

Xing, K., Zhao, X., Ao, H., Chen, S., Yang, T., Tan, Z., et al. (2019). Transcriptome analysis of miRNA and mRNA in the livers of pigs with highly diverged backfat thickness. *Sci. Rep.* 9, 16740.

Xu, L., Yang, X., Wu, L., Chen, X., Chen, L., and Tsai, F. S. (2019). Consumers' willingness to pay for food with information on animal welfare, lean meat essence detection, and traceability. *Int. J. Environ. Res. Public Health* 16, 3616.

Xue, C., Li, F., He, T., Liu, G.-P., Li, Y., and Zhang, X. (2005). Classification of real and pseudo microRNA precursors using local structure-sequence features and support vector machine. *BMC Bioinformatics* 6, 310.

Yang, B., Zhang, W., Zhang, Z., Fan, Y., Xie, X., Ai, H., et al. (2013). Genome-Wide association analyses for fatty acid composition in porcine muscle and abdominal fat tissues. *PLoS One* 8, e65554.

Yip, S. H., Sham, P. C., and Wang, J. (2018). Evaluation of tools for highly variable gene discovery from single-cell RNA-seq data. *Brief. Bioinform.* 20, 1583–1589.

Yones, C., Stegmayer, G., and Milone, D. H. (2018). Genome-wide pre-miRNA discovery from few labeled examples. *Bioinformatics* 34, 541–549.

Yoshizumi, S., Suzuki, S., Hirai, M., Hinokio, Y., Yamada, T., Yamada, T., et al. (2007). Increased hepatic expression of ganglioside-specific sialidase, *NEU3*, improves insulin sensitivity and glucose tolerance in mice. *Metabolism* 56, 420–429.

Yousef, M., Nebozhyn, M., Shatkay, H., Kanterakis, S., Showe, L. C., and Showe, M. K. (2006). Combining multi-species genomic data for microRNA identification using a Naive Bayes classifier. *Bioinformatics* 22, 1325–1334.

Yuan, T., Huang, X., Dittmar, R. L., Du, M., Kohli, M., Boardman, L., et al. (2014). ERNA: A graphic user interface-based tool optimized for large data analysis from high-throughput RNA sequencing. *BMC Genomics* 15, 176.

Zani, F., Breasson, L., Becattini, B., Vukolic, A., Montani, J. P., Albrecht, U., et al. (2013). *PER2* promotes glucose storage to liver glycogen during feeding and acute fasting by inducing Gys2 PTG and GL expression. *Mol. Metab.* 2, 292–305.

Zhang, C., Wang, Z., Bruce, H., Kemp, R. A., Charagu, P., Miar, Y., et al. (2015). Genome-wide association studies (GWAS) identify a QTL close to *PRKAG3* affecting meat pH and colour in crossbred commercial pigs. *BMC Genet.* 16, 33.

Zhang, H., Kolb, F. A., Jaskiewicz, L., Westhof, E., and Filipowicz, W. (2004). Single processing center models for human Dicer and bacterial RNase III. *Cell* 118, 57–68.

Zhang, S., Hulver, M. W., McMillan, R. P., Cline, M. A., and Gilbert, E. R. (2014). The pivotal role of pyruvate dehydrogenase kinases in metabolic flexibility. *Nutr. Metab.* 11, 10.

Zhang, W., Yang, B., Zhang, J., Cui, L., Ma, J., Chen, C., et al. (2016). Genome-wide association studies for fatty acid metabolic traits in five divergent pig populations. *Sci. Rep.* 6, 24718.

Zhang, X., Guo, J., Wei, X., Niu, C., Jia, M., Li, Q., et al. (2018). Bach1: Function, regulation and involvement in disease. *Oxid. Med. Cell. Longev.* 2018, 1347969.

Zhang, Y., Zhang, J., Gong, H., Cui, L., Zhang, W., Ma, J., et al. (2019). Genetic correlation of fatty acid composition with growth, carcass, fat deposition and meat quality traits based on GWAS data in six pig populations. *Meat Sci.* 150, 47–55.

Zhao, W., Liu, W., Tian, D., Tang, B., Wang, Y., Yu, C., et al. (2011). wapRNA: A web-based application for the processing of RNA sequences. *Bioinformatics* 27, 3076–3077.

Zhao, Y., Zhang, Y., Zhou, M., Wang, S., Hua, Z., and Zhang, J. (2012). Loss of mPer2 increases plasma insulin levels by enhanced glucose-stimulated insulin secretion and impaired insulin clearance in mice. *FEBS Lett.* 586, 1306–1311.

Zhou, S. S., Jin, J. P., Wang, J. Q., Zhang, Z. G., Freedman, J. H., Zheng, Y., et al. (2018). MiRNAS in cardiovascular diseases: Potential biomarkers, therapeutic targets and challenges review-article. *Acta Pharmacol. Sin.* 39, 1073–1084.

Zhu, Y. L., Chen, T., Xiong, J. L., Wu, D., Xi, Q. Y., Luo, J. Y., et al. (2018). miR-146b inhibits glucose consumption by targeting *IRS1* gene in porcine primary adipocytes. *Int. J. Mol. Sci.* 19, 783.

Zorc, M., Skok, D., Godnic, I., Calin, G. A., Horvat, S., Jiang, Z., et al. (2012). Catalog of microRNA seed polymorphisms in vertebrates. *PLoS One* 7, e30737.

Zou, Q., Mao, Y., Hu, L., Wu, Y., and Ji, Z. (2014). miRClassify: An advanced web server for miRNA family classification and annotation. *Comput. Biol. Med.* 45, 157–160.

CHAPTER VII. ANNEXES

All Supplementary Tables, Figures and related Documents included and referred in the published papers that form part of the present Ph.D. thesis are available at their corresponding online versions:

Paper I

https://onlinelibrary.wiley.com/doi/full/10.1111/age.12864

Paper II

https://onlinelibrary.wiley.com/doi/full/10.1111/age.12877

Paper III

https://www.nature.com/articles/s41598-019-45108-z

Paper IV

https://www.biorxiv.org/content/10.1101/2020.04.17.038315v1

Paper V

https://www.sciencedirect.com/science/article/pii/S0888754319304884

Paper VI

https://jasbsci.biomedcentral.com/articles/10.1186/s40104-019-0412-z

Additionally, all Supplementary Tables, Figures and related Documents are publicly available and can be downloaded from the following link:

https://figshare.com/projects/PhD_Thesis_Annexes_Emilio_M_rmol_S_nchez_/80777

Acknowledgements

En primer lugar, quisiera agradecer al dr. **Marcel Amills** por confiar en un candidato a PhD desconocido, entusiasta y un poco inocente, que en cierto momento pensó en dedicarse a la investigación. Muchas gracias por tu tiempo y tu paciencia, por tus aportaciones a mis trabajos y por las enseñanzas recibidas. Disculpas por las largas conversaciones con dudas y temores, por las discusiones sobre estrategia, estilo y forma y, sobre todo, por las necesarias correcciones sobre mi tendencia a concatenar y subordinar frases interminables, así como por realizar circunloquios y exposiciones complicadas, como quizás sea ejemplo esta frase. Todo ello ha contribuido a iniciarme en el mundo de la investigación en genética animal, donde he aprendido mucho más de lo que me gustaría admitir, pudiendo cimentar una base sólida para continuar con mi carrera investigadora, allá donde los azares del futuro decidan.

Agradecimientos, por supuesto, extensivo al resto de investigadores del grupo de Animal Genomics del CRAG. A los dres. **Alex Clop**, **Miguel Pérez**, **Sebastián Ramos** y **Josep María Folch**, gracias por vuestros sabios consejos sobre el mundo científico y los posibles futuros postdoctorales, por las charlas de sobremesa, así como por las conversaciones sesudas sobre ciencia.

Al personal de administración y técnicos del CRAG, y en especial a **Tania Sánchez**, gracias por hacernos las farragosas tareas burocráticas mucho más sencillas y amenas, el CRAG perdió una gran profesional con tu marcha.

A los dres. **Victoria Barja** y **Ernesto Llamas**, gracias por acompañarme y enseñarme los pormenores, sinsabores y alegrías de la representación de predoctorales en el CRAG. Vuestra ayuda, consejos y experiencia me ayudaron a desarrollar sin duda una mejor labor. Gracias, por extensión, a **Ferrán**, por tu paciencia y tu sabiduría acompañándome en los momentos difíciles, en las negociaciones y en las disputas. Nuestro esfuerzo mereció la pena compañero.

A los “miguelitos”, y en especial a **Miguel**, muchas gracias por los buenos momentos y las alegrías, por los consejos, por la compañía y por tu amistad. Nuestro siempre postpuesto café seguro que merecerá la pena para recordar aquellos buenos años. Un abrazo.

A todos mis compañeros de fatigas predoctorales en el grupo de Animal Genomics y de Population Statistics: A mis viejos compañeros y mentores en el grupo, dres. **Tainã F. Cardoso** y **Rayner González**, todo agradecimiento es poco por vuestra paciencia, vuestro

tiempo y vuestras enseñanzas. Gracias por las conversaciones, consejos, fiestas y alegrías. Desde la distancia, espero poder seguir compartiendo vuestra amistad, allá donde estemos.

A los nuevos integrantes del grupo, **Dailu** y **María**, mil gracias también por vuestra amistad, vuestro trabajo, consejos, enseñanzas y sabiduría. Seguramente me hayáis enseñado más vosotros a mí de lo que yo pueda haberos enseñado, no os quepa duda. Os deseo fuerza y el mejor de mis deseos con vuestras tesis doctorales, que seguro estarán en el top de mis preferidas. Siempre estaré al otro lado del teléfono o del email.

A los otros miembros pasados del grupo de Animal Genomics, cada uno forjando su propio camino. A los dres. **Daniel Crespo** y **Jordi Leno**, gracias por iniciarme en la bioinformática, en la que entré como elefante en cacharrería. Gracias por vuestra paciencia con un burdo principiante, y gracias, por supuesto, por los buenos momentos, las risas, los chismes y las legendarias fiestas de navidad. A los dres. **Manuel Revilla** y **Marta Gòdia**, mil gracias por vuestra sabiduría y por las largas discusiones sobre ciencia. Con total seguridad me habéis ayudado a ser mejor científico y persona. A la dr. **Antonia Noce**, por tu carisma y alegría, por tu experiencia, tu glamour y por tu ayuda en tiempos difíciles.

A **Lino**, y por extensión a **Moni** y **Abi**, me faltan los adjetivos para agradecer tu compañía, tu amistad, tu tiempo y tu experiencia. El genio sos vos, lo sabés. Mil gracias por las innumerables risas, por las fiestas y barbacoas, por las conversaciones sobre la vida y sobre cualquier otra cosa. Gracias por los abrazos y el cariño, siempre bienvenidos, incluso por las mañanas, aunque no lo parezca. Gracias por los ánimos y la confianza, aún en tiempos difíciles, cuando yo no creía, tú sí lo hiciste, y me ayudaste a superar los baches y dudas. Te debo una botella de whisky del caro, y ya sabes por qué.

A **Lourdes**, por tu incansable trabajo en el laboratorio, gracias por enseñarme buenas prácticas y disciplina fuera del entorno *in silico*. Aunque no lo hayamos comentado tanto como se debería, eres un gran referente de trabajo y esfuerzo para mí, y te admiro por ser tan profesional y aplicada en todo lo que te planteas. Gracias por tu apoyo, por las confidencias y por las conversaciones en los ratos muertos.

A **Elías**, gracias por enseñarme que el inconformismo y los principios éticos y políticos pueden y deben gobernar nuestras decisiones, también en ciencia. Seguiremos en la lucha.

A **Laura**, mil gracias por tu infinita sabiduría argentina, tus ánimos, tu fuerza y tu confianza, aun cuando yo no la tenía. Sos mi ídola, lo sabes y lo entiendes. Gracias por las conversaciones de sobremesa, por los sabios consejos, por tu voz de la experiencia. Nos quedó pendiente un trabajo, ya sabes, siempre dispuesto por mi parte a compartir ciencia contigo, qué mejor regalo que ese.

A las nuevas incorporaciones que tomaron y tomarán el relevo a nuestro trabajo, **Ioanna**, **Alice**, **Magí**, **Jesús**, **Yron** y **Lian**. Gracias particularmente a Jesús, Magí y María, por confiar en mi para preparar las clases, siendo yo igualmente inexperto y poco preparado, aunque con muchas ganas e ilusión por enseñar a nuestros pequeños novatos en la facultad. Many thanks Yron for your trust. Whatever would be our future, I am sure that you will find a successful path, in life and in science. Be yourself, everything gets better. I hope to keep collaborating, you know, I am always open to deal with those freaking microRNAs.

Al dr. **Yuliaxis Ramayo**, por su sabiduría y capacidad de trabajo, mil gracias por ofrecerme la oportunidad de colaborar contigo, así como por abrirme camino a nuevas oportunidades. Gracias también por los buenos momentos, las confidencias y los sabios consejos. A la dra. **María Ballester**, gracias por tu confianza, ánimos y buenas palabras y por abrirme camino a nuevos proyectos. Espero poder seguir colaborando contigo en el futuro, seguro conseguiremos darle una vuelta más a esos datos, que esconden más de lo que parece.

Al dr. **Albert Pla**, por tenerme tanta paciencia y por iniciarme en el mundo del machine learning, con sus más y sus menos. De esos inicios nacieron muchas ideas y proyectos.

A las técnicos de laboratorio, dra. **Anna Castelló**, dra. **Anna Mercadé** y **Betlem Cabrera**, gracias por vuestro trabajo y disponibilidad, vuestra ayuda y vuestras enseñanzas en mis primeros y torpes pasos en el laboratorio. Sin vuestro trabajo, mucho de lo aquí plasmado no habría sido posible.

A mis soportes en la Facultad de Veterinaria durante mis semanas de docencia, dra. **Natàlia Sastre** y **Àngels Domingo**. Gracias por hacerme más fácil mi incursión en la docencia, por vuestra alegría y profesionalidad. Sin vuestro trabajo y vuestra ayuda, todo habría sido mucho más difícil.

A la dra. **Susanna Cirera**, gracias por acogerme en Copenhague y darme la oportunidad de tener mi primera experiencia internacional durante mi tesis doctoral. Gracias también por tu paciencia, por tu alegría y tus sabios consejos, por enseñarme a tomarme las cosas con más perspectiva y por inculcarme disciplina y buen hacer en el laboratorio, aunque fuera a base de horas y de sacarme de mi zona de confort. Thanks also to all other members of the

Department of Animal Genetics at Copenhagen University for their warm welcoming and kindness, especially to drs. **Peter Karlskov-Mortensen** and **Merete Fredholm**.

To dr. **Dominique Rocha**, many thanks for allowing me to stay at INRAE facilities during that particularly hot summer. I had the opportunity to learn a lot about lncRNA biology and how to deal with those tricky non-coding transcripts. I hope to keep collaborating in the near future, and keep pushing forward our knowledge about those intriguing non-coding genes.

To drs. **Marc Friedländer** and **Love Dalen** at Stockholm University, many thanks for the invaluable opportunity to form part of your research teams as a future postdoctoral fellow. I hope we will have a fruitful period during my stay at Sweden and that we will collaborate to make great research together.

A todos los amigos en Barcelona que he ido recogiendo durante estos largos, y cortos, cuatro años de mi vida. Gracias **Miguel**, **Adri**, **Hugo**, **Carles**, **Rubén**, **Oriol**, **Uri**, **Sergio**, **Xavi, Xavi bo**, **Clau**, **Alex**, **Marc**, **Arnau**, **Iván** y tantas otras buenas personas que han pasado por mi vida durante mi estancia en Catalunya. Sin vosotros, nada habría sido lo mismo. Seguiremos en contacto, más cerca o más lejos, no tengo dudas sobre ello. Mil gracias por las risas, las fiestas, los buenos momentos, el cariño, la amistad, el amor y la confianza que me habéis brindado.

Por último, gracias a mis padres. Gracias **papá** por tu sabiduría, tu calma y tu saber estar, por tu apoyo y tus consejos en lo bueno y en lo malo. Y gracias **mamá**, gracias de todo corazón, por estar ahí siempre, por saber esperar, por entender, por querer sin condiciones, por tus consejos y tus palabras, en lo bueno y en lo malo. Por creer en mí más de lo que yo jamás lo hice. Por saberme entender aún sin hablar. Por apoyarme en mis decisiones, más o menos acertadas, siempre adelante.

Escribo estas líneas en tiempos difíciles, en tiempos de incertidumbre, donde el presente parece paralizarse y el futuro se muestra incierto. Pese a los obstáculos, seguiremos adelante, aquí o allá, lejos o cerca, siempre estaremos juntos, con fuerza y con decisión. Vivir, al fin y al cabo, sólo es una historia que escribimos a diario, sin un final predeterminado.