

# **CURSO: LA MEJORA GENÓMICA EN PORCINO**

## **Introducción.** Justificación del curso y objetivos.

El conocimiento del genoma es la puerta de entrada para medir la variación entre individuos a nivel de ADN. El desarrollo en los últimos años de plataformas tecnológicas de genotipado masivo, cada vez más eficientes en términos de rapidez y coste, ha hecho posible considerar la variación del ADN en la selección de los reproductores, a la vez que ha generado nuevas expectativas de mejora en los caracteres en los que es más difícil obtener respuesta genética. En el primer capítulo del curso se presentarán las bases del genotipado masivo, para, en el segundo y tercero, precisar cómo el conocimiento de la variación del ADN puede ayudar a mejorar, respectivamente, la selección y los programas de mejora en porcino, incluyendo el caso del cerdo Ibérico, que por su especificidad se tratará en el cuarto y último capítulo.

# Capítulo 1 - La genómica y las técnicas de genotipado masivo.

*Romi Pena & Joan Estany*

*Departamento de Ciencia Animal, Universidad de Lleida – Agrotecnio Center, Lleida.*

## 1.1. ¿Qué es la genómica?

La **genómica** es el estudio del ADN completo de un individuo, es decir, de su genoma. Los análisis genómicos estudian todos los elementos del genoma en su conjunto, en contraposición a los estudios gen a gen. Este salto cualitativo se inició con la descripción de la secuencia completa del genoma humano, a la que luego siguieron la de otros organismos, entre ellos el cerdo. Este hecho precipitó la descripción y catalogación de los genes presentes en el genoma. Así, por ejemplo, ahora sabemos que el **genoma** de un cerdo contiene algo más 20.000 genes codificantes (es decir, genes que codifican proteínas). Se puede consultar el catálogo de genes conocidos en el proyecto *Ensembl*, desde cuya página web se accede a la versión más actual del genoma porcino ([www.ensembl.org/Sus\\_scrofa](http://www.ensembl.org/Sus_scrofa)).

Los estudios de genómica han sido posibles gracias al desarrollo de nuevas herramientas biotecnológicas capaces de encarar eficientemente el estudio de la variación genética, como por ejemplo los chips de ADN o los microarrays de expresión o, más recientemente, las nuevas generaciones de secuenciadores automáticos de ADN. En consecuencia, en la última década han aparecido nuevas plataformas de análisis global, los llamados estudios -ómicos (**Tabla 1**), que permiten estudiar, en un mismo experimento y para cada uno de los tejidos en estudio, qué genes están activos y cuales no (transcriptómica, epigenómica); qué proteínas se producen y en qué cantidad (proteómica); y cómo cambian globalmente sus metabolitos (metabolómica). Esta capacidad para estudiar la biología de un ser vivo de una manera comprehensiva y a varios niveles está revolucionando no sólo el campo de la medicina personalizada sino también el de la mejora genética animal. Este primer capítulo del curso se destinará a revisar la estructura básica de un genoma y las herramientas tecnológicas que permiten su secuenciación. En especial, se describirá qué es y cómo funciona un chip de ADN, la plataforma más utilizada en la práctica para analizar las diferencias individuales en la secuencia de ADN.

## 1.2. La estructura de un genoma

El primer genoma secuenciado de un organismo superior fue el humano, cuya primera versión se publicó en el año 2001. Sin duda se trata del genoma más estudiado, en constante revisión, y del que ya se dispone de la versión 38. La primera versión del genoma del cerdo se hizo pública en el 2009 y la segunda a finales del 2016, aunque la anotación de su estructura (es decir de los elementos – genes y sus reguladores – que lo integran) no estará disponible hasta mayo de 2017. A grandes trechos, los dos genomas tienen una estructura muy parecida, pero la anotación es mucho más completa en el humano, por lo que ésta se utiliza como referencia de una anotación de calidad. Durante el proceso de anotación se genera un mapa (mapa genómico), donde se localiza la posición relativa de cada elemento (por ejemplo, la posición de los genes).

**Tabla 1. Comparación de los estudios –ómicos más frecuentes en referencia al objeto diana analizado y objetivos perseguidos**

Estudio	Diana	Objetivo
<b>Genómica</b>	ADN genómico	Estudiar la variabilidad en la secuencia del ADN. Describir mutaciones y predecir sus consecuencias.
<b>Transcriptómica</b>	ARN mensajero (o transcrito)	Describir los genes activos e inactivos en cada tejido. Cuantificar esta actividad. Describir los distintos transcritos que genera cada gen activo.
<b>Epigenómica</b>	ADN genómico	Describir la distribución en el genoma de marcas epigenéticas como la metilación del ADN o la acetilación de las histonas.
<b>Proteómica</b>	Proteínas	Describir y cuantificar todas las proteínas producidas por los genes activos en cada tejido.
<b>Metabolómica</b>	Moléculas de pequeño tamaño (metabolitos)	Estudiar la huella bioquímica que deja la actividad de los genes/las proteínas en un tejido. Es una marca de su estado fisiológico.

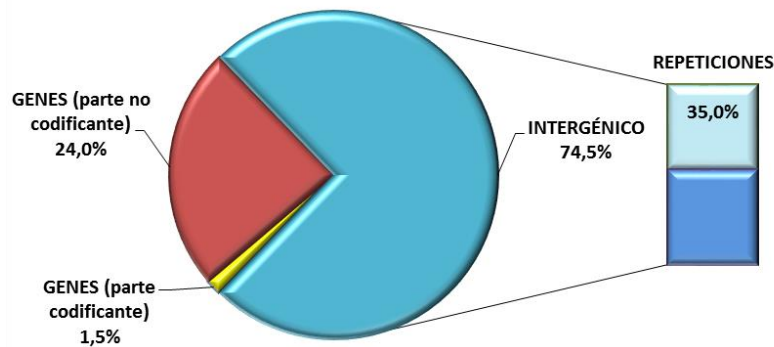
El genoma porcino está formado por  $3,02 \times 10^9$  pares de nucleótidos, repartidos en 18 autosomas más los cromosomas sexuales (X e Y). Cada cerdo tiene dos genomas, uno heredado del padre y otro de la madre. En el genoma del cerdo hay 21.650 genes que codifican para proteínas, aunque este número podría estar algo sobrestimado si se compara con el genoma humano (**Tabla 2**). Sin embargo, las secuencias codificantes (exones) apenas ocupan el 1,5% del genoma (**Figura 1**). Una cuarta parte del genoma la ocupan los intrones, es decir, la parte de los genes que no se traduce finalmente en proteína. El 75% restante se le consideró inicialmente ADN basura, y se postulaba que probablemente fuera una huella de la evolución, una carga de ADN que se había ido arrastrando sin ningún sentido funcional. Es más, la mitad de este ADN basura corresponde a secuencias del denominado ADN repetitivo, es decir, cientos o miles de repeticiones en tándem de un número variable de bases, de 2 a unos 100 pares de nucleótidos. No era de esperar que estas secuencias jugaran ningún papel importante en la fisiología de los animales. No obstante, esta visión del genoma tan centrada en el gen ha sufrido un cambio radical a raíz de las nuevas anotaciones del genoma humano y de su comparación con el genoma de otros organismos en la escala evolutiva.

**Tabla 2. Comparación de los principales elementos del genoma porcino y humano.**

	Genoma	
	Porcino (Ssc11.2)*	Humano (GRCh38)*
<b>Longitud</b> (pares de bases)	3.024.658.544	3.096.649.726
<b>Cromosomas</b>	18 + XY	23 + XY
<b>Número de genes</b>		
Genes codificantes	21.630	20.441
Genes no codificantes	3.124	22.219
Pseudogenes	568	14.606
<b>Número de transcritos</b>		
Transcritos ARN	30.585	198.002
<b>Mutaciones</b>		
Mutaciones de 1 nucleótido (SNP)	60.389.665	156.055.161
Duplicaciones/delecciones	85	5.864.995

(fuente: [www.ensembl.org](http://www.ensembl.org), revisado a 09/03/2017; \*última versión anotada disponibles para cada genoma en esta fecha)

**Figura 1. Composición del genoma porcino.** Sólo el 1,5% del genoma se usa para producir proteínas funcionales. El resto del genoma tiene una función principalmente reguladora de la activación de estos genes codificantes de proteínas.



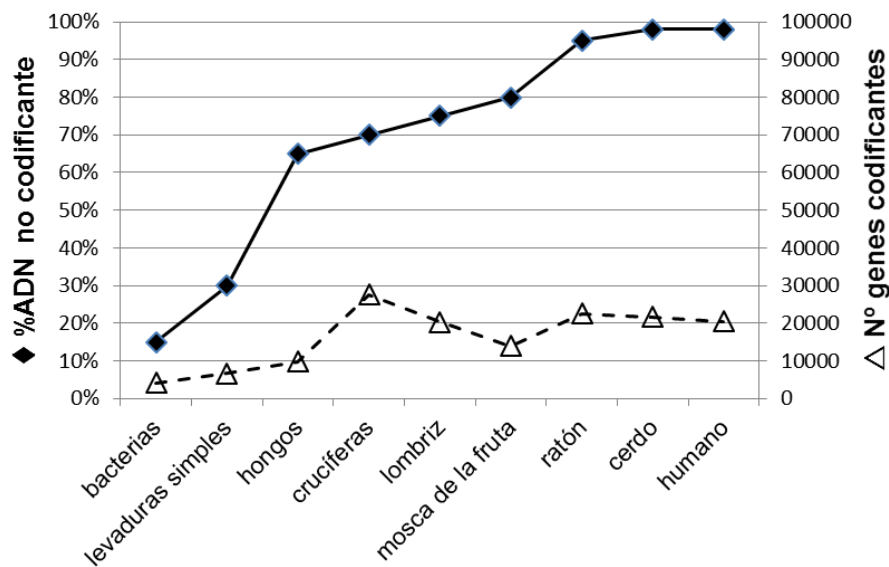
Un hecho sorprendente fue el descubrimiento que el hombre tiene menos genes codificantes que, por ejemplo, la uva (29.971 genes) o que el arroz (35.379 genes). ¿Cómo podía ser que un organismo tan complejo como el hombre tuviera menos genes que un cereal? Es más, otros animales más simples como la mosca de la fruta y las lombrices tienen casi los mismos genes que un humano (13.918 y 20.362, respectivamente). Con estas observaciones queda claro que el número de genes no determina la complejidad de un ser vivo. Sin embargo, al estudiar el ADN no codificante los investigadores vieron que había una relación muy clara entre complejidad de una especie y la cantidad de ADN no codificante de su genoma (**Figura 2**), que llega a ser el 98,5% de todo el genoma en el caso de animales como el hombre o el cerdo.

¿Qué contiene el ADN no codificante que lo hace tan especial? Por una parte, este ADN alberga genes no codificantes, es decir genes que se transcriben en ARN pero que no se traducen nunca en proteínas. En humanos hay tantos genes no codificantes como de codificantes (**Tabla 2**) y es probable que su número en el genoma porcino esté infraestimado. Estos ARNs no codificantes juegan un papel muy importante en la activación e inactivación de los genes codificantes. Además, la mayor parte del ADN restante (el 70-80% del total) son elementos reguladores que también activan e inactivan los genes codificantes. Por lo tanto, el ADN no codificante, lejos de ser “basura”, determina en qué tejido, en qué momento y en qué cantidad se produce cada proteína. Esto explica la relación entre complejidad biológica y ADN no codificante. No importan tanto cuántos genes tiene un organismo sino cómo éstos finamente se utilizan. Utilizando un símil, con 5 hortalizas, en una cocina donde sólo hay un cuchillo disponible, tan sólo se puede preparar una ensalada y poco más. Pero en una cocina mucho más “sofisticada”, que disponga de horno, fogones, robot de cocina, picadora, etc., con las mismas 5 hortalizas se pueden preparar una variedad mucho más amplia de platos y además mucho más elaborados. Esta es la diferencia entre el genoma de una lombriz y el genoma de un cerdo.

**Ejemplo:**

Los genes de las proteínas de la leche, que codifican para las **caseínas**, son genes altamente regulados. Los elementos reguladores del genoma controlan que solo se sintetizen caseínas en las hembras (en qué animales), solo durante la lactación (en qué momento), solo en la glándula mamarias (en qué tejido) y siempre en grandes cantidades (a qué nivel).

**Figura 2.** El porcentaje de ADN no codificante en el genoma, pero no el número de genes codificantes de proteínas, aumenta con la complejidad biológica de las especies (adaptado de Mattick, 2011) (fuente: [www.ensembl.org](http://www.ensembl.org)).



### 1.3. Los chips de genotipado

Una de las consecuencias de la secuenciación del genoma del cerdo es que se han descrito un número elevado de **mutaciones**, es decir, de puntos en la secuencia cuya base (adenina, timina, citosina o guanina) puede ser distinta entre dos cerdos. En la última versión del genoma porcino hay más de 60 millones de mutaciones descritas en las que cambia un sólo nucleótido (**Tabla 2**). Estas mutaciones se denominan SNPs, por las siglas en inglés de *Single Nucleotide Polimorphism*. Esto equivale a una mutación SNP cada 50 nucleótidos. Los SNPs son con diferencia las mutaciones más frecuentes en el genoma porcino. Estas mutaciones suelen presentar dos formas posibles en una población (de ahí la palabra **polimorfismo**): por ejemplo, en una posición determinada unos cerdos pueden tener el nucleótido adenina (A) y otros una guanina (G). Otro tipo de mutaciones son las duplicaciones o las deleciones de fragmentos grandes de ADN, que pueden contener varios genes, y que se conocen como mutaciones estructurales. Estas mutaciones son mucho menos frecuentes que los SNPs y probablemente solo estén parcialmente descritas en el genoma porcino.

Pese a la connotación que tiene la palabra mutación, no todas las mutaciones causan cambios en los cerdos. Es más, se sabe que la mayoría de las mutaciones son silenciosas, esto es, que no tienen efecto, o bien que tienen un efecto casi imperceptible. De hecho, sólo una pequeña minoría provocan cambios en la proteína codificada por un gen o en las zonas que regulan su actividad que, en consecuencia, afectan a un carácter, por ejemplo, el crecimiento, la eficiencia alimentaria o en la susceptibilidad o tolerancia a las enfermedades. Estas mutaciones se conocen como "**mutaciones causales**", ya que causan un cambio relevante en la producción o la salud de un cerdo. Disponer de la lista de todas las mutaciones causales sería de gran utilidad, pues sería suficiente dejar como reproductores aquellos cerdos cuyo ADN portara la combinación más beneficiosa de las distintas versiones de cada mutación. Sin embargo, las mutaciones causales son difíciles de identificar entre todos los polimorfismos presentes en el genoma del cerdo. Por suerte, algunas mutaciones silenciosas pueden utilizarse como marcadores de mutaciones causales si se encuentran próximas a ellas. Esto se debe a que dos mutaciones que están muy cerca en el mismo cromosoma tienden a heredarse juntas, ya que es poco probable que se forme un quiasma de recombinación entre ellas en la formación de los

gametos. Así, el genotipado de muchos marcadores repartidos a lo largo de todo el genoma permitiría controlar la mayoría de los marcadores ligados (unidos) a las mutaciones causales. Las **plataformas de genotipado masivo** (llamados también chips o *arrays* de marcadores) es la tecnología que se ha desarrollado para estudiar de una vez un número elevado de mutaciones.

**Genotipar** una mutación o un marcador significa ser capaz de leer qué base (A, C, T o G) está presente en una posición determinada. Las dos herramientas más utilizadas para genotipar marcadores son los chips de marcadores (SNP-chips) y la secuenciación masiva del genoma.

Aunque cada casa comercial utiliza su propia tecnología, todos los chips de genotipado que actualmente se comercializan constan de un soporte sólido (normalmente derivados de sílice), al que se han unido de forma ordenada unas pequeñas secuencias de ADN (sondas) representativas de cada mutación que se quiere estudiar (**Figura 3**). La información para cada SNP suele ocupar una superficie de aproximadamente 3-5  $\mu\text{m}$  (0,003-0,005 mm) de diámetro en el soporte de sílice. Los chips de ADN pueden llevar información de un elevado número de SNPs. Actualmente hay tres casas comerciales que comercializan chips de SNPs para su uso en porcino (**Tabla 3**). Dos de ellas analizan simultáneamente en torno a 65.000 mutaciones, mientras que la tercera permite obtener información de unos 650.000 SNPs porcinos en una sola reacción. El conjunto de SNPs o marcadores que se analiza conjuntamente se conoce también como un **panel de marcadores**. El orden de cada SNP en el chip está registrado en un fichero de posicionamiento.

**Tabla 3. Chips de genotipado masivo comerciales disponibles actualmente en porcino.**

Proveedor	Nombre comercial	Nº de SNPs	Distancia media entre SNPs (Kb)
<b>Illumina</b>	PorcineSNP60 DNA Analysis Kit v2	64.232	43,4
<b>GeneSeek</b>	GGP Porcine HD	>65.000	43,0
<b>Affymetrix</b>	Axiom® Porcine Genotyping Array	658.692	3,34

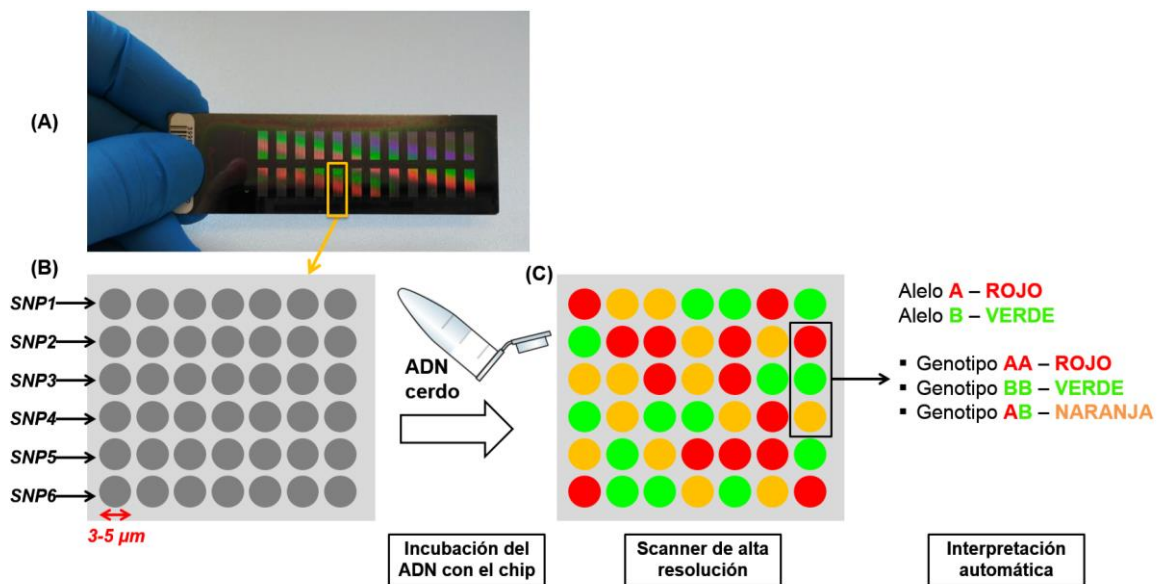
A grandes rasgos, el proceso para genotipar el genoma de un cerdo es parecido entre estas tres plataformas comerciales. Primero, se toma una pequeña muestra de sangre o de tejido de un cerdo. Una muestra de 1mL de sangre en un tubo con anticoagulante (EDTA o citrato) es suficiente. En el laboratorio se aísla ADN genómico a partir de las muestras. Este ADN se incuba sobre el chip de genotipado disponible (**Figura 3**). El pequeño formato de los chips permite procesar 24 chips de cerdo en un mismo soporte de *Illumina* y hasta 96 en el caso de *Affymetrix*. Durante la incubación, el ADN genómico se une a las secuencias que representan cada SNP, proceso que rinde como resultado la emisión de una fluorescencia. Si designamos como A y B las dos variantes de un SNP, entonces, para cada SNP en particular, si el ADN genómico contiene la variante A, se genera una fluorescencia de color rojo y si, por el contrario, éste contiene la variante B, se genera fluorescencia verde. Dada la minúscula superficie que ocupa cada SNP, es necesario tomar una imagen del patrón de colores con un escáner de alta resolución, que se resuelve mediante la utilización de un software asociado que traduce los colores a genotipos, de tal manera que los SNP rojos corresponden a genotipos AA, los verdes a genotipos BB y los naranjas a la mezcla de ambos, es decir, a genotipos heterocigotos AB (**Figura 3**). El resultado final se exporta en un fichero que contiene el genotipo para todo el panel de SNPs del cerdo genotipado, lo que genera una huella genómica única para cada

animal (**Figura 4**). Esta información es la que luego sirve de base para la implementación de la selección asistida por marcadores o la selección genómica (capítulo 2).



**¿Qué se necesita para genotipar un cerdo?**

1. Una pequeña muestra de sangre o tejido (1mL de sangre en un tubo de anticoagulante EDTA o citrato es suficiente) para aislar ADN.
2. Un chip de SNP, cuyo coste actual se mueve entre 80-130€, según la plataforma

**Figura 3. Chips de genotipado masivo. (A)** Imagen del chip PorcineSNP60 de *Illumina*. Cada cuadro (en amarillo) contiene un chip con la información suficiente para analizar 64.000 SNPs en un cerdo. **(B)** Esquema de la estructura interna de un chip representando la posición ordenada de un subgrupo de 42 SNPs (se muestra la posición de los 6 primeros). La información para detectar cada SNP ocupa una superficie de unos 3-5  $\mu\text{m}^2$ ; **(C)** Ejemplo parcial de resultado de genotipado para estos 42 SNPs. Tras incubar con ADN genómico de un cerdo se genera fluorescencia de color rojo o verde según si la variante A o B de cada SNP está presente. Tras escanear el chip con un escáner de alta resolución es posible asignar el genotipo de cada SNP de manera automática mediante la interpretación del color de cada posición.



**Figura 4. Lectura del genotipo de dos cerdos.** Siguiendo el ejemplo de la Figura 3, se resuelve el genotipo para los 6 primeros marcadores del chip en dos animales. Normalmente los genotipos (AA, AB y BB) se simplifican con un código que corresponde al número de alelos A que tiene cada genotipo (2,1 y 0). Así, se muestra el código completo para los primeros 60 SNPs lo que representa una huella genómica única para cerdo (subrayados los correspondientes a los primeros 6 SNPs).

	<b>Cerdo 1</b>	<b>Cerdo 2</b>
		
<b>Genotipo</b>	AA BB AB BB AB AA	BB AB AB BB AA AB
<b>Nº alelos A</b>	2 0 1 0 1 2	0 1 1 0 2 1
<b>Huella genómica</b>	20101221020102102102 21221020202012010200 00200002122111020102...	01102121102010220102 20210202012102012020 22020011111002010201...

## 1.4. La secuenciación del genoma

Otra tecnología que está avanzando con fuerza en estos últimos años consiste en genotipar la secuencia de ADN completa del genoma de un animal. Un secuenciador es un aparato que tiene la capacidad de leer de manera ordenada los nucleótidos de un fragmento de ADN. Los primeros secuenciadores tan sólo leían un fragmento de ADN en cada reacción. En 2005 se comercializó el primer secuenciador de ADN de segunda generación, que ya permitía leer simultáneamente un gran número de fragmentos de ADN. Actualmente los secuenciadores de segunda y tercera generación pueden leer el genoma completo de un cerdo en menos de una semana. Para ello, se fragmenta el ADN genómico en fragmentos de pequeño tamaño (en torno a 200-500 bases) y se secuencian sobre un soporte sólido parecido a un portaobjetos de sílice. Para poder obtener lecturas de cada fragmento por separado, estos se extienden a lo largo del portaobjetos como si de un frotis se tratara. Las lecturas se combinan para reescribir el genoma completo del animal (ensamblaje) y se compara con el genoma de referencia del cerdo, es decir, la última versión actualizada del genoma porcino disponible en [www.ensembl.org/Sus\\_scrofa](http://www.ensembl.org/Sus_scrofa). Así, mediante esta comparación base a base, se pueden describir y genotipar todas las mutaciones del genoma de cada cerdo.

La secuenciación del genoma completo de un cerdo tiene la ventaja de describir de una vez toda la variación de la secuencia de ADN de un animal y no sólo la de los SNPs incluidos en el chip de genotipado. Sin embargo, las técnicas actuales de secuenciación completa presentan todavía numerosas limitaciones, tanto de coste (1.000-3.000 €/cerdo) como de análisis (el ensamblaje de las secuencias y la descripción de nuevas mutaciones no es trivial), lo que las hace inviables para utilizarlas de forma rutinaria. Sin embargo, es posible que en un futuro, con unos precios más asumibles y más experiencia en el análisis de este tipo de datos, esta tecnología represente la nueva manera de integrar la información del genoma de los animales en la mejora de los caracteres productivos. La identificación de todas las variantes del ADN de un cerdo podría hacer posible en un futuro modificarlas directamente con sistemas de edición de genomas (como las herramientas CRISPR-Cas) acumulando así las variantes favorables. La aplicación de estos sistemas está siendo valorada actualmente por algunas de las grandes compañías del sector porcino.

### Lecturas recomendadas.

Mattick JS (2011) The central role of RNA in human development and cognition FEBS Lett 585:1600-1616

Groenen MAM, Schook LB y Archibald AL (2011) Pig genomics. En "The Genetics of the Pig", ed. MF Rothschild y A Rubinsky. CAB International. Pág.179-199



Landi V, Quiroz Valiente J (2011) Los avances de las nuevas tecnologías genéticas y su aplicación en la selección animal. AICA 1:33-43