IDENTIFICACIÓN DE VARIANTES GÉNICAS EXPRESADAS MEDIANTE RNA-Seq EN CERDOS CON FENOTIPOS DIVERGENTES PARA CRECIMIENTO Y ENGRASAMIENTO

Martínez A.M.[1], Pérez-Montarelo D., Rodríguez C., Ibañez N.[2], Folch J.M.[3], Silió L., Alves E., Fernández A.I.

[1]Dpto. Mejora Genética Animal, INIA, Madrid,
[2]IRTA, Genètica i Millora Animal, Lleida
[3]Centre de Recerca en Agrigenòmica (CRAG), Consorci CSIC-IRTA-UAB-UB.
Dpto. de Ciència Animal i dels Aliments, Facultat de Veterinària, UAB
E-mail: martinez.angel@inia.es

INTRODUCCIÓN

Estudios previos llevados a cabo en un cruce experimental de cerdo Ibérico x Landrace han mostrado la existencia de regiones del genoma altamente asociadas con caracteres productivos (Varona et al., 2002; Óvilo et al., 2002; Fernández et al., 2012). A pesar de que se han identificado polimorfismos en genes candidato asociados a estos caracteres, no se han encontrado evidencias concluyentes de su causalidad (Óvilo et al., 2002). Las recientes tecnologías de secuenciación masiva permiten análisis más potentes en la ruta del QTL a la mutación causal. Una de estas tecnologías es el RNA-seq, que permite estudiar el transcriptoma (Wang et al., 2009), de manera más eficiente (Marioni et al., 2008). Esta técnica no se restringe a la búsqueda de genes transcritos y estudios del nivel de expresión, sino que al tratarse de un sistema de secuenciación permite analizar variaciones en la secuencia transcrita. Estudios previos han confirmado su utilidad como potente herramienta para la identificación y análisis de polimorfismos presentes en genes expresados (Salem et al., 2012; Cánovas et al., 2010). Presenta asimismo diversas ventajas frente a las plataformas de microarrays tanto para análisis de expresión diferencial como para la determinación de genotipos.

El objetivo de este trabajo ha sido la identificación de variantes génicas de un solo nucleótido (SNVs) informativas en animales divergentes para crecimiento magro, con el propósito final de identificar potenciales mutaciones con efecto sobre estos caracteres. Para ello se ha llevado a cabo un análisis de RNA-Seq sobre muestras de hipotálamo como la glándula reguladora del crecimiento, consumo de energía y acumulación de grasa, y de hígado como el principal órgano de control del metabolismo.

MATERIAL Y MÉTODOS

Animales y datos fenotípicos. La selección de los individuos se realizó mediante un análisis de componentes principales a partir de los datos fenotípicos de 144 individuos procedentes de un retrocruce Ibérico x Landrace (F1 x Landrace), utilizando cuatro caracteres indicadores de crecimiento y composición corporal. Cuatro machos de cada extremo (grupos A y B), fueron seleccionados para el análisis posterior. Los valores medios de los caracteres incluidos en el análisis para los individuos seleccionados en los grupos A y B fueron: ganancia media diaria 0,78 - 0,92 kg/d, espesor de tocino dorsal 11,6 - 16,2 mm y porcentajes de C18:2 en grasa dorsal 16,7 - 12,6 e intramuscular 11,9 – 8,1.

Obtención de ARN, secuenciación. El ARN de los ocho individuos seleccionados fue aislado a partir de muestras de hígado e hipotálamo utilizando el kit *RibopureTM of High Quality Total RNA* (Ambion), cuantificado y sometido a control de calidad usando el Bioanalizador Agilent 2100. La secuenciación se llevó a cabo en el servicio de secuenciación del CNAG, usando un equipo *Hi-Sec 2000 (Illumina),* en pooles de 3 muestras y lecturas de 2x75pb, con más de 60 M de lecturas por pool.

Mapeo y detección de SNVs. Las secuencias obtenidas fueron mapeadas frente al genoma de referencia de cerdo *Sscrofa10.2* usando el programa *CLC Genomics Workbench*. El mapeo se realizó individualmente por cada una de las muestras usando los parámetros por defecto del programa, permitiendo un máximo de dos inconsistencias. Para la detección de SNVs se utilizó la herramienta *Quality-based Variant Detection* del programa con los parámetros por defecto: calidad mínima central 20 y calidad de la región circundante (5pb) 15, una profundidad mínima de lectura de 10x y una frecuencia mínima de variación al 20%. La distribución de los SNVs identificados por cromosoma se llevó a cabo calculando el porcentaje de SNVs presente en cada cromosoma respecto al número total y se compararon con el porcentaje esperado de variaciones en cada cromosoma a partir de la

información incluida en la base de datos *Ensembl Variation 70.* La anotación funcional y enriquecimiento de términos biológicos de los genes portadores de los SNVs identificados se llevó a cabo utilizando la herramienta FatiGo de Babelomics.

RESULTADOS Y DISCUSIÓN

Se obtuvieron un total de 839 M y 877 M de lecturas pareadas de 75 pb para hipotálamo e hígado, respectivamente. Entre un 66 y un 77% de las lecturas fueron mapeados en el genoma de referencia, de las cuales alrededor del 50% mapeaban sobre genes anotados. Los resultados de la identificación de SNPs respecto a la secuencia de referencia se resumen en la Tabla 1. El número total de SNVs detectados frente a la secuencia de referencia fue similar en hígado e hipotálamo, con más de 125.000 anotados en aproximadamente 10.000 genes. Entre ellos se identificaron un 66% de SNPs, tanto en hígado como en hipotálamo, anotados en más de 7.000 genes diferentes que aparecieron segregando (detección de ambos alelos) en los individuos analizados (Tabla 1). De este modo se identificaron SNVs polimórficos y por lo tanto SNPs reales no atribuibles a errores de anotación en la secuencia de referencia. De éstos, más de 6.000 y de 4.900 fueron no sinónimos en hipotálamo e hígado, respectivamente. Como es de esperar el número de SNPs sinónimos fue superior al de no sinónimos (tasa no sinónimo/ sinónimo = 0,45). Adicionalmente, se identificaron los SNPs informativos en los animales analizados, estableciendo como criterio la detección de al menos dos genotipos a frecuencias distintas entre ambos grupos (Tabla 1). Un 7% del total de SNPs identificados correspondieron a esta clase. En el perfil de la distribución de estos últimos SNPs destacan los cromosomas 1, 7, 10, 11 y 14, donde la proporción de variantes identificadas se mostró por encima o debajo de la esperada en hipotálamo, hígado o en ambos tejidos (Figura 1). Por otro lado, la anotación funcional de los genes portadores de estos SNPs reveló el enriquecimiento de los términos Gene Ontology (p-corregida <0,05) relacionados con la unión de factores de crecimiento, la regulación de procesos metabólicos y el metabolismo de aminoácidos y de lípidos. Además se encontró enriquecimiento en la ruta KEGG (p-corregida<0,05) para el metabolismo de lípidos, tanto en hígado como hipotálamo. Resultados concordantes con los genes que participan de los principales procesos biológicos atribuidos a ambos tejidos.

Finalmente se llevó a cabo un examen más detallado centrado en aquellos SNPs identificados en regiones cromosómicas particularmente relevantes por ser portadoras de QTL con efectos importantes sobre caracteres de crecimiento y engrasamiento (Varona et al., 2002; Fernández et al., 2012). Estas regiones corresponden a los cromosomas SSC2: 148-152 Mb, SSC4: 68-85 Mb, SSC4: 127-132 Mb y SSC6: 134-145Mb. Del total de SNPs identificados, 176 SNPs informativos (43 genes) y 136 SNPs informativos (31 genes) detectados en hipotálamo e hígado respectivamente se localizan en estas regiones. Entre éstos destacan interesantes genes candidatos funcionales como *FGF1* (17 SNPs) para el QTL del SSC2, *VCAM1* (1 SNP) y *AGL* (3 SNPs) para los QTL del SSC4 y *JAK1* (3 SNPs) para el QTL del SSC6.

Como en la mayoría de los estudios que emplean tecnologías de secuenciación masiva, en este estudio se ha obtenido gran cantidad de datos que, tras una etapa de validación, permitirán llevar a cabo nuevos estudios de asociación con el fin de identificar los genes y las mutaciones causales de la variación fenotípica estudiada.

REFERENCIAS BIBLIOGRAFICAS

• Varona, L., Ovilo, C., Clop, A. et al., 2002 Genome Res 80,145-154 • Óvilo, C., Clop, A., Noguera, J.L. et al., 2002 J Anim Sci 80,2801-2808 • Fernández, A.I., Pérez-Montarelo, D., Barragán, C. et al., 2012 BMC Genetics 13, 41 • Óvilo, C., Oliver, A., Noguera, J.L. et al., 2002 Genet Sel Evol 34, 465-479 • Wang, Z., Gerstein, M. & Snyder, M. 2009. Nat Rev Genet.10,57-63 • Marioni, J.C., Mason, C.E., Mane, S.M., et al., 2008 Genome Res 18,1509-1517 • Salem, M., Vallejo, R.L., Leeds, T.D. et al., 2012 PlosOne 7,e36264 • Cánovas, A., Rincon, G., Islas-Trejo, A. et al., 2010 Mamm Genome 21,592-598.

Agradecimientos: Este estudio se ha realizado bajo el marco del proyecto AGL2011-29821-C02. Martínez A.M. y Pérez-Montarelo D. disfrutan de becas FPI (BES-2012-056563 y BES-2009-025417).

Tabla 1. *Resumen de los SNVs identificados por tejido: total de SNVs identificados respecto de la secuencia de referencia, SNVs segregando en los animales analizados y SNVs informativos entre los grupos A y B de divergente crecimiento y composición corporal*

	SNVs	Genes	Región codificante	No sinónimo
Hipotálamo				
Total	125.488	11.328	84.941	9.588
Segregando *	83.640	9.647	51.765	6.088
*Informativos***	10.064	3.133	1.632	306
Hígado				
Total	125.163	9.200	93.559	7.707
*Segregando**	85.491	7.794	61.127	4.975
*Informativos***	8.962	2.118	1.515	192

* SNVs polimórficos= SNPs

** SNPs donde se detectan al menos dos genotipos a frecuencias distintas entre ambos grupos

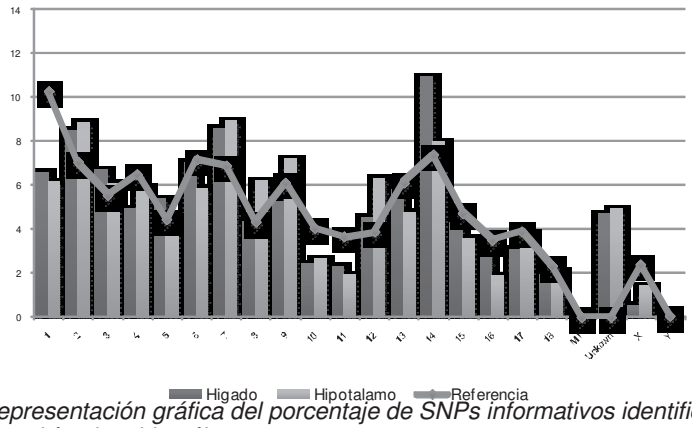


Figura 1. *Representación gráfica del porcentaje de SNPs informativos identificados por cromosoma en hígado e hipotálamo.*

IDENTIFICATION OF EXPRESSED GENE VARIANTS BY RNA-Seq IN PIGS PHENOTIPICALLY DIVERGENT FOR GROWTH AND FATNESS

ABSTRACT: The RNA-Seq technique is not only a great tool for transcriptome characterization and gene expression analyses, but also a powerful tool for analyzing genome variations such as SNVs. In this study we have identified SNPs using RNA-Seq data from liver and hypothalamus samples of two groups of pigs showing divergent phenotypes for growth and fatness coming from an Iberian x Landrace backcross. We have identified more than 83,000 SNPs segregating in these animals. Among them, 10,064 and 8,962 SNPs in hypothalamus and liver, respectively, are potentially informative for further analyses. The distribution of the SNPs along pig chromosomes revealed some interesting data, such as for chromosomes 1, 7, 10, 11 and 14 where the number of informative SNPs differed from the expected. Among the genes carrying the detected informative SNPs highlight *FGF1, VCAM1, AGL* and *JAK1* as powerful biological and positional candidate genes to underlay the QTL for growth and fatness identified in previous studies.

Keywords: RNA-Seq, SNV identification, Growth, Fatness