Genome-wide copy number variation in Hanwoo, Black Angus, and Holstein cattle

Article *in* Mammalian Genome · March 2013

DOI: 10.1007/s00335-013-9449-z · Source: PubMed

Mamm Genome (2013) 24:151–163
DOI 10.1007/s00335-013-9449-z

Genome-wide copy number variation in Hanwoo, Black Angus, and Holstein cattle

Jung-Woo Choi · Kyung-Tai Lee · Xiaoping Liao · Paul Stothard · Hyeon-Seung An · Sungmin Ahn · Seunghwan Lee · Sung-Yeoun Lee · Stephen S. Moore · Tae-Hun Kim

Received: 4 July 2012 / Accepted: 6 February 2013 / Published online: 30 March 2013

Abstract Hanwoo, Korean native cattle, is indigenous to the Korean peninsula. They have been used mainly as draft animals for about 5,000 years; however, in the last 30 years, their main role has been changed to meat production by selective breeding which has led to substantial increases in their productivity. Massively parallel sequencing technology has recently made possible the systematic identification of structural variations in cattle genomes. In particular, copy number variation (CNV) has been recognized as an important genetic variation complementary to single-nucleotide polymorphisms that can be used to account for variations of economically important traits in cattle. Here we report genome-wide copy number variation regions (CNVRs) in Hanwoo cattle obtained by comparing the whole genome sequence of Hanwoo with Black Angus and Holstein sequence datasets. We identified 1,173 and 963 putative CNVRs representing 16.7 and 7.8 Mbp from comparisons between Black Angus and Hanwoo and between Holstein and Hanwoo, respectively. The potential functional roles of the CNVRs were assessed by Gene Ontology enrichment analysis. The results showed that response to stimulus, immune system process, and cellular component organization were highly enriched in the genic-CNVRs that overlapped with annotated cattle genes. Of the 11 CNVRs that were selected for validation by quantitative real-time PCR, 9 exhibited the expected copy number differences. The results reported in this study show that genome-wide CNVs were detected successfully using massively parallel sequencing technology. The CNVs may be a valuable resource for

Jung-Woo Choi and Kyung-Tai Lee contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s00335-013-9449-z) contains supplementary material, which is available to authorized users.

J.-W. Choi · X. Liao · P. Stothard · S. S. Moore
Department of Agricultural, Food and Nutritional Science, University of Alberta, Edmonton, AB T6G 2P5, Canada
e-mail: jwchoi@ualberta.ca

X. Liao
e-mail: xliao2@ualberta.ca

P. Stothard
e-mail: stothard@ualberta.ca

S. S. Moore
e-mail: s.moore3@uq.edu.au

K.-T. Lee · H.-S. An · S. Lee · T.-H. Kim (✉)
Division of Animal Genomics and Bioinformatics, National Institute of Animal Science, Rural Development Administration, Suwon, Republic of Korea
e-mail: thkim63@korea.kr

K.-T. Lee
e-mail: leekt@korea.kr

H.-S. An
e-mail: chivalry78@naver.com

S. Lee
e-mail: slee46@korea.kr

S. Ahn · S.-Y. Lee
Laboratory of Genomics and Genomic Medicine, Lee Gil Ya Cancer and Diabetes Institute, Gachon University of Medicine and Science, Incheon, Republic of Korea
e-mail: smahn@gachon.ac.kr

S.-Y. Lee
e-mail: liketree36@gmail.com

S. S. Moore
Queensland Alliance for Agriculture and Food Innovation, University of Queensland, St Lucia, QLD, Australia

further studies to correlate CNVs and economically important traits in cattle.

Introduction

Hanwoo is a native cattle breed found on the Korean peninsula which has been used as a draft animal for about 5,000 years. Currently, four Hanwoo breeds (Korean Brown, Korean Brindle, Korean Black, and Jeju Black) within Korea are registered by the Food and Agriculture Organization (FAO). Among the breeds, Korean Brown has been widely used for beef, with an emphasis on greater marbling for the improved juiciness and flavor that is in demand in the Korean beef market. Over the past 40 years, the main role of Hanwoo has changed from draft animal to beef cattle. This change is mainly the result of the accelerated industrialization and increased meat consumption in Korea. The first official genetic breeding program for Hanwoo was initiated in 1979 by the Korean government and the productivity of Hanwoo has been substantially improved. For example, between 1998 and 2010, muscle mass and area in Hanwoo steer have been increased from 319 to 418 kg and from 75.4 to 88.9 cm^2, respectively, and marbling scores have gone up from 3.5 to 5.1 on a scale from 1 to 7 (NIAS 2011). In addition, it was reported that Hanwoo and Wagyu have a higher ratio of monounsaturated fatty acids (15–23 %) in their intramuscular fat content at final slaughter age than other beef breeds (Kim et al. 2005; Smith et al. 2009).

After the release of the bovine genome sequence assembly (Elsik et al. 2009; Liu et al. 2009; Zimin et al. 2009), a large number of genetic variations, especially single-nucleotide polymorphisms (SNPs), have become widely available and commercial high-density SNP panels have been developed (Matukumalli et al. 2009). The continued discovery of SNPs in diverse cattle breeds has been further expedited (Eck et al. 2009; Stothard et al. 2011) by the recent availability of massively parallel sequencing technologies called next-generation sequencing (NGS). SNPs and the commercial SNP marker panels have been successfully used to identify genomic regions that potentially shape the current productive cattle (Barendse et al. 2009; Gibbs et al. 2009; Hayes et al. 2009; Jiang et al. 2010; Sherman et al. 2010). Recently, copy number variation (CNV) has been attracting attention as another form of structural variation that can account for diverse, economically important traits in cattle. CNV was defined initially as a DNA segment >1 kb in size that displayed variable copy numbers in comparison with a reference genome (Feuk et al. 2006). The spectrum of CNV has now widened to include shorter segments, even down to 50 bp in length. This new definition is mainly the result of the improved resolution to detect CNV that is provided by the wide availability of NGS (Alkan et al. 2011). In cattle, a limited number of investigations have been performed to detect CNVs using methods that include high-density array comparative genomic hybridization (aCGH), the Bovine 50K SNP panel, and read depth analysis of NGS short reads (Bae et al. 2010; Fadista et al. 2010; Liu et al. 2010; Stothard et al. 2011). In particular, read depth analysis using NGS short reads to identify CNVs has been seen as a promising approach because it can provide higher resolution of putative CNVs than the currently available commercial SNP chip and aCGH methodologies (Alkan et al. 2011; Stothard et al. 2011). In the present study, we present genome-wide CNV regions (CNVRs) that we identified by comparing whole-genome resequenced Hanwoo (Korean Brown cattle) data with resequenced datasets from Black Angus and Holstein, representative beef and dairy breeds in North America. In addition, we performed deep annotation on all the detected CNVRs and downstream functional analysis on selected CNVRs to investigate the potential biological roles (Tables 1, 5).

Materials and methods

DNA sampling and sequencing

We obtained a proven Hanwoo bull (27223, born on 26 April 2007) from the Hanwoo Experiment Station, National Institute of Animal Science, Rural Development Administration, Korea, for whole-genome resequencing. Genomic DNA was extracted from the EDTA-blood using a QIAamp DNA Blood Maxi Kit according to the manufacturer's instructions (Qiagen). DNA was fragmented by Covaris S2 (Covaris) and HydroShear (Genomic Solutions) at proper settings for the targeted sizes. Libraries were prepared according to the "SOLiD™ System Mate-paired Library Preparation" protocol of the Applied Biosystems SOLiD System: Library Preparation Guide (02/2009 and 10/2009 editions). Four mate-pair libraries of three different sizes (600–700, 1–2, and 0.6–2.2 kb) were constructed, and sequencing was performed according to the Applied Biosystems SOLiD System protocol. Templated beads were deposited onto two slides of full-scale per library and sequencing was carried out to 50 bases using SOLiD v3.0 chemistry and following the manufacturer's instructions; an exception was the library prepared from the 0.6–2.2 kb sheared DNA fragments which was used for four slides of full-scale.

Short-reads mapping

The Hanwoo sequence reads were aligned in color space to the bovine reference genome assembly UMD 3.1 (Zimin

Table 1 Gene ontology terms enriched among the genic-CNVRs from HOLvsHAN

GO term	Ontology	Breed of Gain	Description	P-HOL	P-HAN
Biological process					
GO:0032502	P	HOL/HAN	Developmental process	3.9E-118	2.9E-020
GO:0032501	P	HOL/HAN	Multicellular organismal process	2.2E-170	4.5E-030
GO:0016265	P	HOL/HAN	Death	2.8E-029	2.00E-05
GO:0002376	P	HOL/HAN	Immune system process	2.2E-027	3.30E-12
GO:0000003	P	HOL/HAN	Reproduction	4.9E-064	1.30E-08
GO:0016043	P	HOL/HAN	Cellular component organization	1.2E-038	5.70E-05
GO:0022414	P	HOL/HAN	Reproductive process	2.9E-054	5.40E-07
GO:0051234	P	HOL/HAN	Establishment of localization	1.2E-016	1.00E-05
GO:0051179	P	HOL/HAN	Localization	4.9E-022	3.50E-06
GO:0050896	P	HOL/HAN	Response to stimulus	3.1E-076	1.8E-025
GO:0048518	P	HAN	Positive regulation of biological process	–	2.20E-23
GO:0048519	P	HAN	Negative regulation of biological process	–	6.20E-09
GO:0050789	P	HOL	Regulation of biological process	1.10E-09	–
GO:0065007	P	HOL	Biological regulation	5.10E-10	–
GO:0022610	P	HOL	Biological adhesion	1.10E-12	–
GO:0009987	P	HOL	Cellular process	1.40E-04	–
GO:0008152	P	HOL	Metabolic process	1.00E-04	–
GO:0040007	P	HOL	Growth	1.70E-46	–
GO:0044085	P	HOL	Cellular component biogenesis	3.70E-13	–
Molecular function					
GO:0005215	F	HOL	Transporter activity	9.20E-05	–
GO:0030528	F	HOL	Transcription regulator activity	6.20E-04	–
GO:0005488	F	HOL	Binding	1.40E-08	–
GO:0030234	F	HOL	Enzyme regulator activity	1.70E-06	–
Cellular component					
GO:0032991	C	HOL/HAN	Macromolecular complex	3.30E-08	9.50E-03
GO:0005623	C	HOL/HAN	Cell	9.00E-10	4.90E-03
GO:0044464	C	HOL/HAN	Cell part	9.00E-10	4.90E-03
GO:0031974	C	HOL	Membrane-enclosed lumen	7.10E-35	–
GO:0044421	C	HOL	Extracellular region part	1.70E-47	–
GO:0005576	C	HOL	Extracellular region	8.80E-25	–
GO:0045202	C	HOL	Synapse	1.20E-11	–
GO:0044456	C	HOL	Synapse part	1.30E-04	–
GO:0043226	C	HOL	Organelle	1.10E-26	–
GO:0044422	C	HOL	Organelle part	6.40E-31	–

et al. 2009) using LifeScope 2.0 (Life Technologies, Carlsbad, CA, USA) and allowing for up to three mismatches. To compare different breeds with Hanwoo, we remapped the genomic sequences of Black Angus (beef cattle) and Holstein (dairy cattle), which had been mapped previously on Btau4 (Stothard et al. 2011), to the bovine reference genome assembly UMD 3.1 following the same procedures that were used to map the Hanwoo sequences (Table 2). Potential duplicates were removed using SAMtools version 0.1.18 (Li et al. 2009).

Identification of putative CNVRs

Putative CNVs were identified across all 29 autosomes and X chromosome using the CNV-seq application (Xie and Tammi 2009). CNV-seq compares two sets of mapped reads in a sliding window. For each comparison, Black Angus versus Hanwoo (BAvsHAN) and Holstein versus Hanwoo (HOLvsHAN), CNV-seq calls regions that show statistically significant read depth discrepancies CNVs. All the mapped reads for each breed were used to generate

Table 2 Coverage of the genomes of Holstein, Black Angus, and Hanwoo cattle

Genome	Total mapped reads	Assembly coverage (%)[a]	Sequence coverage[b]
Holstein	1,235,682,841	98.78	16.54
Black Angus	836,682,891	97.50	9.89
Hanwoo	3,582,289,560	99.07	57.23

[a] Assembly coverage was calculated as the proportion of bases in the genome assembly that were covered by at least one read

[b] Sequence coverage was computed as the average depth of coverage of the bases that were covered by least one read

best-hit format files which were used as an input file for CNV-seq. To detect CNVs across the Hanwoo genome, we ran a customized Perl script with strict customized threshold values ($P = 0.001$ and log2 threshold $= 0.7$) and a window size of 5 to generate a list of CNVs from the best-hit files. In addition, a minimum-window-required setting of 10 was used to ensure that ten consecutive sliding windows showing a significant read depth discrepancy were required for a region to be annotated as a CNVR.

Annotating CNVRs and gene ontology analysis

The gene content of each CNVR was assessed by searching each CNVR sequence coordinate against the Ensembl gene database (release 67) (Flicek et al. 2011). BioMart (Haider et al. 2009) was used to obtain Ensembl protein IDs for the genes that overlapped with a CNVR, and the Ensembl protein IDs were used to perform the Gene Ontology (GO) analysis. The singular enrichment analysis (SEA) tool in agriGO, a GO analysis tool kit and database, was applied to locate enriched GO terms among the list of Ensembl protein IDs. The significance of term enrichments was assessed using Fisher's exact test according to the author's recommendations, and the default multiple comparison correction (Benjamini-Yekutieli method) was applied (Du et al. 2010).

Quantitative real-time PCR validation of CNVs

Quantitative real-time PCR (qRT-PCR) was performed for CNV validation using the 7500 Real Time PCR system and the SDS 1.4 software package (Life Technologies). Primers and probes (Table 3) were designed for 11 genic-CNVRs (CNVRs that completely or partially overlapped with known Ensembl genes) using Custom TaqMan Copy Number Assays (Life Technologies). All the primers were tested by standard curve analysis using a serial dilution of genomic DNA from Hanwoo 27223, with amplification efficiencies above 90 % and no-template control reactions. All reactions (20 μL) were run in triplicate with 1× TaqMan Genotyping Master Mix (Life Technologies), 1 μL of 20× working stock of TaqMan Copy Number Assays (Life Technologies) for target genes, 100 nM of each primer, 250-nM probe for the reference genes, and 10 ng of genomic DNA. Thermal-cycling conditions were as follows: 95 °C for 10 min followed by 40 cycles of 95 °C for 15 s and 60 °C for 60 s. To calculate the for all genes. The average ΔCt for each sample (from three replicates) was calculated after normalizing to BTF3 (chr 20: 8480505–8487056 on Btau 4.0) (Liu et al. 2010; Stothard et al. 2011). The copy number of each target gene was calculated using the CopyCaller software v2.0 (Life Technologies) based on the assumption that there were two copies of the DNA segment in the calibrator animals. The relative quantification analysis of each target was assayed in seven breeds (Hanwoo: 23 samples, including the Hanwoo 27223 sample that was sequenced in this study; Angus: 10 samples; Holstein: 14 samples; Hereford: 4 samples; Charolais: 10 samples; Limousin: 17 samples; and Simmental:14 samples). Overall, the analysis was performed for 92 animals. To calculate the quantitative copy numbers, the Hanwoo 27223 sample was assumed to represent two copies of each of the target DNA segments.

Results and discussion

Genome sequencing and CNV discovery

One individual bull from each of the three breeds, Hanwoo, Black Angus, and Holstein, were sequenced on the SOLiD™ ver. 3.0 chemistry sequencing platform. In this experiment, only Hanwoo was actually sequenced at higher sequence coverage, while raw sequence data of the other two breeds (Stothard et al. 2011) were reprocessed to be compared with the Hanwoo sequence dataset. Because Black Angus and Holstein sequences had been mapped previously to the bovine reference assembly (Btau4), the sequences were remapped to the *Bos taurus* reference sequence assembly (UMD 3.1) so that they could be directly compared with the Hanwoo sequences that were mapped to the UMD 3.1 assembly. The number of properly mapped reads totaled 3,582,289,560, 836,682,891, and 1,235,682,841, accounting for 57.23, 9.89, and 16.54 sequence coverage for Hanwoo, Black Angus, and Holstein, respectively (Table 2). The coverage difference between the breeds was evenly balanced

Table 3 Selected CNVRs and the primers and probes used for qRT-PCR validation

CNV ID	Entrez gene	Log2 ratio	Forward primer	Reverse primer	Probe
Chr5_CNVR_51_97	*ANKS1B*	−1.1966	TGGCATATTTCCCAGCCTAAAAAGA	GTAGTATTTCACCATTCACAACATTTGATGT	ACCTTGCTGCCTATAGCATTAC
Chr7_CNVR_56	*TBC1D9B*	−1.09395	ACTATGCTCAATGCCTTAATAAATCCCT	CAGTGGTTGAGGGTCACATTAATTC	CACAGTGTGGTTTTTCATTCTTTC
Chr19_CNVR_58	*NOS2*	−1.13523	GCTGTCCTGGGAGAACAATGG	CCAAAGGCCTGTGCTGTGA	CTCTGTCATCCTCCTTCCTCTTCC
Chr23_CNVR_18	*BOLA-DQB*	−1.57762	GGCATGCAGAAATCTGGCATTTG	TCCATGCTGGGAATTCATTTTCAGT	CCAACCCCCTTACTGGAAACA
Chr3_CNVR_26	*ITLN2*	−1.7874	TGTGTGTCCCTCTGCAGAAC	GTTGGGCACATGCCAGATG	CTGGGCCTGGATGTCGTAGTAG
Chr4_CNVR_12	*AGBL3*	−1.63786	CTTTTGGTGAACATGAATTCTCAAAGTAGAA	CATAAAAACATTAGGCCAAGGTGTAACTT	ATGTAGCTCATTTTTCCCAATTTT
Chr14_CNVR_30	*MATN2*	−0.97722	AGCCAGCGACTTGTTTCCT	GGTGGATAGCTTAATGGAGTCAACT	ATTTGCTGGAGCACCCAAGTGA
Chr10_CNVR_21	*EPB41L4A*	−0.96458	GATTAGAGTTGAAATGAACCCTTAGACTTGA	CCCTTCTCTTCAATACATTACCTTCCCTAA	TTGAAACAATAAAGGCAGTA TAAATGG
Chr22_CNVR_10	*FBXL2*	−2.28591	GCCTCTCGGCAGTCTTCTC	AAGCATTTTCTAGAGCGTTCTGTCT	CCTGTCCCAGAGGTAAGCTG
Chr18_CNVR_46	*BOSTAUV1R419*	−2.38822	TCCTCTGTATGCATTTTATGTTTCTGACA	CAGCACAGTACCCTGAATCTCTTAT	AATGGCGCAGCACAAACA
Chr6_CNVR_16_17	*PDE5A*	1.94388	CGGACTGGAAACACTAAAAGAAGCT	TCTCTCCCTTGACCCCTTTACTTT	CTGTGACCTTTCCCAAATAAAT
–			GGTAAGTCAGTCACCCTGAGCAA	AAGCTCTGCCCAAACAATGTG	ACCGCTACTAGCAAAC

Eleven candidates from among the Hanwoo or Holstein gain genic-CNVs were selected for validation. The last row shows the primer and probe sequences used for the BTF3 gene that was used for normalization

across all the chromosomes, with no distinct bias in any one chromosome (Fig. 1). The CNV-seq program was run with the strict customized criteria described in the Materials and methods section to identify putative CNVRs. A total of 1,173 and 963 putative CNVRs were identified from the BAvsHAN and HOLvsHAN comparisons, respectively, across the 29 autosomes and X chromosome (Table 4), and the CNVRs are not evenly distributed throughout the genome (Fig. 2). The putative CNVRs represent about 16.7 and 7.8 Mbp or 0.63 and 0.29 % of the UMD 3.1 reference assembly for BAvsHAN and HOLvsHAN, respectively. The average and median sizes of the CNVRs were 10,026 and 10,721 for BAvsHAN and 7,178 and 7,472 for HOLvsHAN. The size range of the putative CNVRs was 5,770–35,104 bp for BAvsHAN and 4,176–22,398 bp for HOLvsHAN (Table 4). Overall, the number of identified CNVR gains in Hanwoo (more copy numbers than either Black Angus or Holstein) was 147 (2,671,398 bp or 16 % of the total CNVRs) from BAvsHAN and 315 (2,938,742 bp or 38 % of the total CNVRs) from HOLvsHAN, indicating that both Black Angus and Holstein have distinctively more abundance of CNVR gains than Hanwoo. Black Angus and Holstein have a longer history of artificial selection than Hanwoo and this has led to a substantial increase in the productivity in these two breeds. The higher abundance of CNVR gains in Black Angus and Holstein may suggest that their stronger and longer selection history compared with that of Hanwoo has resulted in these two breeds having more copies of the genes involved in their increased productivity traits. In particular, the autosomes BTA6 and BTA14 exhibited an outstandingly higher proportion of CNVR gains in both Black Angus (BTA6: 1,010,603 bp or 96 % of the total; in BTA14: 934,753 bp or 99 % of the total) and Holstein (BTA6: 216,538 bp or 82 % of the total; in BTA14: 171,326 bp or 88 % of the total). BTA6 and BTA14 have been extensively interrogated to identify various quantitative trait loci (QTLs) affecting dairy and carcass traits, as well as proposed functional mutation genes such as ATP-binding cassette, subfamily G, member 2 (*ABCG2*) and secreted phosphoprotein 1 (*SPP1*) on BTA6, and diacylglycerol O-acyltransferase 1 (*DGAT1*) on BTA14. Therefore, we suggest that the CNVs have been influenced by the recent intensive artificial selection that has contributed to improved productivity in the economically important dairy and beef traits. However, because we used only one individual representing each breed in this study and because, so far, only a few studies have looked at a possible association between CNVs and economically important traits in cattle, this suggestion requires further investigation. Further research using more individuals from diverse cattle breeds is necessary to elucidate the possible influence of artificial selection on CNVs. The CNVRs that we identified were not evenly distributed across the genome (Fig. 2). Overall, no CNV was detected from BAvsHAN on

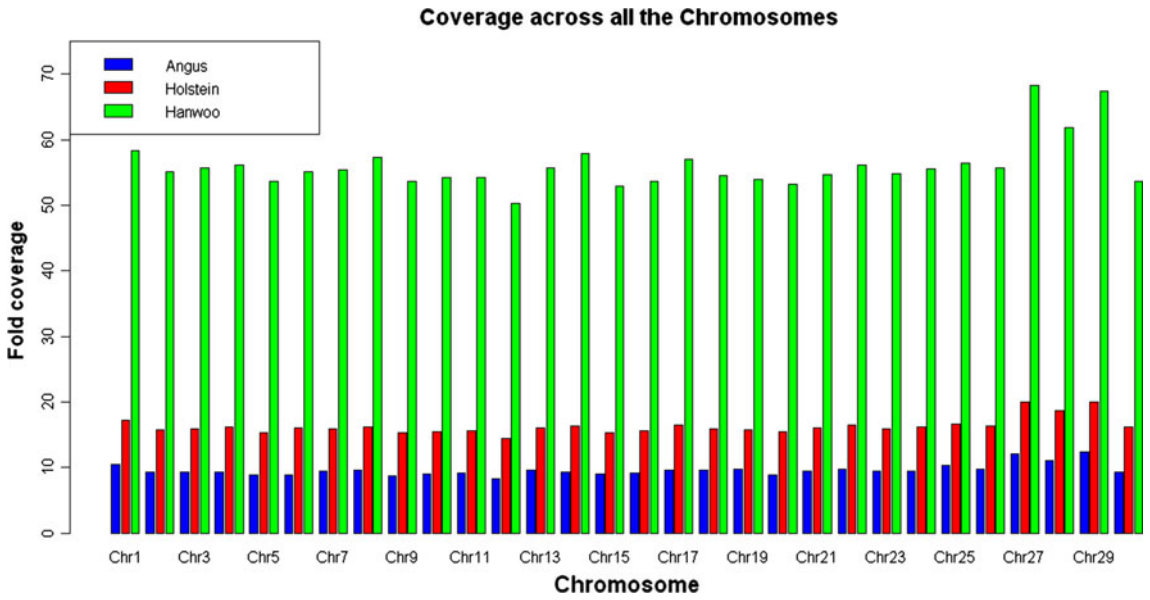

Fig. 1 Distribution of mapped reads in the Hanwoo, Black Angus, and Holstein genome assemblies

BTA1, 3, 4, 5, 7, 17, and 18, and no CNV was detected from HOLvsHAN on BTA9 and 15. As for the proportion of CNVs within each chromosome, CNVs from BAvsHAN were highly enriched (>1 % of each chromosome) on BTA12, 14, 25, 27, 29, and X, while CNVs from HOLvsHAN were highly enriched (>1 %) only on BTA12 (Table 4).

Annotation of the CNVs and gene ontology analysis

GO enrichment analysis of the potential functional roles of the identified CNVs was performed. The Ensembl database searches identified 910 and 436 CNVs that completely or partially overlapped with known genes (genic-CNVs) among CNVs from BAvsHAN and HOLvsHAN, respectively (see the Supplementary Material for details). The genic-CNVs that were annotated with Ensembl protein IDs accounted for 820 and 364 CNVs from BAvsHAN and HOLvsHAN, respectively. The protein IDs were then used to look for statistically significant enriched GO terms. The results indicated that immune system process (GO:0002376), response to stimulus (GO:0050896), and cellular component organization (GO:0016043) were enriched in the set of genic-CNVs from the three breeds in the two comparisons (Tables 1, 5). These enriched terms are well concordant with overrepresented terms for CNVs identified in previous studies on cattle (Stothard et al. 2011), and CNVs particularly enriched for immune response and response to external biotic stimuli have been reported in studies on humans (Conrad et al. 2010; Feuk et al. 2006; Nguyen et al. 2006). In Hanwoo, the overrepresented GO terms for CNVs with higher copy numbers include developmental process (GO:0032502; $P < 0.01$) and multicellular organismal process (GO:0032501; $P < 0.01$) in BAvsHAN (Table 5), and positive regulation of biological process (GO:0048518; $P < 0.01$) and negative regulation of biological process (GO:0048519) in HOLvsHAN (Table 1).

Comparison of cattle CNVRs with previous studies

The two sets of putative CNVRs identified in this study from BAvsHAN and HOLvsHAN were compared with CNVRs detected previously by different platforms, including aCGH and Bovine 50K SNP BeadChip (Hou et al. 2011; Liu et al. 2010). Stothard et al. (2011) observed that substantial numbers of CNVRs overlapped between the multiple cattle CNV studies, even when different detection methodologies and numbers of breeds and individuals had been used to detect the CNVRs. Among the 1,173 CNVRs from BAvsHAN, 341 (6,304,549 bp or 37.7 % of the total CNVRs) and 360 [6,017,588 bp or 36.0 %) of the total CNVRs) overlapped with the CNVR datasets of Liu et al. (2010) and Hou et al. (2011), respectively. Of the overlapping CNVRs, 552 of the 701 were unique CNVRs, accounting for 9,272,845 bp or 55.5 % of the total CNVRs. Among the 963 CNVRs from HOLvsHAN, 350 (3,560,735 bp or 45.8 % of the total) and 408 (3,815,356 bp or 49.0 % of the total) overlapped with the CNVR datasets of Liu et al. (2010) and Hou et al. (2011), respectively. Of the overlapping CNVRs, 539 of the 758 were unique CNVRs, accounting for 5,004,443 bp or 64.3 % of the total CNVRs. In addition, there were distinct differences in the average size of nonoverlapped and overlapped CNVRs. Longer CNVRs tended to have more overlap with CNVRs

Table 4 Distribution and characteristics of the putative CNVRs

BTA	Chr length	Window size	% Length in CNV	CNV length	No. CNV	Mean length	Median length	Max length	Min length
1	158,337,067	1563/817	0.00/0.08	0/130173	0/21	0/6199	0/6121	0/11425	0/4081
2	137,060,424	1542/909	0.49/0.09	673851/124413	51/17	13213/7318	12337/7038	53199/13621	7709/4541
3	121,430,405	1507/900	0.00/0.29	0/355950	0/35	0/10170	0/9675	0/47250	0/4500
4	120,829,699	1543/897	0.00/0.09	0/104848	0/16	0/6553	0/6273	0/11649	0/4481
5	121,191,424	1543/948	0.00/0.40	0/483006	0/62	0/7790	0/9480	0/22752	0/4740
6	119,458,736	1594/906	0.88/0.22	1050455/264106	81/42	12969/6288	14347/7247	60571/19025	7969/4529
7	112,638,659	1511/901	0.00/0.24	0/274535	0/35	0/7844	0/8326	0/18001	0/4501
8	113,384,836	1510/897	0.60/0.21	678754/241050	42/26	16161/9271	15100/9185	47565/34049	7549/4481
9	105,708,250	1573/955	0.16/0.00	172151/0	17/0	10127/0	10612/0	19651/0	7861/0
10	104,305,016	1506/946	0.48/0.25	500750/256366	34/30	14728/8546	12048/7568	73795/29799	7529/4729
11	107,310,763	1456/936	0.86/0.21	920920/226044	67/28	13745/8073	16744/8658	45864/17316	7280/4680
12	91,163,125	1636/1018	2.66/1.82	2423734/1657809	108/149	22442/11126	26994/13743	175870/53445	8180/5089
13	84,240,350	1420/909	0.42/0.16	352870/133944	32/14	11027/9567	11005/6811	51120/51303	7100/4541
14	84,648,390	1498/890	1.12/0.23	944490/194468	74/30	12763/6482	14981/7119	35951/17355	7489/4449
15	85,296,676	1518/955	0.34/0.00	290701/0	25/0	11628/0	11006/0	34915/0	7589/0
16	81,724,687	1472/941	0.86/0.17	702144/142432	63/22	11145/6474	13984/7051	27232/13631	7360/4701
17	75,158,596	1443/882	0.00/0.36	0/273858	0/42	0/6520	0/7055	0/18963	0/4409
18	66,004,023	1403/912	0.00/0.78	0/516648	0/51	0/10130	0/10488	0/49704	0/4560
19	64,057,457	1372/918	0.66/0.29	424634/184058	44/18	9651/10225	11319/9869	21952/25703	6860/4589
20	72,042,655	1517/953	0.32/0.05	233485/37610	21/6	11118/6268	11371/9283	24257/13805	7581/4761
21	71,599,096	1457/914	0.95/0.38	682187/269638	51/36	13376/7490	15289/6855	32033/21937	7281/4569
22	61,435,874	1396/889	0.83/0.18	512332/108353	43/17	11915/6374	13611/6661	42578/12877	6980/4441
23	52,530,062	1450/917	0.98/0.69	514757/360027	35/39	14707/9231	14500/10993	45675/24733	7249/4581
24	62,714,930	1464/908	0.06/0.15	39528/92616	4/9	9882/10291	8052/7945	16104/31780	7320/4994
25	42,904,170	1344/869	1.98/0.27	847392/114596	74/20	11451/5730	13440/6077	55776/9983	6720/4341
26	51,681,464	1430/898	0.48/0.15	249534/75434	22/12	11342/6286	10010/6061	35749/11225	7149/4491
27	45,407,902	1441/717	1.13/0.43	514119/195498	39/30	13183/6517	15121/7340	43201/15395	7201/3581
28	46,312,546	1409/781	0.32/0.50	146443/230916	11/36	13313/6414	11265/7021	21121/17551	7041/3901
29	51,505,224	1338/716	2.04/0.59	1050327/302868	91/51	11542/5939	13380/6444	34119/26134	6689/3580
X	148,823,899	2285/888	1.87/0.29	2785482/429348	144/69	19344/6222	25125/7770	54817/31524	11421/4440
Total	2,660,906,405	1505/896	0.63/0.29	16711040/7780612	1173/963	10026/7178	10721/7472	35104/22398	5770/4176

The CNVs were identified from BAvsHAN and HOLvsHAN comparisons. The data from the comparisons is displayed as BAvsHAN/HOLvsHAN

from previous studies; in BAvsHAN, the average size of CNVRs was 11,978 bp (nonoverlapping) and 16,799 bp (overlapping); in HOLvsHAN, the average size was 6,548 bp (nonoverlapping) and 9,285 bp (overlapping). This phenomenon probably reflects the relatively lower resolution in previous studies which would cause many structural variations to go undetected. The aCGH and Bovine 50K SNP BeadChip that were used suffer from limitations in marker or probe density. Thus, it is likely that a significant number of CNVRs remain undiscovered and further investigations of CNVRs at a fine scale are required.

CNVR candidates potentially affecting traits of interest in cattle

Several genic-CNVRs identified in this study may affect beef or dairy traits of interest. For example, two genic-CNVRs on BTA6 (Chr6_CNVR_16 and Chr6_CNVR_17), which have two times more copies in Holstein (see the Supplementary Material for detail) than in Hanwoo, are likely to be associated with a specific dairy physiological trait. The two CNVRs are closely and consecutively lined up on BTA6 and overlap with the cGMP-specific phosphodiesterase 5A (*PDE5A*) gene. Phosphodiesterase type 5

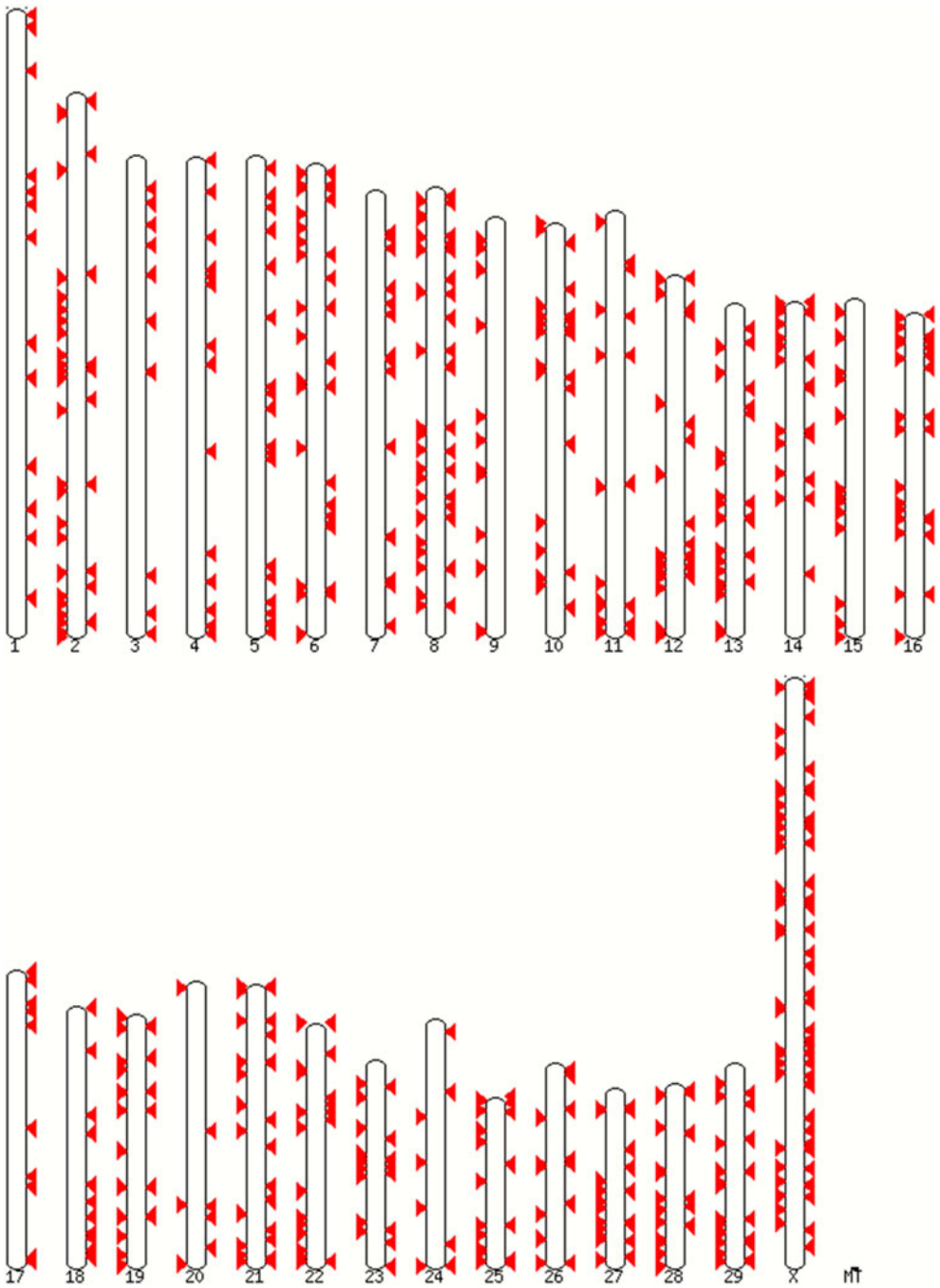

Fig. 2 Distribution of CNVRs across the genome. Putative CNVRs are represented on ideograms of whole chromosomes. *Arrowheads* on the *left* of the chromosome indicate CNVRs from BAvsHAN and *arrowheads* on the *right* represent CNVRs from HOLvsHAN

is widely known as a regulator of nitric oxide (NO)-induced vasodilation, and the vasodilation can be regulated by interactions between NO, PDE5, and guanosine 3′,5′-cyclic monophosphate (cGMP). For proper fetal development and milk production, the uteroplacental blood flow is increased by vasodilation during gestation. It was reported that in sheep, uterine vasodilation is controlled by increases in nitric oxide synthase (NOS) stimulating cGMP and regulating cGMP catabolism by PDE5 (Coppage et al. 2005). The Holstein bull used in this study is one of the top dairy bulls in the world. He is called 'Goldwyn' and his daughters have very high milk production, suggesting that the high copy numbers of *PDE5A* that we detected may confirm its role as a regulator of vasodilation in dairy cattle that are highly developed for high milk production. However, qPCR validation of the *PDE5A* on multiple

Table 5 Gene Ontology terms enriched among the genic-CNVRs from BAvsHAN

GO term	Ontology	Breed of Gain	Description	P-BA	P-HAN
Biological process					
GO:0050789	P	BA	Regulation of biological process	7.90E-27	–
GO:0002376	P	BA/HAN	Immune system process	1.2E-040	8.2e-11
GO:0016043	P	BA/HAN	Cellular component organization	9.4E-140	0.0065
GO:0065007	P	BA	Biological regulation	9.70E-31	–
GO:0022610	P	BA	Biological adhesion	3.50E-24	–
GO:0016265	P	BA	Death	1.70E-100	–
GO:0009987	P	BA	Cellular process	7.10E-14	–
GO:0008152	P	BA	Metabolic process	7.10E-16	–
GO:0051234	P	BA	Establishment of localization	4.00E-57	–
GO:0051179	P	BA	Localization	4.20E-69	–
GO:0050896	P	BA/HAN	Response to stimulus	4.6E-169	9.4e-18
GO:0044085	P	BA	Cellular component biogenesis	3.00E-35	–
GO:0032502	P	HAN	Developmental process	–	1.10E-09
GO:0032501	P	HAN	Multicellular organismal process	–	1.60E-11
Molecular function					
GO:0005215	F	BA	Transporter activity	1.80E-06	–
GO:0030528	F	BA	Transcription regulator activity	2.20E-06	–
GO:0005198	F	BA	Structural molecule activity	4.50E-03	–
GO:0003824	F	BA	Catalytic activity	2.80E-08	–
GO:0005488	F	BA	Binding	2.30E-21	–
GO:0030234	F	BA	Enzyme regulator activity	7.00E-16	–
Cellular component					
GO:0032991	C	BA/HAN	Macromolecular complex	2.8E-024	0.00016
GO:0005623	C	BA	Cell	3.60E-21	–
GO:0044464	C	BA	Cell part	3.60E-21	–
GO:0044421	C	BA	Extracellular region part	1.90E-115	–
GO:0005576	C	BA	Extracellular region	1.20E-48	–
GO:0045202	C	BA	Synapse	1.60E-16	–
GO:0044456	C	BA	Synapse part	6.70E-04	–
GO:0043226	C	BA	Organelle	1.90E-85	–
GO:0044422	C	BA/HAN	Organelle part	4.3E-086	0.0027

individuals failed to show distinctive differences, as estimated (Fig. 3), although there are overall inconclusive higher tendency of signals in Holstein animals than Hanwoo animals. In addition, because we were unable to include the sequenced Black Angus and Holstein samples in the validation, further studies are warranted to clarify the role of the *PDE5A* gene in Holstein. As another example, a CNVR on BTA4 (Chr4_CNVR_12) is within the ATP/GTP binding protein-like 3 (*AGBL3*) gene and it is estimated to have around three times less copies in Holstein than in Hanwoo (see the Supplementary material for details). The Gene ontology term of the *AGBL3* is proteolysis (GO:0006508), and a number of genes related to proteolysis were shown to be decreasingly expressed throughout pregnancy and lactation in the mouse mammary gland (Rudolph et al. 2003). Further qPCR validation on *AGBL3* showed that most Holstein animals do not have any signal observed ($\sim$86 %), as we expected. In spite of the success, the functional role of AGBL3 is still unclear due to the lack of studies to dissect the genetics of the *AGBL3* in Holstein, although it may be speculated that the decline of proteolysis is caused by the lower number of copies in *AGBL3*. In addition, we do not rule out causal roles for other genes encoding proteolytic enzymes.

Validation of CNVs by qRT-PCR on diverse breeds

A total of 11 Hanwoo gain genic-CNVR sequences were selected from the total number of identified CNVRs for qRT-PCR validation on multiple individuals in diverse

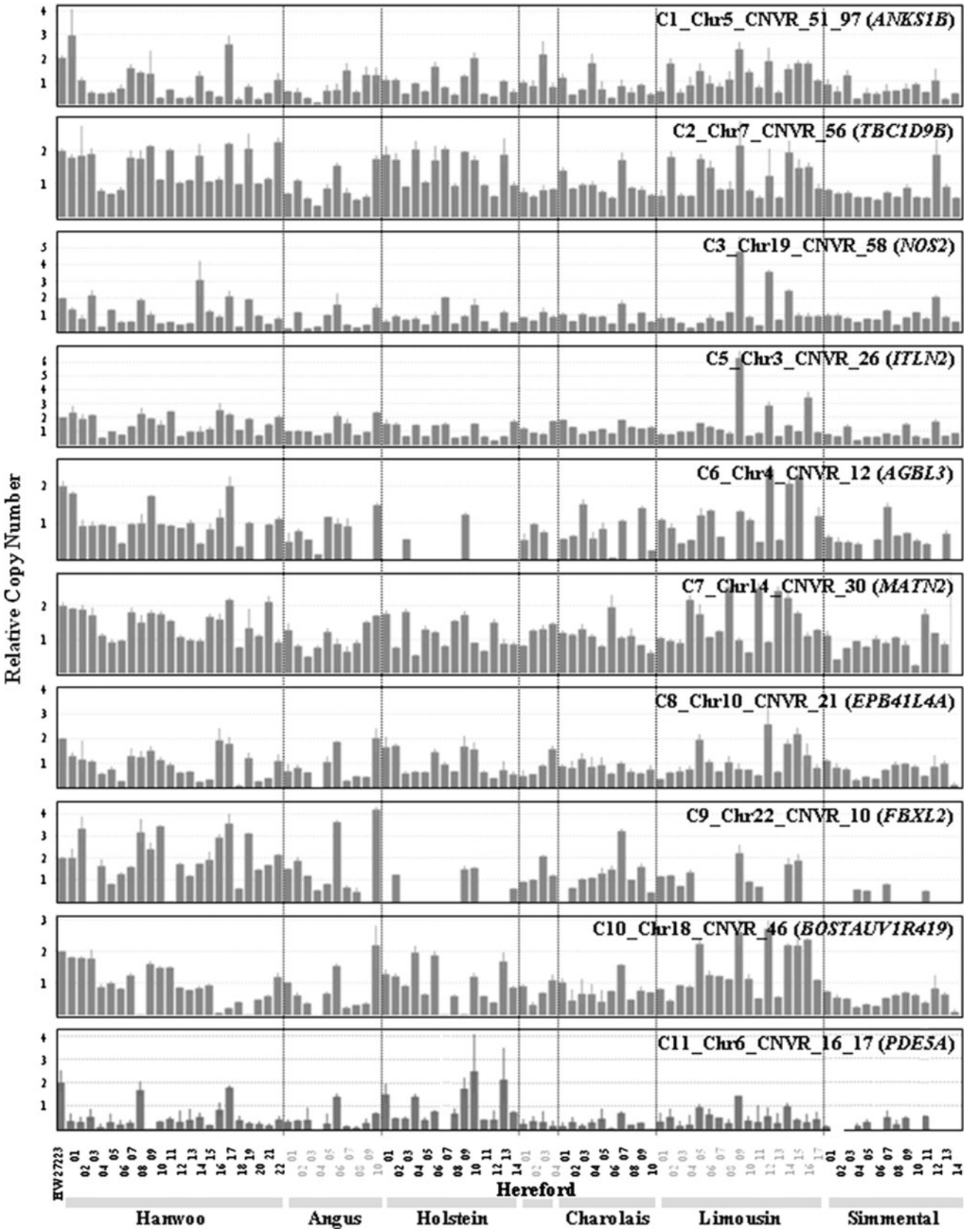

Fig. 3 Validation of selected CNVRs using qRT-PCR on multiple individuals in diverse cattle breeds

cattle breeds (see the Material and methods section for details). Nine of the 11 selected CNVRs successfully exhibited the expected copy number differences when Black Angus/Holstein and Hanwoo were compared (Fig. 3). The one CNVR (C4_Chr23_CNVR_18) that could not be validated overlapped with the *BOLA-DQB* gene, which is in a region of the genome that corresponds to the major histocompatibility complex (MHC). It was therefore difficult to set up a standard curve analysis because of the high similarities in this region. The high success rate of the qRT-PCR validation experiment shows that most of the detected CNVRs in this study are reliable, mainly because of the strict criteria that were set up to detect the CNVRs. While most of the validated CNVs in Hanwoo 27223 closely approximated our expectations, some obvious variations of copy numbers were observed within the larger Hanwoo population. This variable distribution of CNV copy numbers extended to between populations or to within the other populations. Interestingly, the Limousin cattle had higher copy numbers than other breeds of most of the CNVRs, while in Holstein, copy numbers of some CNVRs were distinctly lacking (e.g., C6_Chr4_CNV_12, which overlapped with *AGBL3*, mentioned in the previous subsection, and C9_Chr22_CNV_10, which overlapped with *FBXL2*). The extensive variations within a breed and between breeds are not unexpected because in this study the CNVs were detected based on just one individual bull from each of three breeds. It is well recognized that limited sample size, ethnic diversity of CNV distribution, and technology platforms can substantially affect the identification of potential CNVs, including novel and common CNVs (Pinto et al. 2007). To our knowledge, there is currently a severe lack of CNV research on the cattle genome, particularly with respect to population perspectives. The results of this study clearly support the idea that diverse breeds and multiple individuals should be further investigated to elucidate how and which CNVs can influence economically important traits in cattle. In spite of these limitations, our results provide a reliable and valuable resource for further studies to correlate CNVs and economically important traits in cattle.

Conclusion

In this study, we identified CNVRs throughout the bovine genome by comparing massively parallel Hanwoo sequences with Black Angus and Holstein sequence datasets. A total of 1,173 (16.7 Mbp) and 963 (7.8 Mbp) putative CNVRs were identified from the BAvsHAN and the HOLvsHAN comparisons, respectively. GO enrichment analysis revealed potential functional roles for the identified genic-CNVRs, indicating that response to stimulus, immune system process, and cellular component organization are highly enriched, as observed in previous studies. Substantially more CNV gains were identified in Black Angus and Holstein than were found in Hanwoo. This result may indicate that the intense artificial selection that those two breeds have undergone has led to them having more numbers of copies of CNVs, ultimately affecting the improved productivity traits of interest. To validate the putative CNVRs, 11 selected CNVRs were examined by qRT-PCR. The results showed that 9 of the 11 selected CNVRs exhibited copy number differences, as expected. In conclusion, extensive CNVRs have been successfully identified throughout the bovine genome at relatively higher resolution than in previous studies. We expect that our results can be used as a valuable resource for further investigation of CNVs and may help elucidate how CNVs are correlated with economically important traits in cattle.

Acknowledgments This work was supported by a grant (No. PJ007197) from the BioGreen21 program, Rural Development Administration (RDA), Korea, and through University of Alberta program funds provided to Professor Moore. Xiaoping Liao is funded by the Genome Canada project entitled "Whole Genome Selection Through Genome Wide Imputation in Beef Cattle."

References

Alkan C, Coe BP, Eichler EE (2011) Genome structural variation discovery and genotyping. Nat Rev Genet 12:363–376

Bae JS, Cheong HS, Kim LH, NamGung S, Park TJ, Chun JY, Kim JY, Pasaje CF, Lee JS, Shin HD (2010) Identification of copy number variations and common deletion polymorphisms in cattle. BMC Genomics 11:232

Barendse W, Harrison BE, Bunch RJ, Thomas MB, Turner LB (2009) Genome wide signatures of positive selection: the comparison of independent samples and the identification of regions associated to traits. BMC Genomics 10:178

Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, Aerts J, Andrews TD, Barnes C, Campbell P, Fitzgerald T, Hu M, Ihm CH, Kristiansson K, Macarthur DG, Macdonald JR, Onyiah I, Pang AW, Robson S, Stirrups K, Valsesia A, Walter K, Wei J, Tyler-Smith C, Carter NP, Lee C, Scherer SW, Hurles ME (2010) Origins and functional impact of copy number variation in the human genome. Nature 464:704–712

Coppage KH, Sun X, Baker RS, Clark KE (2005) Expression of phosphodiesterase 5 in maternal and fetal sheep. Am J Obstet Gynecol 193:1005–1010

Du Z, Zhou X, Ling Y, Zhang Z, Su Z (2010) agriGO: a GO analysis toolkit for the agricultural community. Nucleic Acids Res 38:W64–70

Eck SH, Benet-Pages A, Flisikowski K, Meitinger T, Fries R, Strom TM (2009) Whole genome sequencing of a single *Bos taurus* animal for single nucleotide polymorphism discovery. Genome Biol 10:R82

Elsik CG, Tellam RL, Worley KC, Gibbs RA, Muzny DM, Weinstock GM, Adelson DL, Eichler EE, Elnitski L, Guigo R, Hamernik DL, Kappes SM, Lewin HA, Lynn DJ, Nicholas FW, Reymond A, Rijnkels M, Skow LC, Zdobnov EM, Schook L, Womack J, Alioto T, Antonarakis SE, Astashyn A, Chapple CE, Chen HC, Chrast J, Camara F, Ermolaeva O, Henrichsen CN, Hlavina W,

Kapustin Y, Kiryutin B, Kitts P, Kokocinski F, Landrum M, Maglott D, Pruitt K, Sapojnikov V, Searle SM, Solovyev V, Souvorov A, Ucla C, Wyss C, Anzola JM, Gerlach D, Elhaik E, Graur D, Reese JT, Edgar RC, McEwan JC, Payne GM, Raison JM, Junier T, Kriventseva EV, Eyras E, Plass M, Donthu R, Larkin DM, Reecy J, Yang MQ, Chen L, Cheng Z, Chitko-McKown CG, Liu GE, Matukumalli LK, Song J, Zhu B, Bradley DG, Brinkman FS, Lau LP, Whiteside MD, Walker A, Wheeler TT, Casey T, German JB, Lemay DG, Maqbool NJ, Molenaar AJ, Seo S, Stothard P, Baldwin CL, Baxter R, Brinkmeyer-Langford CL, Brown WC, Childers CP, Connelley T, Ellis SA, Fritz K, Glass EJ, Herzig CT, Iivanainen A, Lahmers KK, Bennett AK, Dickens CM, Gilbert JG, Hagen DE, Salih H, Aerts J, Caetano AR, Dalrymple B, Garcia JF, Gill CA, Hiendleder SG, Memili E, Spurlock D, Williams JL, Alexander L, Brownstein MJ, Guan L, Holt RA, Jones SJ, Marra MA, Moore R, Moore SS, Roberts A, Taniguchi M, Waterman RC, Chacko J, Chandrabose MM, Cree A, Dao MD, Dinh HH, Gabisi RA, Hines S, Hume J, Jhangiani SN, Joshi V, Kovar CL, Lewis LR, Liu YS, Lopez J, Morgan MB, Nguyen NB, Okwuonu GO, Ruiz SJ, Santibanez J, Wright RA, Buhay C, Ding Y, Dugan-Rocha S, Herdandez J, Holder M, Sabo A, Egan A, Goodell J, Wilczek-Boney K, Fowler GR, Hitchens ME, Lozado RJ, Moen C, Steffen D, Warren JT, Zhang J, Chiu R, Schein JE, Durbin KJ, Havlak P, Jiang H, Liu Y, Qin X, Ren Y, Shen Y, Song H, Bell SN, Davis C, Johnson AJ, Lee S, Nazareth LV, Patel BM, Pu LL, Vattathil S, Williams RL Jr, Curry S, Hamilton C, Sodergren E, Wheeler DA, Barris W, Bennett GL, Eggen A, Green RD, Harhay GP, Hobbs M, Jann O, Keele JW, Kent MP, Lien S, McKay SD, McWilliam S, Ratnakumar A, Schnabel RD, Smith T, Snelling WM, Sonstegard TS, Stone RT, Sugimoto Y, Takasuga A, Taylor JF, Van Tassell CP, Macneil MD, Abatepaulo AR, Abbey CA, Ahola V, Almeida IG, Amadio AF, Anatriello E, Bahadue SM, Biase FH, Boldt CR, Carroll JA, Carvalho WA, Cervelatti EP, Chacko E, Chapin JE, Cheng Y, Choi J, Colley AJ, de Campos TA, De Donato M, Santos IK, de Oliveira CJ, Deobald H, Devinoy E, Donohue KE, Dovc P, Eberlein A, Fitzsimmons CJ, Franzin AM, Garcia GR, Genini S, Gladney CJ, Grant JR, Greaser ML, Green JA, Hadsell DL, Hakimov HA, Halgren R, Harrow JL, Hart EA, Hastings N, Hernandez M, Hu ZL, Ingham A, Iso-Touru T, Jamis C, Jensen K, Kapetis D, Kerr T, Khalil SS, Khatib H, Kolbehdari D, Kumar CG, Kumar D, Leach R, Lee JC, Li C, Logan KM, Malinverni R, Marques E, Martin WF, Martins NF, Maruyama SR, Mazza R, McLean KL, Medrano JF, Moreno BT, More DD, Muntean CT, Nandakumar HP, Nogueira MF, Olsaker I, Pant SD, Panzitta F, Pastor RC, Poli MA, Poslusny N, Rachagani S, Ranganathan S, Razpet A, Riggs PK, Rincon G, Rodriguez-Osorio N, Rodriguez-Zas SL, Romero NE, Rosenwald A, Sando L, Schmutz SM, Shen L, Sherman L, Southey BR, Lutzow YS, Sweedler JV, Tammen I, Telugu BP, Urbanski JM, Utsunomiya YT, Verschoor CP, Waardenberg AJ, Wang Z, Ward R, Weikard R, Welsh TH Jr, White SN, Wilming LG, Wunderlich KR, Yang J, Zhao FQ (2009) The genome sequence of taurine cattle: a window to ruminant biology and evolution. Science 324:522–528

Fadista J, Thomsen B, Holm LE, Bendixen C (2010) Copy number variation in the bovine genome. BMC Genomics 11:284

Feuk L, Carson AR, Scherer SW (2006) Structural variation in the human genome. Nat Rev Genet 7:85–97

Flicek P, Amode MR, Barrell D, Beal K, Brent S, Chen Y, Clapham P, Coates G, Fairley S, Fitzgerald S, Gordon L, Hendrix M, Hourlier T, Johnson N, Kahari A, Keefe D, Keenan S, Kinsella R, Kokocinski F, Kulesha E, Larsson P, Longden I, McLaren W, Overduin B, Pritchard B, Riat HS, Rios D, Ritchie GR, Ruffier M, Schuster M, Sobral D, Spudich G, Tang YA, Trevanion S, Vandrovcova J, Vilella AJ, White S, Wilder SP, Zadissa A, Zamora J, Aken BL, Birney E, Cunningham F, Dunham I, Durbin R, Fernandez-Suarez XM, Herrero J, Hubbard TJ, Parker A, Proctor G, Vogel J, Searle SM (2011) Ensembl 2011. Nucleic Acids Res 39:D800–806

Gibbs RA, Taylor JF, Van Tassell CP, Barendse W, Eversole KA, Gill CA, Green RD, Hamernik DL, Kappes SM, Lien S, Matukumalli LK, McEwan JC, Nazareth LV, Schnabel RD, Weinstock GM, Wheeler DA, Ajmone-Marsan P, Boettcher PJ, Caetano AR, Garcia JF, Hanotte O, Mariani P, Skow LC, Sonstegard TS, Williams JL, Diallo B, Hailemariam L, Martinez ML, Morris CA, Silva LO, Spelman RJ, Mulatu W, Zhao K, Abbey CA, Agaba M, Araujo FR, Bunch RJ, Burton J, Gorni C, Olivier H, Harrison BE, Luff B, Machado MA, Mwakaya J, Plastow G, Sim W, Smith T, Thomas MB, Valentini A, Williams P, Womack J, Woolliams JA, Liu Y, Qin X, Worley KC, Gao C, Jiang H, Moore SS, Ren Y, Song XZ, Bustamante CD, Hernandez RD, Muzny DM, Patil S, San Lucas A, Fu Q, Kent MP, Vega R, Matukumalli A, McWilliam S, Sclep G, Bryc K, Choi J, Gao H, Grefenstette JJ, Murdoch B, Stella A, Villa-Angulo R, Wright M, Aerts J, Jann O, Negrini R, Goddard ME, Hayes BJ, Bradley DG, Barbosa da Silva M, Lau LP, Liu GE, Lynn DJ, Panzitta F, Dodds KG (2009) Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. Science 324:528–532

Haider S, Ballester B, Smedley D, Zhang J, Rice P, Kasprzyk A (2009) BioMart central portal–unified access to biological data. Nucleic Acids Res 37:W23–27

Hayes BJ, Chamberlain AJ, Maceachern S, Savin K, McPartlan H, MacLeod I, Sethuraman L, Goddard ME (2009) A genome map of divergent artificial selection between *Bos taurus* dairy cattle and *Bos taurus* beef cattle. Anim Genet 40:176–184

Hou Y, Liu GE, Bickhart DM, Cardone MF, Wang K, Kim ES, Matukumalli LK, Ventura M, Song J, VanRaden PM, Sonstegard TS, Van Tassell CP (2011) Genomic characteristics of cattle copy number variations. BMC Genomics 12:127

Jiang L, Liu J, Sun D, Ma P, Ding X, Yu Y, Zhang Q (2010) Genome wide association studies for milk production traits in Chinese Holstein population. PLoS One 5:e13661

Kim KH, Lee JH, Lee SC, Park WY, Oh YG, Kang SW, Ko YD (2005) The optimal TDN levels of concentrates and slaughter age in Hanwoo steers. J Anim Sci Technol 47:731–744

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079

Liu GE, Hou Y, Zhu B, Cardone MF, Jiang L, Cellamare A, Mitra A, Alexander LJ, Coutinho LL, Dell'Aquila ME, Gasbarre LC, Lacalandra G, Li RW, Matukumalli LK, Nonneman D, Regitano LC, Smith TP, Song J, Sonstegard TS, Van Tassell CP, Ventura M, Eichler EE, McDaneld TG, Keele JW (2010) Analysis of copy number variations among diverse cattle breeds. Genome Res 20:693–703

Liu Y, Qin X, Song XZ, Jiang H, Shen Y, Durbin KJ, Lien S, Kent MP, Sodeland M, Ren Y, Zhang L, Sodergren E, Havlak P, Worley KC, Weinstock GM, Gibbs RA (2009) Bos taurus genome assembly. BMC Genomics 10:180

Matukumalli LK, Lawley CT, Schnabel RD, Taylor JF, Allan MF, Heaton MP, O'Connell J, Moore SS, Smith TP, Sonstegard TS, Van Tassell CP (2009) Development and characterization of a high density SNP genotyping assay for cattle. PLoS One 4:e5350

Nguyen DQ, Webber C, Ponting CP (2006) Bias of selection on human copy-number variants. PLoS Genet 2:e20

NIAS (2011) Annual report for Hanwoo genetic evaluation. In: Annual report for livestock improvement in 2010

Pinto D, Marshall C, Feuk L, Scherer SW (2007) Copy-number variation in control population cohorts. Hum Mol Genet 16(2):R168–R173

Rudolph MC, McManaman JL, Hunter L, Phang T, Neville MC (2003) Functional development of the mammary gland: use of expression profiling and trajectory clustering to reveal changes in gene expression during pregnancy, lactation, and involution. J Mammary Gland Biol Neoplasia 8:287–307

Sherman EL, Nkrumah JD, Moore SS (2010) Whole genome single nucleotide polymorphism associations with feed intake and feed efficiency in beef cattle. J Anim Sci 88:16–22

Smith SB, Gill CA, Lunt DK, Brooks MA (2009) Regulation of fat and fatty acid composition in beef cattle. Asian Australas J Anim Sci 22:1225–1233

Stothard P, Choi JW, Basu U, Sumner-Thomson JM, Meng Y, Liao X, Moore SS (2011) Whole genome resequencing of black Angus and Holstein cattle for SNP and CNV discovery. BMC Genomics 12:559

Xie C, Tammi MT (2009) CNV-seq, a new method to detect copy number variation using high-throughput sequencing. BMC Bioinformatics 10:80

Zimin AV, Delcher AL, Florea L, Kelley DR, Schatz MC, Puiu D, Hanrahan F, Pertea G, Van Tassell CP, Sonstegard TS, Marcais G, Roberts M, Subramanian P, Yorke JA, Salzberg SL (2009) A whole-genome assembly of the domestic cow *Bos taurus*. Genome Biol 10:R42