Limitaciones de identificabilidad de la impronta genómica en modelos gaméticos

*David López-Carbonell[*12], Gregor Gorjanc[2], Manuel Sánchez-Díaz[1], Francisco Barceló-Blasco[1], Luis Varona[1]*

[1] Instituto Agroalimentario de Aragón (IA2), Universidad de Zaragoza, Calle Miguel Servet, 177, 50013, Zaragoza
[2] The Roslin Institute and Royal (Dick) School of Veterinary Studies, Easter Bush Campus, Midlothian EH25 9RG, Edinburgh
* Corresponding author: davidlc@unizar.es

Resumen

Los modelos genómicos permiten descomponer la variación en componentes aditivos en función de su origen parental. Las diferencias entre ambos componentes permiten interpretar posibles efectos asociados al progenitor de origen, incluido el fenómeno de impronta. Sin embargo, la estimación de estos efectos no está garantizada en la práctica, ya que depende de manera crítica de la estructura de los datos y de la matriz de relaciones genómicas a nivel gamético. En mejora animal, los animales únicamente transmiten sus alelos vía materna o paterna. No obstante, los modelos con impronta suelen aplicarse sin una evaluación previa de su identificabilidad, lo que puede conducir a inferencias inestables o espurias. En este contexto, es importante disponer de herramientas diagnósticas que permitan determinar si los efectos de impronta están confundidos. El objetivo de este trabajo es desarrollar un marco de diagnóstico para evaluar la identificabilidad y estimabilidad de los efectos de impronta en modelos genómicos gaméticos. Para ello, se propone una aproximación basada en teoría de la información, que permite analizar la identificabilidad estructural mediante el rango de matrices clave y la dimensión del espacio nulo. El procedimiento implica el análisis del rango de Z*GZ*', donde Z* es la diferencia entre las matrices de incidencia asociadas a los efectos gaméticos paterno y materno y G es la matriz de relaciones gaméticas. Si es rango es 0 es estructuralmente no identificable. Si el rango es igual que el rango de las matrices de incidencia paterna o materna la identificabilidad es completa. Valores intermedios implican identificabilidad parcial. En este caso, se puede cuantificar el grado de identificabilidad mediante el cálculo del espacio nulo mediante un Single Value Decomposition de Z*G, que proporciona el número de niveles no estimables. A modo de ilustración, se han simulado 800 observaciones fenotípicas distribuidas en cuatro generaciones asociadas a una matriz de relaciones gaméticas genómicas construida a partir de haplotipos conocidos. Cuando se empleó el conjunto de datos completo, el modelo mostró un porcentaje de efectos estimables del 87.5%. Al restringir el análisis a la primera generación (n=200), el modelo presentó identificabilidad estructural completa, sin espacio nulo. Sin embargo, al incorporar una segunda generación (n=400), la identificabilidad se redujo levemente (99.25%). A pesar de ello, la información de Fisher aumentó progresivamente con el tamaño muestral, reflejando una mayor precisión en la estimación dentro del subespacio identificable cuando la cantidad de datos era mayor. En conjunto, estos resultados muestran que la identificabilidad de los efectos de impronta no depende únicamente del tamaño muestral, sino de la estructura del diseño y de la diversidad efectiva de los gametos representados y que pueden existir redundancias estructurales que pueden comprometer la identificabilidad de los efectos de impronta. Este comportamiento pone de manifiesto la importancia de la profundidad genealógica y la reutilización de haplotipos, y que la mera ampliación del tamaño de los datos no garantiza una mejor estimación de los efectos. Por tanto, la evaluación conjunta de la estructura y de la información disponible resulta esencial para una aplicación robusta de modelos gaméticos.

Keywords: Genomic imprinting, Parameter identifiability, Gametic genomic models