

# ANÁLISIS GWAS PARA RESISTENCIA A NEMATODOS GASTROINTESTINALES MEDIANTE LA IMPUTACIÓN DE GENOTIPOS DEL CHIP DE ALTA DENSIDAD EN EL GANADO OVINO

Chitneedi, P.K., Atlija, M., Arranz, J.J. y Gutiérrez-Gil, B.

Departamento de Producción Animal, Facultad de Veterinaria, Universidad de León, 24071 León. Email: pchi@unileon.es

## INTRODUCCIÓN

Las infecciones por nematodos gastrointestinales (GIN) constituyen un grave problema para las explotaciones ovinas basadas en sistemas extensivos o semi-extensivos, tanto a nivel sanitario y de bienestar animal como a nivel de salud pública (posible riesgo de zoonosis y de residuos de fármacos antiparasitarios). Existe, por tanto, la necesidad de identificar métodos alternativos eficaces frente al método clásico de control basado exclusivamente en el uso de fármacos antiparasitarios. A raíz de diversos estudios que pusieron de manifiesto el componente genético de parte de la variación observada en la resistencia de las ovejas a los parásitos internos (revisado por Raadsma et al., 1997), distintos autores han propuesto la selección genética, combinada con otros métodos, como una estrategia eficaz de control de los GIN. En este contexto, el objetivo de este estudio es la identificación de regiones genómicas asociadas a Resistencia a GIN en una población comercial de Ganado ovino de raza Churra mediante un barrido genómico de alta densidad basado en un protocolo de imputación de los genotipos del *Illumina Ovine HD BeadChip* (HD-chip) en una población inicialmente genotipada con el *Illumina Ovine SNP50 BeadChip* (50K-chip), en base a un subconjunto de animales genotipados con el HD-chip utilizados como referencia. Tras la evaluación de la fiabilidad del proceso de imputación, y en base a los genotipos imputados se ha realizado un estudio de asociación a nivel genómico (GWAS, *Genome-wide Association Study*) con un carácter relacionado con los niveles de infección por GIN, los niveles séricos de Inmunoglobulina A (IgA) frente a *Teladorsagia circumcincta* (L-IV).

## MATERIAL Y MÉTODOS

La población objeto de estudio es una población comercial de ganado ovino lechero de raza Churra que presenta la estructura de una población de 16 familias de medio-hermanas, incluyendo los 16 machos cabeza de pedigrí y 1.670 hijas, todas ellas pertenecientes a rebaños del Núcleo de Selección de Anche (ANCHE). Esta población base ha sido previamente genotipada con el chip de 50K SNPs. Tras el control de calidad realizado según García-Gámez et al. (2012) quedaron disponibles para análisis un total de 43.613 SNPs localizados en los cromosomas autosómicos.

La población de referencia utilizada para la imputación incluyó 240 animales de la población base, con 16 machos y 14 hijas de cada uno, que fueron genotipados con el HD-chip. Inicialmente se obtuvieron genotipos para un total de 606.006 marcadores SNPs, que tras el control de calidad se redujeron a 492.767 SNPs autosómicos, entre los cuales se incluyen la mayoría de los marcadores del 50K-chip. El proceso de imputación de los genotipos del HD-chip ausentes en los animales de la población genotipada con el 50K-chip se realizó con el software *Beagle* 3.3.2 (Browning & Browning et al., 2008). Los genotipos resultantes fueron convertidos a formato *Plink* (Purcell et al., 2007) con el programa fcGENE (Roshyara, et al., 2014). La imputación realizada con *Beagle* se basa en información poblacional (desequilibrio de ligamiento y frecuencias alélicas), obviándose en este caso la información proporcionada por el pedigrí de la población. Es por ello que en algunos casos, se pueden estimar genotipos incorrectamente o no realizar la imputación. Por ello, se estimó la fiabilidad de la imputación en base a la comparación de los genotipos imputados en las hijas de la población de referencia con sus genotipos directamente obtenidos con el HD-chip (genotipos imputados vs genotipos genuinos). Con el fin de evitar sesgos en esta evaluación, se realizaron 14 iteraciones de imputación, enmascarando en cada una de ellas el genotipo de 16 hijas, una de cada familia analizada.

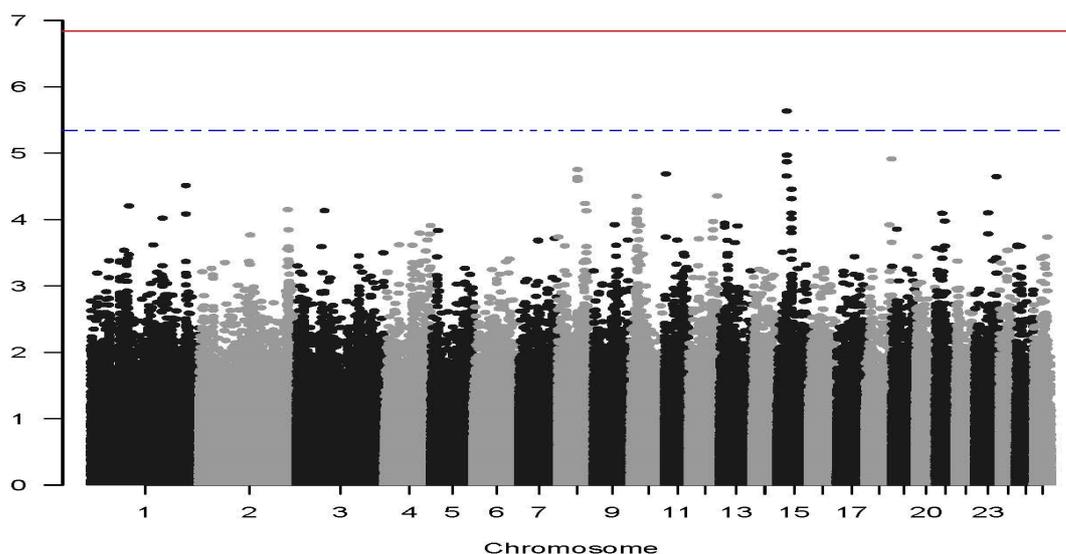
Para un subconjunto de 533 animales de esta población se obtuvieron muestras de sangre para la obtención de medidas fenotípicas de IgA en suero sanguíneo, siendo los genotipos de estos animales los que se utilizaron para el análisis de asociación realizado posteriormente. Tras la normalización de los valores de IgA mediante la transformación box-cox, se estimaron las *Yield Deviations* (datos brutos corregidos para el efecto "rebaño"). El

análisis GWAS se realizó con *ProbABLE* (Aulchenko et al., 2010), un software específicamente desarrollado para realizar análisis GWA a partir de genotipos imputados. Siguiendo el protocolo sugerido por los desarrolladores de este programa, se realizó un primer análisis en *GenABLE* (Aulchenko et al., 2007) con el fin de estimar la matriz de varianzas y covarianzas, lo que proporciona un test de asociación entre el fenotipo a estudiar y los marcadores genéticos en las muestras de individuos relacionados. El segundo paso del análisis de asociación se realizó con *ProbABLE*, utilizando como datos genotípicos para el análisis, los archivos de probabilidad de imputación correcta para el SNP obtenidos en el proceso de imputación de *Beagle*. Los umbrales de significación análisis GWA se determinaron aplicando la corrección de Bonferroni, considerando el número de marcadores analizados para cada cromosoma (umbral 5% *chromosome-wise*) y para todo el genoma (umbral 5% *genome-wise*). Para calcular el número de marcadores “independientes” en cada caso, teniendo en cuenta el desequilibrio de ligamiento existente, se utilizó el programa *SimpleM* (Gao et al., 2008), asumiendo un valor de 0,995 para el parámetro PCA. Todos los análisis, imputación y GWA, se realizaron en un potente servidor Intel (procesador de 2.50GHz, 8 cores y 16GB RAM) en el CPU del Centro de Supercomputación de Castilla y León.

## RESULTADOS Y DISCUSIÓN

La precisión de la imputación realizada por *Beagle* se estimó usando la opción de concordancia de *Plink* (merge-mode 7) en los individuos genotipados para el 50K-chip y el HD-chip. Debido al alto tiempo de computación exigido en el proceso de imputación, la estimación de la fiabilidad se realizó en base a los resultados obtenidos en seis cromosomas de distinta longitud (cromosomas 1, 4, 10, 16, 20, 26). La concordancia promedio estimada a partir de las 14 iteraciones de imputación realizadas fue 90,85%, siendo muy similar entre los seis cromosomas analizados (rango: 90-92 %, SD promedio:  $\pm 1,55\%$ ). Ha de tenerse en cuenta que esta estimación está afectada por la reducción de la potencia en el proceso de imputación debido al enmascaramiento de los animales utilizados como referencia en cada iteración. Así, cuando el proceso se realizó sin eliminar ningún animal de la población de referencia, la precisión de la imputación fue del 94%. Por ello, podemos considerar que la fiabilidad real proporcionada por *Beagle* en la población de Churra considerada varía entre 91-94%, suficientemente alta como para utilizar los genotipos imputados para posteriores análisis de asociación con caracteres de interés productivo. La precisión estimada en este estudio está dentro del rango de precisión de otros estudios realizados en oveja (83-93%, según la raza) estimado por (Hayes et al., 2012) utilizando *Beagle*. En el ganado vacuno, por lo general, los niveles de precisión obtenidos con este programa es más alta. Hozé et al. (2013) estimaron una precisión del 99% tras los análisis realizados en 16 razas de ganado vacuno francés. En ganado vacuno Holstein Chino (Weng et al., 2013) las estimaciones variaron entre el 90-98%, dependiendo de la proporción de individuos de referencia eliminados en las iteraciones.

El análisis de asociación con el fenotipo IgA realizado con *ProbABLE* identificó un único SNP significativo al nivel 5% *chromosome-wise* localizado en el cromosoma 15 ( $p_c$ -value corregido = 5,63), mientras que no se detectó ninguna asociación significativa al nivel 5% *genome-wise*. El marcador que mostró evidencia de influir los niveles de IgA es el *oar3\_OAR15\_24870525*, y se encuentra localizado en la posición 24.870.525 pb del OAR15. El efecto de substitución alélica estimado para este SNP fue de  $0,297 \pm 0,063$  unidades de las YDs analizadas (0,382 desviaciones estándar del fenotipo). De acuerdo a la base de datos *Sheep QTLdb* (<http://www.animalgenome.org/cgi-bin/QTLdb/OA/index>), la posición de este QTL se encuentra dentro del intervalo de confianza estimado (18,3-30,5 Mb) para un QTL con influencia sobre Fecal egg count (FEC), el clásico carácter indicador de los niveles de infección por GINs en una población de retrocruzamiento Red Maasai x Dorper (Silva et al., 2012). La falta de resultados significativos se puede explicar por la escasa potencia del estudio debido al limitado número de animales analizados. Para explotar la información de la estructura de diseño hija de la población en estudio, de forma complementaria a este análisis tipo GWA, se realizarán futuros análisis basados en análisis de ligamiento (LA) y la combinación de desequilibrio de ligamiento con análisis de ligamiento (LDLA).



**Figura 1.** Resultado del análisis GWAS basado en los genotipos del HD-chip (700K) imputados con Beagle en la población comercial de ganado ovino analizada. Para el carácter analizado, IgA, se representan los valores  $\log(1/P)$ . Las líneas horizontales representan el umbral del 5% chromosome-wise promedio para los 26 autosomas (azul), y el umbral del 5% genome-wise (rojo); obtenidos tras la corrección de Bonferroni aplicada (Gao et al., 2011).

#### REFERENCIAS BIBLIOGRÁFICAS

• Atlija et al. (2015). Anim. Genet. submitted • Aulchenko et al. 2007. Bioinformatics 23, 1294-1296. • Aulchenko et al. 2010. Bioinformatics 11,134. • Gao et al., 2011. Genet. Epidemiol. 35, 154-158. • Hayes et al. 2012. Anim. Genet. 43, 72-80. • Hozé et al. 2013. Genet. Sel. Evol. 45, 33. • Purcell et al. 2007. Am. J. Hum. Genet. 81, 559-575. • Radsma et al. 1997. The genetics of sheep. 199-224. (ed. L Piper and A Ruvinsky) • Roshyara et al. 2014. BMC Genet. 15, 88 • Browning & Browning, 2009. Am. J. Hum. Genet. 84, 210-223. • Silva et al. 2012 Anim. Genet. 43, 63-71. CABI Publishing. • Weng et al. 2013. Animal 7, 729-735.

**Agradecimientos:** Este trabajo ha sido financiado por el proyecto Europeo *NematodeSystemHealth* del programa Marie Curie-ITN (Ref. FP7-PEOPLE-2010-ITN #264639) y el proyecto LE245A12-2 financiado por la Junta de Castilla y León. B. Gutierrez-Gil es investigadora contratada a través del programa “Ramón y Cajal” del Ministerio de Economía y Competitividad.

#### GWAS ANALYSIS FOR GASTROINTESTINAL NEMATODES RESISTANCE TRAITS USING IMPUTED HIGH DENSITY CHIP GENOTYPES IN SHEEP

**ABSTRACT:** The aim of this study was to identify genomic regions influencing the serum levels of Immunoglobulin A (IgA), an indicator trait of resistance to gastrointestinal nematode (GIN) infections in Spanish Churra dairy sheep. With this aim, we performed a Genome-wide Association Study (GWAS) based on imputed genotypes for the Ovine High Density (HD)-chip for a population of 1,686 animals belonging to 16 different half-sib families. The whole population, including the 16 sires and their daughters had available genotypes for the *Ovine 50K-chip*. The imputation process was based on the 240 animals of the resource population that had been genotyped for the HD-chip. We estimated the accuracy of imputation, using only the genotypes available for the reference population based on a masking strategy designed to avoid any bias in the accuracy estimation. The GWAS identified a significant SNP at the 5% chromosome-wise level, located on OAR15. As complementary approaches, future analyses will exploit the linkage analysis and combined linkage disequilibrium with linkage analysis to exploit the half-sib structure of the studied resource population.

**Keywords:** sheep, parasite resistance, imputation, GWAS