

Errores mendelianos en paneles de SNPs: el caso del Gochu Asturcelta.

Katherine D. Arias¹, Isabel Álvarez¹, Juan Pablo Gutiérrez², Iván Fernández¹, Juan Menéndez³, Nuria A. Menéndez-Arias¹, Félix Goyache¹

¹ SERIDA-Deva, Camino de Rioseco 1225, 33394-Gijón (Asturias), España.

² Universidad Complutense de Madrid, Puerta de Hierro s/n, 28040-Madrid, España.

³ 3ACGA, C/ Párroco José Fernández Teral nº 5 A, 33403-Avilés (Asturias), España.
kathyah18@gmail.com.

En paneles de SNPs, las variantes incongruentes con las reglas de herencia mendeliana se consideran errores mendelianos (ME). Los ME se deben, más que a mutaciones *de novo*, a errores de genotipado cuyas causas pueden ser la existencia de variación del número de copias (CNV), o la presencia de SINEs o LINEs. El objetivo principal fue caracterizar los errores sistemáticos y no sistemáticos en paneles SNP, así como analizar posibles estrategias en la identificación de ME. Se genotipó con el panel Axiom_PigHDv1(Affymetrix) un pedigrí de cerdo Gochu Asturcelta que incluía 492 individuos y 478 tríos (padre-madre-hijo). No se aplicaron filtros para frecuencias alélicas o test de Hardy-Weinberg (HW). Solo se consideraron individuos y SNPs con *call rate* >0.95 y >0.9, respectivamente, reteniéndose un total de 545,364 SNPs. Los ME se identificaron mediante el programa PLINK v1.9. Se definieron 8 clases de ME de acuerdo a la asignación del error a cada miembro del trío (Trio si no es posible determinar en qué miembro del trío ocurrió el ME) y el alelo en el que se produjo (A o B para los alelos más y menos frecuentes, respectivamente): TrioA/B, PadreA/B, MadreA/B, HijoA/B. Se calculó el F_{IS} para cada locus con ME. Se consideraron posibles errores causados por alelos nulos (ADO) o falsos alelos (ADI_{het} y ADI_{hom}). Las CNV se identificaron mediante el programa PennCNV. Utilizando BedTools, los loci con ME se superpusieron con las CNV identificadas y los SINEs y LINEs anotados en el genoma Sscrofa11.1. Un total de 40,540 SNPs (error medio por locus = 0.074) tuvieron, al menos, un ME en un individuo (de 331 a 2,016 loci por individuo). En total, se identificaron 292,297 errores alélicos (error medio por alelo en la población= 0,0005). La clase TrioB incluyó la mayoría de los loci identificados (67%) pero solo el 22% de los errores alélicos. La mayoría de los errores alélicos se asignaron al Padre (36.9%) o la Madre (38.2%). Estas clases de ME mostraron un claro sesgo por el tamaño familiar. Las clases de error Padre ($F_{IS}=0.27$), Madre ($F_{IS}=0.21$), e Hijo ($F_{IS}=0.6$) tuvieron un marcado déficit de heterocigotos. Un total de 388 loci fueron identificados como ADO, 3,447 loci como ADI_{het} y sólo 17 loci como ADI_{hom} . Nuestros resultados confirman que no todos los ME son equivalentes. Los ME parecen deberse, fundamentalmente, a: a) presencia de alteraciones genómicas (fundamentalmente CNV) que inducen a errores en muchos loci con baja tasa de errores alélicos y en equilibrio HW (TrioB y ADI_{het}) que pueden ser particulares de cada población; y b) incorrecta identificación del alelo (alelos nulos o parcialmente nulos) que afectan a pocos loci y producen muchos errores alélicos (PadreA/B, MadreA/B, HijoA/B, ADO y ADI_{hom}), pudiendo ser problemático para estudios de asociación, ligamiento o identificación de variantes raras en el genoma. El uso de F_{IS} (valores extremos positivos), como criterio de calidad de los paneles SNP cuando el pedigrí no está disponible, puede ser suficiente para eliminar la mayoría de las incongruencias debido a la presencia de ADO y ADI_{hom} .

Palabras clave: CNV, SINEs, LINEs, herencia mendeliana, tamaño familiar.