

QTL CON INFLUENCIA SOBRE EL RECUENTO DE CÉLULAS SOMÁTICAS EN EL GANADO OVINO: ANÁLISIS DE LA SEGREGACIÓN UTILIZANDO DATOS DE SECUENCIACIÓN GENÓMICA

Gutierrez-Gil, B.¹, Esteban-Blanco, C.^{1,2} y Arranz, J.J.¹

¹ Departamento de Producción Animal, Facultad de Veterinaria, Universidad de León, 24071 León. ² Fundación Centro de Supercomputación de Castilla y León, Campus de Vegazana, 24071, León. Email: bgutg@unileon.es

INTRODUCCIÓN

La mastitis subclínica constituye un importante problema a los que se enfrentan las explotaciones de ganado ovino lechero, ya que, además de influir en la productividad global de la explotación, afecta al bienestar animal. El recuento de células somáticas (SCS) se considera un carácter indicador de la resistencia a la mastitis subclínica. La heredabilidad descrita para este carácter en ganado ovino es relativamente baja con valores que oscilan entre 0,03 y 0,11 en Churra (Othmane et al., 2002) y hasta 0.13 en Lacaune (Rupp et al., 2003). Así, la mejora genética del carácter SCS se vería altamente favorecida por la utilización de información genómica, como la derivada de los estudios de detección de QTL, cuyo objetivo final es la identificación de variantes genéticas con efecto directo sobre el fenotipo de interés. En una población comercial de raza Churra, nuestro grupo realizó un primer barrido genómico basado en 181 marcadores microsatélites mediante el cual se identificó un único QTL significativo al nivel *chromosome-wise* para el carácter SCS en el cromosoma ovino (OAR) 20 (Gutiérrez-Gil et al., 2007). Actualmente, la disponibilidad tanto de una secuencia de referencia del genoma ovino de alta calidad (oar_v3.1), como de herramientas genómicas como los Chips de SNPs de media y alta densidad incrementa la eficiencia de los estudios de localización de QTL realizados en esta especie. Además, el progresivo abaratamiento de las nuevas técnicas de secuenciación posibilitan la obtención de la secuencia completa de un genoma a un precio asequible, permitiendo el diseño de estudios que faciliten la identificación de posibles mutaciones causales o QTN que explican los efectos QTL previamente identificados (Ron y Weller, 2007). En este estudio presentamos un análisis de ligamiento realizado con el chip ovino de 50K en una población independiente de la misma raza que la población analizada por Gutierrez-Gil et al. (2007). Para el QTL identificado con mayor apoyo estadístico y que confirma los resultados anteriores, se presentan un análisis preliminar de la secuenciación masiva del genoma de un trío de muestras formado por un padre heterocigoto y dos de sus hijas seleccionadas para estudiar a máxima resolución la segregación de este QTL e identificar los polimorfismos que podrían ser candidatos a QTN.

MATERIAL Y MÉTODOS

Animales, genotipos y fenotipos: La población objeto de estudio incluyó un total de 1.670 ovejas distribuidas en 16 familias de medio-hermanas del Núcleo de Selección de ANCHE (Asociación de Criadores de Ganado Ovino Selecto de Raza Churra), además de los 18 machos cabeza de pedigrí. Todos los animales fueron genotipados con el chip ovino de 50K. Tras el control de calidad de los datos brutos, llevado a cabo según el protocolo descrito por García-Gámez et al. (2013), un total de 43.613 SNP localizados en los 26 autosomas ovinos quedaron disponibles para análisis posteriores. El fenotipo analizado fue el logaritmo del recuento de células somáticas (SCS). Como variable dependiente del análisis de ligamiento, se utilizaron las *Yield deviations* (YDs) estimadas, para el carácter en estudio, como la desviación de la media poblacional del dato fenotípico bruto corregido para los factores fijos que influyen sobre el carácter y el efecto permanente del animal, teniéndose también en cuenta la precisión de la estimación en función del número de medidas repetidas disponibles para cada animal. El mapa genético utilizado para el análisis se basó en la última versión de la secuencia de referencia del genoma ovina (Ovine Assembly v3.1, disponible en Ensembl), utilizando la equivalencia 1 Mb \approx 1 cM para convertir distancias físicas en distancias genéticas.

Detección de QTL: El análisis de ligamiento de los 26 autosomas ovinos a nivel de toda la población (*across-family*) y para familias individuales se realizó con el programa *QTLMap* (Filangi et al., 2010), siguiendo el procedimiento descrito por García-Gámez et al. (2013). A partir de un test de 1.000 permutaciones se determinaron los umbrales de significación a

nivel *chromosome-wise* (p_c -value) y *genome-wise* (p_g -value). Para los QTL significativos identificados se calculó el intervalo de confianza (IC) mediante el método 1-LOD drop-off (Lander y Botstein, 1989). Para una de las familias segregantes del único QTL detectado al nivel *genome-wise*, se seleccionaron tres muestras para ser sometidas a secuenciación completa del genoma (Trio-QTL). Dichas muestras incluyeron el padre segregante (*Qq*) y dos de sus hijas con fenotipos divergentes extremos para el carácter SCS (*QQ vs qq*) en correspondencia con genotipos alternativos en homocigosis para los marcadores incluidos en el intervalo de confianza del QTL. Para dicha selección, se tuvo en cuenta la fase de los marcadores genotipados en la región del IC del QTL a nivel intrafamiliar obtenida con el programa PHASEBOOK (Dualphase; Druet y Georges, 2010).

Análisis de datos de secuenciación masiva: La secuenciación de las muestras del Trio-QTL se llevó a cabo mediante la tecnología *paired-end* de Illumina utilizando un secuenciador HiSeq2500. Brevemente, los pasos del análisis hasta la identificación de las variantes fueron: (1) Control de calidad de las lecturas utilizando el programa *FastQC* (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). (2) Alineamiento de las muestras frente al genoma de referencia utilizando el *Burrows-Wheeler-Aligner* o *BWA* (Li y Durbin, 2009). (3) Extracción de lecturas no mapeadas con *SAMtools* (Li et al. 2009) e indexado del alineamiento generado anteriormente. (4) Ordenamiento de las lecturas por coordenadas, marcado de duplicados y otras manipulaciones realizadas con el programa *Picard* (<http://broadinstitute.github.io/picard>) (5) Estadísticas básicas con *BAMUtil* (BamUtil – Genome Analysis Wiki) y sobre el estado de las lecturas con *SAMtools*. (6) Identificación de variantes a lo largo de todo el genoma: en este paso se utilizaron dos softwares, *SAMtools* (*mpileup*) y *GATK* (HaplotypeCaller) (Genome Analysis Toolkit; McKenna et al. 2010), con el fin de comparar los resultados y seleccionar las variantes identificadas por dos aproximaciones diferentes. Finalmente, se realizó el análisis de anotación funcional con el programa *Ensembl Variant Effect Predictor* (VEP) (McLare et al., 2010), únicamente para las variantes localizadas en el IC estimado para el QTL segregante en el trío, y cuyos genotipos en la tres muestras analizadas se corresponden con el patrón de segregación del QTL (*Qq* en el padre; *QQ* y *qq* en las hijas de menor y mayor valor de SCS, respectivamente).

RESULTADOS Y DISCUSIÓN

El análisis de ligamiento realizado para el carácter SCS en la población de Churra analizada identificó dos QTL a nivel de significación del 5% *chromosome-wise* en OAR5 (max LRT en 77,99 cM; IC: 74,9-81,5 cM), y OAR25 (39,48 cM; IC: 38,6-41,7), y un QTL a nivel 5% *genome-wise* localizado en OAR20 (21,52 cM; IC:20,9-23,8) (Figura 1). Sólo una familia resultó significativa para cada uno de los QTLs identificados a nivel *chromosome-wise*, mientras que tres familias fueron identificadas como segregantes para el QTL en OAR20. De una de estas tres familias se seleccionaron los animales del Trio-QTL para realizar el análisis de secuenciación masiva.

Los ficheros brutos de secuencias contenían 395.439.600, 515.388.126 y 560.017.452 lecturas. Los resultados del control de calidad de todas las secuencias utilizadas fueron considerados como óptimos y no fue necesario realizar el proceso de recorte o *trimming*. La *pipeline* desarrollada para el análisis con *GATK* incluyó un proceso de eliminación de duplicados, que constituyeron como media el 7% de las lecturas iniciales. El número de variantes detectadas a lo largo de todo el genoma (ficheros tipo "VCF") por el *GATK* y *SAMtools* fueron de 22.051.464 y 20.973.384, respectivamente. Tras el primer filtrado de variantes, seleccionando aquellas incluidas en la región entre 14,2 y 24,2 Mb del OAR20 se seleccionaron un total de 69.898 variantes identificadas por *GATK*, y de 67.297 identificadas por *SAMtools*. Posteriormente se realizó una selección por genotipo (seleccionando las variantes que cumplían el patrón de segregación del QTL), eliminándose también aquellas variantes localizadas en las regiones intergénicas. De esta manera se seleccionaron un total de 6.713 y 6.721 con *GATK* y *SAMtools*, respectivamente. Finalmente, de todas las variantes identificadas, se realizó una primera selección de aquellas con consecuencias biológicas importantes (codón de parada, mutaciones que afectan a la pauta de lectura y mutaciones de un solo nucleótido). El número total de mutaciones detectadas por ambos protocolos fue de 15 (10 en común, 3 exclusivas a *GATK* y 2 exclusivas a *SAMtools*), localizadas en un total de 11 genes. En el momento actual se están analizando todas las

variantes genéticas con objeto de verificar el papel funcional de cada una de ellas y las consecuencias de los diferentes polimorfismos sobre el carácter recuento de células somáticas v. por lo tanto de resistencia a la mastitis subclínica en el ganado ovino.

SCS (LA)

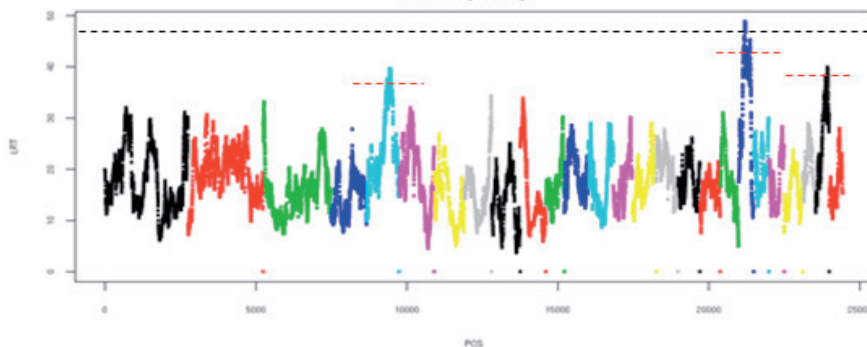


Figura 1. Resultados del análisis de ligamiento (valores LRT) para el carácter SCS realizado en la población comercial de ovejas Churra analizada en este estudio. Las líneas horizontales discontinuas representan los umbrales de significación del 5% chromosome-wise (rojo) y el umbral de significación del 5% genome-wise (negro), obtenidos por permutaciones.

REFERENCIAS BIBLIOGRÁFICAS

- Druet & Georges, 2010. *Genetics* 184, 789-798.
- García-Gámez et al. 2013. *J. Dairy Sci.* 96, 6059-6069
- Gutiérrez-Gil et al. 2007. *J. Dairy Sci.* 90, 422-426.
- Lander & Botstein, 1989. *Genetics.* 121, 185-199.
- Li & Durbin, 2009. *Bioinformatics* 25, 1754-1760.
- Li et al. *Bioinformatics* 25, 2078-2079.
- Mckenna et al. *Genome Res.* 20, 1297-1303.
- McLaren et al. 2010. *Bioinformatics* 26, 2069-2070.
- Othmane et al. 2002. *Genet. Sel. Evol.* 34, 581-96.
- Ron & Weller, 2007. *Anim. Genet.* 38, 429-439.
- Rupp et al. 2003. *J. Dairy Sci.* 86, 1476-81.

Agradecimientos: Este trabajo se incluye dentro del proyecto Nacional AGL2012-34437 financiado por el Ministerio de Economía y Competitividad (MINECO). B. Gutiérrez-Gil es investigadora contratada a través del programa “Ramón y Cajal” del MINECO.

QTLs INFLUENCING SOMATIC CELL COUNT IN SHEEP: SEGREGATION ANALYSIS USING WHOLE GENOME SEQUENCING TRIO ANALYSIS

The present study summarizes the linkage analysis performed in a commercial population of Spanish Churra sheep genotyped with the *Ovine SNP50 BeadChip* to identify QTL influencing SCS. Two QTL were detected at the 5% chromosome-wise level on OAR5 and OAR25, whereas, one other QTL located on OAR20 reached the 5% genome-wise significance. For the most significant QTL, which appears to confirm a QTL previously described in another population of Churra sheep, we selected three samples based on their assumed QTL genotypes (Trio-QTL samples) to be analysed by Whole Genome Sequencing (WGS). The Trio-QTL involved a segregating sire (*Qq*) and two daughters showing alternative marker phases, along the confidence interval of the target QTL, which should show correspondence with extreme divergent phenotypes (the *QQ* daughter has two copies of the chromosome related to mastitis resistance and a very low value for the SCS trait; the *qq* daughter has two copies of the chromosome related to susceptibility, and an extreme high value for the SCS trait). Preliminary results of the analysis of the massive sequencing performed are presented here.

Keywords: SNP-chip, subclinical mastitis, sheep, massive sequencing