CNV: CONSIDERANDO NUEVAS VARIANTES EN EL GENOMA

Ben Sassi[1], N., Fernández, A.I., González-Recio O.
[1]Dpto. Mejora Genética Animal, INIA, 28040 Madrid.
e-mail: naila.bensassi@gmail.com

INTRODUCCIÓN

Las variantes estructurales denominadas 'Copy Number Variants' (CNV) corresponden a deleciones o duplicaciones de segmento del DNA que ocupan 1Kb o más de longitud del genoma. Estudios recientes muestran que estos polimorfismos originan variabilidad genética y podrían ser responsables de una parte de la variabilidad fenotípica de caracteres relevantes tanto en la especie humana como en especies ganaderas (Girirajan *et al*., 2011; Clop *et al*., 2012). A pesar de que varios trabajos han identificado cientos de CNVs en vacuno (Bickhart *et al*., 2012), pocos son los trabajos que los han asociados a caracteres productivos en esta especie (Kramper, 2012).
El objetivo del presente trabajo fue identificar CNVs asociados a los principales caracteres productivos de la raza bovina Holstein mediante el análisis de la señal de hibridación de las sondas del chip de genotipado masivo de SNPs BovineSNP50BeadChip (Illumina).

MATERIALES Y METODOS

Se usaron datos correspondientes a 1900 toros en la raza Holstein cedidos por la Confederación de Asociaciones de Frisona Española (CONAFE) con los cuales se realizó el genotipado masivo de SNPs con el chip bovineSNP50BeadChip (Illumina).
Detección de CNVs: La detección de los CNVs se realizó utilizando el programa PennCNV (Wang, 2008), siguiendo la anotación *UMD3.1* del genoma bovino (cromosoma y posición de los SNPs). Previamente al análisis de detección de CNVs, se llevó a cabo una corrección por el contenido en bases GCs de la región así como un filtrado por la calidad de las muestras basado en las medidas de intensidad de señal de las hibridaciones. Un total de 1724 individuos pasaron el control de calidad. Finalmente, la identificación de CNVs se realizó mediante la opción individual CNV-call de PennCNV. Aquellos CNVs solapantes fueron ensamblados en regiones CNV (CNVRs). Estas regiones fueron filtradas según los siguientes criterios: i) contener 3 o más SNPs consecutivos y ii) ser detectados en un mínimo número de individuos.
Análisis de asociación: Los análisis de asociación se realizaron sobre aquellos toros (1195) que disponían de datos fenotípicos, utilizando como variables dependientes las pruebas MACE de-regresadas de diciembre de 2011 para los caracteres de kilogramos de proteína (PR), kilogramos de grasa (GR), recuento de células somáticas (RCS) y días abiertos (DA). Se utilizaron dos metodologías de análisis: i) regresión linear estándar empleando el programa PLINK y corrección por múltiples test (FDR<0.05) de Benjamini-Hochberg (Wang *et al.,* 2008), ii) análisis bayesiano definiendo dos modelos: lineal y cuadrático. El modelo cuadrático se empleó con el fin de testar asociaciones no lineales entre el carácter y la región CNV. Los CNVs que tras un test de permutación, permitían rechazar la hipótesis nula (no asociación), fueron seleccionados como candidatos al análisis *in-silico.*
Para explorar el contenido de genes en las CNVRs que mostraron asociación con los caracteres analizados se utilizaron las herramientas BioMart de Ensembl (ensembl.org/biomart) y las bases de datos del NCBI (www.ncbi.nlm.nih.gov) y DAVID (david.abcc.ncifcrf.gov/) para explorar la función biológica de los mismos.

RESULTADOS Y DISCUSIÓN

Un total de 2022 CNVRs fueron detectados en los 1724 individuos analizados. El tamaño de estas regiones varió desde 30,128 Kb a 17,165 Mb con un tamaño medio de 616,132 Kb. El porcentaje total del genoma ocupado por CNVR fue del 52,19%. De éstos, se seleccionaron aquellos CNVRs que mostraron una frecuencia superior al 5‰ en los 1195 individuos con datos fenotípicos, que correspondió a un total de 315 CNVRs que fueron finalmente testados en el análisis de asociación, siendo 28 inserciones, 15 deleciones y 269 inserción-deleción (mismo CNVR presenta inserción y deleción en distintos individuos). Las frecuencias oscilaron entre 5‰ y 53%.

Los resultados de los análisis de asociación se resumen en la tabla 1. Estos resultados revelan asociaciones significativas para los cuatro caracteres analizados. Los resultados más consistentes fueron en los CNV800 y 779 detectados en ambos GR y PR. Ambos CNVRs resultaron significativos con ambos métodos lineales tanto para PR como para GR. El que se hayan identificado CNVRs asociados a ambos caracteres es coherente, dado que existe una alta correlación positiva entre la cantidad de grasa y proteína en la leche.

Tabla 1: *CNVs significativos y sus respectivas estimas de los coeficientes según el análisis en el que resultasen significativos. Se muestra además el porcentaje de duplicaciones y delecciones observados en la muestra utilizada*

Car	CNV	RLS		Bayes Lineal		Bayes cuadrático	% Dup	% Del
		β±SE	P-value*	α	β	β_2		
GR	778	5,8±1,4	$6,0x10^{-4}$	1,7	-7,2	-	14,8	11,1
	800	4,6±0,9	$5,6x10^{-3}$	0,4	-5,8	-	29,7	23,1
	739	-	-	-0,9	-15,1	6,9	24,4	22,7
PR	778	6,8±1.2	$1,1x10^{-5}$	3,2	-8,5	-	14,8	11,1
	800	4,6±0.8	$2,0x10^{-5}$	0,5	-5,7	-	29,7	23,1
	427	-	-	16,6	-38,3	15,5	3,0	1,5
	446	-	-	5,3	-20,6	7,9	10,2	11,8
	739	-	-	-1,3	-13,9	6,6	24,4	22,7
	1192	-	-	1,1	-16,5	7,2	18,0	16,0
	1358	-	-	-1,0	-14,5	7,4	20,0	13,2
	1511	-	-	-211,7	307,9	-101,8	0,16	0,75
	1644	-	-	-0,8	-13,7	6,5	21,8	15,7
RCS	82	-	-	123,5	31,4	-30,2	0,1	1,5
	861	-	-	121,7	19,5	-15,8	0,7	1,1
	879	-	-	116,8	30,2	-22,3	0,2	0,4
	1200	-	-	131,5	-12,9	5,4	7,2	3,6
	1675			119,9	22,0	-17,1	0,3	0,4

* P-value corregido (FDR<0.05), RLS: Regresión Linear Estándar, GR: Kg de grasa; PR: Kg de proteína; RCS: Recuento de células somáticas; DA: Días abiertos; RSL: regresión lineal estándar; Dup: Duplicación; Del: Deleción

La anotación de los genes contenidos en los CNVRs asociados (Tabla 2) permite identificar genes funcionalmente relacionados con los caracteres analizados y por tanto constituyen potenciales genes causales. Entre estos destaca el gen de la *hormona de crecimiento (GH)* localizado en el CNV800, previamente asociado con cantidad de proteína y grasa en la leche en ganado bovino Holstein-Frisona (Mullen *et al.,* 2010), y el gen *Low density lipoprotein-related protein 1B LRP1B* localizado en el CNV861, que ha sido asociado al carácter recuento de células somáticas (Cole *et al.,* 2011). Según nuestros resultados, los individuos con deleción en las regiones CNV 778 y 800 producirían más kg de grasa y proteína que la media. Los CNVRs detectados proporcionan un mayor conocimiento de las regiones genómicas asociadas a caracteres de interés en Holstein, y podrían ser útiles en las evaluaciones genómicas de reproductores. Estas utilidades serán objeto de futuros estudios.

Tabla 2: *Genes potencialmente relacionados con los caracteres encontrados en los CNVs significativos*

Carácter	ID_CNVR (nº genes contenidos)	Gen relacionado
RCS	82(0), 861(1), 879(0), 1200(132), 1675(2)	*LRP1B*
GR	800(495), 778(7)	*GH,SREBF, STAT5,ACLY, TNFR,WF1KKN1*
	739(132)	-
PR	800(495), 778(7)	*GH*
	76(70), 80(96), 739(132), 1192(2), 1358(360), 1511(7), 1644(161)	*DGAT1,IGFAL,GFER, PPARA, CPT1B*

DA: Días abiertos, RCS: Recuento de células somáticas; GR: kg de grasa; PR: kg de proteína

REFERENCIAS BIBLIOGRAFICAS

• Bickhart D., Hou Y., Schroeder S., Alkan C. *et al.* 2012 Genome Research 22(4):778–790 • Clop A., Vidal O., Amilis M., 2012 Anim Genet 43(5) • Cole J., Wiggans G., Ma L., *et al.* 2011 BMC Genomics 12:408 • Girirajan S., Rosenfeld J.A., Coe B.P. *et al.* 2012 N Engl J Med 367 • Kramper T., 2012 Tesis Doctoral • Mullen M.P., Lynch C.O., Waters S.M. et al. 2011 GMR 10: 1819-1830 • Wang K., Chen Z., Tadesse M.G. *et al.* 2008 Nucleic Acids Research 36(21).

Agradecimientos: Los autores agradecen a CONAFE, Aberekin S.A., Ascol, Geneticale, Xenetica Fontao, MAGRAMA, LCV Algete por su colaboración en el proyecto. Neila Ben Sassi disfruta de una beca de Master del CIHEAM, y agradece a CONAFE el pago de la inscripción a las jornadas.

CNV: CONSIDERING NEW VARIANTS IN THE GENOME

Abstract: The objective of this study was to identify copy number variants associated to economical traits (fat and protein yield, somatic cell count and days open) in Holstein. CNVs were detected by exploiting hybridization signal intensities from the BovineSNP50BeadChip and using PennCNV software. More than 2000 CNVRs were identified, 315 thereof appeared segregating in the sires with available phenotypic data. Frequentist and Bayesian approaches were used in the association analyses. Both fat and protein yield shared associated CNVRs. We may highlight *GH, SREB, STAT5, TNRF and ACLY* among the genes located within these regions, which have been previously associated to fat yield in different studies. Further, five different genes (*DGAT1, IGFAL, GFER, PPARA, CPT1B*) were detected and previously reported as related to both protein and fat yield pathways in Holstein. Additionally, we reported that six different CNVRs are significantly associated with somatic cell count and among the genes contained within these regions, the *LRP1B*, have been already related to the trait.

Keywords: CNV, association analyses, Holstein.