

## **Evaluación de la imputación de genotipos desde un chip de baja densidad (3K) a media (Chip-50K) y alta (Chip-HD) densidad en el ganado ovino**

Chitneedi<sup>1</sup>, P.K., A., Gutiérrez-Gil, B. y Arranz, J.J.

<sup>1</sup>Departamento de Producción Animal, Facultad de Veterinaria, Universidad de León, 24071 León. pchi@unileon.es

### **INTRODUCCIÓN**

En el ganado ovino la aplicación de la selección genómica depende, en gran medida, del precio asociado al genotipado de los animales. Por ello, la posibilidad de tener un chip con un coste adaptado al valor individual de los animales podría hacer la selección genómica más atractiva para el sector ovino. Sin embargo, el descenso del precio iría encaminado a la utilización de un chip con una baja densidad de marcadores, lo que podría comprometer la precisión de la selección genómica. La imputación de genotipos podría ofrecer una solución práctica y eficiente a este problema, reduciendo el coste total asociado con el genotipado y, al mismo tiempo, mantener una fiabilidad aceptable de las estimaciones genómicas (Erbe et al., 2012). La imputación es una aproximación *in-silico* por la que para una población genotipada con un chip de baja o media densidad se infieren los genotipos faltantes de un chip de más alta densidad en base a la información de una población de referencia que ha sido genotipada con la máxima densidad de marcadores considerada. Distintos autores han probado la eficiencia de la imputación genómica, con una precisión que en ganado ovino varía entre el 71% y el 98% (Hayes et al., 2012; Ventura et al., 2016).

La aproximación de imputación más eficiente se basaría en el genotipado de alta densidad sólo para los animales con mayor contribución genética en la población, como los fundadores o los antecesores, mientras que el resto de la población se genotiparía con un chip de densidad más baja para luego inferir los genotipos faltantes (Cleveland y Hickey, 2013). El tamaño de la población de referencia, la fiabilidad de la información referente al pedigrí, la relación entre la población a imputar y la población de referencia, así como el tipo de información utilizada por el software seleccionado para el análisis, desempeñan un papel crítico para obtener genotipos imputados con alta precisión (Hozé et al., 2013). En este estudio, hemos evaluado el potencial y la precisión de un chip de baja densidad (Chip-LowDensity, Chip-LD) virtual, diseñado *in silico*, para realizar imputación de genotipos a dos niveles superiores de densidades genotípicas, densidad media (Chip-50K) y alta densidad (Chip-500K o Chip-HighDensity, Chip-HD).

### **MATERIAL Y MÉTODOS**

**Población de estudio:** Este trabajo se basa en una población comercial de ganado ovino lechero con un total de 1.775 animales de del Núcleo de Selección de ANCHE. Un total de 1.686 animales (16 machos y 1.670 hijas de los mismos) han sido genotipados con el Chip-50K (37.400 SNPs, tras el control de calidad). Un subconjunto de 240 animales de esa población (los 16 machos y 14 hijas de cada familia), además de 94 de los machos con más hijas en el Núcleo de Selección se genotiparon con el Chip-HD (490.940 SNPs, tras el control de calidad).

**Chip-LD:** El chip de baja densidad a evaluar, Chip-LD, se diseñó mediante la selección de un subconjunto de los marcadores comunes entre los chips ovinos de media (Illumina Ovine SNP50 BeadChip) y de alta densidad (Ovine Infinium® HD SNP BeadChip). Los criterios de selección fueron la localización de los marcadores a lo largo de los 26 autosomas ovinos y su grado de informatividad medida mediante la frecuencia para el alelo menos frecuente (MAF > 0,3). La distancia media entre los marcadores seleccionados para el Chip-LD fue de aproximadamente de 1 Mb y el MAF medio aproximadamente 0,40.

**Imputación de genotipos:** El software utilizado para la imputación ha sido Beagle\_v4.0 (Browning et al., 2008). Los datos de entrada para este análisis incluyeron los datos de genotipado del Chip-LD, los datos del pedigrí de la población y los datos de genotipado del chip de densidad superior en cada caso, el Chip-50K o el Chip-HD. La evaluación de la precisión de imputación llevada a cabo se ha basado en una estrategia de validación cruzada en la que se utilizaron dos porcentajes de enmascaramiento para los animales de la población de referencia, del 10% (10 iteraciones) y 30% (4 iteraciones).

## RESULTADOS Y DISCUSIÓN

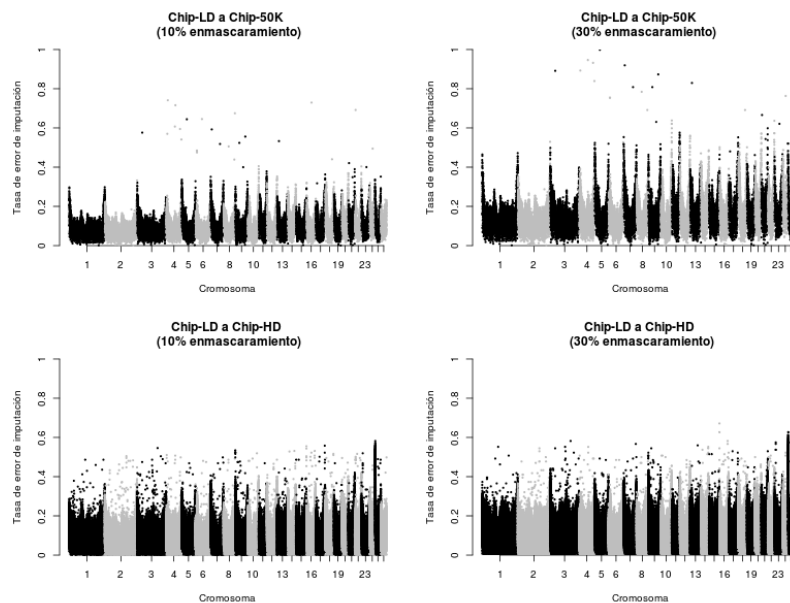
En base a los criterios establecidos para el diseño del Chip-LD *in silico*, un total de 2,935 fueron incluidos en el panel de baja densidad a evaluar en el presente trabajo. El rendimiento del software utilizado fue apropiado, con un rango de tiempo de ejecución de 3-30 minutos (para los cromosomas OAR26 y OAR1, respectivamente, y para la imputación de Chip-LD a Chip-50K y Chip-LD a Chip-HD).

Considerando los 26 cromosomas ovinos, el promedio de la precisión de imputación de los genotipos del Chip-50K a partir del Chip-LD para resultó ser muy similar para los dos niveles considerados de enmascaramiento, el 10% y el 30%, con valores del 93,53% ( $\pm 0,0086$ ) y 93,58% ( $\pm 0,0081$ ), respectivamente. Para la imputación del Chip-LD al Chip-HD, la precisión de la imputación fue del 88,3% ( $\pm 0,021$ ) para el nivel de enmascaramiento del 10% de la población de referencia y del 86,52% ( $\pm 0,012$ ) cuando se enmascaró el 30% de la población. La Figura 1 muestra el promedio, para las iteraciones realizadas, de la tasa de error en la imputación para cada uno de los marcadores imputados en los cuatro escenarios considerados. La menor precisión de imputación observada cuando la imputación de genotipos se hizo del Chip-LD al Chip-HD, comparando con la imputación al Chip-50K (aprox. 5-7% más baja), se podría atribuir a que el porcentaje de SNPs que tienen que ser imputados es mayor que cuando la imputación se hace en dos pasos y también al menor tamaño de la población de referencia (335 vs 1.680 animales) disponible para la imputación a la densidad superior. En ganado vacuno, otros autores han sugerido que, en el caso de la imputación desde la baja a la alta densidad, la mejor estrategia sería la imputación en dos pasos, inicialmente a densidad media y posteriormente a la mayor densidad (Berry et al., 2014). En cualquier caso, una población de referencia de mayor tamaño aumentaría considerablemente la precisión de la imputación (Hozé et al., 2013).

Comparando los resultados para los dos niveles de enmascaramiento considerados en la población de referencia, del 10% y 30%, vemos que este factor no tiene influencia para la imputación del Chip-LD al Chip-50K, mientras que en la imputación del Chip-LD al Chip-HD, la tasa de error de imputación fue superior en un 2% cuando el enmascaramiento afectó al 30% de la población de referencia. Esto se podría atribuir al menor número de iteraciones en los que se basa la estimación para este nivel de enmascaramiento, y/o al mayor número de genotipos a inferir para el mismo tamaño de población de referencia.

Con respecto a la distribución de la tasa de errores de imputación a lo largo de los cromosomas, a pesar de ser bastante homogénea para todos los escenarios, se observó que los marcadores localizados en los extremos de cada cromosoma muestran una tasa de error considerablemente más alta que la media del genoma (ver Figura 1). La identificación de esta condensación de errores en los extremos cromosómicos es relevante y se podría resolver mediante la redistribución de los marcadores en el diseño del Chip-LD, tratando de aumentar la densidad de marcadores en los extremos de los cromosomas. Hay que señalar además que, para la imputación a la densidad más alta, la tasa de errores fue excepcionalmente alta en el extremo proximal del cromosoma 25, llegando casi al 60%. Esto puede ser debido a la inadecuada asignación de SNPs en el Chip-HD, o a la mala calidad del borrador del genoma ovino de referencia en esa región específica.

Como principal conclusión del estudio podemos destacar que la imputación es una potente herramienta para inferir genotipos en poblaciones comerciales de ganado ovino, ofreciendo una tasa de errores relativamente baja, lo que es muy ventajoso en el ahorro de costes de genotipado. El coste de esta estrategia podría ser equivalente en costes al control de paternidad con marcadores microsatélites y, para aquellas poblaciones que utilicen la selección genómica, podría obtenerse con el coste de genotipado de un chip de baja densidad una estimación inicial del valor genético genómico que podría ser utilizado para decidir con qué animales se realiza la reposición.



**Figura 1.** Distribución de la tasa de errores de imputación a lo largo de los 26 cromosomas ovinos en los cuatro escenarios considerados: (i) Imputación del Chip-LD al Chip-50K con el 10% (a) y el 30% (b) de enmascaramiento; (ii) Imputación del Chip-50K al Chip-HD con el 10% (c) y el 30% (d) de animales enmascarados.

## REFERENCIAS BIBLIOGRÁFICAS

- Berry, D.P. et al. 2014. *J. Animal. Breed Genet.* 131: 165-172.
- Browning B.L. & Browning S.R. 2009. *Am. J. Hum. Genet.* 84: 210-223.
- Cleveland, M.A. & Hickey, J.M. 2013. *J. Animal Sci.* 91: 3583-3592.
- Erbe, M., et al. 2012. *J. Dairy Sci.* 95: 4114-4129.
- Hayes, B. et al. 2012. *Animal Gen.* 43: 72-80.
- Hozé, C. et al. 2013. *Genetics Sel. Evol.* 45: 33
- Purcell, S. et al. 2007. *Am. J. Hum. Genet.* 81: 559-575.
- Ventura, R.V. et al. 2016. *Genet. Sel. Evol.* 48: 71.

**Agradecimientos:** Trabajo financiado por el proyecto AGL2012-34437 del Ministerio de Economía y Competitividad España (MINECO). P.K. Chitneedi es beneficiario de una beca predoctoral de la Junta de Castilla y León y el Fondo Social Europeo. B. Gutiérrez-Gil es investigadora contratada del programa “Ramón y Cajal” del MINECO (RYC-2012-10230).

## LOW DENSITY CHIP DESIGN AND GENOMIC IMPUTATION FROM LOW DENSITY (LD-Chip) SNP CHIP TO MEDIUM (50K-Chip) AND HIGH (HD-Chip) DENSITY IN SHEEP

**ABSTRACT:** In this study we designed an *in silico* low density chip (LD-Chip) for a half-sib commercial population of the Selection Nucleus of Churra sheep and investigated the potential of this genomic tool to impute genotypes to medium (50K-Chip) and high (HD-Chip) marker densities. Based on the available genotypes for the ovine Chip-50K ( $n = 1,686$ ) and HD-Chip (335 individuals) a total of 2,935 markers common to the two panels and evenly distributed across the sheep genome were included in the *in silico* LD-Chip. The average error rate for the LD-Chip to 50K-Chip imputation was around 93.5% and ranged from 88,3% to 86,52% for the LD-Chip to HD-Chip imputation (depending on the masking level considered, 10% and 30%, respectively). For all the scenarios, a higher imputation error rate than the average was found at both ends of all the chromosomes. For the imputation to the HD-Chip density, a very high error rate was identified at the proximal end of chromosome 25. Globally, and although the identified limitations need to be solved, the *in silico* LD-Chip described here, together with the implemented imputation strategy, appear as an appropriate and affordable approach to obtain, in addition to the basic genetic tests (e.g. paternity), preliminary estimations of genomic breeding values that could guide breeding decisions.

**Keywords:** Churra sheep, SNP-Chip, imputation, accuracy.