

VARIACIÓN DEL TRANSCRIPTOMA DEL MÚSCULO PORCINO EN ANIMALES CON FENOTIPOS EXTREMOS PARA LA COMPOSICIÓN DE ÁCIDOS GRASOS

Puig-Oliveras, A.^{1,2}, Ramayo-Caldas, Y.^{1,2}, Corominas, J.^{1,2}, Castelló, A.¹, Pérez-Montarelo, D.³, Casellas, J.⁴, Folch, J.M.^{1,2}, Ballester, M.¹

¹Centre de Recerca en Agrigenòmica, CSIC-IRTA-UAB-UB, edifici CRAG, Campus UAB, 08193 Bellaterra, Barcelona. ²Departament de Ciència Animal i dels Aliments, Facultat de Veterinària, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona.

³Departamento Mejora Genética Animal, SGIT-INIA, Ctra. Coruña Km7.5, 28040 Madrid.

⁴Genètica i Millora Animal, IRTA, Av. Rovira Roure 191, 25198 Lleida.

anna.puig@cragenomica.es

INTRODUCCIÓN

Los avances en las técnicas de secuenciación masiva y el correspondiente manejo de gran volumen de datos nos permiten explorar el transcriptoma de forma cada vez más rápida y precisa (Trapnell et al., 2012). La metodología de RNA-Seq permite estimar la expresión génica a partir del número de secuencias mapeadas para cada transcrito y aporta, además, información sobre nuevos genes, procesamiento alternativo del ARNm, utilización de diferentes promotores y señales de terminación de la transcripción. Sin embargo, actualmente existen pocos trabajos donde se haya analizado el transcriptoma de animales de interés productivo como el cerdo (Esteve-Codina et al., 2011; Chen et al., 2011; Ramayo-Caldas et al., 2012b; Yu et al., 2013).

La composición de ácidos grasos en el músculo es especialmente importante en la determinación de la terneza y el sabor de la carne, afectando a su estabilidad oxidativa durante el procesamiento de los productos curados (Wood et al., 2008). Además, ha adquirido recientemente una mayor importancia debido a su impacto sobre la salud humana (Yu et al., 2013). Se trata de un carácter complejo, determinado por factores ambientales, como la dieta, y múltiples factores genéticos. La importancia de esta base genética resulta evidente en la comparación de las diferentes razas porcinas y en la identificación de QTLs para la composición de ácidos grasos.

El objetivo del presente trabajo consiste en analizar el transcriptoma del músculo en cerdos (25% Ibérico, 75% Landrace) con fenotipo extremo para la composición de ácidos grasos en grasa intramuscular. Los resultados obtenidos nos pueden ayudar en la identificación de los genes y redes génicas implicados en la regulación de la composición y el metabolismo de los ácidos grasos en músculo.

MATERIAL Y MÉTODOS

Material animal: El material utilizado en este trabajo procede de un cruce entre 3 machos Ibéricos (Guadyrvas) con 31 madres Landrace (Pérez-Enciso et al., 2000). Cinco animales de la generación F1 fueron cruzados con 26 madres Landrace obteniendo 144 animales (BC1_LD). Se recogieron muestras del músculo *Longissimus dorsi* de estos animales que fueron congeladas con nitrógeno líquido y almacenadas a -80°C.

Se seleccionaron 12 animales BC1_LD pertenecientes a dos grupos extremos, *Low* (L) y *High* (H), a partir del primer componente de un análisis de componentes principales para la composición de ácidos grasos en músculo (Ramayo et al., 2012b). Con el objetivo de eliminar la variabilidad atribuible al sexo, sólo se seleccionaron hembras.

Extracción de ARN y secuenciación: Se aisló el ARN total a partir de las muestras del músculo *Longissimus dorsi* de los 12 animales seleccionados utilizando el kit *RiboPure™* (Ambion). El ARN total se cuantificó en un espectrofotómetro *NanoDrop ND-1000* (NanoDrop products). La pureza y la integridad del ARN se verificaron utilizando un *Bioanalyzer-2100* (Agilent Technologies). Las secuencias se generaron en el *CNAG* (Centro Nacional de Análisis Genómico) utilizando un equipo *Hi-Sec 2000* (Illumina) con una longitud de 75 pb y extremos pareados.

Mapeo y anotación de las lecturas: Se utilizó el programa *FASTQC* para el control de calidad de las lecturas [<http://www.bioinformatics.bbsrc.ac.uk/projects/fastqc/>]. Las lecturas fueron indexadas con *BOWTIE 2.0.0-beta7* (Langmead et al., 2009) y a su vez mapeadas contra la secuencia de referencia del genoma del cerdo (*Sscrofa10.2*), utilizando el programa *TopHat v2.0.1* (Trapnell et al., 2009) con un máximo de tolerancia de dos nucleótidos distintos por cada lectura. Los archivos *bam* resultantes que contenían las

secuencias alineadas, se fusionaron mediante el programa *Samtools* (Li et al., 2009). Se anotaron las lecturas utilizando la opción *intersectBed* del programa *BEDtools* (Quinlan et al., 2010). Los transcritos fueron ensamblados con el programa *Cufflinks* (Trapnell et al., 2010) y *Samtools* fue empleado para determinar los estadísticos relacionados con el mapeo.

Análisis de genes diferencialmente expresados: El número de lecturas que mapean en cada gen fue determinado con el programa *Qualimap v0.5* (García-Alcalde et al., 2012). Para identificar los genes expresados diferencialmente entre grupos se utilizaron los paquetes *DESeq* (Anders et al., 2010) y *EdgeR* (Robinson et al., 2010), seleccionando aquellos genes con un *fold change* $\geq 1,2$ y un *p-value* $\leq 0,01$ (equivalente a un *FDR* $\leq 0,1$). Finalmente, se identificaron las funciones biológicas más relevantes mediante *FatiGO* de *Babelomics 4.3* [<http://babelomics.bioinfo.cipf.es>] (*FDR* $< 0,1$) utilizando como bases de datos *KEGG* (<http://www.genome.jp/kegg>), *Reactome* (<http://www.reactome.org>) y *GO* (<http://www.geneontology.org>).

RESULTADOS Y DISCUSIÓN

Doce muestras del músculo *Longissimus dorsi* de hembras con fenotipos extremos para la composición de ácidos grasos en grasa intramuscular fueron analizadas mediante RNA-Seq en un Hi-Seq de Illumina. Se obtuvieron un total de 787,5 M de lecturas pareadas de 75 pb. El 84,3% (75,3%-86,6%) de los transcritos fueron asignados al genoma de referencia, de los cuales, aproximadamente el 13,8%, no tuvieron una localización única. El porcentaje de lecturas mapeadas fue mayor que el descrito previamente (71,4%-77,7%) por Ramayo-Caldas et al. (2012b) en un análisis del transcriptoma de hígado. Esta diferencia podría deberse al uso de distintas versiones del genoma de referencia, mientras que Ramayo-Caldas et al. (2012b) utilizó la versión *Sscrofa 9.61*, en el presente trabajo se ha utilizado la versión más reciente *Sscrofa 10.2*. Del total de lecturas mapeadas, el 85,9% correspondió a genes anotados, el 93,0% se localizó en exones y sólo el 7,0% en intrones. El 14,1% de las lecturas mapearon en regiones intergénicas, lo que indica que se localizan en genes o exones de genes no anotados en el genoma de referencia.

Una vez ensamblados los transcritos, el 60,8% de las lecturas mapearon sobre exones. El 21,3% se identificaron como posibles nuevas isoformas y un 16,4% como transcritos intergénicos. Se identificaron un total de 314 genes diferencialmente expresados entre grupos con *EdgeR* y 208 con *DESeq*. Para identificar las rutas metabólicas más sobrerrepresentadas se seleccionaron los 131 genes que resultaban diferencialmente expresados por los dos programas (Figura 1). Dieciocho corresponden a genes candidatos localizados en regiones genómicas asociadas con la composición de ácidos grasos en músculo en un estudio GWAS (Ramayo-Caldas et al., 2012a). Las rutas metabólicas más sobrerrepresentadas fueron la ruta de señalización de los receptores activados por proliferadores de peroxisomas (PPAR) y la de la insulina (*KEGG*), ambas vías relacionadas con el metabolismo de ácidos grasos, la ruta del metabolismo de lípidos y lipoproteínas (*Reactome*) y el desarrollo del músculo (*GO*) (Tabla 1). Estos resultados se ajustan con los descritos por Ramayo-Caldas et al. (2012b) analizando el transcriptoma del hígado, donde la señalización del PPAR se identificó como una de las principales rutas metabólicas. Por tanto, estos resultados concuerdan con las diferencias fenotípicas de los animales analizados.

Debido al gran impacto que la composición de ácidos grasos tiene sobre la calidad de la carne, estos resultados pueden ser de gran valor económico en los programas de selección para la industria porcina con el fin de mejorar la calidad de la carne desde el punto de vista sensorial y nutricional. El conocimiento de las rutas metabólicas de los ácidos grasos en cerdo y su regulación, puede servir también como modelo para enfermedades humanas como la obesidad o la diabetes (Chen et al., 2011).

REFERENCIAS BIBLIOGRÁFICAS

- Anders, S. y Huber, W. 2010. *Genome Bio.* 11(10):R106.
- Chen, C., Ai, H., Ren, J. et al. 2011. *BMC Genomics* 12(1):448.
- Esteve-Codina, A., Kofler, R., Palmieri, N. et al. 2011. *BMC Genomics* 12(1):552.
- Langmead, B., Trapnell C., Pop, M. et al. 2009. *Genome Biol.* 10:R25.
- Li, H., Handsaker, B., Wysoker, A. et al. 2009. *Bioinformatics* 25(16):2078-2079.
- Pérez-Enciso M., Clop A., Noguera J. L. et al. 2000 *Journal of Animal*

Science 78: 2525-31. • Quinlan, A.R. y Hall, I.M. 2010. Bioinformatics 26(6):841–842. • Ramayo-Caldas, Y., Mercadé, A., Castelló, A. et al. 2012a. J. Animal Sci 90:1-11. • Ramayo-Caldas, Y., Mach, N., Esteve-Codina, A. et al. 2012b. BMC genomics 13:547. • Robinson, M.D., McCarthy, D.J. y Smyth, G.K. 2010. Bioinformatics 26(1):139-140. • Trapnell, C., Pachter, L. y Salzberg, S.L. 2009. Bioinformatics 25(9):1105-1111. • Trapnell, C., Williams, B.A., Pertea, G. et al. 2010. Nat. Biotechnol. 28:511-515. • Trapnell, C., Roberts, A., Goff, L., et al. 2012. Nature protocols 7(3):562-578. • Wood, J.D., Enser, M., Fisher A.V. et al. 2008. Meat Science 78:343-358. • Yu, K., Shu, G., Yuan, F. et al. 2013. Int. J. Biol. Sci. 9(1):108-118.

Agradecimientos: Este trabajo ha sido financiado por el proyecto AGL2011-29821-C02 (Ministerio de Economía y Competitividad) y por el Programa de Innovación Consolidar-Ingenio 2010 (CSD2007-00036). A. Puig-Oliveras ha sido financiada con una beca de la Universidad Autónoma de Barcelona (PIF, 458-01-1/2011), Y. Ramayo-Caldas con una beca FPU (AP2008-01450) y J. Corominas con una beca FPI (BES-2009-018223).

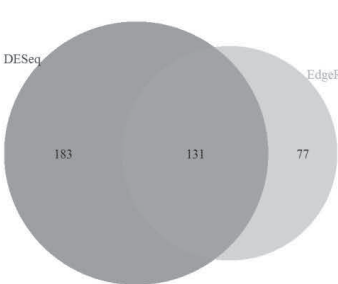


Figura 1. Diagrama de Venn de los genes diferencialmente expresados utilizando los programas DESeq y EdgeR.

Tabla 1. Rutas metabólicas sobrerrepresentadas utilizando el programa Babelomics 4.3 de los genes diferencialmente expresados entre los dos grupos con fenotipos extremos para la composición intramuscular de ácidos grasos.

Ruta Metabólica	Base de datos	Nº genes	Significación (p-value)
Metabolismo de alanina, aspartato y glutamato	KEGG	3	6,2x10 ⁻⁴
Vía de señalización del PPAR	KEGG	4	7,0x10 ⁻⁴
Vía de señalización de la insulina	KEGG	5	1,5x10 ⁻³
Metabolismo de lípidos y lipoproteínas	Reactome	6	1,6x10 ⁻⁴
Hemostasis	Reactome	7	4,8x10 ⁻⁴
Respuesta a sustratos orgánicos	GO	24	3,4x10 ⁻⁷
Desarrollo del órgano muscular	GO	12	3,6x10 ⁻⁵
Biosíntesis de carbohidratos	GO	8	1,4x10 ⁻⁴
Respuesta a estímulos hormonales	GO	16	2,5x10 ⁻⁴

SWINE MUSCLE TRANSCRIPTOME VARIATION IN PHENOTIPICALLY EXTREME ANIMALS FOR FATTY ACIDS COMPOSITION

ABSTRACT: The aim of this study was to explore the muscle (*Longissimus dorsi*) transcriptome and to identify differentially expressed genes among pigs with extreme phenotypes for muscle fatty acid profile. Here, we sequenced the muscle transcriptome of 12 females with divergent phenotypes belonging to an Iberian X Landrace backcross. Differential expression analysis identified 131 genes. Pathway analysis revealed that the most statistically significant metabolic pathways were PPAR signalling, insulin signalling, as well as, metabolism of lipids and lipoproteins. Moreover, 18 of the differentially expressed genes were located in genomic regions associated with muscle fatty-acid composition in a GWAS study. The overrepresented pathways are in concordance with the phenotypic differences observed among the analysed animals. Because fatty acid composition has a great impact on meat quality in terms of taste and nutritional value, the study of muscle transcriptome may be important for both consumers and producers.

Keywords: RNA-Seq, *Longissimus dorsi*, differential expression, fatty acid composition