

IDENTIFICACIÓN DE RNAs LARGOS NO-CODIFICANTES EN RESPUESTA A LA INGESTA DE ALIMENTOS

Mármol-Sánchez, E.¹, Quintanilla, R.², Cardoso, T. F.^{1,3}, Tibau, J.⁴ y Amills, M.¹.

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

²Genètica y Mejora Animal IRTA, Torre Marimon, Caldes de Montbui, 08140.

³Fundación CAPES, Ministerio de Educación de Brasil, Brasilia – DF 70040-020, Brasil. ⁴IRTA-Monells, Finca Camps y Armet s/n 17121 Monells. emilio.marmol@cragenomica.es

INTRODUCCIÓN

Durante los últimos años se ha producido un avance significativo en la identificación de los distintos tipos de RNAs que integran el transcriptoma y en la determinación de sus características funcionales. Los RNAs largos no-codificantes intergénicos (lincRNAs) y transcritos *cis*-antisentido (*cis*-NATs) se caracterizan por tener una longitud superior a los 200 nucleótidos y una estructura exónica similar a los transcritos codificantes (Ullitsky y Bartel, 2013). Los genes que codifican lincRNAs se localizan en regiones intergénicas y algunos de ellos han sido previamente caracterizados en porcino (Esteve-Codina et al., 2011; Zhao et al., 2015; Xia et al., 2016; Kern et al., 2018). Por otra parte, los *cis*-NATs son transcritos cuya secuencia es parcial o totalmente complementaria a la de otra molécula de RNA (p.e. RNA mensajeros). Estudios previos han determinado la presencia de *cis*-NATs en aproximadamente el 30% de los genes anotados en el genoma porcino (Chen et al., 2012). Pese a ello, existe aún un gran desconocimiento de la complejidad y diversidad de funciones del repertorio de RNAs largos no codificantes (lncRNAs) en la mayoría de especies, entre ellas, el porcino.

Con el objetivo de incrementar el catálogo de lncRNAs en la especie porcina y desentrañar el efecto de la nutrición sobre su expresión, hemos realizado un análisis de predicción de nuevos candidatos lincRNAs y *cis*-NATs en músculo *gluteus medius* de porcino. Además, hemos analizado su patrón de expresión en respuesta a la ingesta de alimentos.

MATERIALES Y MÉTODOS

Muestras biológicas y diseño experimental: Para el experimento de análisis de la expresión de nuevos transcritos lncRNA se obtuvieron muestras biológicas del músculo *gluteus medius* de 48 cerdas Duroc alimentadas *ad-libitum* y muestreadas en tres puntos temporales: ALT0 (ayuno, n = 12), ALT1 (5 h post-ingestión, n = 12) y ALT2 (7 h post-ingestión, n = 12), así como otro grupo sometido a restricción calórica en la segunda fase de engorde y muestreado en ayunas (ART0, n = 12), según lo descrito por Cardoso et al. (2017) y Ballester et al. (2018).

Secuenciación y cuantificación de la expresión: De cada muestra de tejido se extrajo la fracción total de RNA mediante el preparado comercial RiboPure RNA Purification kit (Ambion, Austin, TX). Se prepararon librerías mediante el TruSeq Stranded mRNA Library Preparation Kit (Illumina Inc., CA) y fueron secuenciadas mediante un equipo HiSeq 2000 (Illumina Inc., CA) en el Centre Nacional d'Anàlisi Genòmica (CNAG, <https://cnag.cat>). Las secuencias fueron examinadas de acuerdo a su calidad y filtradas mediante el software FASTQC, y se eliminaron los adaptadores con el software Trimmomatic v0.32 (Bolger et al., 2014). Seguidamente fueron alineadas respecto al genoma de referencia porcino (versión 11.1) mediante el software de alineamiento HISAT2 (Kim et al., 2015). Se hizo uso del software Stringtie (Pertea et al., 2015) para la cuantificación de la expresión, permitiendo la reconstrucción de nuevos transcritos carentes de anotación en el genoma porcino de referencia.

Predicción de genes candidatos lncRNA y mapeo de QTLs: Del total de nuevos transcritos identificados, se filtraron aquellos que cumplieran los siguientes requerimientos: (1) longitud de secuencia ≥ 200 nucleótidos para transcritos con 2 o más exones, y (2) ≥ 2000 nucleótidos para transcritos monoexónicos. Por otra parte, se clasificaron como transcritos intergénicos (ITs) aquellos que no solapasen con

ningún gen codificante de proteína (PC) conocido, mientras que se consideraron como transcritos antisentido (ATs) aquellos complementarios a RNAs mensajeros (mRNAs), con un margen de 1 kilobase para ambas clases.

Se analizó el potencial codificante de los transcritos candidatos mediante los softwares CNCI (Sun et al., 2013), CPC2 (Kang et al., 2017) y CPAT (Wang et al., 2013), reteniendo sólo aquellos consistentemente clasificados como no codificantes por los tres programas. Posteriormente, se hizo uso del software LncFinder (Han et al., 2018) para identificar aquellos transcritos clasificados como genes lncRNAs. Finalmente se consideró que los transcritos con valores FPKM > 5 (transcritos monoexónicos) y FPKM > 1 y coverage > 3 (poliexónicos) se hallan realmente expresados en el músculo porcino.

Análisis de expresión diferencial, correlación y variabilidad: Se analizó la expresión diferencial de los transcritos lncRNAs entre los grupos ALT0 vs ALT1, ALT0 vs ALT2 y ALT0 vs ART0, mediante el software DESeq2 (Love et al., 2014). Además, se calculó la correlación lineal entre la expresión de transcritos ATs y mRNAs complementarios, así como su significación estadística, y la correspondiente a ITs y genes PC localizados a 100 kilobases de distancia. Por otra parte, se determinó el coeficiente biológico de variación (BCV) para los transcritos lncRNAs y mRNAs (Robinson et al., 2010) en cada grupo analizado (ALT0, ALT1, ALT2, ART0).

RESULTADOS Y DISCUSIÓN

De un total de 933 ATs y 456 ITs seleccionados a partir de la reconstrucción de nuevos transcritos, 823 ATs (88,21%) y 353 ITs (74,41%) fueron clasificados como no codificantes por los softwares CNCI, CPC2 y CPAT. De estos, un total de 396 ATs (48,11%) y 49 ITs (13,88%) se clasificaron como lncRNAs mediante el software LncFilter. Se detectaron un total de 111 (ALT0), 110 (ALT1), 124 (ALT2) y 80 (ART0) lncRNAs expresados en el músculo *gluteus medius*, 113 de los cuales presentaron expresión en todas las muestras analizadas.

El análisis de expresión diferencial destacó la presencia de 4 transcritos diferencialmente expresados (*fold-change* |FC| > 1,5; *q-value* < 0,05) tal y como se observa en la **Tabla 1**. Por otra parte, al comparar los niveles de expresión y BCV para la *cis*-interacción de genes lncRNAs y mRNAs, se observó una menor expresión y mayor variabilidad en los transcritos no-codificantes, resultados que coinciden con los obtenidos en trabajos previos (Chen et al., 2012; Sigova et al., 2013). Se evidenció la presencia de correlaciones tanto positivas como negativas para cada transcrito lncRNA y sus correspondientes mRNAs complementarios, así como cambios en el signo de la correlación para las mismas interacciones en los diferentes grupos analizados, lo cual podría indicar la existencia de efectos *cis*-reguladores. Por ejemplo, el transcrito MSTRG.7723.1 mostró un aumento significativo de la expresión tras la ingesta de alimento (grupos ALT1 y ALT2 respecto a ALT0), como puede observarse en la **Tabla 1**. Este transcrito fue mapeado en la región 3'-UTR del gen *PPP1R3D*, que ha sido relacionado con la regulación de la concentración de glucosa en estados postprandiales, el metabolismo de lípidos, y la resistencia a la insulina asociada a obesidad (Morton et al., 2011; Zhang et al., 2014). La correlación positiva detectada entre MSTRG.7723.1 y *PPP1R3D* ($r = 0,6626$; $P\text{-value} = 0,0189$) en ALT2 podría ser indicativa de la unión de dicho transcrito a la región 3'-UTR de *PPP1R3D*, enmascarando su interacción con otros elementos reguladores inhibidores, aunque dicha hipótesis debe analizarse experimentalmente. Por otra parte, cabe destacar el transcrito intergénico MSTRG.6412.1, localizado en el cluster *HOXD*, y cuyo ortólogo, en la especie humana, es el lncRNA *HAGLROS*. Este lncRNA ha sido asociado a la inhibición de la función reguladora de miR-100-5p sobre la expresión del gen *MTOR* y la activación del eje PI3K/AKT/NF- κ B en la proliferación celular y estados proinflamatorios (Chen et al., 2018; Liu et al., 2018). A modo de resumen, nuestros resultados indican la existencia de cambios en la expresión de lncRNAs en respuesta a la ingesta de alimentos.

REFERENCIAS BIBLIOGRÁFICAS

- Ullitsky, I. & Bartel, D.P. 2013. *Cell*. 154(1): 26-46. • Esteve-Codina, A. et al. 2011. *Heredity*. 107(3): 256-64. • Zhao, W. et al. 2015. *Sci Rep*. 5: 8957. • Xia, J. et al. 2016. *Sci Rep*. 6: 30709. • Kern, C. et al. 2018. *BMC Genom*. 19: 684. • Chen, C. et al. 2012. *PLoS One*. 7(12): e52433. • Cardoso, T. F. et al. 2017. *BMC Genom*. 18(1): 603. • Ballester, M. et al. 2018. *BMC Genom*. 19(1): 682. • Bolger, A. et al. 2014. *Bioinformatics*. 30(15): 2114-2120. • Kim, D. et al. 2015. *Nat Meth*. 12: 357-360. • Pertea, M. et al. 2015. *Nat Biotechnol*. 33(3): 290-295. • Sun, L. et al. 2013. *Nucl Acid Res*. 41(17): e166. • Kang, Y. J. et al. 2017. *Nucl Acid Res*. 45(W1): W12-W16. • Wang, L. et al. 2013. *Nucl Acid Res*. 41(6): e74. • Han, S. et al. 2018. *Brief Bioinform*. bby065. • Love, M. et al. 2014. *Genome Biol*. 15: 550. • Robinson, M. D. et al. 2010. *Bioinformatics*. 26(1): 139-140. • Sigova, A. A. et al. 2013. *Proc Natl Acad Sci USA*. 110(8): 2876-81. • Morton, N. M. et al. 2011. *PLoS One*. 6(9): e23944. • Zhang, Y. et al. 2014. *Mol Endocrinol*. 28(1): 116-126. • Chen, J. F. et al. 2018. *Mol Cancer*. 17(1):6. • Liu, M. et al. 2018. *Biochem Biophys Res Comm*. 500(3): 589-596.

Agradecimientos: Este trabajo ha sido financiado con los proyectos AGL2013-48742-C2-1-R and AGL2013-48742-C2-2-R concedidos por el Ministerio de Economía y Competitividad (MINECO). Emilio Mármol-Sánchez es beneficiario de un contrato predoctoral FPU15/01733. Los autores agradecen el apoyo del MINECO por la acreditación del Centre de Recerca en Agrigenòmica como Centro de Excelencia Severo Ochoa 2016-2019 (SEV-2015-0533). Nuestro reconocimiento a la empresa Batallé por su apoyo en la generación del material animal.

Tabla 1. *lncRNAs diferencialmente expresados ($|FC| > 1.5$; q -value < 0.05) en los grupos ALT0 vs ALT1, ALT0 vs ALT2 y ALT0 vs ART0. *mean = expresión media medida en conteos normalizados.*

ALT0 vs ALT1					
ID	FC	P-value	q-value	Mean (ALT0)	Mean (ALT1)
MSTRG.3527.1	-1,9093	0,0009	0,0343	199,1875	75,2540
MSTRG.7723.1	1,7078	0,0005	0,0343	1139,2332	2028,9169
MSTRG.14178.1	-1,8519	0,0009	0,0343	283,5238	177,7682
ALT0 vs ALT2					
ID	FC	P-value	q-value	Mean (ALT0)	Mean (ALT2)
MSTRG.7723.1	1,6188	0,0008	0,0497	1139,2332	2004,5894
MSTRG.1736.1	1,9624	0,0007	0,0497	116,7764	231,7182
ALT0 vs ART0					
ID	FC	P-value	q-value	Mean (ALT0)	Mean (ART0)
MSTRG.1736.1	-2,1213	0,00001	0,0011	116,7764	48,1149

PROFILING MUSCULAR LONG NON-CODING RNAs IN RESPONSE TO FOOD INTAKE

The landscape of the long non-coding RNA transcriptome remains to be fully characterized in domestic animals. We aimed to identify and characterize the expression of long intergenic non-coding RNAs (lincRNAs) and natural antisense transcripts (NATs) in *gluteus medius* skeletal muscle samples from 48 Duroc pigs. We implemented a bioinformatics pipeline to accurately detect expressed non-coding lincRNAs and, moreover, we evaluated whether they are differentially expressed in feeding-fasting and food-restricted pigs. Our results indicate the presence of novel lincRNAs in the porcine skeletal muscle with known roles in glucose metabolism and cell proliferation.

Keywords: lincRNA, RNA-sequencing, food intake.