

ANÁLISIS GENÓMICO DE LA EXPRESIÓN DIFERENCIAL DE ISOFORMAS DE RNA MENSAJERO EN CERDOS CON PERFILES LIPÍDICOS DIVERGENTES

Cardoso, T.^{1,2}, Quintanilla, R.³, Amills, M.¹, Canela, O.³, González-Prendes, R.¹ y Cánovas, A.¹

¹Departament de Genètica Animal, Centre de Recerca en Agrigenòmica (CSIC-IRTA-UAB-UB), Campus de la Universitat Autònoma de Barcelona, Bellaterra 08193. España.

²Fundación CAPES, Ministerio de Educación de Brasil, Brasilia - DF 70040-020, Brasil.

³Genètica y Mejora Animal, IRTA, Torre Marimon, Caldes de Montbui, 08140. España.

INTRODUCCIÓN

El metabolismo lipídico muscular porcino está influenciado por una multitud de factores genéticos y ambientales (Rosenvold et al., 2003). Una de las aproximaciones para comprender la fisiología genómica de dicho carácter consiste en comparar los patrones de expresión de individuos con perfiles divergentes para fenotipos de engrasamiento (Cánovas et al., 2010a; Puig-Oliveras et al., 2014). Sin embargo, se ha analizado muy poco si la expresión diferencial afecta por igual a todos los transcritos, generados por procesamiento (*splicing*) alternativo, a partir de un RNAm, o bien únicamente afecta a una isoforma concreta del mismo. Esta cuestión es relevante puesto que en los genomas de mamíferos la gran mayoría de genes están sometidos a *splicing* alternativo, generando una amplia variedad de transcritos que pueden tener efectos funcionales diferentes (Witten et al., 2011). Los patrones de *splicing* pueden estar afectados por polimorfismos nucleotídicos (SNPs) exónicos localizados en secuencias potenciadoras (*enhancer*) del *splicing* o incluso en los mismos lugares consenso de *splicing*. El advenimiento de la tecnología de “*RNA-Sequencing*”, ha permitido cuantificar, por primera vez, la expresión diferencial de isoformas de RNAm a una escala genómica así como identificar las variaciones estructurales en el transcriptoma (Piskolet et al., 2013). Por otra parte, el análisis conjunto de datos genómicos estructurales y de expresión (genética/genómica) con las técnicas de modelización de sistemas complejos (biología de sistemas) ofrece la oportunidad de estudiar la regulación del transcriptoma muscular desde una perspectiva más completa y profunda. En este trabajo se ha analizado el transcriptoma del músculo *Gluteus medius* en dos grupos de cerdos con fenotipos lipídicos divergentes (ALTO y BAJO) mediante la tecnología de “*RNA-Sequencing*”. Nuestro objetivo principal consiste en identificar isoformas mRNA diferencialmente expresadas entre ambos grupos. Por otra parte, se pretende averiguar si hay polimorfismos génicos que segreguen específicamente en los grupos ALTO y BAJO, y correlacionar dicha información con los patrones de *splicing*.

MATERIAL Y MÉTODOS

Muestras biológicas y Diseño Experimental: El material animal utilizado en el presente estudio procede de muestras del músculo *gluteus medius* (GM) de una población comercial Duroc de 350 individuos con alto contenido en grasa intramuscular (Gallardo et al., 2008). Se ha realizado un análisis de componentes principales (PCA) para seleccionar los animales con perfiles lipídicos divergentes para una combinación de 13 caracteres relacionados con el depósito de grasa y los niveles de lípidos plasmáticos (Cánovas et al., 2010a). Finalmente, se han seleccionado 52 animales distribuidos en dos grupos ALTO (n=26) y BAJO (n=26) de acuerdo a su perfil lipídico.

RNA-Sequencing: Con la finalidad de medir la expresión génica con una elevada resolución y analizar la estructura de los transcritos, se ha secuenciado mediante la tecnología *RNA-Sequencing*, el transcriptoma de 52 músculos GM con la plataforma HiSeq2000 (Illumina, Inc.). Las lecturas han sido mapeadas y anotadas de acuerdo al genoma de referencia porcino (versión 78; ftp://ftp.ensembl.org/pub/release-78/genbank/sus_scrofa/) utilizando el software CLC Bio Genomics (Qiagen).

Variación estructural: Los datos de expresión se han normalizado calculando RPKM (lecturas por kilobase por cada millón de lecturas mapeadas) para cada gen y se han transformado para seguir una distribución normal. La expresión diferencial de los transcritos y la selección de los genes que presentan más de una isoforma *splicing*, se ha realizado mediante el análisis empírico “*digital gene expression (DGE)*” descrito por Robinson y Smyth (2008). La expresión de cada una de las variantes de *splicing* alternativo ha sido

normalizada mediante el cálculo del número de lecturas. La detección de SNPs se ha llevado a cabo de acuerdo a un modelo estadístico que permite discriminar errores de secuenciación de variantes que segregan a baja frecuencia. La identificación y filtrado de SNPs artefactuales se ha realizado aplicando los parámetros descritos en Cánovas et al. (2010b).

Análisis de vías metabólicas: Se ha integrado la información estructural y funcional derivada de los datos genómicos con el fin de identificar los genes que actúan como reguladores del depósito de la grasa en el músculo porcino. Para identificar las rutas metabólicas más representadas en la lista de genes con variantes de *splicing* diferencialmente expresadas en los grupos ALTO y BAJO, se ha utilizado el software IPA (Qiagen).

RESULTADOS Y DISCUSIÓN

Mediante la tecnología de *RNA-Sequencing* se ha obtenido un total de 70 millones de lecturas *paired-end*, de las cuales el 80% ha mapeado en el genoma de referencia porcino. Se ha identificado 2.937 genes diferencialmente expresados (DE) entre los animales con ALTO y BAJO nivel de engrasamiento ($p < 0,05$). En un segundo análisis más estricto, $p < 0,01$ y Fold-change (FC) $> 1,5$, se ha identificado 96 genes DE. El análisis de rutas metabólicas ha evidenciado que los genes DE entre los grupos ALTO y BAJO están involucrados en la ruta de la glucólisis y gluconeogénesis, señalización mediante el factor de transcripción PPAR, fosforilación oxidativa y biosíntesis de ácidos grasos.

En el estudio de la variación estructural del transcriptoma del músculo porcino, se ha identificado 181 isoformas generadas por *splicing* (correspondientes a 141 genes) diferencialmente expresadas en los grupos ALTO y BAJO ($p < 0,05$). Aplicando criterios más estrictos ($p < 0,01$ y $FC > 1,5$), se ha detectado 30 isoformas (correspondientes a 28 genes) con expresión diferencial. La Tabla 1 muestra algunas de las variantes *splicing* con una expresión diferencial más significativa. Un total de 21 isoformas mRNA ha presentado una expresión aumentada en el grupo ALTO (p.e. *ITGA5*, *HP*, *SCD*, *TIMP1*), mientras que 9 han observado una menor expresión (p.e. *UBD*, *RXRG*, *IGJ*, *SPINK4*). Por ejemplo, el gen *UBD*, relacionado con la diabetes, codifica dos isoformas mRNA de 1.281 pb y 534 pb, de las cuales sólo la isoforma de 1.281 pb presenta expresión diferencial. De modo similar, el gen *SCD*, que tiene un papel clave en la síntesis de ácidos grasos monoinsaturados, codifica 4 isoformas distintas (5.585 pb, 1.791 pb, 742 pb y 690 pb), presentando una expresión diferencial restringida a la isoforma de 5.585 pb. En definitiva, nuestros resultados indican que una fracción importante de la expresión diferencial entre grupos afecta, de manera selectiva, a ciertas isoformas mRNA. El conocimiento funcional de estas variantes *splicing* permitiría comprender como la variación estructural de los transcritos generada por procesamiento alternativo del RNAm afecta al metabolismo lipídico porcino y a la calidad de la carne.

Tabla 1. Isoformas mRNA diferencialmente expresadas en cerdos con fenotipos divergentes para caracteres relacionados con el metabolismo lipídico.

Gen	SSC	Tránsito (pb)	Expresión	P-Value	RPKM ALTO	RPKM BAJO
HP	6	1.340	x 4,02	0,00001	0,79	3,19
IGJ	8	1.327	x -6,83	0,00120	42,35	6,20
IL4R	3	1.579	x 1,84	0,00066	4,27	7,87
IL4R	3	2.574	x 1,90	0,00479	2,63	4,93
LITAF	3	2.190	x 1,92	0,00010	2,56	3,24
RXRG	4	544	x -1,68	0,00041	5,46	3,24
SCD	14	5.585	x 1,96	0,00324	7,53	14,77
UBD	1	1281	x -1,70	0,0010	2,37	4,05

*RPKM: Lecturas por kilobase de transcripción por millón de lecturas mapeadas

Por otra parte, se ha observado la existencia de polimorfismos nucleotídicos (SNPs) del transcriptoma que segregan de forma específica en los grupos ALTO o BAJO. Más concretamente, se ha identificado 2.955 (16,8% no-sinónimos) y 1.467 (12% no-sinónimos)

SNP específicos de los grupos ALTO y BAJO, respectivamente. Algunos de estos SNPs están localizados en loci relevantes desde un punto de vista metabólico p.e. ciertos SNPs situados en los genes *DHRS7C* (deshidrogenasa/reductasa expresada en cardiomiocitos), *MYO18B* (regulador de genes específicos del músculo) e *IRS1* (receptor 1 de la insulina), segregan únicamente en los cerdos del grupo ALTO. Por otra parte, SNPs localizados en los genes *PYGM* (una de las enzimas que participan en la glucogénesis muscular), *CPE* (carboxipeptidasa) y *PARVB* (organización del citoesqueleto), segregan únicamente en los animales pertenecientes al grupo BAJO. Es posible que una fracción de los SNPs específicos de grupo explique parte de las diferencias fenotípicas que existen entre cerdos con perfiles lipídicos divergentes.

Asimismo, en algunos casos se ha identificado genes que presentan expresión diferencial a nivel de variantes *splicing* y que también poseen SNPs que segregan específicamente en los grupos ALTO o BAJO. Por ejemplo, la isoforma de 814 pb del gen *CFL1* (Cofilin 1), el cual afecta a la proporción de ácidos grasos poli-insaturados n-6/n-3 en vacuno (Sevane et al., 2014), está sobre-expresada en los cerdos del grupo ALTO. Por otra parte, el gen *CFL1* posee un SNP no-sinónimo en la posición g.5578687(T>A), que segrega específicamente en el grupo ALTO. Resultaría de gran interés determinar si existe algún vínculo funcional entre la expresión diferencial de isoformas mRNA y la segregación grupo-específica de SNPs que pudieran afectar al mecanismo de *splicing*.

REFERENCIAS BIBLIOGRÁFICAS

• Cánovas et al., 2010a. BMC Genomics. 11:372. • Cánovas et al., 2010b. Mamm. Genome, 21:592-598. • Gallardo et al., 2008. Physiol. Genomics. 35,3:199-209. • Piskol et al., 2013. Am. J. Hum. Genet. 3,93:641-51. • Puig-Oliveras et al., 2014. PloS One. 9: e99720. • Robinson & Smyth. 2008. Biostatistics. 2:321-32. • Rosenvold & Andersen. 2003. Meat Sci. 64:219–237. • Sevane et al., 2014. J. Anim. Sci. & Biotech. 5:20. • Witten & Ule. 2011. Trends Genet. 27:89-97

Agradecimientos: Financiado por el proyecto AGL2010-22208-C02-01. A. Cánovas disfruta de un contrato financiado por el programa Juan de la Cierva (JCI-2011-10804). Los autores agradecen la colaboración de Selección Batallé y del personal de IRTA-Monells en la obtención del material animal y los datos fenotípicos.

A GENOME-WIDE ANALYSIS OF DIFFERENTIALLY EXPRESSED MESSENGER RNA ISOFORMS IN PIGS WITH DIVERGENT LIPID PROFILES

ABSTRACT: We have used *RNA-Sequencing* to examine the structural variation of the *gluteus medius* (GM) muscle transcriptome of commercial Duroc pigs with divergent phenotypes (HIGH and LOW) for 13 fatness traits. The whole transcriptome of 52 GM samples from HIGH (n=26) and LOW (n=26) pigs was sequenced with a HiSeq2000 platform. In this way, 30 splice variants were differentially expressed between pigs with divergent fatness traits (p<0.01 and Fold-change>1.5). Pathway analysis evidenced that most of differentially expressed genes belong to the glycolysis and gluconeogenesis, oxidative phosphorylation, insulin signaling and fatty acid biosynthesis metabolic routes. We can conclude that in many instances differential expression targets specific mRNA isoforms rather than the global set of transcripts produced by a given gene. Besides, several SNP variants segregated specifically in either the HIGH (2,995 SNPs) or LOW (1,467 SNPs) fatness profiles. Among them, 479 (16.8% of HIGH group-specific SNPs) and 177 (12.0% of LOW group-specific SNPs) were associated with amino acid changes. It would be interesting to correlate SNP frequencies and isoform differential expression data to identify polymorphisms with effects on the mechanism of *splicing*.

Keywords: *RNA-Sequencing*, Alternative *splicing*, SNPs, Intramuscular fat.