

# CARACTERIZACIÓN DE FRAGMENTOS DE ARN NO CODIFICANTE LARGOS (lncRNAs) EN LAS CÉLULAS SOMÁTICAS DE LA LECHE EN EL GANADO OVINO EL DÍA 50 DE LACTACIÓN

Suárez-Vega<sup>1</sup>, A., Gutiérrez-Gil, B. y Arranz, J.J.

<sup>1</sup> Departamento de Producción Animal, Facultad de Veterinaria, Universidad de León, 24071 León. [asuav@unileon.es](mailto:asuav@unileon.es)

## INTRODUCCIÓN

El desarrollo de la secuenciación masiva paralela del transcriptoma (RNA-Seq) ha permitido la detección de una gran cantidad y diversidad de transcritos procedentes de regiones genómicas consideradas hasta ahora como no-codificantes. Entre ellos, se encuentran los denominados RNAs no codificantes largos (lncRNAs), que constituyen un grupo heterogéneo de transcritos cuyas características en común son una longitud superior a 200 pb y un bajo o nulo potencial de codificación para proteínas (Milligan y Lipovich, 2015).

La identificación y caracterización de los lncRNAs ha adquirido gran importancia en los últimos años debido a dos factores principales: (i) constituyen una alta proporción del transcriptoma (hasta dos tercios del transcriptoma en humanos (Marques y Ponting, 2014)) y (ii) su función como reguladores de la expresión génica (Milligan y Lipovich, 2015; Engreitz et al., 2016) actuando como elementos clave en procesos de desarrollo, diferenciación celular y patológicos (Schmitz et al., 2016). Además, en el caso de las especies domésticas, la anotación de los lncRNAs resulta especialmente interesante, ya que la identificación de estos elementos puede ser de gran ayuda para la valoración de mutaciones en partes del genoma que no codifican para proteínas y que pueden estar asociadas con caracteres productivos de interés.

Así pues, y dado que en análisis previos del transcriptoma de la oveja durante la lactación se había identificado un alto porcentaje de transcritos intergénicos (>60%) (Suárez-Vega et al., 2015), nos hemos planteado, como objetivo de este estudio, la identificación de los lncRNAs presentes en la glándula mamaria del ganado ovino en el día 50 de lactación.

## MATERIAL Y MÉTODOS

Para este estudio se han utilizado las muestras de RNA-Seq de un experimento previo realizado por nuestro grupo de investigación correspondientes al día 50 de la lactación. Estas muestras están depositadas en la base de datos *Gene Expression Omnibus* (GEO) con el número de acceso GSE74825. De forma breve, los datos de RNA-Seq proceden de la secuenciación del RNA extraído a partir de las células somáticas de la leche de ocho ovejas, cuatro de raza Churra y cuatro de raza Assaf. Los protocolos de toma de muestras, extracción y secuenciación del RNA se detallan en el artículo de Suárez-Vega et al. (2015).

Las lecturas fueron mapeadas contra el genoma ovino. (OAR\_v.3.1) utilizando el programa *STAR v.2.3.0e* (Dobin et al., 2013). Una vez realizado el mapeo, utilizamos las herramientas de *Samtools v.1.3.1* (Li et al., 2009) para eliminar los duplicados de PCR (*Samtools rmdup*) en cada una de las muestras y para unirlos en un solo archivo sam (*Samtools merge*). El ensamblado de las lecturas mapeadas se realizó con el paquete bioinformático *Cufflinks v.2.2.1* (Trapnell et al., 2010).

La identificación de los lncRNAs se realizó con el programa *FEELnc* (Wucher et al., 2017). Básicamente, este software consta de tres módulos: “*FEELnc\_filter*”, “*FEELnc\_codpot*” y “*FEELnc\_classifier*”. El primer módulo filtró los transcritos inferiores a 200 pb y/o que concordasen con elementos codificantes presentes en la anotación del genoma ovino OAR: v.3.1. El segundo módulo estimó el potencial codificante (CPS) de cada transcrito y calculó un CPS límite para la identificación de los lncRNAs. El “*FEELnc\_classifier*” clasificó los lncRNAs en función de su localización comparado con el gen anotado más próximo, para este análisis se fijó un rango de ventanas entre 10.000-100.000 pb.

Una vez identificados los lncRNAs y clasificados, se realizó una cuantificación de la expresión de los mismos con *HTSeq-count v.0.6.1p2* (Anders et al., 2014) y la normalización con *DESeq2 v. 1.14.1* (Love et al., 2014).

## RESULTADOS Y DISCUSIÓN

En el ensamblado con Cufflinks de las lecturas alineadas a la versión OAR\_v.3.1 del genoma ovino se detectaron un total de 85.712 transcritos. Tras el primer filtrado con “*FEELnc\_filter*”, se obtuvo un archivo con 6.951 transcritos para los cuales se evaluó su CPS con “*FEELnc\_codpot*”. Finalmente, nuestro análisis del transcriptoma de la glándula mamaria de la oveja en plena lactación detectó un total de 4.508 lncRNAs. La herramienta “*FEELnc\_classifier*” identificó 6.126 posibles interacciones de lncRNAs con genes próximos y 548 lncRNAs que no tendrían interacción con ningún gen. Además, el 60,27% de las interacciones resultaron ser intergénicas frente al 39,73% que fueron intragénicas. Esta observación concuerda con estudios previos de los transcriptomas humano y aviar en los que la mayoría de los lncRNAs detectados son intergénicos (Derrien et al., 2012; Muret et al., 2017). La distribución de los lncRNAs por cromosoma, así como la de los RNAs codificantes (mRNAs) anotados en la versión v3.1 del genoma ovino y con expresión en la glándula mamaria de la oveja se muestra en la Figura 1. Aunque la densidad de lncRNAs es menor que la de transcritos codificantes de proteínas en todos los cromosomas; al igual que ocurre en hígado y tejido adiposo en la especie aviar (Muret et al., 2017), se puede apreciar que existe una alta correlación ( $r = 0,82$ ) entre los mRNAs y los lncRNAs detectados en el mismo tejido.

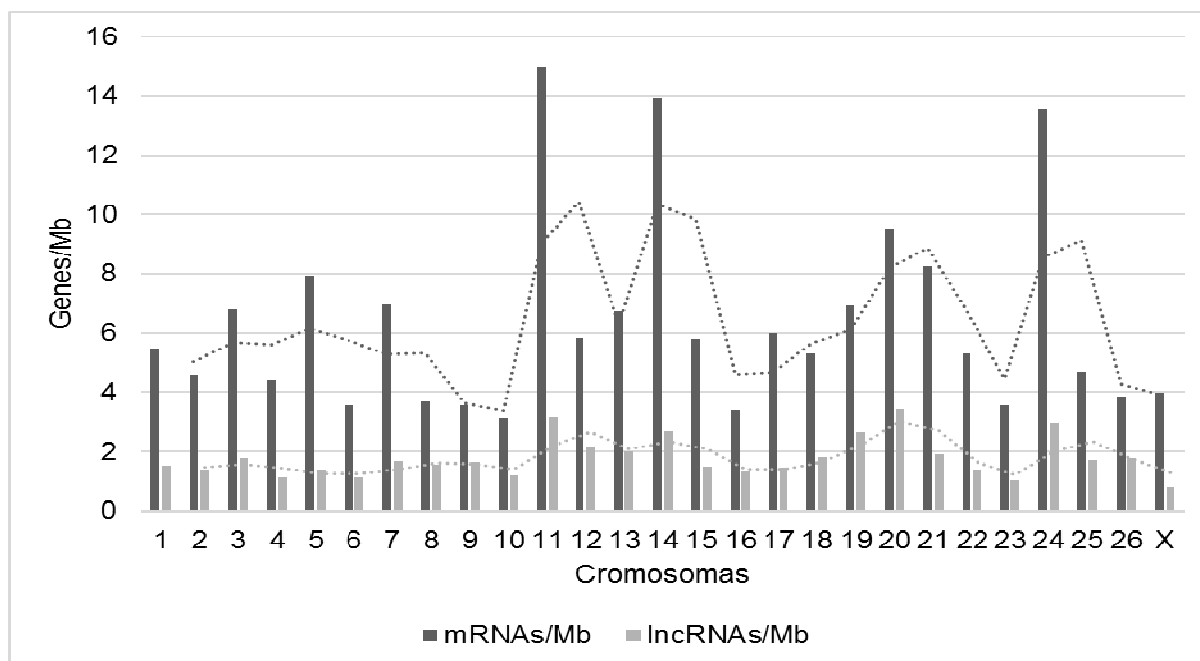
Los cinco lncRNAs más expresados se recogen en la Tabla 1 en la que también se indican la clase a la que pertenecen y los genes con los que podrían tener una posible interacción. El lncRNA más expresado (*CUFF.56836*) está relacionado con la caseína  $\alpha$ -S1, importante constituyente de la leche. El gen *CSN1S1* es a su vez uno de los genes más expresados a lo largo de toda la lactación ( $> 135,000$  FPKMs) en el transcriptoma mamario en análisis (Suárez-Vega et al., 2015). Los lncRNAs *CUFF.37245*, *CUFF.29134* están relacionados con genes del metabolismo lipídico, en concreto con el gen *SCD* que codifica para la esteroil-coA desaturasa y el gen *LPL* que codifica para la lipoprotein lipasa. La expresión de ambos genes en el transcriptoma mamario de la oveja ha sido analizada en un artículo previo que hacía referencia a los genes que podrían influir el rendimiento quesero, teniendo, ambos genes, una expresión elevada ( $> 500$  FPKMs) (Suárez-Vega et al., 2016). El lncRNA *CUFF.64999* está relacionado con el gen *SAT1*, que codifica para la espermidina N1-acetiltransferasa 1. Esta enzima cataliza la acetilación de poliaminas, que son elementos bioactivos de la leche implicados en funciones de diferenciación celular, crecimiento y regulación de la inflamación (Galitsopoulou et al., 2015). El lncRNA *CUFF.23782* se encuentra en relación con el gen *SDS* que codifica para la proteína serina deshidratasa. En relación a esta proteína, Wang et al. (2012) demostraron que la serina deshidratasa es una de las enzimas cuya expresión disminuye significativamente en el ganado vacuno con mastitis.

En resumen, según el presente estudio, la distribución, en términos de densidad de genes, de los lncRNAs y de los RNAs codificantes en la glándula mamaria de la oveja en plena lactación parece ser muy similar. Además, en el análisis de los cinco lncRNAs más expresados se observa la co-expresión de los mismos con genes que codifican para proteínas con funciones relevantes dentro de la glándula mamaria en lactación, lo que corroboraría el papel regulatorio de los lncRNAs en funciones específicas de tejido.

## REFERENCIAS BIBLIOGRÁFICAS

- Anders, S. et al. 2014. *Bioinformatics*. 31 (2): 166-169.
- Derrien, T. et al. 2012. *Genome Res*. 22 (9): 1775-89.
- Dobin, A. et al. 2013. *Bioinformatics*. 29 (1): 15-21.
- Engreitz, J.M. et al. 2016. *Nat Rev Mol Cell Biol*. 17 (12): 756-770.
- Galitsopoulou, A. et al. 2015. *Food Chem*. 173 : 80-5.
- Li, H. et al. 2009. *Bioinformatics*. 25 (16): 2078-9.
- Love, M.I. et al. 2014. *Genome Biol*. 15 (12): 550.
- Marques, A.C. & Ponting, C.P. 2014. *Curr Opin Genet Dev*. 27: 48-53.
- Milligan, M.J. & Lipovich, L. 2015. *Front Genet*. 5: 476.
- Muret, K. et al. 2017. *Genet Sel Evol*. 49: 6.
- Schmitz, S.U. et al. 2016. *Cell Mol Life Sci*. 73 (13): 2491-509.
- Suárez-Vega, A. et al. 2015. *Sci Rep*. 5: 18399.
- Suárez-Vega, A. et al. 2016. *J Dairy Sci*. 99 (8): 6381-90.
- Trapnell, C. et al. 2010. *Nat Biotechnol*. 28 (5): 511-515.
- Wang, C. et al. 2012. *Mol Biol Rep*. 39 (7): 7311-8.
- Wucher, V. et al. 2017. *Nucleic Acids Res*. pii: gkw1306.

**Agradecimientos:** Este trabajo ha sido financiado por el proyecto AGL2015-66035-R del Ministerio de Economía y Competitividad. B. Gutiérrez-Gil es investigadora contratada a través del programa “Ramón y Cajal” del Ministerio de Economía y Competitividad.



**Figura 1.** Densidad por megabase (Mb) y cromosoma de los RNAs codificantes de proteínas (mRNAs) y RNAs no codificantes largos (lncRNAs) expresados en las células somáticas de la leche del ganado ovino el día 50 de lactación.

**Tabla 1.** Características y nivel de expresión de los cinco RNAs no codificantes largos (lncRNAs) con mayor número de lecturas normalizadas en las células somáticas de la leche del ganado ovino el día 50 de lactación.

lncRNA	Gen relacionado	Dirección	Tipo	Subtipo	Media	SD
CUFF.56836	CSN1S1	antisentido	intragénico	contenido	$4,7 \times 10^6$	$2,6 \times 10^6$
CUFF.37245	SCD	antisentido	intragénico	contenido	$6,7 \times 10^4$	$4,2 \times 10^4$
CUFF.64999	SAT1	antisentido	intragénico	contenido	$2,5 \times 10^4$	$2,1 \times 10^4$
CUFF.23782	SDS	antisentido	intragénico	contenido	$2,4 \times 10^4$	$5,1 \times 10^4$
CUFF.29134	LPL	antisentido	intragénico	contenido	$2,0 \times 10^4$	$1,2 \times 10^4$

### CHARACTERIZATION OF LONG NON-CODING RNAs IN OVINE MILK SOMATIC CELLS AT DAY 50 OF LACTATION

**ABSTRACT:** The development of massive parallel sequencing technologies (RNA-Seq) has allowed the discovery of thousands of previously unannotated noncoding functional elements. Among them, the identification and characterization of long noncoding RNAs (lncRNAs) has become a major challenge over the last years due to their putative role as regulators of gene expression. To predict lncRNAs within the sheep lactating transcriptome, we have used RNA-Seq samples from milk somatic cells (MSCs) of eight ewes at mid lactation. In this study, a total of 4.508 lncRNAs and 6.126 interactions with nearby genes were identified. We have observed a high correlation between the densities per megabase and chromosome of the protein-coding RNAs and the lncRNAs detected within the MSCs transcriptome. Moreover, we found that the five highest expressed lncRNAs were related to genes coding for proteins with significant functions in the lactating mammary gland.

**Keywords:** sheep, milk, transcriptome, lncRNAs.