

Cómo traducir las lecturas genotípicas del chip SNP ovino Affymetrix (ThermoFisher) al formato Illumina

Z.G. Vitezica^{1*}, C. Pineda-Quiroga², J.M. Astruc³, I. Granado-Tajada², E. Ugarte², A. Zubiri-Gaitán¹, S.T. Rodríguez-Ramilo¹; I. Aguilar⁴

¹ GenPhySE, Université de Toulouse, INRAE, ENVT, 31326 Castanet-Tolosan, Francia

² NEIKER-BRTA, Campus Agroalimentario de Arkaute, 01192, Vitoria-Gasteiz, España

³ Institut de l'Élevage, F-75595, Paris, Francia

⁴ Instituto Nacional de Investigación Agropecuaria (INIA), 11500, Montevideo, Uruguay

* Corresponding author: zulma.vitezica@inrae.fr

Resumen

En las razas ovinas lecheras Manech (en Francia) y Latxa (en España) se puede realizar una evaluación genética conjunta de las poblaciones rubias (Manech Tête Rousse y Latxa Cara Rubia) o bien negras (Manech Tête Noire y Latxa Cara Negra de Navarra).

La evaluación genética conjunta exige la puesta en común de datos entre países: registros, pedigrí y genotipos SNP. Los genotipos provienen de diferentes tecnologías: Illumina y Affymetrix. Los animales Manech (10.955 individuos) fueron genotipados mediante chips Illumina de baja densidad (18K) y de densidad media (50K), mientras que los animales de la raza Latxa (2.751 individuos) se genotiparon utilizando chips Affymetrix e Illumina de densidad media en proporciones cercanas al 75-25. Integrar datos de SNP procedentes de diferentes tecnologías supone un reto, y la dificultad suele subestimarse. El objetivo de este trabajo es mostrar cómo crear un conjunto de SNP comunes.

En principio, ambos chips Affymetrix e Illumina de densidad media tienen una gran cantidad de loci genotipados en común. Sin embargo, es difícil decir para un locus del chip de Illumina cuál es el locus correspondiente de Affymetrix y viceversa. Un estudio preliminar en 2022, usando la localización (cromosoma y posición) de los SNP encontró solo 22.827 SNPs comunes a ambos chips. El uso de la posición de los SNP es inadecuado. Los SNP pueden tener ubicaciones diferentes si la versión del ensamblaje del genoma difiere, incluso dentro de la misma tecnología (por ejemplo, el mismo chip genotipado en momentos diferentes). Una concordancia más adecuada se basa en el “nombre del SNP” (también llamado el identificador del SNP o “SNP ID”). Para ello es necesaria una lista de concordancia “nombre del locus en Illumina” y “nombre del mismo locus en Affymetrix”. Sin embargo, esta información básica de concordancia no es fácilmente accesible.

La información se consigue de la siguiente manera. El software *Axiom Analysis Suite* proporciona una columna “*custom ID*” que es básicamente la misma identificación que en Illumina, con algunas ligeras diferencias de formato (por ejemplo, OAR1_107588571_01 en Affymetrix es OAR1_107588571.1 en Illumina).

Una dificultad adicional es que el conjunto de loci genotipados por *Axiom Analysis Suite* varía según la cantidad de animales genotipados, con lo cual es mejor analizar todos los animales genotipados disponibles con este software. Los nombres de los SNP se deben extraer únicamente de la categoría “*Best and Recommended*”, dado que un SNP se incluye en esta categoría si hay tres grupos de buena resolución (los dos homocigotos y los heterocigotos). Además, hay que usar la versión más reciente del archivo de anotación de Axiom, en este caso la R5.

Siguiendo estas pautas, se analizaron todos los animales de raza Latxa genotipados. Los nombres de los SNP se obtuvieron cotejando el “name” en el Manifest File de Illumina OvineSNP50 Beadchip v3 con el “*custom ID*” del informe de Axiom. Se usó un script en Python para ajustar los nombres (por ejemplo, OAR1_107588571_01 en Affymetrix es OAR1_107588571.1 en Illumina). Como resultado, 38516 SNPs concordaban entre ambos chips.

Keywords: SNP, chips, Affymetrix, Illumina

Agradecimientos: Este estudio ha recibido financiación del proyecto ARDI2 (EFA 032/01) de los fondos POCTEFA. Agradecemos a la plataforma bioinformática Genotoul Toulouse Occitanie (Bioinfo Genotoul, <https://doi.org/10.15454/1.5572369328961167E12>) la provisión de recursos informáticos y

de almacenamiento. Los autores agradecen a los ganaderos por los datos.