



## Discovery of Gene Sources for Economic Traits in Hanwoo by Whole-genome Resequencing

Younhee Shin, Ho-jin Jung, Myunghee Jung, Seungil Yoo, Sathiyamoorthy Subramaniam, Kesavan Markkandan, Jun-Mo Kang<sup>1</sup>, Rajani Rai<sup>1</sup>, Junhyung Park\*, and Jong-Joo Kim<sup>1,\*</sup>

Codes Division, Insilicogen Inc., Yongin 446-908, Korea

**ABSTRACT:** Hanwoo, a Korean native cattle (*Bos taurus coreana*), has great economic value due to high meat quality. Also, the breed has genetic variations that are associated with production traits such as health, disease resistance, reproduction, growth as well as carcass quality. In this study, next generation sequencing technologies and the availability of an appropriate reference genome were applied to discover a large amount of single nucleotide polymorphisms (SNPs) in ten Hanwoo bulls. Analysis of whole-genome resequencing generated a total of 26.5 Gb data, of which 594,716,859 and 592,990,750 reads covered 98.73% and 93.79% of the bovine reference genomes of UMD 3.1 and Btau 4.6.1, respectively. In total, 2,473,884 and 2,402,997 putative SNPs were discovered, of which 1,095,922 (44.3%) and 982,674 (40.9%) novel SNPs were discovered against UMD3.1 and Btau 4.6.1, respectively. Among the SNPs, the 46,301 (UMD 3.1) and 28,613 SNPs (Btau 4.6.1) that were identified as Hanwoo-specific SNPs were included in the functional genes that may be involved in the mechanisms of milk production, tenderness, juiciness, marbling of Hanwoo beef and yellow hair. Most of the Hanwoo-specific SNPs were identified in the promoter region, suggesting that the SNPs influence differential expression of the regulated genes relative to the relevant traits. In particular, the non-synonymous (ns) SNPs found in CORIN, which is a negative regulator of Agouti, might be a causal variant to determine yellow hair of Hanwoo. Our results will provide abundant genetic sources of variation to characterize Hanwoo genetics and for subsequent breeding. (**Key Words:** Single Nucleotide Polymorphism [SNP], Next Generation Sequencing, Quantitative Trait Locus, Hanwoo, UMD 3.1, Btau 4.6.1)

### INTRODUCTION

Many cattle breeds have been domesticated for milk or beef all over the world since Neolithic age. Ancient populations used cattle breeds of *Bos taurus* (non-humped) or *Bos indicus* (humped), mainly for draught work, clothes, meat and manures. After the industrial revolution, the cattle breeds were intensively selected for meat protein. Based on the geographical distribution, native or local breeds can be classified as for either meat or dairy purposes, based on such characteristics as body size, milking potential, age at puberty,

temperature adaptability, fleshing ability, muscle expression, cutability, and marbling.

Hanwoo (*Bos taurus coreana*) is one of the economically and biologically important ruminants in South Korea. Hence, characterization of cattle genomes may afford an opportunity to gain a deeper understanding of particular breed genetics (Decker et al., 2009). Korean beef consumers prefer Hanwoo meat rather than imported meat, because of its high intramuscular fat (IMF or marbling), flavor and tenderness. However, large volumes of imported beef are gradually occupying Korean beef market, due to the competitive beef price and FTA (free trade agreement) (Paredi et al., 2012). The Hanwoo breed has characteristic yellow hair and doesn't have more than two colors. Besides, Hanwoo is known to be strongly disease-resistant, implying that a distinct level of activation exists in their immune-related pathway (Kim et al., 2013).

Single nucleotide polymorphism (SNP) and copy number

\* Corresponding Authors: Junhyung Park. Tel: +82-31-278-0061, Fax: +82-31-278-0062, E-mail: [jhpark@insilicogen.com](mailto:jhpark@insilicogen.com) / Jong-Joo Kim. Tel: +82-53-810-3027, Fax: +82-53-801-3027, E-mail: [kimjj@ynu.ac.kr](mailto:kimjj@ynu.ac.kr)

<sup>1</sup> School of Biotechnology, Yeungnam University, Gyeongsan 712-749, Korea.

Submitted Sept. 10, 2015; Revised Jan. 6, 2016; Accepted Feb. 16, 2016

variation are the most abundant and common type of genetic variations existing in mammalian genomes. High-throughput genotyping techniques have allowed the use of molecular marker sets to identify the evolution, imprints of domestication and breed development of many species, including humans (Novembre and Ramachandran, 2011), sheep (Kijas et al., 2009), pigs (Groenen et al., 2012), dogs (Vonholdt et al., 2010) and horses (McCue et al., 2012). In Hanwoo, association studies between genes and meat quality have been reported and also genetic information of the cattle genome would provide abundant resources of candidate genes for economic traits and enable characterizing the functionality of the genes (Elsik et al., 2009; Zimin et al., 2009; Lee et al., 2013; Kōks et al., 2014).

In this study, we carried out whole-genome sequencing of Hanwoo bulls using Illumina Genome Analyzer IIX platform to discover SNP variation that may be related to economic traits and to characterize the genetics and molecular biology of Hanwoo, which would enable efficient Hanwoo breeding.

## MATERIALS AND METHODS

### Sample and DNA extraction

Ten Hanwoo bull semen straws were obtained from the Hanwoo Improvement Center of the National Agricultural Cooperative Federation, Seosan, South Korea, to represent Hanwoo breed. Many Hanwoo calves were born between the late 1980s and 2000 from the ten sires. The ten samples were chosen, so as not to be genetically related, and to reflect genetic diversity of the Hanwoo population. The genomic DNA was extracted using standard phenol/chloroform extraction method.

### Sequencing, assembly and annotation

A genomic DNA library was constructed using a paired-end DNA sample prep kit following the manufacturer's protocols (Illumina Inc., San Diego, CA, USA). Purity and yield were confirmed using the 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA).

Sequencing results were analyzed using IncoAnalysis service at Insilicogen, Inc., YongIn, South Korea. For assembly and annotation of the sequence reads, two bovine genome assemblies named UMD3.1 (Zimin et al., 2009) and Btau4.6.1 (Elsik et al., 2009) were used as alternate reference sources. Using the CLC Assembly Cell Software 4.0.6, the pooled reads were aligned separately to the Btau 4.6.1 and UMD 3.1 using the parameters set as: Mismatch cost = 2; gap cost = 3; deletion cost = 3; high-scoring segment pairs (HSP) coverage (length fraction) = 100%; Identity (similarity) = 98%. Sequence data were deposited in the NCBI database. Genomic (pseudo-molecules) sequences representing all Hanwoo chromosomes were constructed and annotated with

corresponding bovine reference genomes, separately. The results were blasted against the UniProtKB database for gene ontology (GO) annotation. Each resultant pool-specific assembly was used to find SNP allele frequency differences and to calculate accumulated distribution per 10-kb interval on each chromosome.

### SNP discovery

For SNP identification, sequence reads were mapped to the two bovine genome assemblies, Btau 4.6.1. and UMD 3.1, using the 'find\_variations' program in CLC Assembly Cell along with stringent parameters as follows: minimum depth = 5; minimum mismatch count (show information when more than a given number of reads is different from the consensus) = 3; limit fraction (Show information when more than a given fraction of reads is different from the consensus)  $\geq 20\%$ . In order to compare our findings with dbSNP database, we carried out the liftover from dbSNP set to Btau 4.6.1 and UMD 3.1, respectively, since the dbSNP set is not yet mapped to the Btau 4.6.1 and UMD 3.1 assemblies. We thus compared our SNPs with dbSNPs (Btau 4.6.1 and UMD 3.1) mapped by SNP location of dbSNP. Heterozygous and homozygous SNPs were determined with 50% cut-off of aligned reads identifying the SNP, but homozygous alleles which be constituted exclusively same bases were applied in order to unravel Hanwoo-specific SNPs. The study for promoter and downstream SNPs was conducted in upstream 2 kb and downstream 1 kb of each gene, respectively.

### Ontology analysis of SNP-containing genes

Ontology analysis was conducted SNP-containing genes through GeneCodis3 programme (<http://genecodis.cnb.csic.es/>). This analysis carried out based on GO, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways, interpro motifs, and Panther pathways.

### Genes related to economic traits

Five traits were focused to characterize genetic variants in the Hanwoo genome; meat quality including juiciness, tenderness, marbling (or IMF), milk and yellow hair. Pathways/groups were collected with entities and relations extracted from searching PubMed database together with MedScan reader v5.0 based on literature-search algorithm. The results were imported into Pathway Studio 9 (Elsevier Research Solutions, Amsterdam, Netherlands), in which pathway-based analysis was performed using mammalian ResNet database and customized bovine ResNet database. Conversion of genes having human Gene ID into bovine Gene ID was performed using HomoloGene, which is a database to automatically detect homologs including orthologs among the annotated genes of 20 completely sequenced eukaryotic genomes.

**Table 1.** Statistics of reference assemblies according to the two bovine genomes

Reference bovine	No. of Mapped reads	Depth	All mapped nucleotide (bp)	Coverage (%)	3+ mapped nucleotide (bp)*	Coverage (%)
UMD 3.1	332,688,239	9.63	2,595,276,825	97.20	2,383,313,321	89.26
Btau 4.6.1	329,753,017	8.58	2,731,967,419	91.97	2,457,655,129	82.74

\* Each nucleotide from more than 3 reads is identical in accordance with each nucleotide of reference genome.

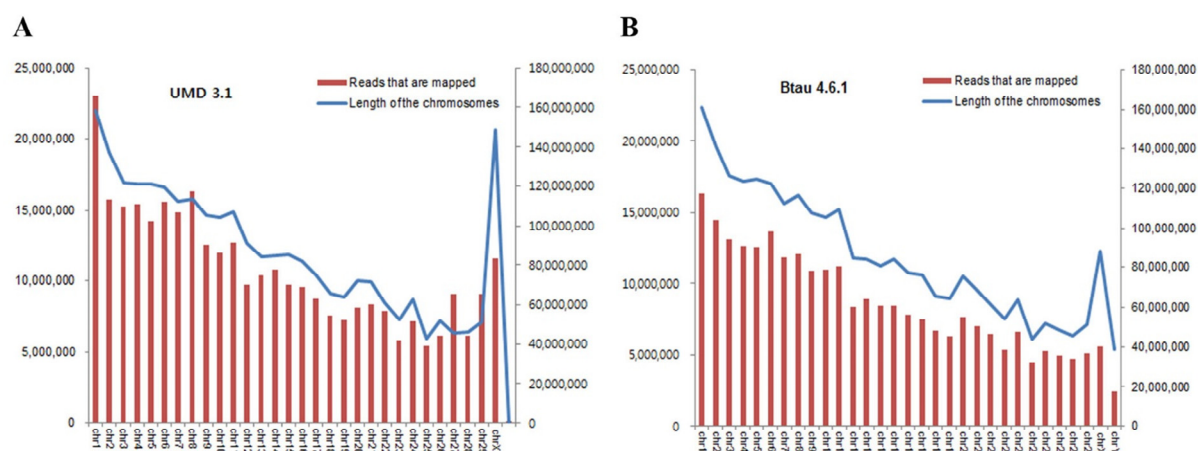
## RESULTS AND DISCUSSION

### Sequencing, assembly and mapping

Sequencing reads from the ten samples were pooled to perform the whole-genome sequencing analysis followed by SNP discovery. Sequencing generated 27 Gbp of high quality sequences on paired-end libraries (76-bp reads and 84-bp reads) with insert size of 500 bp. The genome sequence traces were submitted to the NCBI Short Read Archive (SRA) as accession SRP060306. Preprocessed clean reads were mapped on two reference genome assemblies, giving an average depth of 9.64-fold (UMD 3.1) and 8.58-fold (Btau 4.6.1) coverage, resulting in mapping of 97.20% and 91.97% to the reference genomes, UMD 3.1 and Btau 4.6.1, respectively (Table 1). Overall, the number of reads mapped to each bovine chromosome (Chr) was unbiased, except for Chr1, Chr8, Chr27, and Chr29 in the UMD 3.1 reference genome (Figure 1). The coverage of the mapped reads was fairly even on individual chromosomes (Supplementary Figure 1). This result is consistent with the results of whole-genome resequencing of Japanese native cattle Kuchinoshima-Ushi (Kawahara-Miki et al., 2011).

In the previous study, whole-genome sequencing revealed 98% next generation sequencing (NGS) reads (at 15.8-fold read depth) of Fleckvieh genome and 93% (at 7.4-fold read depth) of Kuchinoshima-Ushi genome were mapped to the NCBI bovine genome (Btau 4.6.1). The relatively lower genome coverage in this study may be caused by two reasons. Firstly, although mapping with

default parameter (identity 80% and HSP coverage 50%) increased mapping coverage of 98.73% (UMD 3.1) and 93.79% (Btau 4.6.1), we produced reliable sequencing data with the optimized parameter, which caused the diminished coverage. Secondly, compared to Fleckvieh breed and Kuchinoshima-Ushi breed, Hanwoo is evolutionally more distant to the breed of the reference genome (Hereford). The different reference genome is attributable to different coverage, because Hanwoo genome had more similarity with the UMD 3.1 reference genome than the Btau 4.1.6 genome. Thus, the UMD 3.1 version was used for further analysis of Hanwoo genome. Substantially, UMD 3.1 and Btau 4.1.6 genome assemblies are significantly different in terms of several aspects (Kawahara-Miki et al., 2011). Direct alignment of two genomes against each other showed diverse discrepancies in ten chromosomes and most remarkable differences were observed in X chromosome (Chr X), though Btau was homologous to the human Chr X (Zimin et al., 2009). For another parameter of assembly completeness, the number of full-length cow mRNA mapping to the two reference genomes was different in accordance with levels of coverage, which reveals that UMD contains more mRNAs than Btau (Kawahara-Miki et al., 2011). Therefore, our suggestion is critical because, obviously, the mapped reads generated by the reference assembly method can vary depending on different reference genomes. Approximately 2.80% and 8.03% of the Hanwoo sequence reads were unmapped to each reference assembly (UMD 3.1 and Btau 4.6.1), possibly due to complexity of the breed or individual



**Figure 1.** The number of mapped reads according to each chromosome. In each chromosome, the reads that were mapped to the reference genome sequence of (A) UMD 3.1 and Btau 4.6.1 (B) is shown in red column, respectively. Length of the chromosomes is marked as blue line.

**Table 2.** Genotypes and variation types of Hanwoo SNPs

Reference	Total SNP	Genotypes		Variation types		
		Homozygous SNP	Heterozygous SNP	Substitution	Insertion	Deletion
UMD 3.1	2,473,884 (100%)	239,370 (9.7%)	2,234,514 (90.3%)	2,400,354 (97.0%)	55,100 (2.2%)	18,430 (0.7%)
Btau 4.6.1	2,402,997 (100%)	240,087 (10.0%)	2,162,910 (90.0%)	2,332,070 (97.0%)	45,792 (1.9%)	25,135 (1.0%)
		----- Common SNPs -----		----- Novel SNPs -----		
UMD 3.1	2,473,884	1,377,962 (55.7%)		1,095,922 (44.3%)		
Btau 4.6.1	2,402,997	1,420,323 (59.1%)		982,674 (40.9%)		

SNP, single nucleotide polymorphism.

uniqueness, sequencing errors, or the current incompleteness of the Hanwoo sequencing.

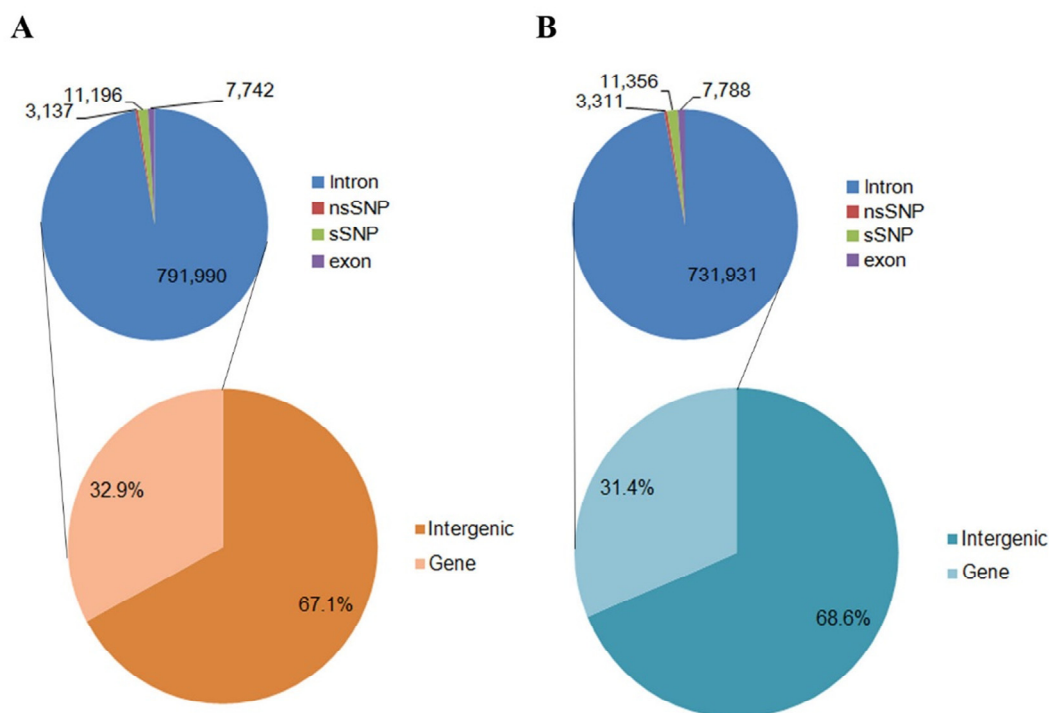
### High-quality SNP detection and novel SNPs

In order to reduce the false-discovery rate, stringent criteria for SNP calling were applied (materials and methods). As a result, 2,473,884 (UMD 3.1) and 2,402,997 (Btau 4.6.1) SNPs were identified. Among the SNPs, 2,234,514 (90.3%) were heterozygous and 239,370 (9.7%) were homozygous for UMD 3.1 (cut-off: 50% alleles), whereas 2,162,910 (90.0%) were heterozygous and 240,087 (10.0%) were homozygous for Btau 4.1.6 (Table 2). This result suggests that substantial variations exist between the ten individuals. Especially, 791,990 (32.9%) and 731,931 (31.4%) SNPs that were localized in genic regions. Synonymous SNPs (sSNPs) and nonsynonymous SNPs (nsSNPs) in exon region was

11,196 (1.38%) and 3,137 (0.39%) for UMD 3.1, while 11,356 (1.50%) and 3,311 (0.44%) for Btau 4.6.1 were defined in the exon regions for the sSNPs and nsSNPs, respectively (Figure 2). To discover novel SNPs, all the SNPs were compared to the locations where SNPs were already known from NCBI dbSNP database (Supplementary Figure 2). Amongst, 1,095,922 (44.3%) and 982,674 (40.9%) SNPs were novel for UMD 3.1 and Btau 4.1.6, respectively, (Table 2). Considering SNP density, 927 SNPs for UMD 3.1 and 809 SNPs for Btau 4.6.1 were discovered per 1 Mb.

### Hanwoo-specific SNPs

Hanwoo-specific SNPs were determined by comparison of genomic position of each SNP with those of NCBI dbSNPs. We isolated 100% homozygous SNPs for discovering SNPs associated with Hanwoo-specific



**Figure 2.** Functional class membership of identified SNPs. Based on the (A) UMD 3.1 and (B) Btau 4.6.1, SNPs on the Hanwoo pseudomolecules were classified as gene and intergenic region, and locations within the gene were annotated. The number of SNPs is shown in each class. SNPs, single nucleotide polymorphisms.

phenotypes. As a result, 46,301 SNPs (1.9%) for UMD 3.1 and 28,613 SNPs (1.2%) for Btau 4.6.1 were identified, respectively, (Table 3). Among the SNPs, 46,301 Hanwoo-specific SNPs that were based on UMD 3.1 were chosen for further analysis.

**Functional classification of SNP-containing genes**

Non-synonymous SNPs (nsSNPs) may lead to phenotype change. To elucidate the effects of nsSNPs in Hanwoo, GO analysis was conducted in nsSNP-containing genes. A set of 2,329 nsSNP-containing genes was significantly enriched in GO including olfactory receptor activity (molecular function), zinc ion binding, G protein-coupled receptors (GPCR), rhodopsin-like superfamily (Figure 3). Majority of these genes are involved in signal transduction pathways that transfer the signals from outside of a cell through cell surface receptors to the inside of the cell including hormone and neurotransmitter. A similar result was observed not only in cow but also in human, implying that the genetic variation in

**Table 3.** Hanwoo-specific SNPs among homozygous SNPs

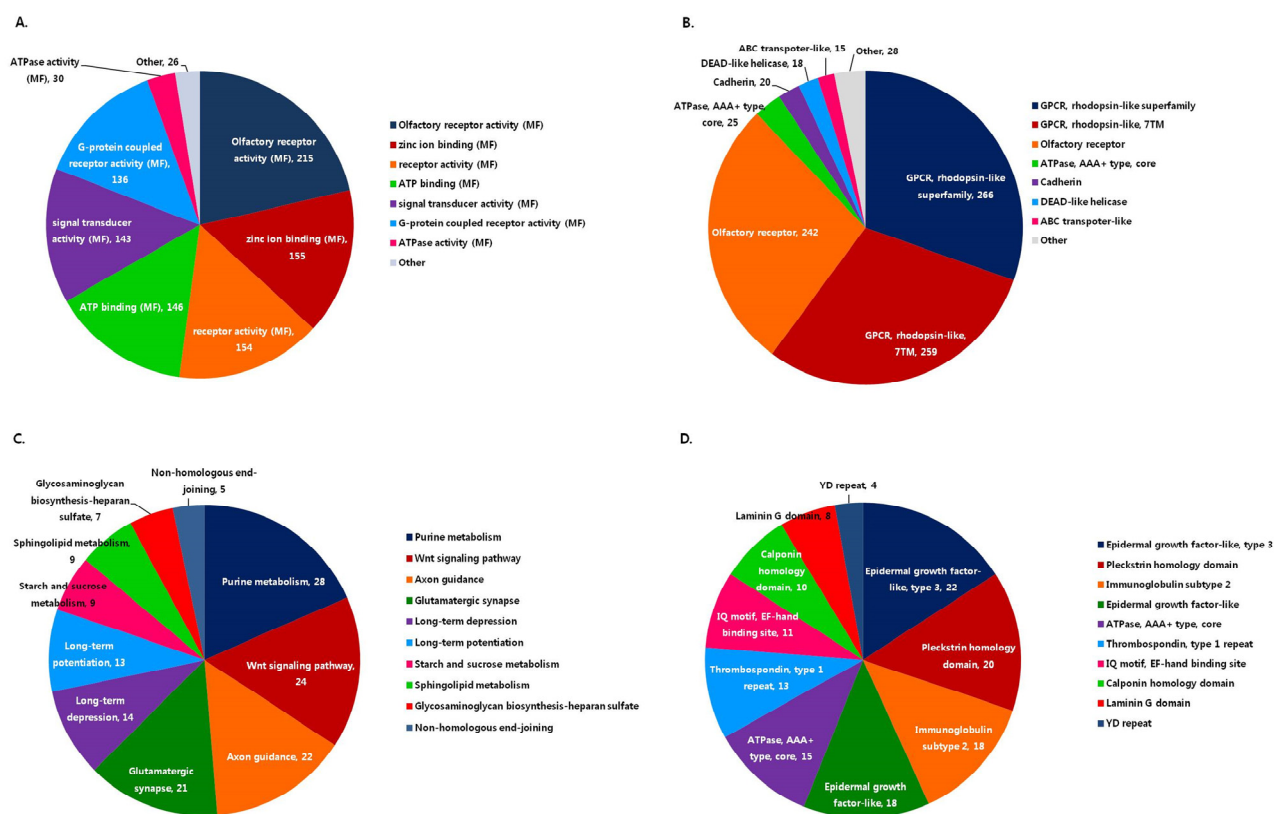
Reference	Total SNP	100% the same alleles	
		Hanwoo-specific SNPs	Common SNPs (found in dbSNP)
UMD 3.1	2,473,884	46,301 (1.9%)	193,001 (7.8%)
Btau 4.6.1	2,402,997	28,613 (1.2%)	211,460 (8.8%)

SNP, single nucleotide polymorphism.

signal transduction genes may be one of main causes of individual phenotypic diversity (Nusse and Varmus, 2012).

**Hanwoo-specific SNP containing genes**

A total of 5,022 genes were characterized as Hanwoo-specific SNP containing genes. Those genes were enriched in purine metabolism, Wnt signaling pathway, axon guidance, and glutamatergic synapse pathway (Figure 3). Axon guidance pathway represents a key stage in the formation of neuronal network, and glutamatergic synapse pathway is also the major excitatory neurotransmission in the mammalian



**Figure 3.** The functional classifications of genes containing nsSNP (using UMD 3.1) and Hanwoo-specific containing genes. (A) Genes with nsSNPs were enriched for functions associated with olfactory receptor activity and zinc ion binding in molecular function category of GO (B) In interpro domains showed enrichment for GPCR, rhodopsin-like superfamily, a group of signal transduction pathways in gene with nsSNPs (C) KEGG classification of Hanwoo-specific SNP containing genes. Hanwoo-specific SNP containing genes were observed in immune diversity related gene groups. Purine metabolism, Wnt signaling pathway and Axon guidance show significant enrichment in KEGG pathway. (D) Epidermal growth factor-like domain (EGF-like domain), Pleckstrin homology domain (PH domain) and immunoglobulin subtype 2 have immune activation function from external stimuli. These domains show features of Hanwoo-specific SNPs containing genes as contributors to immune diversity. SNPs, single nucleotide polymorphisms; GO, gene ontology; GPCR, G protein-coupled receptors; KEGG, Kyoto Encyclopedia of Genes and Genomes.

central nervous system. These neuronal pathways can modulate immune-related system by releasing neurotransmitter (Kim et al., 2009). Especially, 581 genes that have Hanwoo-specific SNPs at least five are associated with immune activation by external stimuli, through epidermal growth factor-like domain (EGF-like domain), pleckstrin homology domain (PH domain), immunoglobulin subtype 2, and thrombospondin (Figure 3) (Chen et al., 2015). The Hanwoo-specific SNP containing genes might contribute to diversity of immune system.

#### SNPs-containing genes for economic traits

We collected the genes that are putatively associated with five traits of economic importance, i.e. milk production, marbling (or IMF), juiciness, tenderness, and yellow hair, by using MedScan Reader v5.0 (Ariadne Genomics, USA). The MedScan Reader is a natural processing language engine with text-mining algorithm that automatically extracts and displays information of interest from the biomedical literature in the PubMed and Google Scholar database. Using PubMed and Google Scholar server, we collected the research articles related with the traits of interest using relevant keywords. The information of entities and relations was extracted from the selected papers, but entities having no relations were used for this analysis. To obtain the accurate results, we checked whether the genes were involved in physiological processes associated with the five traits (Supplementary Tables 1 and Table 2).

#### SNPs-containing genes for yellow hair follicles

The pigmentation in mammals is based on the presence or absence of melanin in hair or skin. There are two types of pigments, i.e. eumelanins and pheomelanins which are responsible for black/brown and yellow/red colors, respectively. The major genes affecting the synthesis of these pigments are tyrosinase-related protein 1 (*TYRP1*), dopachrometautomerase (*DCT*), melanocortin receptor 1 (*MC1R*), tyrosinase (*TYP*), agouti signaling protein (*ASIP*) and corin, serine peptidase (*CORIN*). Especially, in the rodents, the change between black and yellow pigment is arbitrated by the interaction between *MC1R* and its antagonist *ASIP*. We analyzed the genetic variations of the genes associated with hair color traits including above genes using MedScan reader. It has been revealed that the T-to-C substitution in *MC1R* gene caused amino acid change in position 99, Leu to Pro in Angus black cattle. This shift allowed the expression of the constitutive activation of *MC1R* receptor, leading to a decrease in the Agouti concentration, and finally to determine the black color in Angus. Whereas, in the absence of Corin, a negative regulator of Agouti, Agouti activity is prolonged and determines the yellow color (Van den Bossche et al., 2012).

Hanwoo has yellow hair follicles over the whole body, where no SNP in *MC1R* was observed, a set of 227 SNPs were identified by including the nsSNP in *CORIN*. The nsSNP (NW\_003103974:4359766C>T) of *CORIN* leads to change of amino acid Glu (negative charge) >Lys (positive charge) in position 826. In addition, six SNPs were found in upstream promoter region of the gene, implying that those variations may affect *CORIN* gene expression. In previous studies, variations in *MC1R* and microphthalmia-associated transcription factor (*MITF*) were reported to be related with cattle follicles. However, the polymorphism of *CORIN* in Hanwoo was reported for the first time in this study. *MITF*, which is a master regulator of melanocyte development, is known to regulate expression of *TYP*, *TYRP1*, and *DCT*, by direct binding to cis-elements in their promoter sequences. *MITF* mutations associated with a dominant white phenotype were not found in Hanwoo. In other cattle breeds, there are many SNPs reported in color related genes such as *TYP*, *TYRP1*, *ASIP*, *DCT*, melanophilin (*MLPH*), c-Kit (*KIT*), c-Kit ligand (*KITLG*), which caused phenotypic variations in hair follicles (Van den Bossche et al., 2012). Also, many nsSNPs were found in those genes in Hanwoo. However, it is not likely that the SNPs cause significant changes in protein structure enabling it to cause Hanwoo hair coloring (Table 4).

#### SNPs-containing genes for beef palatability

Tenderness, juiciness, and marbling are key components of beef palatability. A set of genes were determined for tenderness (94 genes) and juiciness (121 genes) (Supplementary Table 2). Many SNPs from our study were previously reported in other cattle breeds (Table 4). For example, Leptin (*LEP*) that is related with tenderness contains one nsSNP (Cys>Arg) of 269 SNPs. Leptin receptor (*LEPR*) genes were associated with fat content in carcass meat in Nellore, Holstein-Friesian, Angus, Charolais, Hereford, and Simmental (Pinto et al., 2011). In this study, many nsSNPs were also observed in *LEP* and *LEPR* genes in Hanwoo (Table 4). Based on a mutation similar to the mKIAA0657 protein (titin, *TTN*), we hypothesized that the pressure of the strong selection for high marbling in Japanese Black breed increases the frequencies of the T, among T and A alleles at the *TTN* g.231054C > T, *RPL27A* g.3109537C > T and ankirin 2 genes c.\*188G > A SNPs, respectively (Watanabe et al., 2011). Many nsSNPs were also found in meat quality related genes such as ankyrin 1 (*ANK1*), stearoyl-CoA desaturase (*SCD*), microtubule-associated protein 1B (*MAP1B*), NCK-associated protein 1-like (*NCKAP1L*), insulin-like growth factor binding protein 5 (*IGFBP5*) and mitochondrial poly (A) polymerase (*MTPAP*) (Table 4). Some of the SNPs were localized in promoter region, which can affect gene expression of the genes

**Table 4.** Summary of genes containing SNPs associated with traits in other breeds of cattle

Traits	Gene	nsSNP (#)	Promoter (#)	hwSNP (#)	Reference (PubMed id)
Yellow hair	<i>CORIN</i>	1	6	3	21098273
Milk	<i>STAT5A</i>	0	3	1	18218766
	<i>ABCG2</i>	1	2	0	15998908
	<i>NOD2</i>	0	1	0	21310019
Marbling	<i>HMCN1</i>	3	0	5	24476087
	<i>EGF</i>	2	2	1	19855844
	<i>SLC2A2</i>	0	6	0	PMC2726579
	<i>ASAP1</i>	0	2	28	22682072
	<i>ME1</i>	0	3	5	22221029
	<i>RUNX1</i>	0	2	9	21912507
Tenderness /juiciness	<i>ANK1</i>	2	2	1	21159195
	<i>LEP</i>	1	1	1	21159195, 14652360
	<i>LEPR</i>	1	3	1	20300864
	<i>TTN</i>	12	3	3	21251324
	<i>SCD</i>	1	4	0	20163629
	<i>MAP1B</i>	2	0	1	22921614
	<i>NCKAP1L</i>	2	3	0	23424621
	<i>IGFBP5</i>	1	1	1	19941052
	<i>MTPAP</i>	1	4	2	22297614

SNPs, single nucleotide polymorphisms.

(Supplementary Table 2). Calcium activated neural protease (CAPN) on BTA29 containing a quantitative trait locus for meat tenderness activated by increased muscle calcium is responsible for meat tenderness through proteolysis. Calpastatin (CAST) is an inhibitory protein of CAPN and its inhibitory activity is an indicator of tenderness (Carnagey et al., 2008). The calpain-calpastatin enzyme complex regulates the rate of protein degradation for tenderness (Cafe et al., 2010). In our study, a total of 3,935 and 5,144 SNPs, including 58 and 82 Hanwoo-specific SNPs, were found in UMD and Btau, respectively (Supplementary Table 2). Particularly, the Hanwoo-specific SNPs will be a valuable genetic resource to characterize beef quality of Hanwoo.

### SNPs-containing genes for marbling

Marbling (or IMF) is a multi-factorial trait, i.e. determined by multiple quantitative trait loci (QTLs), interacted with environmental factors. Marbling is one of the important indicators for palatability of beef (Lim et al., 2011). Certainly, IMF accumulation is attributable to multiple pathways regulating the balance between uptake, synthesis and degradation of triglycerides, rather than the involvement of a single pathway (Canovas et al., 2010). Marbling can be redistributed from other fat deposits including subcutaneous fat and visceral fat, which gave rise to inefficient meat production. Thus, genomic studies on marbling have been performed to discover marbling-specific genes including ribosomal protein L27a (*RPL27A*), sphingosine-1-phosphate receptor 1 (*S1PR1*), ADAM metalloproteinase with

thrombospondin type 1 motif 4 (*ADAMTS4*), hormone-sensitive lipase (*LIPE*), peroxisome proliferator-activated receptor alpha (*PPAR*), CCAAT/enhancer binding protein (*C/EBP*), alpha (*CEBPA*), runt-related transcription factor 1 (*RUNX1*), and insulin (*INS*) in animal breeding researches (Lee et al., 2008). In this study, *LEP*, calpain 1, ( $\mu$ /I) large subunit (*CAPN1*), *CAST*, adiponectin, *C1Q* and collagen domain containing (*ADIPOQ*), peroxisome proliferator-activated receptor gamma (*PPARG*), insulin-like growth factor 1 (somatomedin C) (*IGF1*), fibronectin ED-B+ region (*FN1*), sorbin and SH3 domain containing 1 (*SORBS1*), myosin, light chain 12A, regulatory, non-sarcomeric (*MYL12B*), SH3-domain kinase binding protein 1 (*SH3KBP1*), acyl-CoA synthetase long-chain family member 1 (*ACSL1*), myosin light chain kinase (*MYLK*), ATP-binding cassette, sub-family A, member 1 (*ABCA1*), insulin receptor (*INSR*), protein kinase C, alpha (*PRKCA*) and similar to Retinoblastoma-associated protein (*RB1*) were newly identified as marbling-related (Supplementary Table 2). As an inhibitor to the growth of muscle tissue, myostatin is involved in muscle marbling during fattening period by inhibiting myogenesis and promoting adipogenesis (Shibata et al., 2006). A disintegrin and metalloproteinase with thrombospondin motifs 4 (*ADAMTS4*), which takes part in connective tissue degradation, is expected to be involved in a pivotal biological pathway for building up marbling. Also, as regulators of *ADAMTS4*, *IL-17A*, *TNF $\alpha$* , and *TGF $\beta$ 1* were connected with marbling. Peroxisome proliferator-activated receptor  $\gamma$  (*PPAR $\gamma$* ), *C/EBP $\alpha$* , and *RUNX1* were

significantly and positively related to protein increase in intramuscular fat content (Lee et al., 2010). In this study, we identified a set of 232 genes and found 22 nsSNP in hemicentin 1 (*HMCN1*, 437 sSNPs, 3 nsSNPs) and EGF (124 sSNPs, 2 nsSNPs). Also, significant SNPs were found in the promoter region of *RUNX1*, *ME1*, and *ASAP1*. A previous study revealed that *RUNX1* was significantly up-regulated as intramuscular fat content increased (Bagnato et al., 2008). Mutations in *ME1* and *ASAP1* gene were associated with meat quality traits in Angus and Nelore, respectively (Lee et al., 2013). Then Hanwoo-specific SNPs in above genes are a valuable genetic source to improve meat quality (Table 4).

### SNPs-containing genes for milk production

A set of 202 genes were found for milk production (Supplementary Table 2). Hanwoo is known to have low milking ability, and there are few studies about genes for milk production in Hanwoo. A list of genes for milking metabolism are growth hormone (GH), such as *GH1* (1 sSNP), *IGF1* (39 sSNPs), and *IGF2* (2 sSNPs), prolactin (*PRL*, 3 sSNPs), *LARGE* like-glycosyltransferase (687 sSNPs), diglyceride acyltransferase (*DGAT1*, 2 sSNPs), alpha-lactalbumin (*LALBA*, 4 sSNPs), lactoferrin (*LTF*, 53 sSNPs) and signal transducer and activator of transcription 5A (*STAT5A*, 13sSNPs) (Supplementary Table 2). *GH1* has significant positive effect on milk yield throughout lactation (Chanchai et al., 2010). The short isoform of prolactin (*PRL*) receptor has a suppressive role in *PRL* signaling to production of milk protein (Berlanga et al., 1997). The *DGAT1* gene, involved in triglyceride synthesis, is suggested to be the causative gene for the QTL affecting milk fat on bovine chromosome 14 (Ron et al., 2007). The *LALBA* is necessary for lactose synthesis, which is important for milk production as an osmotic regulator of milk secretion. The *STAT5A* gene, as a member of the interferon-tau, plays major role in milk production (Khatib et al., 2008). Although all of these genes are known to be associated with milk production, the genes with sSNPs are required for additional investigation within diverse genomic sources encompassing several cattle breeds. Moreover, many nsSNPs were found in a set of 26 genes including *FAT* tumor suppressor homolog 3 (*FAT3*, 903 SNPs), family with sequence similarity 13, member A (*FAM13A*, 94 SNPs), ATP-binding cassette, sub-family G (*ABCG2*, 92 SNPs) and SREBP cleavage activating protein (*SCAP*, 69 SNPs), in which the relationship of the genes with milk production was not previously reported (Supplementary Table 2).

### CONCLUSIONS

In this study, we investigated the genetic variations related with economically important traits by using whole-

genome resequencing of Korean native cattle, Hanwoo. A total of 2.47 million SNPs were identified, of which more than 1.1 million SNPs (44%) were found to be novel. Among the SNPs, many of the nsSNPs were in the genes that were reported as candidates for QTL affecting economically important traits such as meat quality, milk production, and exterior traits in other cattle breeds. Since Hanwoo has been selected for growth, disease resistance and beef quality for the last several decades, there should be genetic variations between Hanwoo individuals, e.g. SNPs in the genes that are functionally related to the respectively traits, which was supported in this study. In particular, a set of 136 Hanwoo-specific genes including 446 SNPs were found in the genes related to the above traits. Among the genes, *CAPN1*, *CAST*, *LEP*, *ANK1*, *ADIPOQ*, *IGF1*, *NR3C1*, *PPARG*, and *FNI* were included, which were known to be related with marbling, tenderness and juiciness. We identified Hanwoo specific genes, for yellow hair, meat quality, and milk production. Our results will provide valuable genetic resources for genome study of other cattle breeds as well as Hanwoo, and will facilitate the characterization of the genetic architecture of economically important traits in cattle.

### CONFLICT OF INTEREST

We certify that there is no conflict of interest with any financial organization regarding the material discussed in the manuscript.

### ACKNOWLEDGMENTS

This work was supported by a grant (715003-07) from the Research Center from Production Management and Technical Development for High Quality Livestock Products through Agriculture, Food and Rural Affairs Research Center Support Program (2016), and the Technology Development Program for Agriculture and Forestry (No. 311016-3), Ministry of Agriculture, Food, and Rural Affairs (2013), Republic of Korea.

### REFERENCES

- Bagnato, A., F. Schiavini, A. Rossoni, C. Maltecca, M. Dolezal, I. Medugorac, J. Sölkner, V. Russo, L. Fontanesi, and A. Friedmann, et al. 2008. Quantitative trait loci affecting milk yield and protein percentage in a three-country Brown Swiss population. *J. Dairy Sci.* 91:767-783.
- Berlanga, J. J., J. P. Garcia-Ruiz, M. Perrot-Appanat, P. A. Kelly, and M. Ederly. 1997. The short form of the prolactin (*PRL*) receptor silences *PRL* induction of the beta-casein gene promoter. *Mol. Endocrinol.* 11:1449-1457.
- Cafe, L. M., B. L. McIntyre, D. L. Robinson, G. H. Geesink, W. Barendse, D. W. Pethick, J. M. Thompson, and P. L. Greenwood. 2010. Production and processing studies on calpain-system gene



- markers for tenderness in Brahman cattle: 2. Objective meat quality. *J. Anim. Sci.* 88:3059-3069.
- Canovas, A., R. Quintanilla, M. Amills, and R. N. Pena, 2010. Muscle transcriptomic profiles in pigs with divergent phenotypes for fatness traits. *BMC Genomics* 11:372.
- Carnagey, K. M., E. J. Huff-Lonergan, S. M. Lonergan, A. Trenkle, R. L. Horst, and D. C. Beitz. 2008. Use of 25-hydroxyvitamin D3 and dietary calcium to improve tenderness of beef from the round of beef cows. *J. Anim. Sci.* 86:1637-1648.
- Chanchai, W., S. Chanpongsang, and N. Chaiyabutr. 2010. Effects of misty-fan cooling and supplemental rbST on rumen function and milk production of crossbred Holstein cattle during early, mid and late lactation in a tropical environment. *Anim. Sci. J.* 81:230-239.
- Chen, N., W. Fang, J. Zhan, S. Hong, Y. Tang, S. Kang, Y. Zhang, X. He, T. Zhou, T. Qin, Y. Huang, X. Yi, and L. Zhang. 2015. Upregulation of PD-L1 by EGFR activation mediates the immune escape in EGFR-driven NSCLC: Implication for optional immune targeted therapy for NSCLC patients with EGFR mutation. *J. Thorac. Oncol.* 10:910-923.
- Decker, J. E., J. C. Pires, G. C. Conant, S. D. McKay, M. P. Heaton, K. Chen, A. Cooper, J. Vilkki, C. M. Seabury, and A. R. Caetano, et al. 2009. Resolving the evolution of extant and extinct ruminants with high-throughput phylogenomics. *Proc. Natl. Acad. Sci. USA.* 106:18644-18649.
- Elsik, C. G., R. L. Tellam, and K. C. Worley, 2009. The genome sequence of taurine cattle: a window to ruminant biology and evolution. *Science* 324:522-528.
- Groenen, M. A. M., A. L. Archibald, H. Uenishi, C. K. Tuggle, Y. Takeuchi, M. F. Rothschild, C. Rogel-Gaillard C. Park, D. Milan, and H. J. Megens, et al. 2012. Analyses of pig genomes provide insight into porcine demography and evolution. *Nature* 491:393-398.
- Köks, S., E. Reimann, R. Lilleoja, F. Lättekiivi, A. Salumets, P. Reemann, and Ü. Jaakma. 2014. Sequencing and annotated analysis of full genome of Holstein breed bull. *Mamm. Genome* 25:363-373.
- Kawahara-Miki, R., K. Tsuda, Y. Shiwa, Y. Arai-Kichise, T. Matsumoto, Y. Kanesaki, S. I. Oda, S. Ebihara, S. Yajima, and H. Yoshikawa, et al. 2011. Whole-genome resequencing shows numerous genes with nonsynonymous SNPs in the Japanese native cattle *Kuchinoshima-Ushi*. *BMC Genomics* 12:103.
- Khatib, H., R. L. Monson, V. Schutzkus, D. M. Kohl, G. J. M. Rosa, and J. J. Rutledge. 2008. Mutations in the *STAT5A* gene are associated with embryonic survival and milk composition in cattle. *J. Dairy Sci.* 91:784-793.
- Kijas, J. W., D. Townley, B. P. Dalrymple, M. P. Heaton, J. F. Maddox, A. McGrath, P. Wilson, R. G. Ingersoll, R. McCulloch, and S. McWilliam, et al. 2009. A genome wide survey of SNP variation reveals the genetic structure of sheep breeds. *PLoS ONE* 4:e4668.
- Kim, H., S. K. Lee, M. W. Hong, S. R. Park, Y. S. Lee, J. W. Kim, H. K. Lee, D. K. Jeong Y. H. Song, and S. J. Lee. 2013. Association of a single nucleotide polymorphism in the *akirin 2* gene with economically important traits in Korean native cattle. *Anim. Genet.* 44:750-753.
- Kim, J. I., Y. S. Ju, H. Park, S. Kim, S. Lee, J. H. Yi, J. Mudge, N. A. Miller, D. Hong, and C. J. Bell, et al. 2009. A highly annotated whole-genome sequence of a Korean individual. *Nature* 460:1011-1015.
- Lee, K. T., W. H. Chung, S. Y. Lee, J. W. Choi, J. Kim, D. Lim, S. Lee, G. W. Jang, B. Kim, and Y. H. Choy, et al. 2013. Whole-genome resequencing of Hanwoo (Korean cattle) and insight into regions of homozygosity. *BMC Genomics* 14:519.
- Lee, S. H., Y. M. Cho, S. H. Lee, B. S. Kim, N. K. Kim, Y. H. Choy, K. H. Kim, D. Yoon, S. K. Im, S. J. Oh, and E. W. Park. 2008. Identification of marbling-related candidate genes in *M. longissimus dorsi* of high- and low marbled Hanwoo (Korean native cattle) steers. *BMB Reports* 41:846-851.
- Lee, S. H., C. Gondro, J. van der Werf, N. K. Kim, D. J. Lim, E. W. Park, S. J. Oh, J. P. Gibson, and J. M. Thompson. 2010. Use of a bovine genome array to identify new biological pathways for beef marbling in Hanwoo (Korean Cattle). *BMC Genomics* 11:623.
- Lim, D., N. K. Kim, H. S. Park, S. H. Lee, Y. M. Cho, S. J. Oh, T. H. Kim, and H. Kim, 2011. Identification of candidate genes related to bovine marbling using protein-protein interaction networks. *Int. J. Biol. Sci.* 7:992-1002.
- McCue, M. E., D. L. Bannasch, J. L. Petersen, J. Gurr, E. Bailey, M. M. Binns, O. Distl, G. Guérin, T. Hasegawa, and E. W. Hill, et al. 2012. A high density SNP array for the domestic horse and extant Perissodactyla: Utility for association mapping, genetic diversity, and phylogeny studies. *PLoS Genet.* 8:e1002451.
- Novembre, J. and S. Ramachandran. 2011. Perspectives on human population structure at the cusp of the sequencing era. *Annu. Rev. Genomics Hum. Genet.* 12:245-274.
- Nusse, R. and H. Varmus. 2012. Three decades of Wnts: a personal perspective on how a scientific field developed. *EMBO J.* 31:2670-2684.
- Paredi, G., S. Raboni, E. Bendixen, A. M. de Almeida, and A. Mozzarelli, 2012. "Muscle to meat" molecular events and technological transformations: The proteomics insight. *J. Proteomics* 75:4275-4289.
- Pinto, L. F., J. B. Ferraz, V. B. Pedrosa, J. P. Eler, F. V. Meirelles, M. N. Bonin, F. M. Rezende, M. E. Carvalho, D. C. Cucco, and R. C. Silva. 2011. Single nucleotide polymorphisms in *CAPN* and *leptin* genes associated with meat color and tenderness in Nellore cattle. *Genet. Mol. Res.* 10:2057-2064.
- Ron, M., G. Israeli, E. Seroussi, J. I. Weller, J. P. Gregg, M. Shani, and J. F. Medrano. 2007. Combining mouse mammary gland gene expression and comparative mapping for the identification of candidate genes for QTL of milk production traits in cattle. *BMC Genomics* 8:183.
- Shibata, M., K. Matsumoto, K. Aikawa, T. Muramoto, S. Fujimura, and M. Kadowaki. 2006. Gene expression of myostatin during development and regeneration of skeletal muscle in Japanese Black Cattle. *J. Anim. Sci.* 84:2983-2989.
- Van den Bossche, J., B. Malissen, A. Mantovani, P. De Baetselier, and J. A. Van Ginderachter. 2012. Regulation and function of the E-cadherin/catenin complex in cells of the monocyte-macrophage lineage and DCs. *Blood* 119:1623-1633.
- Vonholdt, B. M., J. P. Pollinger, K. E. Lohmueller, E. Han, H. G. Parker, P. Quignon, J. D. Degenhardt, A. R. Boyko, D. A. Earl, and A. Auton, et al. 2010. Genome-wide SNP and haplotype analyses reveal a rich history underlying dog domestication. *Nature* 464:898-902.
- Watanabe, N., Y. Satoh, T. Fujita, T. Ohta, H. Kose, Y. Muramatsu, T. Yamamoto, and T. Yamada. 2011. Distribution of allele

frequencies at *TTN* g. 231054C> T, *RPL27A* g. 3109537C> T and *AKIRIN2* c.\*188G> A between Japanese Black and four other cattle breeds with differing historical selection for marbling. BMC Res. Notes 4:10.

Zimin, A. V., A. L. Delcher, L. Florea, D. R. Kelley, M. C. Schatz, D. Puiu, F. Hanrahan, G. Pertea, C. P. Van Tassell, and T. S. Sonstegard, et al. 2009. A whole-genome assembly of the domestic cow, *Bos taurus*. Genome Biol. 10:R42.