

CONSTRUCCIÓN DE REDES GÉNICAS PARA EL METABOLISMO LIPÍDICO EN PORCINO

N. Aznárez¹, A. Cánovas¹, J. Hernández¹, M. Amills², R.N. Pena¹, R. Quintanilla¹

¹Institut de Recerca i Tecnologia Agroalimentàries (IRTA), Genètica i Millora Animal, 25198 Lleida, Spain.

²Universitat Autònoma de Barcelona, Dept. Ciència Animal i dels Aliments, 08193 Bellaterra, Spain.
nitdia.aznarez@irta.cat

INTRODUCCIÓN

El metabolismo lipídico porcino es un sistema complejo de gran importancia en producción animal puesto que agrupa numerosos caracteres relacionados con la salud animal, el rendimiento de la canal, y la calidad tecnológica, sensorial y nutricional de la carne (Wood et al., 2008). El estudio de las redes génicas a partir de los datos masivos de expresión permite profundizar en el conocimiento de los mecanismos moleculares reguladores de estos fenotipos complejos. Hasta la fecha, son escasos los estudios de redes génicas en especies ganaderas (*e.g.* Hudson et al., 2009; García-Gámez et al., 2011) y en particular en porcino (*e.g.* Gao et al., 2010; Li et al., 2010). En todos los casos, estos trabajos abordan el estudio de redes para un número limitado de genes que muestran asociaciones funcionales con algún fenotipo. Adicionalmente, tal y como destacan Rosa et al. (2011), las redes fenotípicas establecidas a partir de datos de expresión pueden resultar de gran ayuda para entender el comportamiento de este sistema complejo ante intervenciones externas como pudiera ser un proceso de selección.

En el presente trabajo se aborda la construcción y el estudio de las redes genéticas y fenotípicas implicadas en la variación conjunta de diez caracteres relacionados con el metabolismo lipídico y la deposición de grasa. Para ello, se han utilizado datos masivos de expresión génica en hígado obtenidos mediante la tecnología de *microarrays*.

MATERIAL Y MÉTODOS

Material animal y fenotipos. El material animal utilizado en el presente estudio procede de una población de 350 machos castrados de una línea comercial Duroc con alto contenido en grasa intramuscular, distribuidos en 4 grupos contemporáneos. Para cada individuo, se registraron medidas de diez caracteres asociados con la deposición de grasa (grasa dorsal y porcentaje de magro), las concentraciones de lípidos séricos (colesterol total, LDL, HDL y triglicéridos), el contenido de grasa intramuscular (IMF) y el perfil de ácidos grasos en músculo (SFA, MUFA y PUFA). Las muestras de hígado (200 g) se conservaron a -80°C inmediatamente después del sacrificio.

Datos de expresión y normalización. Los datos de expresión de hígado de 100 individuos se obtuvieron utilizando la herramienta *GeneChip Porcine Genome Arrays* (Affymetrix). La normalización de los datos de expresión se realizó con el algoritmo gcRMA (Wu et al., 2004) y, posteriormente, se aplicó el filtro “minimum fold-change” implementado en el BRB-ArrayTools (Simon et al., 2007) para eliminar las sondas con expresión poco variable. Finalmente se seleccionaron 6919 sondas en las que más de un 20% de los niveles de expresión alcanzaban ± 1.5 veces la expresión media de la sonda.

Construcción de la red fenotípica. La asociación de los datos de expresión y los fenotipos estudiados fue estimada, para cada combinación de sonda-carácter, con una regresión lineal siguiendo el modelo:

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{X}\mathbf{f} + \mathbf{b}_i\mathbf{t}_i + \mathbf{e}$$

donde \mathbf{y} es el vector que contiene los fenotipos de cada carácter, μ es la media, \mathbf{f} es el vector de efectos fijos correspondiente al grupo contemporáneo (4 niveles), \mathbf{t}_i es el vector que contiene los datos de expresión (niveles de mRNA) de cada individuo para la i -ésima sonda

(6919 sondas), b_i es el coeficiente de regresión de los fenotipos sobre los datos de expresión de la sonda i -ésima, y e es el vector de efectos residuales

Posteriormente se construyó una matriz de correlaciones “transcriptómicas” entre los diez caracteres en base a los coeficientes de regresión b_i , y se usó el algoritmo PCIT (Partial Correlation and Information Theory) desarrollado por Reverter y Chan (2008) para filtrar las correlaciones indirectas entre caracteres (*i.e.* que se explican por una fuerte correlación con un tercer carácter). Finalmente, la red fenotípica se construyó en base a las correlaciones retenidas en la matriz usando el *software* Cytoscape (Shannon et al., 2003).

Construcción de la red génica. Se construyó una red ponderada de co-expresión génica (Weighted Gene Coexpression Network o WGCNA) siguiendo el procedimiento descrito en Zhang y Horvath (2005), utilizando para ello el paquete WGCNA (Langfelder y Horvath, 2008) desarrollado en el entorno de programación R (R Development Core Team, 2008). Brevemente, se define una “matriz de adyacencia” $A=[a_{ij}]$, cuyos elementos se obtienen mediante una función exponencial de la similitud entre los patrones de expresión de las sondas, $a_{ij} = |r(m_i, m_j)|^\beta$, siendo m_i y m_j los niveles de mRNA de los genes i y j respectivamente. El valor de β ha sido escogido considerando que las redes biológicas presentan aproximadamente una topología libre de escala (Jeong et al., 2000; Khanin y Wit, 2006). En este tipo de aproximación se penalizan las correlaciones débiles mientras que se enfatizan las fuertes para disminuir el ruido y el efecto de un posible tamaño de muestra reducido, preservando la naturaleza continua de los datos de co-expresión.

A continuación se identificaron los módulos de genes altamente relacionados uniendo información de superposición topológica (Ravasz et al., 2002) y de agrupación jerárquica siguiendo el protocolo de Zhang y Horvath (2005). Este procedimiento crea un dendrograma en el que aparecen agrupados en distintas ramas aquellos genes con un mismo patrón de expresión. Seguidamente, se obtuvo el primer Componente Principal de cada uno de los módulos (module eigengene), determinando la asociación entre los módulos y los caracteres estudiados a partir de la correlación entre el componente principal y los fenotipos.

RESULTADOS Y DISCUSIÓN

Red fenotípica. La figura 1 muestra la red fenotípica obtenida a partir de los datos de expresión global de hígado. Esta red ilustra las complejas relaciones existentes entre los caracteres estudiados: la deposición de grasa de cobertura y grasa intramuscular (IMF), junto con la proporción de MUFA y SFA y las concentraciones de lípidos séricos, muestran en su conjunto relaciones positivas, mientras que el contenido en PUFA y el porcentaje de magro de la canal están negativamente asociados a todos ellos. No obstante, la ausencia de asociación entre algunos de estos caracteres con elevadas correlaciones fenotípicas (Quintanilla et al., 2011), como por ejemplo los contenidos de SFA y MUFA o PUFA, quedan explicadas por el filtro establecido en función del algoritmo PCIT, que descarta aquellas correlaciones que pueden explicarse por una fuerte correlación con un tercer carácter, que en este caso sería la IMF.

Estos resultados muestran que, de modo general, la red fenotípica establecida a partir de la covariación con los perfiles de expresión en un tejido clave para el metabolismo lipídico como es el hígado presenta una topología similar a la que se establece a partir de las correlaciones fenotípicas y genéticas entre estos caracteres observadas en otros estudios de la bibliografía (Suzuki et al., 2005; Solanes et al., 2009; Quintanilla et al., 2011). En este sentido podemos afirmar que la red fenotípica construida a partir de los datos del transcriptoma proporciona una buena descripción de las relaciones fisiológicas entre estos caracteres, si bien cabe destacar las elevadas correlaciones obtenidas, que presentan valores absolutos entre 0,47 a

0,97, muy superiores a las correlaciones genéticas e incluso fenotípicas entre estos caracteres descritas hasta el momento.

Red génica. El estudio de co-expresión mediante el uso de información de superposición topológica y de agrupación jerárquica sobre la red ponderada de co-expresión permitió identificar 17 módulos (clústers) conteniendo genes con un patrón de expresión similar en el hígado, representados con colores diferentes en la Figura 2A. Cada uno de estos módulos contiene un número variable de sondas, que fluctuó entre 1384 y 37 sondas, con una media de 390 sondas por módulo. La agrupación jerárquica permitió asignar 6632 de las sondas a los diferentes módulos, quedando 287 sondas sin clasificar.

Posteriormente se analizó el grado de asociación de cada uno de estos módulos con los fenotipos. Los resultados obtenidos (Figura 2B) permitieron identificar un módulo (color canela, *tan*) que destacaba sobre el resto por presentar correlaciones moderadas pero altamente significativas con todos los caracteres excepto la concentración sérica de triglicéridos. De forma acorde a la relación entre caracteres en la red fenotípica, la variación conjunta de la expresión de las sondas del clúster resultó positivamente asociada a los caracteres de deposición de grasa y lípidos en sangre (IMF, HDL, colesterol total, SFA, MUFA y grasa dorsal) y negativamente relacionada con el contenido magro de la canal y los PUFA.

El módulo color canela (*tan*) está formado por 85 sondas correspondientes a 59 genes anotados. Un análisis algo más pormenorizado de los genes incluidos en este módulo revela la presencia de algunos genes ontológicamente relacionados con el metabolismo lipídico, como son *LIPN1* y *LIPN2*, fosfatidato fosfatasa que catalizan la reacción de conversión de ácido fosfatídico a diacilglicerol. También se encuentran genes asociados a la homeostasis y el transporte de los ácidos grasos como *GOT1* y *GOT2*. Cabe asimismo destacar la presencia de algunos factores de transcripción como *NFIL* y *ETS2*, así como del factor de crecimiento *IGF1*, que pueden estar jugando un papel importante en la regulación de alguno de estos caracteres. Por último, y a pesar de tratarse de tejidos con funcionalidades distintas, cabe mencionar que alguno de los genes comprendidos en este clúster canela (*e.g.* *EGFR*, *FST*, *GNAS* y *MED1*) ya se habían identificado previamente como genes diferencialmente expresados en músculo esquelético cuando se compararon dos grupos de animales con perfiles divergentes de deposición de grasa (Cánovas et al., 2010). Mencionar también la presencia de seis genes de la superfamilia de los transportadores de solutos (*SLC*), uno de los cuales también se había hallado diferencialmente expresado en músculo.

A modo de conclusión, los resultados obtenidos en el presente trabajo avalan el interés de la aproximación propuesta para la identificación de grupos de genes con un patrón de co-expresión que a su vez estén asociados a los caracteres estudiados. Por otro lado, este procedimiento permite evitar el descarte de información que supone la corrección por testeo múltiple, a la vez que la agrupación en módulos altamente co-expresados permite inferir nuevas funciones de los genes de un organismo para el que todavía se dispone de una anotación incompleta, como es el caso del porcino. Este trabajo nos ha permitido ir más allá en el conocimiento de la asociación entre la expresión génica en un tejido altamente lipogénico, como es el hígado, y diversos fenotipos relacionados con el metabolismo lipídico y la deposición de grasa. Actualmente estamos llevando a cabo estudios de eQTL para identificar zonas del genoma implicadas en la variación de la expresión de los genes identificados y en modular la topología de la red de co-expresión.

Agradecimientos: Este trabajo ha sido financiado por los proyectos AGL2007-66707-C02-01 y AGL2010-22208-C02-01 (Ministerio de Ciencia e Innovación). Gracias a Selección Batallé S.A. por proporcionar el material animal, y a J. Reixach, J.L. Noguera, I. Díaz y D. Almuzara por su colaboración en el protocolo experimental. N. Aznárez es beneficiaria de una beca FPI (BES-2011-043348).

Figura 1. Red fenotípica para diez caracteres¹ relacionados con el metabolismo lipídico obtenida a partir de los datos de expresión global del hígado.

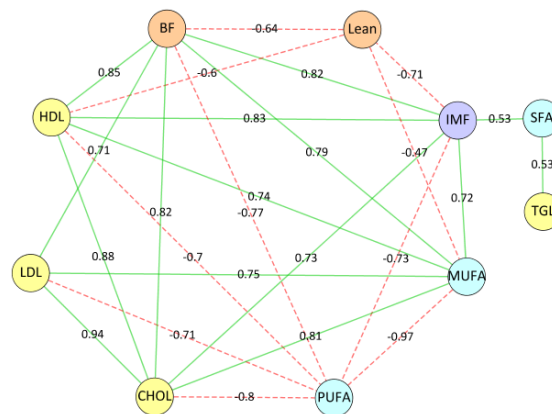
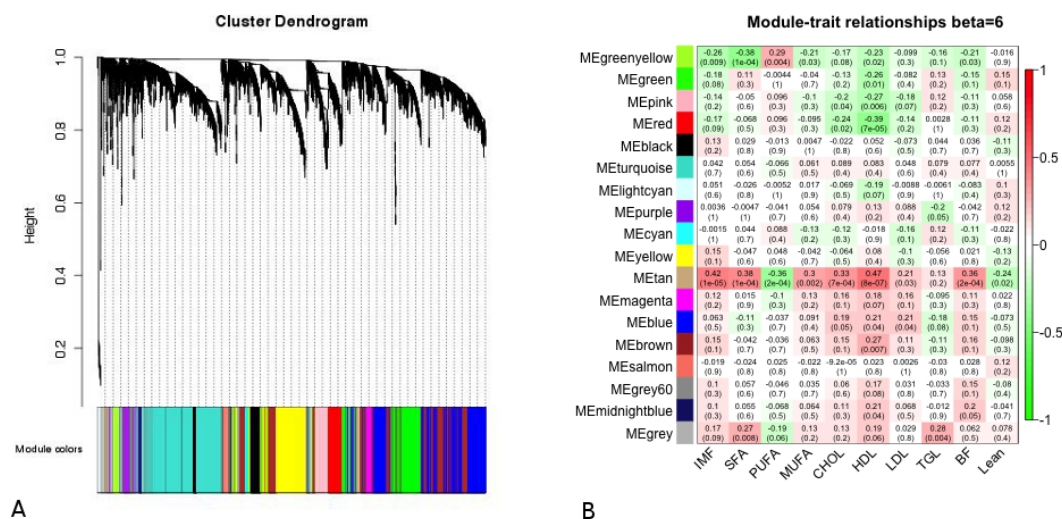


Figura 2. A) Dendrograma del agrupamiento jerárquico de los genes basado en superposición topológica. Los genes se han asignado a módulos altamente interconectados (identificados por los colores de la parte inferior) mediante el método de *dynamic tree cut*. **B)** Asociaciones módulo-carácter. Las filas corresponden a los eigengenes y las columnas a los caracteres¹. Cada celda contiene el valor de la correlación y el p-valor.



¹ Lean-porcentaje de magro; BF-grasa dorsal; IMF-grasa intramuscular; SFA-ácidos grasos saturados; MUFA- ácidos grasos monoinsaturados; PUFA-ácidos grasos poliinsaturados; CHOL-colesterol sérico total; TGL-concentración sérica de triglicéridos; LDL- lipoproteínas de baja densidad; HDL- lipoproteínas de alta densidad.

REFERENCIAS BIBLIOGRÁFICAS

- Cánovas, A., Quintanilla, R., Amills, M., Pena, R.N., 2010. Muscle transcriptomic profiles in pigs with divergent phenotypes for fatness traits. *BMC genomics*. 11, 372.
- Gao, Y., Flori, L., Lecardonnel, J., Esquerré, D., Hu, Z.-L., Teillaud, A., Lemonnier, G., Lefèvre, F., Oswald, I.P., Rogel-Gaillard, C., 2010. Transcriptome analysis of porcine PBMCs after in vitro stimulation by LPS or PMA/ionomycin using an expression array targeting the pig immune response. *BMC genomics*. 11, 292.
- García-Gómez, E., Reverter, A., Whan, V., McWilliam, S.M., Arranz, J.J., Kijas, J., 2011. Using regulatory and epistatic networks to extend the findings of a genome scan: identifying the gene drivers of pigmentation in merino sheep. *PloS one*. 6, e21158.
- Hudson, N.J., Reverter, A., Wang, Y., Greenwood, P.L., Dalrymple, B.P., 2009. Inferring the transcriptional landscape of bovine skeletal muscle by integrating co-expression networks. *PloS*

- one. 4, e7249.
- Jeong, H., Tombor, B., Albert, R., Oltvai, Z.N., Barabási, A.L., 2000. The large-scale organization of metabolic networks. *Nature*. 407, 651-4.
- Khanin, R., Wit, E., 2006. How scale-free are biological networks. *Journal of computational biology : a journal of computational molecular cell biology*. 13, 810-8.
- Langfelder, P., Horvath, S., 2008. WGCNA: an R package for weighted correlation network analysis. *BMC bioinformatics*. 9, 559.
- Li, M., Li, X., Zhu, L., Teng, X., Xiao, H., Shuai, S., Chen, L., Li, Q., Guo, Y., 2010. Differential expression analysis and regulatory network reconstruction for genes associated with muscle growth and adipose deposition in obese and lean pigs . *Progress in Natural Science*. 18, 387-399.
- Ravasz, E., Somera, A.L., Mongru, D.A., Oltvai, Z.N., Barabási, A.L., 2002. Hierarchical organization of modularity in metabolic networks. *Science (New York, N.Y.)*. 297, 1551-5.
- Reverter, A., Chan, E.K.F., 2008. Combining partial correlation and an information theory approach to the reversed engineering of gene co-expression networks. *Bioinformatics* 24, 2491-7.
- Rosa, G.J.M., Valente, B.D., de los Campos, G., Wu, X.-L., Gianola, D., Silva, M.A., 2011. Inferring causal phenotype networks using structural equation models. *Genetics, selection, evolution : GSE*. 43, 6.
- Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., Ideker, T., 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research*. 13, 2498-504.
- Simon, R., Lam, A., Li, M.-C., Ngan, M., Meneses, S., Zhao, Y., 2007. Analysis of gene expression data using BRB-ArrayTools. *Cancer informatics*. 3, 11-7.
- Wood, J., Enser, M., Fisher, A., Nute, G., SHEARD, P., RICHARDSON, R., HUGHES, S., WHITTINGTON, F., 2008. Fat deposition, fatty acid composition and meat quality: A review. *Meat Science*. 78, 343-58.
- Wu, Z., Irizarry, R.A., Gentleman, R., Martinez-Murillo, F., Spencer, F., 2004. A Model-Based Background Adjustment for Oligonucleotide Expression Arrays. *Journal of the American Statistical Association*. American Statistical Association. 99, 909-917.
- Zhang, B., Horvath, S., 2005. A general framework for weighted gene co-expression network analysis. *Statistical applications in genetics and molecular biology*. 4, Article17.

CONSTRUCTING GENE NETWORKS UNDERLYING LIPID METABOLISM IN PIGS

ABSTRACT: Lipid metabolism in pigs represents a complex system of paramount interest for the pig industry, as it gathers a number of traits related to animal health, carcass performance, and nutritional and technological properties of pork. In the present study, global liver expression data was used to infer gene and phenotype networks between ten lipid metabolism traits, including fatness (backfat thickness and lean percentage), serum lipid concentrations (total cholesterol, LDL, HDL and triglycerides), intramuscular fat content (IMF) and muscle fatty acid profile (MUFA, PUFA and SFA). Global liver mRNA expression levels were obtained for 100 individuals of a commercial Duroc line using *GeneChip Porcine Genome arrays (Affymetrix)*. Phenotype network inferred from global liver expression data provided a thoughtful description of physiological relationships between these traits. Besides, a weighted gene coexpression network (WGCNA) between 6919 microarray probes was constructed on the basis of soft thresholding using a power function and scale free topology. The topological overlap measure, together with hierarquical clustering, of this WGCNA allowed creating a dendrogram that clustered similarly expressed genes. Among the highly connected clusters, one of the identified modules was selected for further analyses given its highly significant association with lipid phenotypes. This cluster comprised 59 genes, some of them ontologically related with lipid metabolism and reported in the literature. Results from additional analysis and eQTL studies, currently in progress, will allow going far beyond by detecting genome regions modulating the expression of these genes and the gene network topology.

Key words: gene network, lipid metabolism, fat deposition, microarray, co-expression, pig.