

# ESTUDIOS DE CO-ASOCIACIÓN GENÓMICA Y REDES GÉNICAS APLICADOS A LA DISECCIÓN GENÉTICA DEL CONTENIDO INTRAMUSCULAR DE ÁCIDOS GRASOS

Ramayo-Caldas, Y.<sup>1,2</sup>, Reverter, A.<sup>3</sup>, Ballester, M.<sup>1</sup>, Fortes, M.R.S.<sup>4</sup>, Esteve-Codina, A.<sup>1</sup>, Noguera, J.L.<sup>5</sup>, Fernández, A.I.<sup>6</sup>, Folch, J.M.<sup>1,2</sup>

<sup>1</sup>Centre de Recerca en Agrigenòmica (CRAG), Consorci CSIC-IRTA-UAB-UB, Campus UAB, Bellaterra, 08193. <sup>2</sup>Departament de Ciència Animal i dels Aliments, Facultat de Veterinària, Universitat Autònoma de Barcelona, 08193, Bellaterra. <sup>3</sup>Commonwealth Science and Industrial Research Organisation, division of Animal, Food and Health Sciences, Brisbane, Qld 4067, Australia. <sup>4</sup>The University of Queensland, Queensland Alliance for Agriculture and Food Innovation, Center for Animal Science, Gatton, Qld 4343, Australia. <sup>5</sup>Genètica i Millora Animal, IRTA, Av. Rovira Roure 191, 25198 Lleida, <sup>6</sup> Departamento Mejora Genética Animal, SGIT-INIA, Ctra. Coruña Km7.5, 28040 Madrid.  
E-mail: [yulixaxis.ramayo@uab.es](mailto:yulixaxis.ramayo@uab.es)

## INTRODUCCIÓN

El cerdo (*Sus scrofa domestica*) constituye una de las principales fuentes de carne para la humanidad, y es también un excelente modelo animal para el estudio de enfermedades metabólicas en humanos como la obesidad, la diabetes tipo 2 y la arteriosclerosis (Bendixen et al., 2010). Los ácidos grasos (AG) juegan un papel importante actuando como moléculas de señalización celular en diversas rutas metabólicas asociadas a este tipo de patologías (Wakil and Abu-Elheiga 2009), y además en la determinación de la calidad organoléptica y nutricional de la carne porcina (Wood et al., 2008). Como consecuencia, descifrar la base molecular del metabolismo de los AG podría contribuir al conocimiento de la base genética de las enfermedades metabólicas, además de favorecer el desarrollo tecnológico para mejorar la calidad de la carne porcina. El metabolismo de los AG es un proceso complejo determinado por factores ambientales e interacciones entre muchos *loci*. Por lo tanto, estudiar la base genética del metabolismo de los AG, en el contexto de las interacciones génicas y propiedades topológicas de las redes biológicas que los determinan, parece una alternativa interesante a los clásicos métodos reduccionistas empleados hasta la actualidad, como por ejemplo el análisis de ligamiento y/o de asociación (GWAS).

El objetivo fundamental de este trabajo fue estudiar la co-asociación entre genes y/o SNPs a lo largo de varios fenotipos e identificar redes génicas, rutas metabólicas y los factores de transcripción (TF) más relevantes de los fenotipos considerados mediante un enfoque holístico denominado *Association Weight Matrix* (AWM) (Fortes et al., 2010) basado en la biología de sistemas.

## MATERIAL Y MÉTODOS

**Material animal, fenotipos y genotipos.** Se analizaron un total de 144 animales procedentes de un retrocruce experimental (BC1\_LD) entre Ibérico (25%) y Landrace (75%). A partir de un estudio previo de componentes principales (Ramayo-Caldas et al., 2012) se seleccionaron los 15 fenotipos más informativos de esta población relacionados con el contenido e índices de AG en el músculo *Longissimus dorsi*. Todos los animales (incluidos los fundadores del cruce) fueron genotipados con el *Porcine SNP60K BeadChip* (Illumina). Después del control de calidad se obtuvieron un total de 48.119 SNPs informativos con una tasa de genotipado superior al 99%. Las coordenadas genómicas de los SNPs corresponden a la última versión de la secuencia del genoma porcino (*Sscrofa10.2*).

### Análisis bioinformático

La construcción del AWM comienza con la selección de SNPs a partir de un GWAS. Los elementos de la matriz AWM  $\{i, j\}$  corresponden al efecto aditivo normalizado del SNP en la fila 'i' en el fenotipo de la columna 'j' (Fortes et al., 2010). El GWAS se realizó con el programa *Qxpak5* (Perez-Enciso and Misztal 2011) bajo un modelo mixto considerando efectos genéticos aditivos. Para la construcción del AWM se desarrolló un programa en el entorno de programación R siguiendo el procedimiento descrito por Fortes, et. al (2010), pero con algunas modificaciones. De forma concreta, en lugar de seleccionar arbitrariamente el umbral  $P < 0,05$  para la selección inicial de los SNPs, este fue elegido en base a la sensibilidad de los datos al comparar diferentes umbrales de distribución ( $P=1$  hasta  $P= 10^{-5}$ ) de los TF de mayor relevancia biológica en el metabolismo de los AG *versus* el resto de TF. Por tanto, se seleccionó el umbral  $P < 0,035$  que maximizó la diferencia entre ambos grupos de TF. Posteriormente, todos los SNPs asociados con el ácido palmitoleico (seleccionado como fenotipo principal) y/o con tres o más caracteres fenotípicos ( $P < 0,035$ ),

y que además mapean a menos de 2,5 Kb o más de 850 Kb de un gen, fueron finalmente incluidos en la matriz AWM (total 1.096). El programa *PermutMatrix* (Caraux and Pinloche 2005) fue empleado para visualizar los *clusters* fenotípicos (AWM columnas) y génicos (AWM filas) obtenidos. Con el objetivo de identificar y describir interacciones significativas (gen-gen y/o SNP-gen) se utilizó el algoritmo *PCIT* (Reverter and Chan 2008). La red génica resultante fue visualizada con el programa *Cytoscape* (Shannon et al., 2003). La identificación de Ontologías Génicas (GO) enriquecidas se realizó con la herramienta *BiNGO* (Maere et al., 2005). Mientras que las rutas metabólicas sobre representadas fueron identificadas utilizando el programa *ClueGO* (Bindea et al., 2009). Finalmente, con el objetivo de identificar los 3 TF más relevantes, la red fue explorada hasta identificar dentro de todas las posibles combinaciones el trío de TF co-asociados con el mayor número de genes y con un menor solapamiento entre ellos.

## RESULTADOS Y DISCUSIÓN

La Figura 1 muestra en forma de dendrograma jerárquico la distribución de los *clusters* fenotípicos obtenidos a partir de las estimas normalizadas del efecto aditivo. En el gráfico, la proximidad entre los ejes factoriales representa correlaciones fuertes y positivas, mientras que las distancias grandes representan correlaciones negativas. De acuerdo a sus relaciones fisiológicas se observa proximidad entre el ácido palmítico (C16:0) y la suma de AG saturados (SFA), el ácido oleico (C18:1 n-9) y la suma de AG mono-insaturados (MUFA), y el ácido linoleico (C18:2 n-6) y la suma de AG poli-insaturados. En sentido general se observa una estructura similar a la obtenida mediante las estimas de las correlaciones fenotípicas y genéticas, confirmando, como había sido descrito previamente (Fortes et al., 2010), que el AWM puede ser utilizado en la estimación de las correlaciones genéticas.

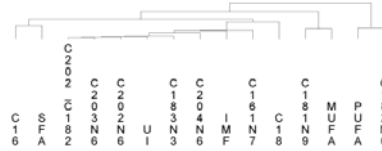
La red génica obtenida está compuesta por 1.096 nodos y 111.198 ejes. En la misma, cada nodo representa un gen/SNP, mientras que los ejes que los unen representan interacciones significativas entre ellos. El análisis de GO muestra un enriquecimiento de procesos biológicos relacionados entre otros con: *Estructura Anatómica y Morfogénesis*, *Metabolismo Lipídico*, *Actividad de Quinasas y Enzimas Reguladoras* (FDR < 0,05). Además, se observa una sobrerrepresentación de rutas metabólicas como la del *Factor Inducible por la Hipoxia-1* y *Diabetes Mellitus Tipo 2*. La figura 2A muestra la topología de la red génica una vez identificado el trío de TF co-asociados con el mayor número de genes, los cuales fueron: *Nuclear receptor coactivator 2 (NCOA2, alias TIF2)*, *E1A binding protein p300 (EP300, alias p300)* and *four and a half LIM domains 2 (FHL2, alias SLIM-3)*. Cabe destacar que los 3 TF tienen en su región promotora sitios de unión para algunos de los TF más relevantes en la regulación del metabolismo de los lípidos y los carbohidratos como son: *SREBP-1*, *PPARG*, *PPAR- $\alpha$* , *HNF1A*, *HNF4 - $\alpha$* , *ER- $\alpha$*  y *GR- $\alpha$* . Además, según la base de datos de interacciones entre proteínas *STRING* (<http://string-db.org/>), existen datos experimentales que confirman la interacción proteína-proteína entre los 3 TF identificados en este estudio y algunos de los TF previamente mencionados (Figura 2B). En la red se observa un total de 730 genes predichos como genes diana de los 3 TF (Figura 2A). En dos meta-análisis realizados en humanos (Teslovich et al., 2010; Asselbergs et al., 2012) se ha descrito la relevancia biológica de 39 de estos genes en el control genético de la composición de lípidos plasmáticos. Esta observación constituye una evidencia adicional de la utilidad del cerdo como biomodelo para el estudio de enfermedades metabólicas humanas. Sin embargo, muchos de estos genes nunca hubieran sido considerados bajo un enfoque reduccionista debido a la falta de significación a nivel genómico (por ejemplo, FDR > 0,05). Los resultados obtenidos muestran la utilidad de los enfoques holísticos en el estudio de caracteres complejos y la capacidad del AWM para la identificación de módulos génicos que de forma independiente no serían considerados, a pesar de tener un papel biológico importante en la determinación de los caracteres estudiados.

## REFERENCIAS BIBLIOGRÁFICAS

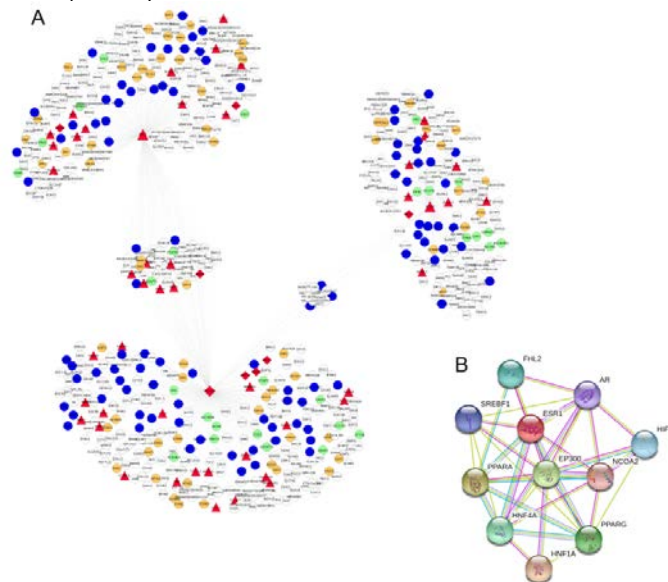
- Asselbergs et al.2012.The American Journal of Human Genetics **91**(5): 823-838. • Bendixen et al.2010.Brief Funct Genomics **9**(3): 208-19. • Bindea et al.2009. Bioinformatics **25**(8): 1091-1093. • Caraux et al. 2005. Bioinformatics **21**(7): 1280-1. • Fortes et al.2010. Proc Natl Acad Sci U S A **107**(31): 13642-7. • Maere et al.2005 Bioinformatics **21**(16): 3448-

9. ● Pérez-Enciso et al.2011.BMC Bioinformatics **12**: 202. ● Ramayo-Caldas et al.2012. BMC Genomics **13**: 547. ●Reverter and Chan 2008. Bioinformatics **24**(21): 2491-7. ● Shannon et al. 2003. Genome Res **13**(11): 2498-504. ● Teslovich et al.2010. Nature **466**(7307): 707-713. ● Wakil et al.2009. Journal of Lipid Research **50**(Supplement): S138-S143. ● Wood et al.2008. Meat Sci **78**(4): 343-58

**Figura 1.** Dendrograma jerárquico que muestra la distribución de los *clusters* fenotípicos obtenidos a partir de las estimas normalizadas del efecto aditivo.



**Figura 2. (A)** Topología de la red génica una vez identificado el trío de TF. El color de los nodos representa la función biológica de los genes: metabolismo lipídico (azul), metabolismo de carbohidratos (verde), morfogénesis (naranja), TF (rojo). **(B)** Evidencias extraídas de *STRING* de interacciones proteicas entre el trío de TF y otros TF relevantes en la regulación del metabolismo de lípidos y/o carbohidratos. El color de los ejes corresponde al tipo de evidencia que soporta la interacción entre los genes: experimentos funcionales (morado), evidencias bibliográficas (verde).



**Agradecimientos:** Este trabajo ha sido financiado por el proyecto AGL2011-29821-C02 (Ministerio de Economía y Competitividad) y por el Programa de Innovación Consolider-Ingenio 2010 (CSD2007-00036). Yulixaxis Ramayo Caldas es becario del programa de Formación de Personal Universitario (AP2008-01450).

## CO-ASSOCIATION AND GENE NETWORK ANALYSIS FOR PIG INTRAMUSCULAR FATTY ACID COMPOSITION

Pig (*Sus scrofa*) is both a source of food and an animal model for the study of metabolic diseases. Fatty acids play a role in pathways of metabolic diseases and its composition influences pork meat quality. In this study, GWAS results were used in an *Association Weight Matrix* (AWM) approach to predict gene networks for intramuscular fatty acid composition in pigs. Three key transcription factors were identified, which are network hubs: *NCOA2*, *EP300*, *FHL2*. An overrepresentation for GO and pathways related to lipid metabolism was observed. The relevance of 39 of the AWM-predicted target genes controlling plasma lipids in human is already well documented. Our results suggest a cooperative role for the three key TF in the transcriptional regulation of intramuscular fatty acid composition and the control of energy homeostasis in pigs. In addition, we provide further evidences that support the usefulness of pig as an animal model for the study of metabolic diseases in humans.

**Keywords:** Gene Networks, pig, fatty acids, transcription factors