



UNIVERSITAT  
POLITÈCNICA  
DE VALÈNCIA

**Universitat Politècnica de València**

Escuela Técnica Superior de Ingeniería Agroalimentaria y del Medio Natural  
(ETSIAMN)

# OPTIMIZACIÓN DE LA EFICIENCIA REPRODUCTIVA EN CERDO IBÉRICO

---

TRABAJO DE FIN DE GRADO DE INGENIERIA  
AGROALIMENTARIA Y DEL MEDIO RURAL

Autora: **Dña. Zulema Rodríguez Hernández**

Tutora: Dra. Dña. Noelia Ibáñez-Escriche

Directora experimental: Dña. Cristina Casto Rebollo

Valencia, julio 2019



# Optimización de la eficiencia reproductiva en cerdo ibérico

## RESUMEN

El cerdo ibérico es una raza porcina autóctona de la península ibérica, especialmente adaptada a las condiciones geoclimáticas del suroeste español y con gran capacidad adipogénica. De hecho, es precisamente esta última característica la que permite la obtención de productos curados de máxima calidad y hace que sean animales muy valorados en el mercado. El aumento de la demanda de estos productos ha hecho aumentar el número de explotaciones intensivas. Sin embargo, en comparación con el porcino blanco la eficiencia reproductiva del cerdo ibérico es sensiblemente menor debido, principalmente, a su baja prolificidad. La prolificidad es un carácter regulado genéticamente y de gran importancia para la viabilidad económica del porcino, especialmente en ibérico. Hasta hace poco, en ibérico no había programas de mejora genética sistemáticos en los que la prolificidad fuese un objetivo claro de selección. Por tanto, esta raza sufre un retraso comparada con otras razas comerciales. Además, otro problema del cerdo ibérico es el bajo número de tetinas (casi fijado a diez) que dan importantes problemas de manejo en granjas de reproductoras donde se pretende aumentar la eficiencia reproductiva.

En la última década los grandes avances en tecnologías de genotipado han permitido la posibilidad de realizar análisis de asociación del genoma completo (GWAS) en relación con caracteres fenotípicos complejos como, en este caso, los reproductivos. Este tipo de análisis permite la identificación de las variantes genéticas asociadas con el carácter de interés, sin restricciones en cuanto a la localización genómica de las mutaciones. La identificación de variantes asociadas a los caracteres reproductivos puede posibilitar su uso para seleccionar los progenitores de manera más eficiente y aumentar la respuesta a la selección de estos caracteres considerablemente. El objetivo de este trabajo es identificar regiones genómicas asociadas a los caracteres prolificidad y número de tetinas en cerdo ibérico a través de un análisis de asociación del genoma completo.

**Palabras clave:** cerdo ibérico, GWAS, número de tetinas, eficiencia reproductiva, prolificidad.

Autora: **Dña. Zulema Rodríguez Hernández**

Tutora: Dra. Dña. Noelia Ibáñez-Escriche

Directora experimental: Dña. Cristina Casto Rebollo

**Valencia, julio 2019**

# Optimization of reproductive efficiency in Iberian pig

## ABSTRACT

The Iberian pig is an autochthonous breed of the Iberian Peninsula, specially adapted to the geoclimatic conditions of the Spanish Southwest and with great adipogenic capacity. In fact, this last characteristic allows to obtain the cured products of the highest quality and makes them highly valued animals in the market. The increasing demand of these products has caused an increase on the number of reproductive intensive farms. However, in contrast to white pigs, the reproductive efficiency of the Iberian pig is significantly lower, mainly due to its low prolificacy. The prolificacy is a genetically regulated trait and of great importance for the economic viability of the pig production, especially in Iberian. Until recently, in Iberian pig there were no systematic genetic selection programs in which prolificacy was a clear objective of selection. Therefore, this breed suffers a delay compared to other commercial breeds. In addition, another problem of the Iberian pig is the low number of teats (almost set to ten) which gives important management problems in breeding farms with a better reproductive efficiency.

In the last decade, the great advances in technologies of genotyping have allowed the possibility of carrying out genome wide association analysis (GWAS) with complex phenotypic traits as, in this case, the reproductive ones. This kind of analysis allows the identification of the genetic variants associated with the trait of interest, without restrictions regarding the genomic location of the mutations. The identification of variants associated with reproductive traits can make possible to use them for selecting parents more efficiently, so increasing considerably the response to the selection of traits. The objective of this work is to identify genomic regions associated with the prolificacy and number of nipple in Iberian pigs through genome wide association analysis.

**Key words:** Iberian pigs, GWAS, número de tetinas, eficiencia reproductiva, prolificidad, litter size.

Autora: **Dña. Zulema Rodríguez Hernández**

Tutora: Dra. Dña. Noelia Ibáñez-Escriche

Directora experimental: Dña. Cristina Casto Rebollo

**Valencia, julio 2019**

Cuatro años después...

Por todos los que han confiado en mí, incluso cuando yo dudaba de mi misma.

Noelia y Cristina no sé por dónde empezar... en fin, sois todo un referente a seguir, GRACIAS.

A mi L<sup>3</sup> porque siempre has estado ahí.

A MT porque tus “mandamientos” siempre estarán conmigo.

# ÍNDICE GENERAL

1. INTRODUCCIÓN.....	1
2. OBJETIVOS.....	5
3. MATERIAL Y MÉTODOS .....	6
3.1. Material biológico .....	6
3.2. Procesado de datos .....	7
3.2.1. Control de calidad y filtrado de muestras y SNPs .....	7
3.2.2. Imputación de genotipos faltantes .....	7
3.3. Análisis de asociación del genoma completo (GWAS).....	7
3.4. Búsqueda de genes candidatos .....	8
4. RESULTADOS .....	9
4.1. Análisis de asociación del genoma completo (GWAS).....	9
4.2. Búsqueda de genes candidatos. ....	11
5. DISCUSIÓN.....	14
6. CONCLUSIONES.....	19
7. REFERENCIAS.....	20
8. DATOS SUPLEMENTARIOS.....	25

## ÍNDICE TABLAS

Tabla 1. Estructura de la población. ....	6
Tabla 2. Análisis descriptivo de los caracteres número de tetinas (NT), tamaño de camada al primer parto (TC <sub>1</sub> ) y tamaño de camada en el segundo parto (TC <sub>2</sub> ). ....	6
Tabla 3: SNPs significativos asociados al tamaño de camada en el primer parto (TC <sub>1</sub> ) y al número de tetinas (NT). ....	10
Tabla 4: Genes localizados en las regiones genómicas de interés para el carácter TC <sub>1</sub> y NT. ....	12
Tabla S1. Genes localizados en las regiones candidatas de TC1 implicados en la reproducción y el desarrollo embrionario. ....	25

## ÍNDICE FIGURAS

Figura 1. A) Manhattan plot para el carácter tamaño de camada en el primer parto ( $TC_1$ ). B) Manhattan plot para el carácter número de tetinas. ....	9
Figura 2. Efecto de T3 sobre la prolificidad, regulada por el eje hipotálamo-hipofisiario.....	15
Figura 3. Tendencia de las frecuencias genotípicas del SNP WU_10.2_13_13314717.....	16
Figura 4. Mecanismos reguladores de la formación de las mamas. ....	18
Figura S1. Manhattan plot para el carácter tamaño de camada en el segundo parto ( $TC_2$ ). ....	25

# 1. INTRODUCCIÓN

España es el cuarto país productor de carne de cerdo a nivel mundial después de China, Estados Unidos y Alemania (datos 2018, Fuente: SG Estadísticas). El sector porcino español juega un papel fundamental en la economía del país, representando hasta el 36,4% (en el año 2017) de la Producción Final Ganadera según el informe anual emitido por el Ministerio de Agricultura, Alimentación y Medio Ambiente (MAPAMA). Dentro de este sector, España destaca por la producción de cerdo ibérico, cuyos productos derivados son de alta calidad y bien valorados por el consumidor (Martínez et al., 2000). El término cerdo ibérico, catalogado por el MAPAMA como una raza autóctona de fomento, define una agrupación racial de cerdos nativos de la península ibérica originados a partir de *Sus Mediterraneus* (Lopez-Bote, 1998). Su producción está limitada, tradicionalmente, al sudoeste de la península ibérica en consonancia con la superficie de la dehesa arbolada. Su mayor representación se encuentra en las provincias españolas de Salamanca, Cáceres, Badajoz, Ciudad Real, Toledo, Sevilla, Córdoba, Huelva, Cádiz y Málaga. El sector de porcino ibérico se ha caracterizado por ser un sector muy tradicional, ligado a una producción extensiva debido a la morfología del animal, que lo hace resistente a las insolaciones y altas temperaturas estivales (Fabuel et al., 2004). No obstante, en los últimos años la demanda de porcino ibérico se ha incrementado considerablemente provocando una profunda transformación hacia una producción más intensiva e industrializada. En la actualidad, la industria transformadora se surte mayoritariamente de cerdos provenientes del cruce de una madre ibérica con un macho de la raza Duroc y cebados en sistemas intensivos (Noguera e Ibáñez-Escriche, 2017). El cebo de ejemplares ibéricos puros se limita casi exclusivamente a aquellos que aprovechan la montanera (cebo extensivo a base de bellotas y pastos) (Lopez-Bote, 1998).

Al contrario de lo que ha ocurrido con la mayoría de las razas porcinas comerciales, el cerdo ibérico apenas ha sufrido presión selectiva organizada salvo aquella vinculada a potenciar determinadas características morfológicas, que han concretado su actual patrón racial, o fisiológicas relacionadas con su rusticidad y capacidad de engrasamiento (AECERIBER, 2012). En la actualidad, hay cinco variedades de cerdo ibérico (Retinto, Entrepelado, Torbiscal, Lampiño y Manchado de Jabugo) reconocidas en el libro genealógico oficial gestionado por la Asociación Española de Criadores de Ganado Porcino Selecto Ibérico Puro y Tronco Ibérico (AECERIBER, 2012). La estirpe Retinto (RR), de pelaje rojo, es la principal variedad de ibérico debido a su densidad de población y amplia distribución. Lampiño es una variedad de color negro que se caracteriza, como su nombre indica, por no presentar pelo. La estirpe Torbiscal (TT), de pelaje negro, proviene del cruce de dos líneas negras de Lampiño y otras dos rojas: Retinto y una autóctona portuguesa (Martínez et al., 2000). El Entrepelado (EE), híbrido fijado del cruce entre Lampiño con Retinto, es una cepa peluda negra que presenta un color rojo al nacimiento (Martínez et al., 2000; AECERIBER, 2012). Por último, se encuentra el Manchado de Jabugo, con manchas diseminadas por todo el cuerpo e introgresión de

genes asiáticos (Fang y Andersson, 2006). No obstante, al igual que en el cerdo blanco existe variabilidad inter e intrarracial tanto a nivel genético como fenotípico. Estudios en caracteres como el peso de los lechones al nacimiento (Barba et al., 2002; Ibáñez- Escriche et al., 2014) o el tamaño de camada (Suárez et al., 2002; Casellas et al., 2019) han encontrado diferencias genéticas relevantes entre líneas, así como heredabilidades similares al cerdo blanco, lo que indica la posibilidad de mejorar los caracteres en porcino ibérico a través de la selección genética.

En producción porcina uno de los caracteres económicos más importantes, debido a su relación directa con la cantidad de lechones producidos por cerda y año, es el tamaño de camada. Nótese, que el número creciente de lechones destetados reduce los costes de amortización por hembra (Distl, 2007; Ma et al., 2018). Este carácter, aunque es de difícil selección por su baja heredabilidad ( $h^2 \approx 0.1$ ), es uno de los principales objetivos de mejora en las líneas maternas de cerdo blanco (Ma et al., 2018). De hecho, en los últimos 60 años, las líneas maternas de cerdo blanco han aumentado considerablemente el tamaño de camada gracias a la selección genética llevada a cabo (Chalkias, 2013). No obstante, en algunos casos la mejora del tamaño de camada no ha venido acompañada por un aumento del número de tetinas. Este hecho puede conllevar problemas de manejo (adopciones) e incluso una mayor mortalidad debido a la competencia entre lechones después del nacimiento (Rutherford et al., 2013). La competencia se produce, principalmente, por el acceso de los lechones a las tetas y esta puede existir incluso cuando el número de pezones funcionales es igual al número de crías, ya que los lechones con una alta capacidad competitiva pueden monopolizar más de una teta (Andersen et al., 2011). Por tanto, las crías que no pueden acceder a una teta funcional se enfrentan a una situación crítica, puesto que los lechones nacen con una inmunidad baja dependiendo totalmente de la ingesta de anticuerpos a través del calostro, y pueden morir de hambre (Rutherford et al., 2013; Chalkias, 2013).

Hasta ahora, el cerdo ibérico apenas se había seleccionado por tamaño de camada, lo que se ha traducido en un retraso considerable en la eficiencia reproductiva de esta raza respecto al cerdo blanco (Ibáñez-Escriche et al., 2016). En la actualidad, existe un creciente interés por el aumento del tamaño de camada en ibérico y, diversos programas de mejora genética de esta raza han definido este carácter como un objetivo principal de selección (Noguera, 2011). Sin embargo, además de la dificultad que conlleva la mejora del tamaño de camada, el número de tetinas en cerdo ibérico (10) es menor que en cerdo blanco y presenta muy poca variabilidad (Fernández et al., 2004). Por tanto, para optimizar la eficiencia reproductiva en cerdo ibérico sería deseable un incremento conjunto de ambos caracteres: el tamaño de camada y el número de tetas.

Tradicionalmente, la mejora genética en porcino se ha realizado a través de la aplicación del Modelo Animal que utiliza metodología BLUP (Best Linear Unbiased Predictor) y en el que se asume que los

caracteres están regulados por muchos genes de pequeño efecto (Phocas y Colleau, 1996). Este tipo de modelos utilizan principalmente la información de los fenotipos y genealogía de la población y su aplicación sistemática ha supuesto importantes tasas de respuesta genética (Noguera, 2011). Sin embargo, su aplicación todavía tiene limitaciones en caracteres de baja heredabilidad y en aquellos que solo se miden en un solo sexo como son los caracteres reproductivos (Ibáñez-Escriche y Gonzalez-Recio, 2011). Por otra parte, el desarrollo de la genómica en los últimos años ha revolucionado los métodos y herramientas de investigación en el campo de la genética y la biología. Además, la genómica no solo ha permitido descifrar la secuencia completa de la mayoría de los genomas de animales de producción, sino que también ha permitido obtener una mejor comprensión de los mecanismos genéticos que regulan los caracteres complejos (Noguera, 2011). La disponibilidad de grandes densidades de marcadores moleculares (cada vez a menor coste) a lo largo del genoma acompañada de avances estadísticos y bioinformáticos han permitido realizar estudios de asociación de todo el genoma de los individuos con su fenotipo (GWAS). Los GWAS generalmente relacionan los fenotipos de interés con las variaciones entre polimorfismos de un solo nucleótido (SNPs por sus siglas en inglés) (Bush y Moore, 2012). Los SNPs son cambios de un par de bases en la secuencia de ADN, que ocurren con alta frecuencia en el genoma y, que se utilizan normalmente como *marcadores* de una región genómica (Bush y Moore, 2012). En humanos, gracias a esta herramienta, se han identificado asociaciones estadísticamente significativas entre cientos de genes y enfermedades complejas comunes, lo cual ha contribuido a incrementar el conocimiento de las bases moleculares implicadas en la predisposición a padecerlas (Anaya-Chávez y Martínez, 2010). En producción animal, los GWAS también se han empleado para localizar regiones genómicas asociadas a caracteres complejos (Xie et al., 2012; Jang et al., 2018; Jardim et al., 2018). Normalmente, los métodos estadísticos utilizados para identificar las regiones genómicas (QLTs) asociadas a un carácter aprovechan el desequilibrio de ligamiento (LD) entre el marcador y la mutación causal. El análisis de regresión sobre un único marcador (SMR) constituye uno de los métodos estadísticos más empleados en los GWAS y asume que los QLTs son unidades independientes con gran efecto sobre el carácter (Yang et al., 2014). En el análisis SMR se consideran los marcadores SNPs como efectos fijos y se analizan uno por uno (Yang et al., 2014). Uno de los inconvenientes del método SMR a tener en cuenta, es la asociación errónea de SNPs con el fenotipo (falso positivo o error de tipo I) debido a factores de confusión como la estratificación de la población. Por ello, normalmente el modelo SMR también se corrige por la matriz de relaciones de parentesco genómica (GRM), en la que se incluye todos los marcadores a excepción del SNP candidato y los marcadores del cromosoma del SNP a estudiar (Yang et al., 2014).

A partir de la secuenciación completa del genoma de cerdo en 2009 y la comercialización del primer chip de SNPs específicos para esta especie, el descubrimiento de regiones genómicas que contribuyen a la variabilidad genética de caracteres complejos se ha visto incrementado (Ramos et al., 2009). En

cerdo blanco, estudios GWAS y de expresión diferencial han identificado genes que podrían estar regulando caracteres reproductivos como el tamaño de camada y el número de tetinas. Para el tamaño de camada se han encontrados genes que participan en el desarrollo uterino, la implantación del embrión en útero y la ovulación (Zhang et al., 2013; Rohrer y Nonneman, 2017; Ma et al., 2018;). Respecto al número de tetinas se han encontrado genes que intervienen en el desarrollo de las mamas y de las vértebras durante la embriogénesis (Rohrer y Nonneman, 2017; van Son et al., 2019; Verardo et al., 2016 a). No obstante, a pesar de los numerosos estudios de asociación realizados en porcino, todavía no se han realizado GWAS para caracteres reproductivos en cerdo ibérico. El presente estudio tiene como objetivo identificar las regiones genómicas asociadas con el carácter número de tetinas y tamaño de camada en el primer y segundo parto. Otra de las finalidades de este trabajo, es identificar genes candidatos en estas regiones que pudieran realizar funciones y/o estar implicados en rutas metabólicas asociadas a estos caracteres de interés. Este estudio, además de ser pionero tiene una gran importancia para el porcino ibérico ya que es un primer paso para establecer estrategias de mejora en caracteres económicos tan importantes como es el tamaño de camada. La mejora de este carácter junto con el del número de tetinas, podría suponer una ventaja tanto a nivel productivo como de bienestar animal, en un sector que es económicamente importante en España y valorado internacionalmente por sus productos de alta calidad.

## **2. OBJETIVOS**

1. Identificar regiones genómicas asociadas a los caracteres número de tetinas y tamaño de camada en el primer y segundo parto.
2. Identificar genes candidatos en esas regiones que estén asociados con los caracteres que se estudian.

### 3. MATERIAL Y MÉTODOS

#### 3.1. Material biológico

Los datos utilizados en este estudio corresponden a un proyecto de investigación (CGL2016-80155-R) en colaboración con la empresa Inga Food S.A. y provienen de cuatro granjas de cerdo ibérico (Morantes, el Muerto, Chorrero y Risco). En total se utilizaron 751 animales (724 hembras y 27 machos) de distintas estirpes de cerdo ibérico (Torbiscal [TT], Retinto [RR], y Entrepelado [EE]) y de sus respectivos cruces (Tabla 1).

Los animales fueron genotipados con el chip de Illumina GGP Porcine HD Array para un total de 68.516 SNPs. La extracción del ADN de los individuos se realizó a partir de muestras de sangre y de tejido de la cola.

**Tabla 1. Estructura de la población.**

		Línea Materna		
		E <sup>(a)</sup>	R <sup>(b)</sup>	T <sup>(c)</sup>
Línea Paterna	E	145	181	5
	R	140	257	11
	T	11	1	0

<sup>(a)</sup>Entrepelado; <sup>(b)</sup> Retinto; <sup>(c)</sup> Torbiscal.

Los caracteres estudiados fueron el número de tetinas (anotadas al destete de los animales) y el tamaño de camada (número de nacidos totales) en los dos primeros partos. En la tabla 2 se presentan los estadísticos descriptivos de los caracteres de interés.

**Tabla 2. Análisis descriptivo de los caracteres número de tetinas (NT), tamaño de camada al primer parto (TC<sub>1</sub>) y tamaño de camada en el segundo parto (TC<sub>2</sub>).**

CARACTER	N <sup>(a)</sup>	$\mu$ <sup>(b)</sup>	Me <sup>(c)</sup>	CV <sup>(d)</sup>	Min <sup>(e)</sup>	Max <sup>(f)</sup>
NT	730	10,04	10	0,02	9	12
TC <sub>1</sub>	701	7,72	8	0,24	4	14
TC <sub>2</sub>	680	8,66	9	2,23	4	17

<sup>(a)</sup>Número total de datos; <sup>(b)</sup>media; <sup>(c)</sup>mediana; <sup>(d)</sup>coeficiente de variación; <sup>(e)</sup>mínimo y <sup>(f)</sup>máximo del rango de datos.

### 3.2. Procesado de datos

Los datos fenotípicos y genotípicos de los animales pertenecientes a las cuatro granjas se organizaron y agruparon mediante el software Rstudio (<https://www.r-project.org/>). Con este mismo software, también se obtuvieron los estadísticos presentados anteriormente y se crearon todos los archivos que se usaron posteriormente para llevar a término el estudio de asociación del genoma completo (GWAS).

#### 3.2.1. Control de calidad y filtrado de muestras y SNPs

Los controles de calidad y el filtrado de los SNPs y de las muestras genotipadas se llevaron a cabo mediante el software PLINK v1.9 (Chang et al., 2015). Los individuos con más del 10% de genotipos faltantes fueron eliminados del análisis. También se excluyeron los SNPs con una frecuencia del alelo minoritario (MAF) menor del 3% (18.186) y con más del 10% de genotipos faltantes por SNP (4.724). Además, se eliminaron del estudio los SNPs localizados en los cromosomas sexuales X e Y (11.082). Finalmente 34.524 SNPs y 748 individuos superaron los controles de calidad.

#### 3.2.2. Imputación de genotipos faltantes

Los genotipos faltantes por SNP se imputaron mediante el software Beagle v.4.1 (Browning and Browning, 2016) imputándose un total de 36.437 SNPs. Ninguno de los SNPs imputados fueron descartados, al ser la calidad de la imputación ( $R^2$ ), en todos los casos, superior a 0.90.

### 3.3. Análisis de asociación del genoma completo (GWAS)

El análisis de asociación del genoma completo (GWAS) se realizó mediante el método de regresión marcador por marcador (SMR) con el software GCTA v1.91.beta (Yang et al., 2011). El modelo lineal mixto utilizado fue:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{z}_j\alpha_j + \mathbf{Z}\mathbf{g} + \mathbf{e} \quad (1)$$

Donde  $\mathbf{y}$  es el vector de fenotipos del carácter a estudiar;  $\mathbf{X}$  es la matriz de incidencia del vector de efectos fijos  $\mathbf{b}$  que incluye el tipo de monta (TP; 2 niveles), línea macho (LM; 3 niveles) y el efecto granja y año-estación (GAE; 40 niveles);  $\mathbf{z}_j$  es el vector de genotipos del SNP<sub>j</sub> codificado según el número de copias del alelo de referencia (0,1,2);  $\alpha_j$  es el efecto de sustitución alélica del SNP<sub>j</sub>;  $\mathbf{Z}$  es la matriz de incidencia de los efectos genéticos  $\mathbf{g}$  que sigue una distribución  $N(\mathbf{0}, \mathbf{GRM}\sigma_g^2)$ ;  $\mathbf{e}$  es el vector de los efectos residuales. La **GRM** es la matriz de relaciones de parentesco genómica en la que

se excluye el cromosoma del SNP<sub>j</sub> y se reestima el valor de  $\sigma_g^2$  en cada cálculo de la **GRM**. Esta matriz ha sido calculada usando el método desarrollado por Yang et al. (2011) y se incluyó en el modelo para corregir los posibles efectos de estratificación de la población (Yang et al., 2014). A partir de este modelo el programa obtiene el valor del efecto de sustitución alélica para cada SNP<sub>j</sub> ( $\alpha_j$ ) y el p-valor asociado a la estima de  $\alpha_j$ . El umbral de relevancia estadística para decidir si un SNP está asociado al carácter en estudio se estableció en un p-valor de 1e-4, umbral sugerido para datos de producción animal.

Además del p-valor se calculó el porcentaje de varianza genética explicada por aquellos SNPs con un p-valor menor de 1e-4. El porcentaje de varianza explicada se calculó como la ratio entre la varianza genética explicada por un SNP calculada como  $2pqb^2$  y la varianza genética total. Nótese, que este porcentaje de varianza explicada por un SNP puede estar sobreestimado ya que el método aplicado analiza SNP por SNP y asume que los SNPs son independientes. Sin embargo, este parámetro sí que sería un indicador de la importancia de la región genómica en la que está localizado el SNP.

### **3.4. Búsqueda de genes candidatos**

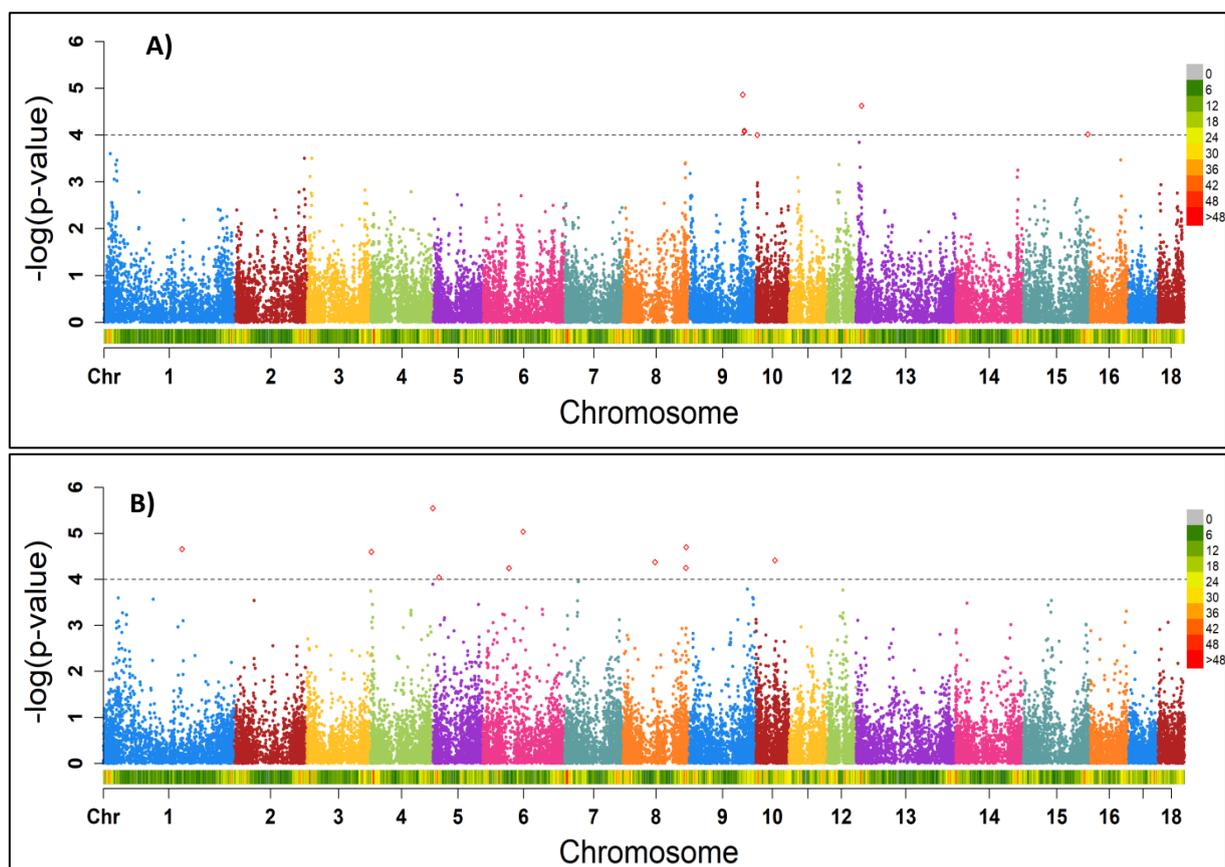
La búsqueda de genes candidatos en las regiones genómicas asociadas a los caracteres estudiados se realizó con la base de datos Ensembl (Cunningham et al., 2018) usando como genoma de referencia el cerdo (*Sus Scrofa* v11.1). Las regiones genómicas asociadas fueron aquellas situadas a  $\pm 0.5$ MB de los SNPs más significativos ( $p\text{-valor} \leq 1e-4$ ). Los genes localizados en esas regiones se propusieron como genes candidatos y se revisaron los procesos biológicos en los que estaban implicados. Las bases de datos Enrichr v1.0 (Jawaid, 2017) y DAVID v6.8 (Huang et al., 2009) se utilizaron para conocer las rutas metabólicas más relevantes en las que están implicadas estos genes. La ontología génica de cada gen se validó con la base de datos GeneCards (Rappaport et al., 2017).

Posteriormente, se localizaron en el genoma del cerdo los SNPs que mostraron una mayor relevancia en el análisis GWAS ( $p\text{-valor} \leq 1e-0.4$ ). El objetivo fue comprobar si estos SNPs se encontraban dentro de la secuencia genética de los genes candidatos y si su localización era en una zona intrónica o exónica. Este procedimiento se realizó con la base de datos Ensembl (Cunningham et al., 2018).

## 4. RESULTADOS

### 4.1. Análisis de asociación del genoma completo (GWAS)

Los análisis de asociación del genoma completo identificaron regiones asociadas al tamaño de camada en el primer parto, ( $TC_1$ ) y al número de tetinas (NT) (Figura 1), pero no para el tamaño de camada en el segundo parto  $TC_2$  (Figura S1). Para el primer carácter,  $TC_1$ , se identificaron un total de nueve SNPs relevantes localizados en los cromosomas 9, 10, 13 y 15 (Tabla 3). De todos ellos, el SNP H3GA0028062 situado en la posición 112181287 pb del cromosoma 9 es el de mayor relevancia, con un p-valor de  $1.37e-05$ . Destaca también el SNP WU\_10.2\_13\_13314717 situado en el cromosoma 13, ya que además de tener un valor de significancia alto (p-valor de  $2.36e-05$ ) explicaría casi un 40% de la varianza genética del  $TC_1$ .



**Figura 1. A) Manhattan plot para el carácter tamaño de camada en el primer parto ( $TC_1$ ). B) Manhattan plot para el carácter número de tetinas (NT).** El eje y representa el logaritmo negativo en base diez del p-valor de asociación para cada SNP. En el eje x representa la posición física de cada SNP en su cromosoma correspondiente. La línea horizontal discontinua indica el umbral de significancia sugerido para datos de producción animal  $-(\log_{10}(p\text{-value}))=4$ . La leyenda de colores de la derecha indica la densidad de SNPs a lo largo del genoma del cerdo de 0% (gris) a mayor de 48% (rojo). Los SNPs con mayor relevancia están resaltados como puntos rojos.

Para el NT, se identificó un total de 14 SNPs, localizados en los cromosomas 1, 4, 5, 6, 7, 8 y 10 (Tabla 3). Los SNPs más relevantes son el WU\_10.2\_5\_110558175 (cromosoma 5) y el ALGA0114347 (cromosoma 6), con un p-valor de 2.82e-06 y 9.20e-06, respectivamente. Las regiones localizadas en torno a los SNPs H3GA0020625 y MARC0033365 (cromosoma 7) fueron también destacadas debido a que estos SNPs presentaron un p-valor cercano al valor umbral (1,12e-04) y sus regiones (7:28 - 30) mostraban una clara tendencia de asociación (Figura 1).

**Tabla 3: SNPs significativos asociados al tamaño de camada en el primer parto (TC<sub>1</sub>) y al número de tetinas (NT).**

Carácter	Chr <sup>(a)</sup>	SNP <sup>(b)</sup>	A1 <sup>(c)</sup>	A2 <sup>(d)</sup>	pb <sup>(e)</sup>	Freq <sup>(f)</sup>	b <sup>(g)</sup>	se <sup>(h)</sup>	p <sup>(i)</sup>	Vs <sup>(i)</sup>
TC <sub>1</sub>	9	H3GA0028062	G	A	112181287	0,19	-0,56	0,13	1,37E-05	38,61
	9	WU_10.2_9_127122055	A	G	115478190	0,16	-0,58	0,15	8,47E-05	36,17
	9	WU_10.2_9_127092511	C	A	115507783	0,16	-0,58	0,15	8,47E-05	36,17
	9	ASGA0044392*	C	G	115554264	0,13	-0,63	0,16	8,06E-05	35,91
	9	WU_10.2_9_127040028*	A	C	115559818	0,13	-0,63	0,16	8,06E-05	35,91
	10	WU_10.2_10_4963544*	C	A	3146772	0,21	-0,5	0,13	9,94E-05	33,18
	13	WU_10.2_13_13314717	G	A	11767174	0,43	0,45	0,11	2,36E-05	39,71
	15	H3GA0054422*	G	A	135627733	0,05	-0,9	0,23	9,58E-05	30,78
	15	H3GA0055520*	G	A	135630321	0,05	-0,9	0,23	9,58E-05	30,78
NT	1	ALGA0006774	A	G	164766550	0,03	0,17	0,04	2,19E-05	5,61
	4	WU_10.2_4_1897808	G	A	2387092	0,11	0,10	0,02	2,52E-05	6,53
	5	WU_10.2_5_110558175*	A	G	222125	0,08	0,12	0,03	2,82E-06	7,07
	5	H3GA0015724*	G	A	12963910	0,07	0,11	0,03	9,10E-05	5,25
	6	ASGA0028206*	A	G	54827092	0,06	0,11	0,03	5,73E-05	4,55
	6	M1GA0008537*	G	A	54909745	0,06	0,11	0,03	5,73E-05	4,55
	6	ASGA0028216	C	A	54961295	0,06	0,11	0,03	5,73E-05	4,55
	6	ALGA0114347*	A	G	84478202	0,07	0,11	0,03	9,20E-06	5,25
	7	H3GA0020625*	A	G	28404749	0,06	0,12	0,03	1,12E-04	5,41
	7	MARC0033365*	G	A	29256819	0,06	0,12	0,03	1,12E-04	5,41
	8	H3GA0024939	A	G	67647137	0,03	0,15	0,04	4,23E-05	4,37
	8	MARC0012400	A	G	132375375	0,14	0,08	0,02	5,65E-05	5,14
	8	WU_10.2_8_142092491*	A	G	132705623	0,06	0,13	0,03	2,00E-05	6,35
	10	ALGA0117721	A	C	40418554	0,18	0,07	0,02	3,86E-05	4,82

<sup>(a)</sup>Cromosoma en el que se encuentra el SNP; <sup>(b)</sup>nombre del SNP significativo; <sup>(c)</sup>alelo de referencia; <sup>(d)</sup>alelo alternativo; <sup>(e)</sup>posición (en pares de bases) en la que se encuentra el SNP en el cromosoma; <sup>(f)</sup>frecuencia alélica; <sup>(g)</sup>efecto de sustitución alélica del SNP; <sup>(h)</sup>error estándar; <sup>(i)</sup>p-valor asociado al efecto del SNP; <sup>(j)</sup>porcentaje de varianza que explican los SNPs para el carácter a estudiar (%); \*SNPs dentro de la secuencia génica de un gen.

## 4.2. Búsqueda de genes candidatos.

En este estudio se asumieron como regiones genómicas asociadas a los caracteres de interés aquellas situadas a  $\pm 0.5$ MB de los SNPs más relevantes ( $p$ -valor  $\leq 1e-4$ ) (Tabla 4). Un total de 11 genes se identificaron en las regiones genómicas asociadas al TC<sub>1</sub>: 6 genes en el cromosoma 9, 1 en el cromosoma 10, 1 en el cromosoma 13 y 3 en el cromosoma 15 (Tabla 4). De estos genes, a pesar de que 6 de ellos están implicados en el desarrollo embrionario y la espermatogénesis (Tabla S1), solo 1 (*THRB*) destaca por su rol significativo en la ovulación y la implantación del embrión. El gen *THRB*, localizado en el cromosoma 13, codifica el receptor de la hormona tiroidea Beta (THR $\beta$ ) y participa en la regulación del eje hipotálamo-hipofisiario para mantener constantes los niveles de la hormona tiroidea (HT) (Ortiga-Carvalho et al., 2014). Por su parte, la HT regula el metabolismo reproductivo de los seres vivos participando en la estimulación de las células sexuales, el crecimiento folicular y la secreción de hormonas masculinas y femeninas. También, esta hormona interviene en el crecimiento y desarrollo fetal a través de la placenta de la madre, la receptibilidad femenina y el engrosamiento del endometrio (Kilby et al., 2004; Aghajanova et al., 2011).

La localización de los SNPs más relevantes para TC<sub>1</sub> en el genoma del cerdo reveló que los SNPs WU\_10.2\_9\_127040028 y ASGA0044392 (cromosoma 9) se encontraban situados en la región del gen *TNFSF4*, que participa en el desarrollo de la respuesta inflamatoria y defensiva del individuo (Tabla 4). Además, WU\_10.2\_9\_127040028 también se localiza dentro de la secuencia de un gen, que situado en la región 9:115551565-115590431 (Tabla 4), está asociado temporalmente con el nombre de ENSSSCG00000015488 (función que todavía se desconoce). En el cromosoma 15, los SNPs H3GA0054422 y H3GA0055520 se encontraron en la región del gen *AGAPI* (Tabla 4). Este gen codifica una proteína activadora de GTPasa que regula procesos relacionados con el citoesqueleto de actina y la endocitosis.

Para el carácter NT se identificaron, en las regiones genómicas asociadas, un total de 151 genes (Tabla 4). De estos, destacaron 34 al presentar relación con el desarrollo mamario y expresarse en las mamas. Solo se seleccionaron aquellos genes que estuviesen involucrados en los procesos y rutas metabólicas que regulan el desarrollo mamario y la formación de las vértebras durante la embriogénesis. Este criterio se fijó porque estudios recientes demostraron que el aumento de número de vértebras, estaba correlacionado positivamente con el número de tetinas (Rohrer y Nonneman, 2017). De esta forma, solo 6 genes se seleccionaron como posibles candidatos asociados al NT (*SMAD6*, *PIM3*, *MED25*, *PTOV1*, *PRIM2* y *MAPK10*).

**Tabla 4: Genes localizados en las regiones genómicas de interés para el carácter TC<sub>1</sub> y NT.**

<b>Carácter</b>	<b>Chr<sup>(a)</sup></b>	<b>Regiones<sup>(b)</sup></b>	<b>Genes Candidatos</b>
TC <sub>1</sub>	9	115 - 116	<i>SUCO, TNFSF18, TNFSF4, SLC9C2, TEX50, KLHL20</i>
	10	3- 4	<i>BRINP3</i>
	13	11 – 12	<i>THRB</i>
	15	135 - 136	<i>AGAP1, GBX2, ASB18</i>
NT	1	164 – 165	<i>MEGF11, DIS3L, TIPIN, SNAPC5, RPL4, ZWILCH, LCTL, SMAD6, SMAD3, AAGAB, IQCH</i>
	4	2 - 3	<i>PTP4A3, GPR20, SLC45A4, DENND3, PTK2</i>
	5	0 – 1	<i>SHANK3, MAPK8IP2, CPT1B, SYCE3, ODF3B, LMF2, MIOX, SBF1, PPP6R2, DENND6B, PLXNB2, MAPK11, MAPK12, HDAC10, TUBGCP6, TRABD, PANX2, MOV10L1, MLC1, TTLL8, PIM3, CRELD2, ALG12, ZBED4, BRD1</i>
	6	12 - 13	<i>FBXO7, BPIFC, ASCL4, PRDM4, PWP1, BTBD11, CRY1, MTERF2, TMEM263</i>
	6	54 - 55	<i>PPFIA3, HRC, TRPM4, CD37, TEAD2, DKKL1, CCDC155, PTH2, GFY, SLC17A7, PIH1D1, ALDH16A1, FLT3LG, RPS11, FCGRT, RCN3, NOSIP, PRRG2, PRR12, RRAS, SCAF1, BCL2L12, PRMT1, CPT1C, TSKS, AP2A1, FUZ, MED25, PTOV1, PNKP, AKT1S1, TBC1D17, IL4I1, NUP62, ATF5, VRK3, ZNF473, IZUMO2, MYH14, KCNC3, NAPSA, POLD1, MYBPC2, FAM71E1, EMC10, JOSD2, LRRC4B, SYT3</i>
	6	54 - 55	<i>ARID1A, PIGV, ZDHHC18, GPATCH3, NR0B2, KDF1, TRNP1, TENT5B, WDTC1, TMEM222, SYTL1, MAP3K6, FCN3, CD164L2, WASF2, AHDC1, FGR, IFI6, FAM76A, PPP1R8</i>
	7	28 - 30	<i>PRIM2, RAB23, BAG2, ZNF451, BEND6, DST, VPS52, B3GALT4, WDR46, TAPBP, DAXX, KIFC1, PHF1, CUTA, SYNGAP1</i>
	8	67 - 68	<i>CABS1, AMTN, JCHAIN, RUFY3, GRSF1, DCK, SLC4A4</i>
	8	132 - 133	<i>AFF1, SLC10A6, PTPN13, MAPK10, ARHGAP24</i>
	10	40 - 41	<i>WAC, MAP3K8, MTPAP, JCAD</i>

<sup>(a)</sup> Cromosoma en el que se encuentran las regiones candidatas; <sup>(b)</sup> posición en MB que ocupan las regiones.

El gen *SMAD6* en el cromosoma 1, regulan la ruta metabólica de BMP y TGF- $\beta$  interactuando con *RUNX2*, que actúa en la maduración de los condrocitos y la formación de los osteoclastos en las etapas embrionarias. BMP son proteínas de crecimiento multifuncionales que pertenecen a la familia de TGF- $\beta$ , citoquinas involucradas en procesos esenciales durante la embriogénesis, como la angiogénesis, diferenciación y proliferación celular, entre otros (Chen et al., 2004; Monks, 2007). En el cromosoma 6, se encuentran *MED25* y *PTOVI* que regulan la señalización del ácido retinoico (AR), el cual se necesita para activar la expresión de *TBX3*, que participa en la formación de la yema mamaria durante el desarrollo embrionario (Cowin y Wysolmerski, 2010). El gen candidato *PIM3* (cromosoma 5), regula la actividad transcripcional de *MYC*, que está implicado en el mantenimiento de la población de las células madre en el epitelio mamario (Moumen et al., 2012).

La posición física de los SNPs más relevantes para el NT mostró que 8 de ellos (Tabla 3) se encontraban en la secuencia de los genes *NCAPH2*, *SCO*, *BTBD11*, *FUZ*, *IL4I*, *WDTC*, *PRIM2*, *DST* y *MAPK10*. De los 8 SNPs mencionados, solo se destacan H3GA0020625 en el cromosoma 7 (posición 28404749 pb) y WU\_10.2\_8\_142092491 en el 8 (posición 132705623 pb) que se encuentran dentro de las regiones de los genes candidatos *PRIM2* y *MAPK10*, respectivamente. Ambos genes han sido asociados con el NT en otros estudios de GWAS y también, en estos, SNPs significativos se localizaron en sus secuencias génicas (Verardo et al., 2016 b; Peng et al., 2017). Esto podría indicar, que estos SNPs están capturando la mutación causal y, por consiguiente, estos genes están realmente asociados con el carácter que se estudia.

## 5. DISCUSIÓN

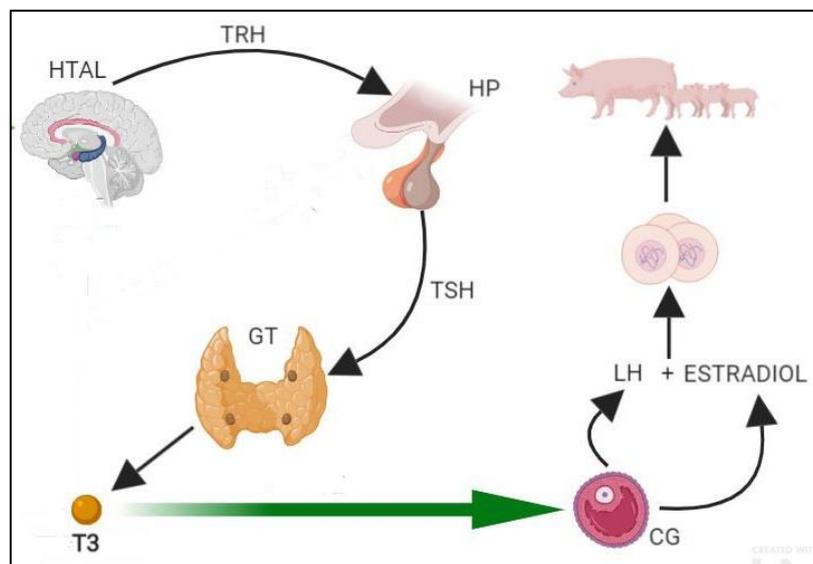
En producción porcina, el tamaño de camada (TC) es uno de los principales objetivos de selección, ya que tiene un fuerte impacto en la economía del sector (Distl, 2007). Sin embargo, se ha visto que además de aumentar el TC, es necesario incrementar el número de tetinas (NT) de la hembra para así evitar la competencia entre los lechones por alimentarse (Andersen et al., 2011). Es por ello que el estudio de las bases genéticas que regulan tanto la prolificidad como el NT en cerdos está generando mucho interés en la industria del porcino ibérico. A su vez, la reducción en los costes de genotipado y la implementación de chips de alta densidad ha facilitado la identificación de regiones genómicas asociadas a caracteres de interés en producción animal.

Este estudio identificó regiones genómicas asociadas tanto a NT como a TC<sub>1</sub> en cerdo ibérico a través de un análisis de regresión marcador a marcador (SMR). Para evitar factores de confusión debido al uso de diferentes estirpes y cruces de cerdo ibérico se incluyó en el modelo la matriz de relaciones genómicas (GRM). De esta manera se corrigieron los posibles efectos de estratificación de la población. La matriz GRM construida sin tener en cuenta el cromosoma del SNP a testar, también, permite aumentar la potencia del análisis para la detección de más variantes con efecto (Yang et al., 2014). Además, el modelo consideró como efectos fijos el tipo de monta (inseminación artificial o monta natural), la línea macho y el efecto granja y año-estación eliminando, así, otros posibles factores de confusión en el análisis.

En este estudio se identificaron regiones genómicas asociadas con el TC<sub>1</sub> (cromosomas 9, 10, 13 y 15) pero no para TC<sub>2</sub> (Figura S1). Las diferencias encontradas entre TC<sub>1</sub> y TC<sub>2</sub> indicarían que los genes que regulan este carácter se expresan o interaccionan de forma distinta en hembras nulíparas y multíparas. Estos resultados concuerdan con el estudio de Noguera et al. (2002) donde se encontraron correlaciones genéticas diferentes de cero entre el primer y segundo parto en cerdas blancas. De hecho, en otro estudio se demostró que las hembras primíparas tienen un menor número de lechones que las hembras multíparas debido a una baja tasa de ovulación y una alta mortalidad embrionaria (Ek-Mex et al., 2014). Esto podría deberse a que el espacio del útero es limitado en las primíparas debido a que aún no está completamente maduro (Sell-Kubiak et al., 2018). Además, en un estudio GWAS reciente en cerdo blanco, también encontraron distintas regiones genómicas asociadas para el TC<sub>1</sub> y el TC<sub>2</sub> (Sell-Kubiak et al., 2015).

En las regiones genómicas identificadas para el TC<sub>1</sub> (Tabla 4) se localizaron un total de 11 genes. El gen *THRB* (localizado en el cromosoma 13 entre las megabases 11-12) se propuso como gen candidato debido a su relación con los procesos de ovulación e implantación del embrión en el útero. *THRB* codifica el receptor THRβ de la forma más activa de la hormona tiroidea (T3). Este gen es el

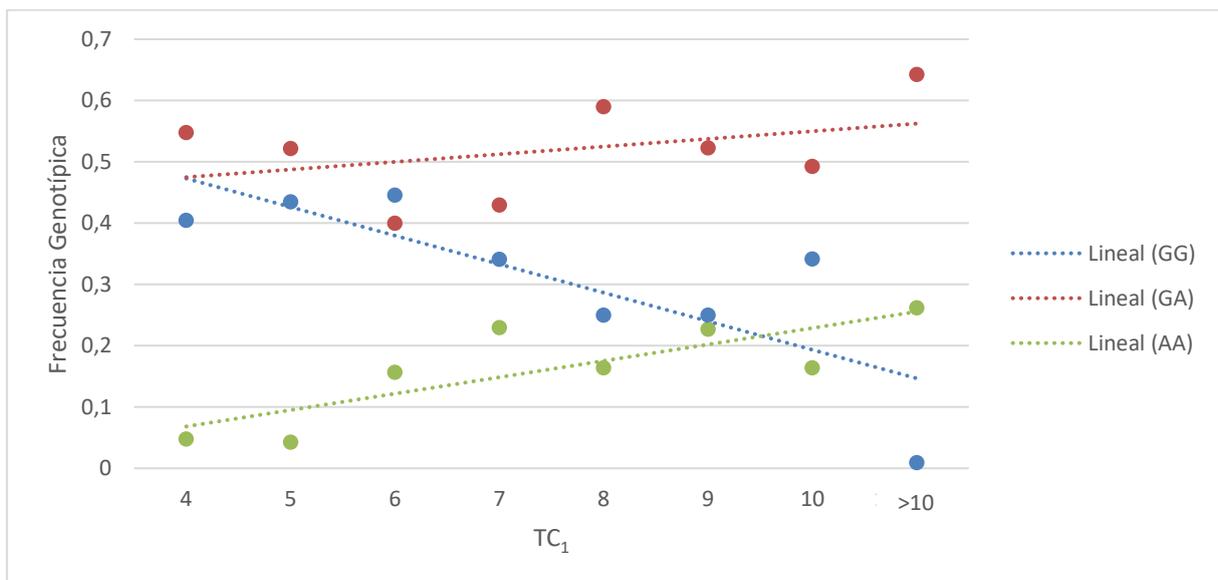
encargado de controlar los niveles de T3 en sangre a través del eje hipotálamo-hipófisis (Figura 2) (Ortiga-Carvalho et al., 2014). En un estudio de Osorio et al. (2013) se observó que la acción conjunta de T3 con la FSH (hormona folículo estimulante) permite controlar la diferenciación de las células de la granulosa en porcino. Además, también estimulan la producción de otras hormonas reproductivas como el estradiol y la hormona luteinizante (LH) (Figura 2) (Osorio et al., 2013; Aghajanova et al., 2011). Es por ello que la regulación de los niveles de HT podría controlar procesos relacionados con la ovulación, el desarrollo del endometrio, la implantación y el desarrollo del embrión (Aghajanova et al., 2011; Cooke et al., 2004). De hecho, todos estos procesos se han visto que son importantes para determinar el tamaño de camada en los animales (Marco-Jiménez y Vicente-Antón, 2012). Por tanto, *THR*B podría estar regulando la prolificidad en cerdo ibérico a través de un control de la homeostasis de la hormona tiroidea T3.



**Figura 2. Efecto de T3 sobre la prolificidad, regulada por el eje hipotálamo-hipofisiario.** La activación de la hormona tiroidea T3, se inicia con la liberación de TRH (Hormona liberadora de tirotrina) por parte del hipotálamo (HTAL). TRH a su vez, actúa sobre los receptores de TRH de la hipófisis (HP), estimulando la liberación TSH (Hormona estimulante de la tiroides). La TSH estimula a la glándula tiroidea (GT) para la producción de T3. Esta actúa sobre las células de la granulosa (CG), estimulando la liberación de LH (Hormona luteinizante) y estradiol, lo que induce la ovulación.

Además de sus funciones biológicas, la importancia de este gen es respaldada por los resultados obtenidos en el estudio GWAS. En primer lugar, el SNP WU\_10.2\_13\_13314717, que determina la región candidata donde se encuentra *THR*B, muestra una asociación relevante con un p-valor de 2,36e-05 (Tabla 3). En segundo lugar, este SNP explica un importante porcentaje de la varianza genética (~40%) para TC<sub>1</sub>. En tercer lugar, la frecuencia del alelo de mínima frecuencia de este SNP en la población es elevada (0,43) (Tabla 3), lo que aumenta la potencia de detección de asociación con el TC<sub>1</sub>. En la figura 3 se observa como el genotipo GG tiende a disminuir mientras que el AA y el GA aumenta a medida que incrementa el tamaño de camada. Por último, en la región candidata asociada a

este SNP solo se encuentra el gen *THRB*. Todos estos resultados indicarían que este SNP está, realmente, recogiendo la mutación causal para TC<sub>1</sub> y que esta mutación causal podría estar asociada a *THRB*, por ser el único gen que se encuentra en la región (Tabla 4). No obstante, hay que tener en cuenta que, aunque este análisis sirve para identificar regiones genómicas asociadas al carácter debido al desequilibrio de ligamiento entre los SNPs y las mutaciones causales (Lopez de Maturana et al., 2014), el efecto estimado del SNP WU\_10.2\_13\_13314717 podría estar sobrestimado ya que el modelo solo incluye un marcador en cada análisis y no tiene en cuenta el desequilibrio de ligamiento entre marcadores (Bush y Moore, 2012).



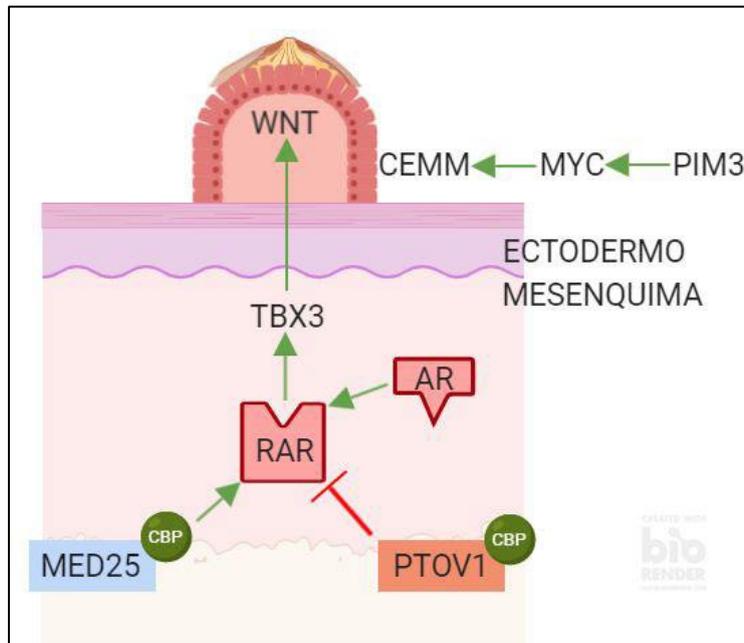
**Figura 3. Tendencia de las frecuencias genotípicas del SNP WU\_10.2\_13\_13314717.** El eje x representa el número de nacidos totales en el primer parto (TC<sub>1</sub>). Las hembras muy prolíficas (>10) están agrupadas. El eje y representa las frecuencias genotípicas para el SNP WU\_10.2\_13\_13314717. Las líneas discontinuas representan la tendencia de las frecuencias genotípicas con el aumento de TC<sub>1</sub>: la línea verde representa el genotipo AA, la roja el GA y la azul el GG.

En este estudio también se identificaron regiones genómicas asociadas con NT, localizadas en los cromosomas 1, 4, 5, 6, 7, 8 y 10. Los genes candidatos de estas regiones fueron *MED25*, *PTOVI*, *PIM3*, *MAPK10*, *PRIM2* y *SMAD6* (Tabla 4). Los genes *MED25*, *PTOVI*, *PIM3*, *MAPK10* y *PRIM2* fueron seleccionados debido a su relación con la formación de las glándulas mamarias, mientras que *SMAD6*, por su relación con el desarrollo de las vértebras. A simple vista, el desarrollo vertebral y el NT no presentan una relación evidente. Sin embargo, Rohrer y Nonneman, (2017) encontraron que el número de vértebras está positivamente correlacionando (0.35) con el NT. Además, diversos GWAS para NT en cerdos también identificaron genes (p.ej. *VRTN*) relacionados con el desarrollo de las vértebras (Lopes et al., 2014; Yang et al, 2016; Rohrer y Nonneman, 2017).

Entre los genes candidatos asociados a la formación de las mamas, *MED25* y *PTOVI* regulan la expresión del receptor del ácido retinoico (RAR) a través de un dominio común que se une a un

coactivador de RAR. Dependiendo de qué gen forme el complejo con el coactivador, se regulará positiva o negativamente la expresión de RAR (Figura 4) (Youn et al., 2011). Existen evidencias que demuestran que los receptores RAR (como el RAR- $\beta$ ) se expresan en diversas estructuras de la glándula mamaria de ratón durante la etapa embrionaria (Cho et al., 2012). De hecho, Cho et al., (2012) en un estudio con ratones, demostraron que se requiere de la señalización del ácido retinoico (AR) para la expresión de *TBX3*, un gen esencial para la formación de las mamas durante la embriogénesis. Se ha observado que ratones knockout para *TBX3* carecen de glándulas mamarias (Davenport et al., 2003). Por otra parte, el gen *PIM3* participa en la regulación de la transcripción de *MYC* (Figura 4). *MYC* es un gen que también participa el desarrollo mamario, al igual que *TBX3*. Ratones knockout demostraron que la supresión de *MYC* de las células epiteliales basales mamarias de los ratones, también impide el correcto desarrollo de la glándula mamaria en estos animales (Moumen, et al., 2012). Por tanto, *MED25*, *POTVI* y *PIM3* podrían estar regulando el desarrollo mamario a través del control de las rutas de señalización de *TBX3* y *MYC* (Figura 4). El gen *MAPK10* pertenece a la familia de las MAP quinasa e interviene en la mayoría de rutas de señalización celular. Este gen se ha visto que se expresa en las glándulas mamarias humanas y que también participa en la morfogénesis de las mismas (Jardim et al., 2018). Finalmente, se vio que SNPs relevantes (Tabla 3) se localizaban dentro de la secuencia génica de *MAPK10* y *PRIM2*. Esto podría indicar que estos SNPs están capturando la mutación causal de estos genes que estarían asociados con NT. Además, estudios de asociación realizados para NT observaron también la presencia de SNPs relevantes dentro de la secuencia génica de estos mismos genes (Peng et al., 2017; Verardo et al., 2016 b). Todos estos resultados evidenciarían la importante relación entre los genes *MAPK10* y *PRIM2* y el NT.

En relación con los procesos involucrados en el desarrollo de las vértebras, el gen *SMAD6* es un regulador de *RUNX2* (Shen et al., 2005). *RUNX2* es un factor de transcripción que controla la maduración de los condrocitos y es esencial en el desarrollo del sistema esquelético de los fetos (Wang et al., 2010). Además, diversos estudios observaron que *RUNX2* se expresa en las vértebras y en el borde de las mismas en los fetos (Wang et al., 2010; Iwasaki et al., 2012). La regulación de *RUNX2* vía *SMAD6* apoyaría los resultados de estudios GWAS para NT que identificaron genes candidatos relacionados con el desarrollo de las vértebras (Lopes et al., 2014; Yang et al, 2016; Rohrer y Nonneman, 2017). No obstante, a pesar de que en este estudio se han encontrado regiones genómicas asociadas a NT, se debe tener en cuenta que los GWAS están diseñados para detectar asociaciones de los fenotipos con variantes comunes en las poblaciones, no para variantes raras (Bush and Moore, 2012; McCarthy et al., 2008). En este caso las frecuencias alélicas de las regiones genómicas asociadas son bajas y cercanas a 0.05 (Tabla 3) lo que limita la potencia de detección (error tipo II).



**Figura 4. Mecanismos reguladores de la formación de las mamas.**  
 El complejo *MED25-CBP* y el *PTOV1-CBP* regulan la expresión del receptor del *RAR* (Receptor del ácido retinoico) favoreciendo a *TBX3*, que participa en el desarrollo mamario. El gen *PIM3* regula la expresión de *MYC*, que favorece las células madre epiteliales mamarias (CEMM).

## 6. CONCLUSIONES

- El estudio de asociación realizado ha revelado regiones genómicas asociadas al NT y al TC<sub>1</sub>, pero no para el TC<sub>2</sub>.
- El gen *THRB* relacionado con procesos como la ovulación y la implantación del embrión a través de la hormona tiroidea T3 es el único gen candidato identificado para el carácter TC<sub>1</sub>.
- Se identificaron 6 genes candidatos que podrían afectar al NT en ibérico debido a su relación con la formación de las vértebras (*SMAD6*) y la formación mamaria (*PIM3*, *MED25*, *PTOVI*, *PRIM2* y *MAPK10*).
- La identificación de genes candidatos para NT y TC<sub>1</sub> podría permitir una mejora conjunta de ambos caracteres, esencial para optimizar la eficiencia reproductiva del cerdo ibérico. Sin embargo, estudios adicionales de los genes candidatos serían necesarios para corroborar y reafirmar la asociación encontrada de estos genes con los caracteres de interés.

## 7. REFERENCIAS.

- AECERIBER. (2012). *Raza porcina ibérica: Reglamentación libro genealógico y programa de mejora*. Recuperado de <http://www.aeceriber.es/noticias/publicado-raza-porcina-ibrica-reglamentacin-especifica-del-libro-genealogico-y-programa-de-mejora.html>
- Aghajanova, L., Stavreus-Evers, A., Lindeberg, M., Landgren, B. M., Sparre, L. S., y Hovatta, O. (2011). Thyroid-stimulating hormone receptor and thyroid hormone receptors are involved in human endometrial physiology. *Fertility and sterility*, 95(1), 230-237. doi: 10.1016/j.fertnstert.2010.06.079
- Anaya-Chavez, Y., y Martinez, B. (2010). Genome-wide association studies (gwas) and its contribution to the genetic of asthma. *Revista Salud Uninorte*, 26(2), 223-231. Recuperado de [http://www.scielo.org.co/scielo.php?script=sci\\_arttext&pid=S0120-55522010000200010](http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S0120-55522010000200010)
- Andersen, I. L., Nævdal, E., y Bøe, K. E. (2011). Maternal investment, sibling competition, and offspring survival with increasing litter size and parity in pigs (*Sus scrofa*). *Behavioral ecology and sociobiology*, 65(6), 1159-1167. doi: 10.1007/s00265-010-1128-4
- Barba, C., Camacho, M. E., Sereno, J. R. B., Dieguez, E., y Delgado, J. V. (2002). Caracterización productiva de las variedades del cerdo ibérico en el periodo predestete. *Archivos de Zootecnia*, 51(194), 229-233. Recuperado de <https://www.redalyc.org/pdf/495/49519427.pdf>
- Browning, B. L., y Browning, S. R. (2016). Genotype Imputation with Millions of Reference Samples. *The American Journal of Human Genetics*, 98(1), 116-126. doi: 10.1016/j.ajhg.2015.11.020
- Bush, W. S., y Moore, J. H. (2012). Genome-wide association studies. *PLoS computational biology*, 8(12), e1002822. doi: 10.1371/journal.pcbi.1002822
- Casellas, J., Ibáñez-Escriche, N., Varona, L., Rosas, J. P., y Noguera, J. L. (2019). Inbreeding depression load for litter size in Entrepelado and Retinto Iberian pig varieties. *Journal of animal science*, 97(5), 1979-1986. doi: 10.1093/jas/skz084
- Chalkias, H. (2013). *Genetic and clinical studies of teat traits in the pig* (tesis doctoral). Swedish University of Agricultural Sciences, Uppsala.
- Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., y Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*, 4(1), 7. doi: 10.1186/s13742-015-0047-8
- Chen, D., Zhao, M., y Mundy, G. R. (2004). Bone morphogenetic proteins. *Growth Factors*, 22(4), 233-241. doi: 10.1080/08977190412331279890
- Cho, K.-W., Kwon, H.- J., Shin, J. -O., Lee, J.- M., Cho, S.-W., Tickle, C., y Jung, H. -S. (2012). Retinoic acid signaling and the initiation of mammary gland development. *Developmental biology*, 365(1), 259-266. doi: 10.1016/j.ydbio.2012.02.020
- Cooke, P. S., Holsberger, D. R., Witorsch, R. J., Sylvester, P. W., Meredith, J. M., Treinen, K. A., y Chapin, R. E. (2004). Thyroid hormone, glucocorticoids, and prolactin at the nexus of physiology, reproduction, and toxicology. *Toxicology and applied pharmacology*, 194(3), 309-335. doi: 10.1016/j.taap.2003.09.016

- Cowin, P., y Wysolmerski, J. (2010). Molecular Mechanisms Guiding Embryonic Mammary Gland Development. *Cold Spring Harbor Perspectives in Biology*, 2(6), a003251. doi: 10.1101/cshperspect.a003251
- Cunningham, F., Achuthan, P., Akanni, W., Allen, J., Amode, M. R., Armean, I. M.,...Flicek, P. (2018). Ensembl 2019. *Nucleic acids research*, 47(D1), D745-D751. doi: 10.1093/nar/gky1113
- Davenport, T. G., Jerome-Majewska, L. A., y Papaioannou, V. E. (2003). Mammary gland, limb and yolk sac defects in mice lacking Tbx3, the gene mutated in human ulnar mammary syndrome. *Development*, 130(10), 2263-2273. doi: 10.1242/dev.00431
- De Maturana, E. L., Ibáñez-Escriche, N., González-Recio, Ó., Marenne, G., Mehrban, H., Chanock, S. J.,...Malats, N. (2014). Next generation modeling in GWAS: comparing different genetic architectures. *Human genetics*, 133(10), 1235-1253. doi: 10.1007/s00439-014-1461-1
- Distl, O. (2007). Mechanisms of Regulation of Litter Size in Pigs on the Genome Level. *Reproduction in Domestic Animals*, 42(2), 10-16. doi: 10.1111/j.1439-0531.2007.00887.x
- EL SECTOR DE LA CARNE DE CERDO EN CIFRAS: Principales Indicadores Económicos 2018, Subdirección General de Productos Ganaderos, Dirección General de Producciones y Mercados Agrarios. (2019). Recuperado de [https://www.mapa.gob.es/va/ganaderia/temas/produccion-y-mercados-ganaderos/indicadoreseconomicossectorporcinoano2018\\_tcm39-379728.pdf](https://www.mapa.gob.es/va/ganaderia/temas/produccion-y-mercados-ganaderos/indicadoreseconomicossectorporcinoano2018_tcm39-379728.pdf)
- Ek-Mex, J. E., Segura-Correa, J. C., Batista-Garcia, L., y Alzina-López, A. (2014). Factores ambientales que afectan los componentes de producción y productividad durante la vida de las cerdas. *Tropical and Subtropical Agroecosystems*, 17(3), 447-462. Recuperado de <http://www.redalyc.org/articulo.oa?id=93935728002>
- Fabuel, E., Barragán, C., Silió, L., Rodríguez, M. C., y Toro, M. A. (2004). Analysis of genetic diversity and conservation priorities in Iberian pigs based on microsatellite markers. *Heredity*, 93(1), 104-113. doi: 10.1038/sj.hdy.6800488
- Fang, M., y Andersson, L. (2006). Mitochondrial diversity in European and Chinese pigs is consistent with population expansions that occurred prior to domestication. *Proceedings of the Royal Society B: Biological Sciences*, 273(1595), 1803-1810. doi: 10.1098/rspb.2006.3514
- Fernández, A., Toro, M., Rodríguez, C., y Silió, L. (2004). Heterosis and epistasis for teat number and fluctuating asymmetry in crosses between Jiaying and Iberian pigs. *Heredity*, 93(2), 222-227. doi: 10.1038/sj.hdy.6800498
- Gjorevski, N., y Nelson, C. M. (2011). Integrated morphodynamic signalling of the mammary gland. *Nature Reviews Molecular Cell Biology*, 12(9), 581-593. doi:10.1038/nrm3168
- Huang, D. W., Sherman, B. T., y Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols*, 4(1), 44-57. doi: 10.1038/nprot.2008.211
- Ibáñez-Escriche, N., y Gonzalez-Recio, O. (2011). Review.Promises, pitfalls and challenges of genomic selection in breeding programs. *Spanish Journal of Agricultural Research*, 9(2), 404-413. Recuperado de <http://revistas.inia.es/index.php/sjar/article/view/1730>
- Ibáñez-Escriche, N., Forni, S., Noguera, J. L., y Varona, L. (2014). Genomic information in pig breeding: Science meets industry needs. *Livestock Science*, 166, 94-100. doi: 10.1016/j.livsci.2014.05.020

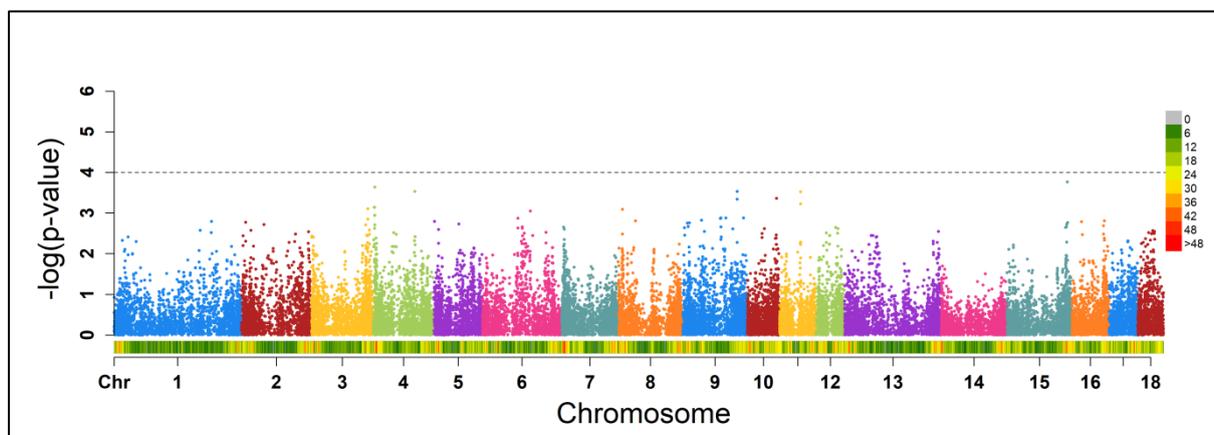
- Ibáñez-Escriche, N., Magallón, E., Gonzalez, E., Tejeda, J. F., y Noguera, J. L. (2016). Genetic parameters and crossbreeding effects of fat deposition and fatty acid profiles in Iberian pig lines. *Journal of Animal Science*, *94*(1), 28-37. doi: 10.2527/jas.2015-9433
- Iwasaki, M., Piao, J., Kimura, A., Sato, S., Inose, H., Ochi, H.,...Takeda, S. (2012). Runx2 Haploinsufficiency Ameliorates the Development of Ossification of the Posterior Longitudinal Ligament. *PLoS One*, *7*(8), e43372. doi: 10.1371/journal.pone.0043372
- Jang, D., Yoon, J., Taye, M., Lee, W., Kwon, T., Shim, S., y Kim, H. (2018). Multivariate genome-wide association studies on tenderness of Berkshire and Duroc pig breeds. *Genes & Genomics*, *40*(7), 701-705. doi: 10.1007/s13258-018-0672-6
- Jardim, J. G., Gulbrandsen, B., Lund, M. S., y Sahana, G. (2018). Association analysis for udder index and milking speed with imputed whole-genome sequence variants in Nordic Holstein cattle. *Journal of Dairy Science*, *101*(3), 2199-2212. doi: 10.3168/jds.2017-12982
- Jawaid, W. (2017). Provides an R Interface to 'Enrichr'
- Kilby, M. D., Barber, K., Hobbs, E., y Franklyn, J. A. (2005). Thyroid hormone action in the placenta. *Placenta*, *26*(2-3), 105-113. doi: 10.1016/j.placenta.2004.08.004
- Lopes, M. S., Bastiaansen, J. W. M., Harlizius, B., Knol, E. F., y Bovenhuis, H. (2014). A genome-wide association study reveals dominance effects on number of teats in pigs. *PLoS One*, *9*(8), e105867. doi: 10.1371/journal.pone.0105867
- Lopez-Bote, C. J. (1998). Sustained utilization of the Iberian pig breed. *Meat Science*, *49*(1), S17-S27. doi: 10.1016/s0309-1740(98)90036-5
- Ma, X., Li, P. H., Zhu, M. X., He, L. C., Sui, S. P., Gao, S.,...Huang, R. H. (2018). Genome-wide association analysis reveals genomic regions on Chromosome 13 affecting litter size and candidate genes for uterine horn length in Erhualian pigs. *Animal*, *12*(12), 2453-2461. doi: 10.1017/s1751731118000332
- Marco-Jiménez, F., Vicente-Antón, J. S. (2012). FUNDAMENTOS Y TÉCNICAS DE LA REPRODUCCIÓN. *Colección Académica. Editorial UPV*.
- Martinez, A. M., Delgado, J. V., Rodero, A. y Vega-Pla, J. L. (2000). Genetic structure of the Iberian pig breed using microsatellites. *Animal Genetics*, *31*(5), 295-301. doi: 10.1046/j.1365-2052.2000.00645.x
- McCarthy, M. I., Abecasis, G. R., Cardon, L. R., Goldstein, D. B., Little, J., Ioannidis, J. P. A., y Hirschhorn, J. N. (2008). Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nature Reviews Genetics*, *9*(5), 356-369. doi: 10.1038/nrg2344
- Monks, J. (2007). TGFβ as a Potential Mediator of Progesterone Action in the Mammary Gland of Pregnancy. *Journal of Mammary Gland Biology and Neoplasia*, *12*(4), 249-257. doi: 10.1007/s10911-007-9056-2
- Moumen, M., Chiche, A., Deugnier, M.-A., Petit, V., Gandarillas, A., Glukhova, M. A., y Faraldo, M. M. (2012). The Proto-Oncogene Myc Is Essential for Mammary Stem Cell Function. *Stem Cells*, *30*(6), 1246-1254. doi: 10.1002/stem.1090
- Noguera, J. L., Varona, L., Babot, D., y Estany, J. (2002). Multivariate analysis of litter size for multiple parities with production traits in pigs: I. Bayesian variance component estimation. *Journal of Animal Science*, *80*(10), 2540-2547. doi: 10.1093/ansci/80.10.2540
- Noguera, J. L. (2011). Mejora genética de los caracteres reproductivos en porcino. *Suis*, *78*, 22-32.

- Noguera, J. L., e Ibáñez-Escriche, N. (2017). Mejora del cerdo ibérico. *Suis*, 142, 28-34.
- Ortiga-Carvalho, T. M., Sidhaye, A. R., y Wondisford, F. E. (2014). Thyroid hormone receptors and resistance to thyroid hormone disorders. *Nature Reviews Endocrinology*, 10(10), 582-591. doi: 10.1038/nrendo.2014.143
- Osorio, J. H., Rodríguez, J. V., y Bedoya, P. A. (2013). Actualización en el metabolismo de las hormonas tiroideas y su relación con la reproducción en el cerdo. *Biosalud*, 12(2), 77-78. Recuperado de <http://www.scielo.org.co/pdf/biosa/v12n2/v12n2a07.pdf>
- Peng, W.-F., Xu, S.-S., Ren, X., Lv, F.-H., Xie, X.-L., Zhao, Y.-X.,...Li, M.-H. (2017). A genome-wide association study reveals candidate genes for the supernumerary nipple phenotype in sheep (*Ovis aries*). *Animal Genetics*, 48(5), 570-579. doi: 10.1111/age.12575
- Phocas, F., y Colleau, J. J. (1996). Genetic steady-state under BLUP selection for an infinite and homogeneous population with discrete generations. *Theoretical and Applied Genetics*, 93(1-2), 287-294. doi: 10.1007/bf00225759
- Rappaport, N., Fishilevich, S., Nudel, R., Twik, M., Belinky, F., Plaschkes, I.,...Lancet, D. (2017). Rational confederation of genes and diseases: NGS interpretation via GeneCards, MalaCards and VarElect. *BioMedical Engineering OnLine*, 16(1). doi: 10.1186/s12938-017-0359-2
- Ramos, A. M., Crooijmans, R. P. M. A., Affara, N. A., Amaral, A. J., Archibald, A. L., Beever, J. E.,...Groenen, M. A. M. (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(8), e6524. doi: 10.1371/journal.pone.0006524
- Rohrer, G. A., y Nonneman, D. J. (2017). Genetic analysis of teat number in pigs reveals some developmental pathways independent of vertebra number and several loci which only affect a specific side. *Genetics Selection Evolution*, 49(1), 1-11. doi: 10.1186/s12711-016-0282-1
- Rutherford, K. M. D., Baxter, E. M., D'Eath, R. B., Turner, S. P., Arnott, G., Roehe, R.,...Lawrence, A. B. (2013). The welfare implications of large litter size in the domestic pig I: biological factors. *Animal Welfare*, 22(2), 199-218. doi: 10.7120/09627286.22.2.199
- Sell-Kubiak, E., Duijvesteijn, N., Lopes, M. S., Janss, L. L. G., Knol, E. F., Bijma, P., y Mulder, H. A. (2015). Genome-wide association study reveals novel loci for litter size and its variability in a Large White pig population. *BMC Genomics*, 16(1) [1049]. doi: 10.1186/s12864-015-2273-y
- Sell-Kubiak, E., Knol, E. F., y Mulder, H. A. (2019). Selecting for changes in average “parity curve” pattern of litter size in Large White pigs. *Journal of Animal Breeding and Genetics*, 136(2), 134-148. doi: 10.1111/jbg.12372
- Shen, R., Chen, M., Wang, Y.-J., Kaneki, H., Xing, L., O’Keefe, R. J., y Chen, D. (2006). Smad6 Interacts with Runx2 and Mediates Smad Ubiquitin Regulatory Factor 1-induced Runx2 Degradation. *Journal of Biological Chemistry*, 281(6), 3569-3576. doi: 10.1074/jbc.m506761200
- Suárez, M. V., Barba, C., Forero, J., Sereno, J. R. B., Diéguez, E., y Delgado, J. V. (2002). Caracterización reproductiva de varias razas porcinas de origen ibérico. I. Análisis descriptivo. *Archivos de Zootecnia*, 51(194), 245-248. Recuperado de <http://www.unavarra.es/rmga/ppor/Comun/web%20razas/iberico.pdf>
- Tang, J., Zhang, Z., Yang, B., Guo, Y., Ai, H., Long, Y.,... Nengshui, D. (2017). Identification of loci affecting teat number by genome-wide association studies on three pig populations. *Asian-Australasian journal of animal sciences*, 30(1), 1-7. doi: 10.5713/ajas.15.0980

- Van Son, M., Lopes, M. S., Martell, H. J., Derks, M. F. L., Gangsei, L. E., Kongsro, J.,...Harlizius, B. (2019). A QTL for Number of Teats Shows Breed Specific Effects on Number of Vertebrae in Pigs: Bridging the Gap Between Molecular and Quantitative Genetics. *Frontiers in Genetics*, *10*,[272]. doi: 10.3389/fgene.2019.00272
- Verardo, L. L., Lopes, M. S., Wijga, S., Madsen, O., Silva, F. F., Groenen, M. A. M.,...Guimarães, S. E. F. (2016 a). After genome-wide association studies: Gene networks elucidating candidate genes divergences for number of teats across two pig populations. *Journal of Animal Science*, *94*(4), 1446–1458. doi: 10.2527/jas.2015-9917
- Verardo, L. L., Silva, F. F., Lopes, M. S., Madsen, O., Bastiaansen, J. W. M., Knol, E. F.,...Guimarães, S. E. F. (2016 b). Revealing new candidate genes for reproductive traits in pigs: combining Bayesian GWAS and functional pathways. *Genetics Selection Evolution*, *48*(1), 1-13. doi: 10.1186/s12711-016-0189-x
- Wang, S., Qiu, Y., Ma, Z., Xia, C., Zhu, F., y Zhu, Z. (2010). Expression of Runx2 and Type X Collagen in Vertebral Growth Plate of Patients with Adolescent Idiopathic Scoliosis. *Connective Tissue Research*, *51*(3), 188–196. doi: 10.3109/03008200903215590
- Xie, L., Luo, C., Zhang, C., Zhang, R., Tang, J., Nie, Q.,...Zhang, X. (2012). Genome-wide association study identified a narrow chromosome 1 region associated with chicken growth traits. *PLoS One*, *7*(2), e30910. doi: 10.1371/journal.pone.0030910
- Yang, J., Lee, S. H., Goddard, M. E., y Visscher, P. M. (2011). GCTA: a tool for genome-wide complex trait analysis. *The American Journal of Human Genetics*, *88*(1), 76-82. doi: 10.1016/j.ajhg.2010.11.011
- Yang, J., Zaitlen, N. A., Goddard, M. E., Visscher, P. M., y Price, A. L. (2014). Advantages and pitfalls in the application of mixed-model association methods. *Nature Genetics*, *46*(2), 100-106. doi: 10.1038/ng.2876
- Yang, J., Huang, L., Yang, M., Fan, Y., Li, L., Fang, S.,...Ren, J. (2016). Possible introgression of the VRTN mutation increasing vertebral number, carcass length and teat number from Chinese pigs into European pigs. *Scientific Reports*, *6*(1), 1-8 [19240] . doi: 10.1038/srep19240
- Youn, H.-S., Park, U.-H., Kim, E.-J., y Um, S.-J. (2011). PTOV1 antagonizes MED25 in RAR transcriptional activation. *Biochemical and Biophysical Research Communications*, *404*(1), 239-244. doi: 10.1016/j.bbrc.2010.11.100
- Zhang, H., Wang, S., Liu, M., Zhang, A., Wu, Z., Zhang, Z., y Li, J. (2013). Differential gene expression in the endometrium on gestation day 12 provides insight into sow prolificacy. *BMC Genomics*, *14*(1), 1-10. doi: 10.1186/1471-2164-14-45

## 8. DATOS SUPLEMENTARIOS.

**Figura S1.** Manhattan plot para el carácter tamaño de camada en el segundo parto (TC<sub>2</sub>).



El eje y representa el logaritmo negativo en base diez del p-valor de asociación para cada SNP. En el eje x representa la posición física de cada SNP en su cromosoma correspondiente. La línea horizontal discontinua indica el umbral de significancia sugerido en para datos de producción animal  $-(\log_{10}(\text{p-value}))=4$ . La leyenda de colores de la derecha indica la densidad de SNPs a lo largo del genoma del cerdo de 0% (gris) a mayor de 48% (rojo). Los SNPs con mayor relevancia están resaltados como puntos rojos

**Tabla S1.** Genes localizados en las regiones candidatas de TC1 implicados en la reproducción y el desarrollo embrionario.

GEN	FUNCIONES
SUCO	Factor de osificación del dominio SUN Relacionado con procesos de meiosis, movilidad de la cromatina, etc Formación de espermatozoides Requerido para embriogénesis tardía
THRB	Receptor de la hormona tiroidea Regulación negativa de la receptibilidad femenina Comportamiento de cortejo femenino
GBX2	Factor de transcripción para la pluripotencia y diferenciación celular del embrión Interviene en la gastrulación
KLHL20	Proteína Kelch también llamada BTB Kelch Mediador de VEGF (Factor de crecimiento endotelial vascular) Presente en capilares pulmonares y embrionario Relacionado con HIF-alfa
TEX50	Proteína expresada en testículos
SLC9C2	Intercambiador sodio-hidrógeno (NHE) Relacionado con el transporte de glucosa, sales biliares y ácidos orgánicos Interviene en la fertilidad masculina y la motilidad de los espermatozoides