

# Priorización de variantes identificadas por secuenciación del genoma completo para la validación de asociaciones con caracteres de interés

E. Molinero, R.N. Pena, J. Estany y R. Ros-Freixedes

Departamento de Ciencia Animal, Universidad de Lleida - Agrotecnio-CERCA Center, 25198 Lleida, Catalunya, España  
[eduard.molinero@udl.cat](mailto:eduard.molinero@udl.cat)

## Resumen

La secuenciación del genoma completo es una herramienta muy potente para identificar las variantes genéticas que subyacen los caracteres complejos. Esto se debe a que esta tecnología permite identificar variantes raras y específicas de la población estudiada, incluyendo variantes causales. Las variantes causales pueden incrementar la precisión de las predicciones genéticas, así como la persistencia de la predicción entre generaciones y tipos genéticos. Sin embargo, la elevada densidad de variantes con un alto grado de desequilibrio de ligamiento dificulta la identificación de variantes relevantes. El objetivo de este estudio es determinar si una estrategia de priorización secuencial de las variantes obtenidas tras la secuenciación del genoma completo de un grupo reducido de animales es efectiva para identificar variantes que puedan ser relevantes para caracteres complejos. Para ello, se secuenció el genoma completo de 205 cerdos Duroc y se identificaron 19 millones de variantes, que se priorizaron según información de estudios de asociación, índice de fijación, consecuencia funcional predicha de la variante y la ontología del gen, utilizando como caracteres modelo el contenido y composición de la grasa (12 caracteres). Se priorizaron las variantes (i) que presentaron un índice de genotipado superior a 0.9 y una frecuencia alélica igual o superior a 0.2, de tal manera que todos los genotipos estuvieran suficientemente representados en los 205 animales; y (ii) que estuvieran localizadas en regiones transcritas de los genes, incluyendo intrones y 500 bases alrededor de las regiones UTR, ya que es difícil de predecir el efecto de las variantes en regiones no codificantes utilizando únicamente la secuencia de ADN. De esta manera, se seleccionaron 240 SNPs en 193 genes que fueron genotipados en 960 cerdos de la misma población mediante un chip de genotipado personalizado. La asociación de las 240 variantes se analizó mediante un modelo lineal que incluía los efectos del lote de sacrificio (14 niveles), los genotipos de genes con un efecto mayor sobre los caracteres modelo (*LEPR* y *SCD*, con 3 niveles cada uno), y la edad de sacrificio. Resultados preliminares con 628 cerdos mostraron que 133 de las 240 variantes presentaban una asociación significativa ( $p < 0.05$ ) con al menos uno de los caracteres testados. Una vez aplicada la corrección de Bonferroni para test múltiples ( $p < 2.5 \cdot 10^{-4}$ ), 10 de estas variantes presentaron asociación significativa con al menos uno de los caracteres testados. Estas variantes correspondían a seis regiones que contienen nueve genes candidatos. En una de estas regiones se encuentra el gen *DGAT2* (SSC9, 10-12Mb), cuyo efecto ya se había detectado previamente en esta población. Las otras cinco regiones (cinco genes candidatos) no se habían detectado todavía en esta población. Este estudio nos permitirá evaluar una estrategia para la detección masiva de variantes asociadas a caracteres complejos para su posterior uso en programas de mejora, que puede ser especialmente efectiva si se complementa con un esquema de validación apropiado.

*Palabras clave: secuenciación del genoma completo, análisis de asociación, validación*