IDENTIFICACIÓN DE VARIANTES RELACIONADAS CON CARACTERES DE RENDIMIENTO QUESERO DE LA LECHE OVINA UTILIZANDO UN CHIP DE SNPs DISEÑADO A PARTIR DE DATOS DE RNA-Seq

Marina[1], H., Gutiérrez-Gil[1], B., Sánchez-Mayor[1], M., Esteban-Blanco[1], C., Suárez-Vega[1], A., A. Garzón[2], A[2] y Arranz[1], J.J.

[1] Dpto. de Producción Animal, Facultad de Veterinaria, Universidad de León, Campus de Vegazana s/n, León 24071, España; [2]Dpto. de Producción Animal, Universidad de Córdoba, Córdoba 14071, España.; hmarg@unileon.es

INTRODUCCIÓN

La leche de oveja, dado su alto contenido en elementos sólidos, se utiliza principalmente para la producción de queso de alta calidad, siendo España el país que más queso puro de oveja produce en la Unión Europea (MAGRAMA, 2015). Por ello, el estudio de los factores que influyen en el rendimiento quesero es un reto importante para la industria lechera ovina. Los programas de mejora de ovino de leche utilizan como criterio de selección el porcentaje de proteína y, en algunos casos, también el porcentaje de grasa, ambos relacionados con el rendimiento quesero. A nivel genético, los caracteres de producción de lechera son el resultado de la acción de un gran número de genes. La disponibilidad de chips de SNPs distribuidos de forma más o menos uniforme a lo largo del genoma ovino, ya sean comerciales o desarrollados "a la carta", ha facilitado en gran medida la identificación de regiones genómicas asociadas a estos caracteres complejos, mediante el uso de análisis de asociación a nivel genómico o GWAS (García-Gámez et al., 2012; Tosser-Klopp et al., 2014). Por otra parte, el gran desarrollo que han sufrido las tecnologías de secuenciación en los últimos años, permite a día de hoy secuenciar genomas (WGSeq) y transcriptomas (RNA-Seq) completos a un precio asequible. Los datos así generados son adecuados para detectar nuevos polimorfismos a lo largo del genoma y del transcriptoma y aplicarlos al desarrollo de chips de SNPs personalizados. En este sentido, un estudio anterior de nuestro grupo describe el análisis del transcriptoma de la glándula mamaria ovina mediante RNA-Seq como método para para la detección de variantes en los genes expresados durante la lactación (Suárez-Vega et al., 2017). Utilizando la información de las variantes detectadas se diseñó un Chip a la carta de SNPs (MilkProteinChip). El presente trabajo describe los resultados del correspondiente GWAS para caracteres de rendimiento quesero en una población comercial de ganado ovino lechero de raza Assaf.

MATERIAL Y MÉTODOS

Animales y fenotipos: El estudio incluyó una población de 982 ovejas de raza Assaf pertenecientes a tres ganaderías de Castilla y León, inscritas en ASSAFE. De cada oveja se recogió, antes del ordeño de la mañana, una muestra de leche de 50 ml que fue procesada posteriormente para determinar dos caracteres relacionados con el rendimiento quesero: el rendimiento quesero en el laboratorio (ILCY) y el extracto seco de la cuajada (ILDCY). Para el análisis de asociación, se utilizaron como variables dependientes los EBVs calculados a partir de los fenotipos de los caracteres anteriores, utilizando como efectos fijos del modelo: los días en lactación, como covariable, y los días de control del rebaño, la edad al parto y el número de corderos vivos, como factores.

Diseño del *MilkProteinChip*: Este chip incluyó, por una parte, un total de 55.789 SNP incluidos previamente en un chip comercial ovino de media densidad; y por otra, un total de 3.194 SNPs seleccionados a partir de las 212.742 variantes identificadas por Suárez-Vega et al. (2017) a partir del análisis de RNA-Seq del transcriptoma de la glándula mamaria ovina en lactación. En esa selección se incluyeron aquellas variantes que, tiendo un MAF > 0,125, se localizaban (i) dentro de QTLs (*Quantitative Traits Loci*) o de genes candidatos para caracteres de composición lechera, y habían sido clasificadas como variantes de moderado o alto impacto funcional; o (ii) en regiones de baja densidad de marcadores del chip comercial.

Análisis de asociación y estudio de las regiones identificadas: Los genotipos brutos obtenidos para los 58.983 SNPs incluidos en el *MilkProteinChip* fueron sometidos a un control de calidad con el programa PLINK (Purcell et al., 2007), utilizando los siguientes parámetros (--mind 0.1 --geno 0.05 --maf 0.01 --hwe 0.0001). Además, el mismo programa se utilizó para realizar un análisis de desequilibrio de ligamiento (LD) con el fin de estimar el número de marcadores con $r^2 < 0,8$ (--r2) en cada cromosoma. El GWAS para los dos caracteres objeto

de estudio se realizó con el programa GeneABEL (Aulchenko et al., 2007), con el que los valores de significación obtenidos (*P-value*), se corrigieron utilizando el factor de inflación λ estimado, con el fin de subsanar la posible estratificación de la población estudiada. Siguiendo a Gao et al. (2010), los resultados del análisis de asociación se corrigieron para el múltiple número de test realizados, considerando el número de SNPs independientes obtenido en el análisis de LD para cada cromosoma (*chromosome-wise*) y para el total del genoma (*genome-wise*). Finalmente, se definió el intervalo de interés, considerando una ventana de ±250 kb centrada en cada SNP significativo. A partir de estos intervalos, se extrajeron los genes candidatos posicionales utilizando la herramienta BioMart de *Ensembl* (http://www.ensembl.org) y se los QTLs previamente descritos en ganado ovino para caracteres de producción lechera a partir de la base de datos *SheepQTLdb* (http://www.animalgenome.org/cgi-bin/QTLdb/OA/index)

RESULTADOS Y DISCUSIÓN

En base al control de calidad realizado se eliminaron 12.608 marcadores (de los cuales se eliminaron: 4.141 por tener la tasa de genotipado menor al 95%; 2.218 por no ajustarse al equilibrio Hardy-Weinberg (P < 0,0001); 6.249 por tener un MAF menor a 0,01). También se eliminaron del análisis los marcadores localizados en el cromosoma X. A partir de los 45.331 SNPs restantes el análisis de LD identificó un total de 32.198 marcadores independientes. El análisis GWAS realizado identificó un total de tres marcadores significativos a nivel "*chromosome-wise"*, ninguno alcanzó el nivel de significación "*genome-wise"* (Figura 1). En la caracterización presentada en la Tabla 1 para los SNPs significativos, se detallan los genes localizados dentro del intervalo de interés definido en cada caso. Respecto a la correspondencia de nuestros resultados con QTLs previamente identificados, nos parece destacable que en la región del SNP significativo para el carácter ILDCY en el cromosoma 17 se han descrito anteriormente dos QTL relacionados con dos caracteres de producción de leche (QTL57755: Cantidad de leche producida y QTL57756:Cantidad de grasa producida) en una población de raza Churra (García-Gámez et al., 2013). Ninguno de los genes candidatos posicionales identificados está, a priori, relacionado directamente con los caracteres objeto de estudio.

Tabla 1*. Caracterización de los SNPs asociados significativamente con los caracteres rendimiento quesero en el laboratorio (ILCY) y extracto seco de la cuajada (ILDCY) según el análisis GWAS presentado en este estudio.*

Carácter	Crom.	Marcador	Intervalo de interés (Mb)	Valor P	Genes candidatos posicionales
ILCY	1	AX-123214783	209.713 - 210.213	4.31E-06	*NAALADL2*
	7	AX-123241825	27.996 - 28.496	3.47E-05	*DPH6* *ENSOARG00000026703*
ILDCY	17	AX-124366306	15.834 - 16.334	4.16E-05	*ENSOARG00000023056* *ENSOARG00000015086* *ENSOARG00000025657* *ZNF330*

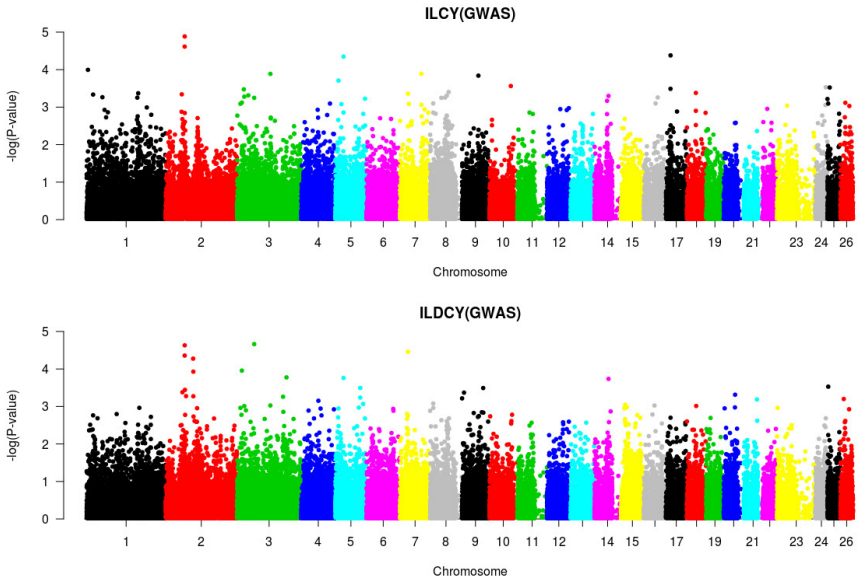


***Figura 1**. Manhattan plot del análisis GWAS realizado para el genoma autosómico ovino en el presente trabajo para los caracteres rendimiento quesero en el laboratorio (ILCY) y extracto seco de la cuajada (ILDCY).*

REFERENCIAS BIBLIOGRÁFICAS

• Aulchenko, Y.S. 2007. Bioinforma. Appl. 23: 1294-1296. • Gao, X. 2010. Genet. Epidemiol. 34: 100–5. • García-Gámez, E. 2012. PLoS One 7: e47782. • Garcia-Gámez, E. 2013. J. Dairy Sci. 96: 6059–69. • Purcell, S. J. Hum. Genet. 81: 559–575. • Suárez-Vega, A. 2017. BMC Genomics 18: 170. • Tosser-Klopp, G. 2014. PLoS One 9: e86227.

Agradecimientos: Esta investigación se ha financiado gracias al proyecto AGL2015-66035-R del MINECO, cofinanciado por el Fondo Europeo de Desarrollo Regional. H. Marina, es beneficiario de una beca del Programa F.P.U. del Ministerio de Educación, Cultura y Deporte (Ref. FPU16/01161). Gutiérrez-Gil, B. disfruta de un contrato del programa Ramón y Cajal (Ref. RYC-2012-10230) del MINECO. Sánchez-Mayor, M. está financiada por un contrato F.P.I. del MINECO (Ref. BES-2013-063614).

IDENTIFICATION OF VARIANTS RELATED TO THE COAGULATION CHARACTERS OF CHEESE YIELD OF SHEEP MILK USING A CHIP OF SNPs DESIGNED FROM RNAseq DATA

ABSTRACT: Sheep milk is mainly used for the production of high quality cheese, so the study of the factors that influence the cheese yield of sheep milk is an important challenge for this industry. In this work, we conduct a genome-wide association study (GWAS) for 982 Spanish Assaf breed ewes using the information of two traits related with cheese yield: individual laboratory curd yield (ILCY) and individual laboratory dried curd yield (ILDCY). The EBV used in the association analysis were obtained using as fixed effects the days in milk (covariate) and the age at parturition, the flock test day, the number of born lamb (factors). In order to carry out GWAS, we use a chip designed from the results of an analysis of the mammary gland transcriptome performed by RNA-Seq. The GWAS has identified several regions significantly associated with the coagulation characteristics of sheep milk have been identified for the traits.

Keywords*:* rna-seq, gwas, dairy sheep, milk coagulability.