

## USO DEL APRENDIZAJE REPRESENTATIVO PARA LA PRIORIZACIÓN DE GENES ASOCIADOS A EFECTOS PLEIOTRÓPICOS EN GANADO OVINO DE LECHE

Fonseca<sup>1,2\*</sup>, P.A.S., Suárez-Vega<sup>2</sup>, A., Gutiérrez-Gil<sup>2</sup>, B., Marina<sup>2</sup>, H., Pelayo<sup>2</sup>, R. y Arranz<sup>2</sup> J.J.

<sup>1</sup>Instituto de Ganadería de Montaña (CSIC-Univ. de León), Finca Marzanas, 24346 Grulleros, León;

<sup>2</sup>Dpto. Producción Animal, Facultad de Veterinaria, Universidad de León, 24007 León;

\*p.fonseca@csic.es

### INTRODUCCIÓN

El estudio de la pleiotropía ofrece oportunidades para entender mejor los mecanismos genéticos subyacentes a los caracteres complejos y puede contribuir a diseñar programas de selección más precisos. En el ganado ovino de leche se seleccionan caracteres que presentan correlaciones genéticas negativas, como son la cantidad de leche y los porcentajes de grasa y proteína. Comprender estas correlaciones, a menudo impulsadas por la pleiotropía, es importante para la selección simultánea de diferentes caracteres. Identificar y clasificar funcionalmente los genes asociados con los efectos pleiotrópicos es complejo pero esencial para comprender la base molecular de los caracteres productivos. La integración de datos ómicos, apoyada por modelos de aprendizaje automático como el *network embeddin* y el aprendizaje representativo, ofrece herramientas para identificar los genes responsables de los efectos pleiotrópicos. Este trabajo propone un flujo informático para integrar datos funcionales y multi-ómicos para mejorar la priorización de genes con efecto pleiotrópico.

### MATERIAL Y MÉTODOS

En una población de 3.459 animales de raza assaf genotipada con un chip de 50K e imputada para un total de 515.544 SNPs se ha analizado la asociación en el genoma completo para cinco caracteres de producción lechera que incluyen: la producción de leche, los porcentajes de proteína y grasa de la leche, el índice combinado de producción y un carácter de salud, recuento de células somáticas. Los fenotipos utilizados han sido los valores genéticos para una lactación normalizada de 150 días y facilitados por la asociación de criadores (ASSAFE). El efecto pleiotrópico para todos los marcadores incluidos en los análisis GWAS se estimó utilizando la estadística multi-carácter propuesta por Bolormaa et al. (2014). Por otro lado, se estimaron las redes de co-expresión génica (*weighted gene co-expression networks* WGCNs) a partir del transcriptoma de la leche de 28 ovejas assaf. Además, se realizó un enriquecimiento funcional y se construyeron redes utilizando los genes contenidos en cada WGCN y los términos de ontología génica, rutas metabólicas y *quantitative trait loci* (QTL). Los QTL se anotaron utilizando la información disponible en Sheep QTLdb a partir de las coordenadas de los genes significativos (100 Kb upstream and downstream) para el efecto pleiotrópico. Con la información de las redes construidas, se obtuvieron los *embedding vectors* para los genes en cada red (valor  $p < 0,05$ ) mediante aprendizaje representativo. La probabilidad posterior de asociación (PPA) para efectos pleiotrópicos se estimó mediante una adaptación del pipeline REGENT (Wu et al., 2018), integrando los *embedding vector* y los valores  $p$  a nivel de gen.

### RESULTADOS Y DISCUSIÓN

En total, 111 genes fueron significativos para el efecto pleiotrópico y se ubicaron en el 1% superior de la PPA. La evaluación funcional de estos genes indicó una asociación con tres procesos biológicos: 1) transporte entre membranas; 2) metabolismo de lípidos; 3) desarrollo neural y estímulo nervioso. Además, entre los genes priorizados para el control del efecto pleiotrópico entre caracteres de producción y salud en ovejas lecheras se incluyeron genes responsables del desarrollo de la glándula mamaria (*SRF*), el control de la síntesis de proteínas y lípidos durante la lactación (*SLC7A5*, *CSN3*, *DAGLB* y *DEGS1*) y la modulación de la respuesta inmune (*RUNX3* y *TSC2*), entre otros.

### CONCLUSIÓN

Los resultados obtenidos sugieren que la integración de datos multi-ómicos e información funcional mediante un flujo de trabajo basado en *network embedding* y aprendizaje representativo, constituye una estrategia prometedora y potencialmente eficiente para priorizar los genes reguladores clave de efectos pleiotrópicos de interés entre caracteres de producción y salud en ovejas lecheras.

### REFERENCIAS BIBLIOGRÁFICAS

Wu, M. et al. 2018. *Methods* 145: 41-50 □ Liu, H. et al. 2009. *Computational Statistics and Data Analysis* 53: 853-856 □ Bolormaa, S. et al. 2014. *PLoS genetics* 10(3): e1004198.

**Agradecimientos:** PASF ha sido beneficiario de un contrato María Zambrano en la Universidad de León financiada por el Ministerio de Universidades (España) y la Unión Europea-Next Generation EU.