

## SNP detection using RNA-sequences of candidate genes associated with puberty in cattle

M.M. Dias<sup>1</sup>, A. Cánovas<sup>2</sup>, C. Mantilla-Rojas<sup>3</sup>, D.G. Riley<sup>3</sup>, P. Luna-Nevarez<sup>4</sup>, S.J. Coleman<sup>5</sup>, S.E. Speidel<sup>5</sup>, R.M. Enns<sup>5</sup>, A. Islas-Trejo<sup>6</sup>, J.F. Medrano<sup>6</sup>, S.S. Moore<sup>7</sup>, M.R.S. Fortes<sup>8</sup>, L.T. Nguyen<sup>8,9</sup>, B. Venus<sup>7</sup>, I.S.D.P. Diaz<sup>1</sup>, F.R.P. Souza<sup>10</sup>, L.F.S. Fonseca<sup>1</sup>, F. Baldi<sup>1</sup>, L.G. Albuquerque<sup>1</sup>, M.G. Thomas<sup>5</sup> and H.N. Oliveira<sup>1</sup>

<sup>1</sup>Departamento de Zootecnia, Faculdade de Ciências Agrárias e Veterinárias, Universidade Estadual Paulista, Jaboticabal, SP, Brasil

<sup>2</sup>Centre for Genetic Improvement of Livestock, Department of Animal Bioscience, University of Guelph, Guelph, ON, Canada

<sup>3</sup>Department of Animal Science, Texas A&M University, College Station, TX, USA

<sup>4</sup>Departamento de Ciencias Agronómicas y Veterinarias, Instituto Tecnológico de Sonora, Ciudad Obregón, SON, México

<sup>5</sup>Department of Animal Sciences, Colorado State University, Fort Collins, CO, USA

<sup>6</sup>Department of Animal Science, University of California, Davis, CA, USA

<sup>7</sup>Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, Brisbane, Australia

<sup>8</sup>School of Chemistry and Molecular Biosciences, The University of Queensland, Brisbane, St Lucia, Australia

<sup>9</sup>Faculty of Biotechnology, Vietnam National University of Agriculture, Vietnam

<sup>10</sup>Departamento de Ecologia, Zoologia e Genética, Universidade Federal de Pelotas, Pelotas, RS, Brasil

Corresponding author: M.M. Dias  
E-mail: marina.mortati@gmail.com

Genet. Mol. Res. 16 (1): gmr16019522

Received November 7, 2016

Accepted January 10, 2017

Published March 22, 2017

DOI <http://dx.doi.org/10.4238/gmr16019522>

Copyright © 2017 The Authors. This is an open-access article distributed under the terms of the Creative Commons Attribution ShareAlike (CC BY-SA) 4.0 License.

**ABSTRACT.** Fertility traits, such as heifer pregnancy, are economically important in cattle production systems, and are therefore, used in

genetic selection programs. The aim of this study was to identify single nucleotide polymorphisms (SNPs) using RNA-sequencing (RNA-Seq) data from ovary, uterus, endometrium, pituitary gland, hypothalamus, liver, longissimus dorsi muscle, and adipose tissue in 62 candidate genes associated with heifer puberty in cattle. RNA-Seq reads were assembled to the bovine reference genome (UMD 3.1.1) and analyzed in five cattle breeds; Brangus, Brahman, Nellore, Angus, and Holstein. Two approaches used the Brangus data for SNP discovery 1) pooling all samples, and 2) within each individual sample. These approaches revealed 1157 SNPs. These were compared with those identified in the pooled samples of the other breeds. Overall, 172 SNPs within 13 genes (*CPNE5*, *FAM19A4*, *FOXN4*, *KLF1*, *LOC777593*, *MGC157266*, *NEBL*, *NRXN3*, *PEPT-1*, *PPP3CA*, *SCG5*, *TSG101*, and *TSHR*) were concordant in the five breeds. Using Ensembl's Variant Effector Predictor, we determined that 12% of SNPs were in exons (71% synonymous, 29% nonsynonymous), 1% were in untranslated regions (UTRs), 86% were in introns, and 1% were in intergenic regions. Since these SNPs were discovered in RNA, the variants were predicted to be within exons or UTRs. Overall, 160 novel transcripts in 42 candidate genes and five novel genes overlapping five candidate genes were observed. In conclusion, 1157 SNPs were identified in 62 candidate genes associated with puberty in Brangus cattle, of which, 172 were concordant in the five cattle breeds. Novel transcripts and genes were also identified.

**Key words:** Bovine; Gene discovery; SNP; Fertility; Transcript discovery

## INTRODUCTION

Puberty is an important physiological process for fertility traits, such as age at first calving and heifer pregnancy. This process is a gradual maturation that begins during fetal development, continues in the pre- and peripubertal phases, and is initiated in the brain following activation of the hypothalamic-pituitary-gonadal (HPG) axis (Day and Nogueira, 2013). In heifers, the events leading to puberty are similar among the two bovine sub-species, *Bos taurus* and *Bos indicus*, but these occur later in *B. indicus* heifers (Rodrigues et al., 2002; Day and Nogueira, 2013). Sexual maturity of the heifer has a direct influence on reproductive efficiency and is considered economically more relevant and important than growth traits (Brumatti et al., 2011).

Since publication of the bovine reference genome in 2009 (Elsik et al., 2009), development and increased knowledge of the “-omics” technologies have accelerated the investigation of the genetic regulation of bovine fertility traits (Fortes et al., 2012; Snelling et al., 2013; Cánovas et al., 2014, Nascimento et al., 2016). Moreover, it is important to delineate candidate genes and to determine whether polymorphisms exist within genes that influence these traits. The advantage of detecting polymorphisms using transcriptome data is the possibility that a variant may influence the translation of mRNA.

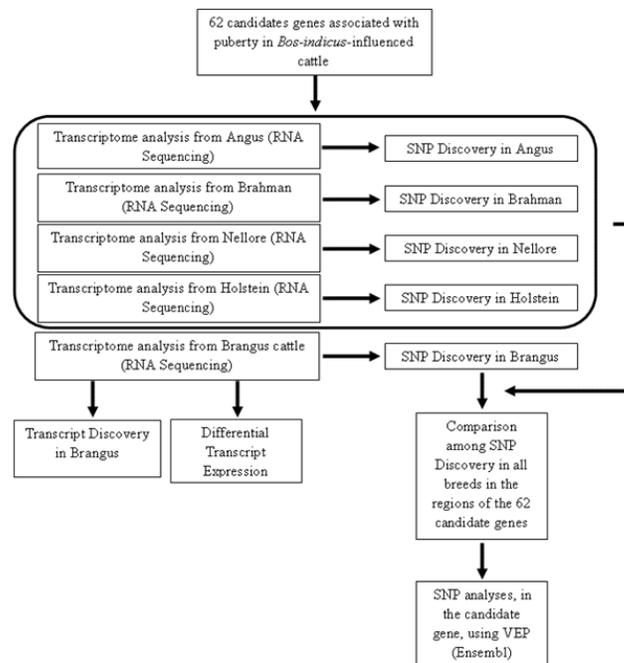
The aim of this study was to identify single nucleotide polymorphisms (SNPs) in

the transcriptome of eight tissues that are involved in the physiological process of puberty (i.e., hypothalamus, pituitary gland, uterus, endometrium, and ovary) and in the metabolism required for growth and development (i.e., liver, adipose, and longissimus dorsi muscle). Variant discovery occurred within regions of the genome corresponding to 62 genes associated with puberty in *B. indicus*-influenced cattle. In addition, the SNPs detected in Brangus cattle were also found to exist in other breeds (Angus, Brahman, Holstein, and Nellore). Finally, alternative splicing of the RNA was investigated to identify genes that overlap at the loci of the candidate genes. The aims of this study were achieved and the results demonstrated the value of using RNA-Seq data to identify trait-associated SNPs. These data also helped improve annotation of the bovine genome.

## MATERIAL AND METHODS

### Populations

The animals and data used in this study are part of collaboration aiming to identify polymorphisms associated with reproductive traits in sub-species of cattle: *B. taurus*, *B. indicus*, and their crosses. The base population used in this study was called the Discovery population, which was composed of Brangus heifers. The SNPs identified in the Discovery population were validated across the Angus, Brahman, Nellore, and Holstein breeds. These breeds are collectively referred to as the Comparison population. Figure 1 briefly describes the analyses conducted in each population.



**Figure 1.** Workflow of the analytical steps used to identify single nucleotide polymorphisms (SNPs) and transcripts within 62 candidate genes associated with puberty in *Bos indicus*-influenced cattle using RNA sequencing.

## Discovery population

Heifers were handled and managed as per approval of the Institutional Animal Care and Use Committee of New Mexico State University (protocol #2010-013; Cánovas et al., 2014). Eight heifers representing the pedigree diversity of the population used in genome-wide association study (GWAS) of fertility traits were selected from a Brangus (3/8 Brahman x 5/8 Angus) cattle breeding program (Cánovas et al., 2014). Heifers were randomly assigned to a physiological state-group PRE (N = 4) or POST (N = 4) puberty. Eight tissues were collected from each heifer, including: hypothalamus (HYP), pituitary gland (PIT), liver (LIV), uterus (UTE), endometrium (END), ovary (OVA), adipose (FAT), and longissimus dorsi muscle (LDM). During the tissue sampling and laboratory processing, procedures failed for two END samples.

RNA extraction, library construction, and sequencing were performed for each tissue and animal as described by Cánovas et al. (2014). Additionally, libraries were multiplexed, six libraries per lane, and sequenced with an Illumina HiSeq 2000 analyzer. Single read sequences (100 bp) from Angus, Holstein, and Brangus breeds and paired-end read (100 bp) sequences from Brahman breeds were mapped to the annotated bovine reference genome (UMD3.1; release annotation 78; [ftp://ftp.ensembl.org/pub/release-77/genbank/bos\\_taurus/](ftp://ftp.ensembl.org/pub/release-77/genbank/bos_taurus/)). The CLC Genomics workbench software (CLC Bio, Aarhus, Denmark) was used with default parameters for alignment and quality control analyses, as previously described (Cánovas et al., 2010; Fortes et al., 2016). All samples passed quality control with the exception of two samples from hypothalamus tissue in Brangus heifers, which were removed from the analyses. The same hypothalamic and pituitary gland samples were used for peptide extraction and neuropeptide identification (DeAtley, 2012). Sixty samples were analyzed using RNA-Seq (Figure 1). Thirty million reads, on average, were obtained for each sample in all tissues from Brangus heifers as well as Angus steers, Brahman heifers, and Holstein bull calves as described below.

## Comparison population

### *Angus*

Tissues from 10 Angus steers that were phenotyped to have low or high pulmonary arterial pressures (LPAP and HPAP, respectively) were harvested for the study of pulmonary hypertension. Six tissues were collected and included from each steer: left ventricle (heart), right ventricle (heart), pulmonary artery, aorta, longissimus dorsi muscle, and lung (Cánovas et al., 2016; Li et al., 2016).

### *Brahman*

Twelve Australian Brahman heifers were assigned to the physiological state of PRE (N = 6) or POST (N = 6) puberty as described by Fortes et al. (2016). Six tissues were harvested from each heifer, including hypothalamus, pituitary, liver, uterus, left ovary, and right ovary.

### *Nellore*

Forty-eight Brazilian Nellore bulls were slaughtered and phenotyped for their meat fatty acid profile. These animals were selected to have a maximum of 90 days between

slaughter ages. These bulls were the progeny of six sires. One tissue, longissimus thoracis muscle, was collected from each animal. RNA-seq data were generated on the HiScanSQ (Illumina, San Diego, CA, USA), which yielded 100-bp paired-end reads. These samples had an average of 24 million reads/sample.

## Holstein

Twelve neonatal Holstein calves were assigned to a physiological state of healthy (N = 6) or hypertensive (N = 6) in response to hypoxia. Fibroblasts from the pulmonary artery were harvested and cultured, and transcriptome analysis was performed using RNA from these cells (Li et al., 2016).

## Candidate genes

Based on the results of previous studies, we selected 62 candidate genes from GWAS and transcriptome results (Cánovas et al., 2014), neuropeptidome results (DeAtley, 2012), and from a study based on genotype-to-phenotype associations of SNPs from a high-density chip (Dias et al., 2015). Specifically, the 62 genes included: 25 genes harboring SNPs associated with fertility traits also expressed in the fertility tissue transcriptome (Table 1), 20 genes that encoded relevant transcription factors expressed in the fertility tissue transcriptome (Table 2), 19 neuropeptide genes related to reproduction (DeAtley, 2012) (Table 3), and two genes, *PPP3CA* and *FABP4*, from an association study with the trait early pregnancy probability (P16) in Nellore cattle (Dias et al., 2015). Three genes were the same among the group of candidate genes as described below. In total, we studied 62 candidate genes in our SNP discovery analyses.

**Table 1.** Genes harboring SNPs associated with heifer fertility either differentially expressed or with tissue-specific expression in RNA-Seq analysis in pre- and post-pubertal Brangus heifers (adapted from Cánovas et al., 2014).

Gene	Tissue	SNP	Chr.	Position (bp)	Distance (bp)	SNP location in gene
<i>DPPA4</i>	HYP	rs41571491	1	54,523,895	5,517	Downstream
<i>TP63</i>	ADIPOSE	rs109034747	1	78,088,393	0	Intron
<i>INHA</i>	OVA	rs136318374	2	108,231,291	1,460	Downstream
<i>IL22RA1</i>	END	rs29019426	2	129,348,719	2,978	Upstream
<i>RHEBL1</i>	PIT	rs1107702	5	30,912,029	7,398	Upstream
<i>LYSB</i>	END	rs110181782	5	44,556,197	0	Intron
<i>ADH6</i>	END	rs41612964	6	26,799,276	0	Intron
<i>MEPE</i>	PIT	rs41650773	6	38,286,952	0	Intron
<i>TECRL</i>	LDM	rs109759346	6	81,577,343	0	Intron
<i>LOC777593</i>	END	rs109621209	7	61,683,533	0	Intron
<i>MGC157266</i>	END	rs136828274	8	62,927,382	8,411	Upstream
<i>C10H11ORF46</i>	HYP	rs109705635	10	4,882,326	0	Intron
<i>NRXN3</i>	END	rs43651752	10	91,751,584	0	Intron
<i>TSHR</i>	END	ss117964119	10	93,528,030	0	Intron
<i>NEBL</i>	END	ss61500113	13	22,476,667	0	Intron
<i>MOS</i>	OVA	rs110243083	14	24,973,324	2,465	Upstream
<i>PENK</i>	PIT	rs134428213	14	25,218,861	0	Exon
<i>ELF5*</i>	UTE	rs133233558	15	65,840,403	0	Intron
<i>POU4F2*</i>	END	rs41838669	17	11,828,433	1,818	Upstream
<i>FAM19A4</i>	LIV	rs136177962	22	32,820,574	0	Intron
<i>ITIH1</i>	OVA	rs110723544	22	48,663,639	0	Exon
<i>CPNE5</i>	END	rs109264326	23	10,690,330	0	Intron
<i>MMD2</i>	UTE	rs135797510	25	39,622,260	0	Intron
<i>DKK1</i>	UTE	rs29024937	26	6,860,529	4,762	Downstream
<i>SYCE1</i>	LIV	rs41629412	26	50,517,416	0	Intron

HYP (hypothalamus); PIT (pituitary gland); UTE (uterus); OVA (ovary); FAT (adipose); LIV (liver); LDM (longissimus dorsi muscle). \*Genes also are 20 relevant transcription factors.

**Table 2.** Twenty relevant transcription factors and the tissue of maximum expression in pre- and post-pubertal Brangus heifers (adapted from Cánovas et al., 2014).

Gene	Tissue of maximum expression	
<i>OVGP1</i>	Ovary	
<i>VAX2</i>	Hypothalamus and/or pituitary gland	
<i>FOXE1</i>		
<i>FOXB2</i>		
<i>LHX4</i>		
<i>PITX2</i>		
<i>POU4F2</i>		
<i>PROP1</i>		
<i>SIX6</i>		
<i>NHLH1</i>		
<i>DACH2</i>		
<i>CDX2</i>		Uterus and/or endometrium
<i>ELF5</i>		
<i>FOXP4</i>		
<i>VGLL1</i>		
<i>TSG101</i>		
<i>MYOCD</i>		
<i>KLF1</i>	Adipose	
<i>GATA1</i>		
<i>E2F3</i>	Longissimus dorsi muscle	

**Table 3.** Neuropeptides in the hypothalamus and pituitary gland of pre- and post-pubertal Brangus heifers reported by DeAtley (2012).

Gene name	Peptide isoform
<i>CART</i>	1
<i>CBLN4</i>	1
<i>CBLN1</i>	4
<i>CBLN2</i>	2
<i>CCK</i>	1
<i>CHGA</i>	3
<i>NTS</i>	1
<i>PCSK1N</i>	9
<i>PEPT-1</i>	2
<i>Penk-B (isoform)</i>	1
<i>Penk-A (isoform)</i>	4
<i>POMC</i>	7
<i>TAC1</i>	3
<i>SCG5</i>	1
<i>CHGB</i>	7
<i>SCG2</i>	2
<i>SCG3</i>	1
<i>SST</i>	1
<i>STMN1</i>	1

## RNA-sequence assembly

Transcripts from RNA-Seq data derived from the five breed groups were analyzed separately using CLC Genomics Workbench 8.0.1 (CLC Bio, Aarhus, Denmark). For the Brangus sequence data, two assemblies were performed. The first assembly included all samples pooled and the second assembly included each individual sample. A single assembly was performed for the sequence data of each of the other breed groups, Angus, Brahman, Holstein, and Nellore, where sequence data from all tissues were pooled within a breed group.

## SNP discovery within candidate genes

Following the assembly of RNA-Seq data, SNP discovery was executed by CLC Genomics Workbench 8.0.1 (CLC Bio, Aarhus, Denmark) using the tool “Fixed Ploidy Variant Detection”. This tool detects germline variants and discards variants when representation in reads is due to sequencing errors or mapping artifacts. Quality and significance filters were used as described by Cánovas et al. (2010).

Two approaches were used for the Brangus samples. Initially, we performed variant detection for all pooled samples, which allowed full-use of aligned reads and maximized variant discovery. The second approach involved detecting the variants in each individual sample. After the two approaches were performed, the results were combined and SNPs that were observed using both approaches were validated in the other breed groups, thereby decreasing the error of false positives in SNP detection. In the other breeds, we performed the analysis using pooled samples of each breed. Then, we verified whether each identified SNP existed in all breeds.

Resources of Ensembl (Cunningham et al., 2015) were used to query SNPs in the gene regions of the 5'UTR, exon, and the 3'UTR (**Tables S1-S4**). The SNPs within dbSNP (<http://www.ncbi.nlm.nih.gov/SNP/>) for cattle were also compared with those found in the RNA-Seq data to verify how many SNPs were novel discoveries and how many were already annotated variants. In addition, the Variant Effect Predictor (VEP) from Ensembl was used to predict whether those SNPs were synonymous or non-synonymous, as well as the potential biochemical structure of resulting peptides and proteins.

## Transcript discovery within genes and gene discovery

Large gap mapping analysis was performed in CLC Genomics Workbench 8.0.1 (CLC Bio, Aarhus, Denmark) to map RNA-Seq data that spanned introns without the need for prior transcript annotations. Transcript discovery was also performed with this software. This tool identifies likely regions within genes that are splice sites. This analysis included existing annotations; therefore, we could enrich existing gene and mRNA annotations. Default parameters of the software were used in both analyses.

## RESULTS AND DISCUSSION

### SNP discovery and comparison across breeds

In the Ensembl SNP database, 231,770 SNPs were already annotated in the 5'UTR, exon, intron, and 3'UTR regions of the 62 candidate genes as reported in **Tables S1-S4**. Of these, 1061 SNPs were in the 5'UTR, 218,036 were in introns, 10,493 were in exons, and 2180 were in the 3'UTR. Within exons, 8023 nonsynonymous and 2470 synonymous mutations were identified. The expectation was that the number of synonymous mutations would be larger than the number of nonsynonymous mutations because synonymous mutations do not influence the primary structure of the protein. A possible reason for the observed discrepancy is that cattle breeds are experiencing selection, and the nonsynonymous mutations may potentially be associated with traits of interest, such as heifer pregnancy. To clarify this issue, it was necessary to estimate the synonymous and nonsynonymous

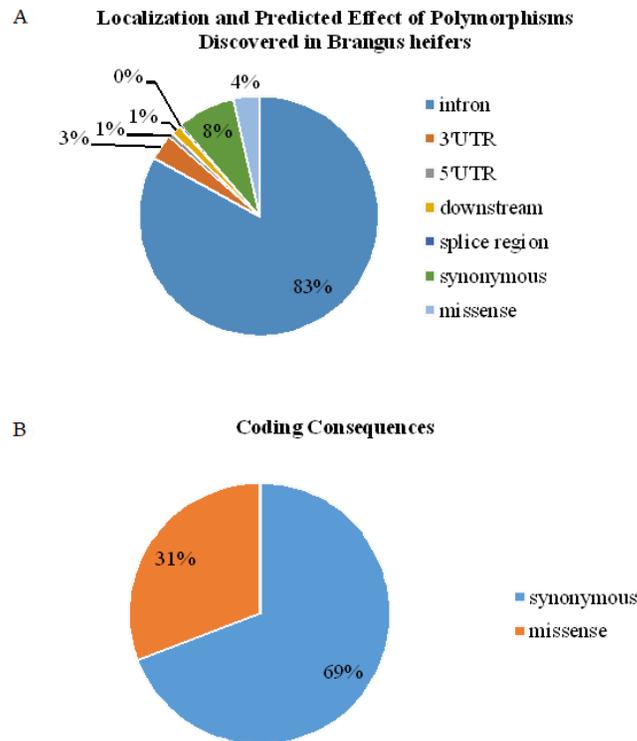
substitution rates; however, we were not able to do this estimation, once it is necessary the number of substitutions per site between as additional sequences and animals are needed to complete these types of analyses (Yang and Nielsen, 2000).

The first method of variant detection applied to sequence Brangus heifers was the pooled samples analyses, and it revealed 9357 SNPs in 56 of the 62 candidate genes. No SNPs were detected in *CDX2*, *FOXB2*, *FOXE1*, *LYSB*, *MOS*, or *NHLH1*. The second analysis assembled each sample individually and 2175 SNPs within 50 candidate genes were observed. However, no SNPs were detected in *CDX2*, *DACH2*, *DPPA4*, *FOXB2*, *FOXE1*, *GATA1*, *IL22RA1*, *LYSB*, *MEPE*, *MOS*, *NHLH1*, or *RHEBL1*. The minimum number of SNPs observed in a single sample was 64, and the maximum number was 424. The average number of SNPs per gene was  $186.1 \pm 11.3$ . The observed differences in the number of SNPs in each sample could be because some genes are tissue specific, transcript-specific, and/or the animals were non-related as per the diversity of the pedigree. In this study, 840 SNPs were observed that were sample specific. Nine variations that appeared in the 60 samples were homozygotes. These observations were expected, as the software used the bovine reference genome in its algorithms. This reference genome is from an inbred *B. taurus*-Hereford cow named Dominette; therefore, the likelihood of detecting breed-specific SNPs in this study was very high.

After performing the two analyses in Brangus heifers, the results were merged and SNPs detected using both approaches were used to compare breeds. Therefore 1157 SNPs within 46 candidate genes were usable in additional analyses involving breeds. Genes without a detected SNP were *ADH6*, *CDX2*, *DACH2*, *DPPA4*, *FOXB2*, *FOXE1*, *GATA1*, *IL22RA1*, *LYSB*, *MEPE*, *MOS*, *NHLH1*, *NTS*, *PCSKIN*, *RHEBL1*, and *VAX2*.

After combining the results from the two approaches using sequences from Brangus heifers, the VEP tool was used to predict the effect of the variation on the protein as per the intra-gene locus of the 1157 SNPs (Figure 2). Overall, 699 (60.4%) SNPs were identified as novel, and 458 (39.6%) were identified as existing variants within the dbSNP database. These SNPs were detected in 48 genes in the present study. There were two genes detected that were not in candidate genes initially described in our study, *LYSMD2* and *OBSL1*. These SNPs were also located in the genes *SCG5* and *SCG2*. A surprising result of this study was the observation that 78% of the 1157 SNPs resided in introns. There is ongoing discussion concerning the accuracy of the sequence and annotation of the bovine reference genome, which is from a Hereford cow named Dominette (Elsik et al., 2009). Bohmanova et al. (2010) reported that some SNPs on the BovineSNP50 bead-chip were erroneously mapped and Zhou et al. (2015) reported discordance in both publically available genome reference assemblies of Dominette, UMD 3.1, and Btau 4.6. Currently, a new and more advanced bovine assembly is under development that is expected to resolve these discrepancies. It should also be noted that the *B. taurus* bovine reference genome was used in the present study; therefore, development of a *B. indicus* reference genome could greatly enhance results from SNP discovery studies.

Because we detected SNPs in the RNA sequence that the current annotations described as being in introns, we conducted a transcript discovery analysis. Specifically, the VEP tool of Ensembl assigned 130 SNPs to exons. Of these, 78 already existed in SNP databases and 52 were novel. Only 40 of these SNPs in the exons were non-synonymous. Table 4 describes the amino acid changes and changes in the biochemical properties resulting from these SNPs in potential proteins.



**Figure 2.** Descriptive statistics results from the Variant Effect Predict (VEP) tool of Ensembl using the 1157 SNPs identified in Brangus cattle.

The VEP tool classified some SNPs as splice variants. This means that polymorphisms occurred within the region of the splice site and potentially influenced the length of the RNA transcript (Cunningham et al., 2015). SNPs located within the non-coding region (3'UTR) of the candidate gene may be located in a binding site for miRNA and/or the transcriptional machinery (Jin and Lee, 2013). Moreover, SNPs in 3'UTR region can also influence mRNA stability (Matoulkova et al., 2012). RNA transcripts often contain a poly-A tail, which could have a role in gene expression as well as post-transcriptional stability (Kuersten and Goodwin, 2003). It should be noted that SNPs located within an intron may also affect miRNA or transcription factor binding sites, possibly affecting the transcription of other genes (Jin and Lee, 2013). SNPs located within the upstream intergenic region of a gene may also be within, or influence, a binding site for transcription factors (Vinsky et al., 2013).

Analysis of SNPs discovered in Angus cattle samples revealed 2742 SNPs associated with puberty within 38 genes. Genes containing the fewest numbers of SNPs were *PCSKIN* and *PENK*, with one polymorphism detected. The *NRXN3* gene contained the largest number of mutations, with 441 SNPs. The average number of SNPs per gene was  $72.16 \pm 16.33$ . Angus genes with no SNPs were *CARTPT*, *CBLN1*, *CCK*, *CDX2*, *DACH2*, *DKK1*, *E2F3*, *FOXB2*, *FOXE1*, *GATA1*, *IL22RA1*, *ITIH1*, *LHX4*, *LYSB*, *MEPE*, *MOS*, *NHLH1*, *OVGP1*, *PITX2*, *POU4F2*, *PROPI*, *SIX6*, *SST*, and *TAC1*.

**Table 4.** Non-synonymous mutations detected in the candidate genes associated with puberty in Brangus cattle.

CHR	Position	Gene	Amino acid change	Change in the biochemistry properties of the amino acid	Codon	Existing variant
1	80251288	<i>SST</i>	Lys/Glu	PP/PN	AAA/GAA	rs17870997
2	112550874	<i>SCG2</i>	Ser/Pro	PNeu/NP	TCG/CCG	rs43316973
3	32058726	<i>OVGP1</i>	His/Arg	PP/PP	CAC/CGC	rs41570791
3	32058831	<i>OVGP1</i>	His/Arg	PP/PP	CAC/CGC	rs41570792
4	14921388	<i>TAC1</i>	Val/Leu	NP/NP	GTG/TTG	rs378353665
4	14922541	<i>TAC1</i>	Ile/Leu	NP/NP	ATT/CTT	rs211390627
7	13786612	<i>KLF1</i>	Ser/Gly	PNeu/NP	AGC/GGC	-
7	13786627	<i>KLF1</i>	Ala/Pro	NP/NP	GCA/CCA	-
7	13786637	<i>KLF1</i>	Arg/Thr	PP/PNeu	AGG/ACG	-
7	13786644	<i>KLF1</i>	Leu/Pro	NP/NP	CTG/CCG	-
7	13786663	<i>KLF1</i>	Pro/Ala	NP/NP	CCG/GCG	-
7	13786679	<i>KLF1</i>	Ser/Thr	PNeu/PNeu	AGC/ACC	-
7	13788755	<i>KLF1</i>	Ala/Pro	NP/NP	GCC/CCC	-
7	13788756	<i>KLF1</i>	Ala/Asp	NP/PN	GCC/GAC	-
7	13788762	<i>KLF1</i>	Thr/Lys	PNeu/PP	ACG/AAG	-
7	13788769	<i>KLF1</i>	Asp/Glu	PN/PN	GAC/GAA	-
7	13788782	<i>KLF1</i>	Arg/Gly	PP/NP	CGG/GGG	rs456791503
7	13788783	<i>KLF1</i>	Arg/Gln	PP/PNeu	CGG/CAG	-
10	93571223	<i>TSHR</i>	Gln/Arg	PNeu/PP	CAG/CGG	rs207552184
13	48446641	<i>CHGB</i>	Asn/Asp	PNeu/PN	AAT/GAT	rs110261417
13	48446975	<i>CHGB</i>	Thr/Met	PNeu/NP	ACG/ATG	rs109400236
13	48447215	<i>CHGB</i>	His/Arg	NP/PP	CAC/CGC	rs110013334
13	48447590	<i>CHGB</i>	Pro/Arg	NP/PP	CCG/CGG	rs111002720
13	48447685	<i>CHGB</i>	Arg/Gly	PP/NP	CGT/GGT	rs380061510
13	48448222	<i>CHGB</i>	Met/Val	NP/NP	ATG/GTG	rs110931509
17	11830384	<i>POU4F2</i>	Ile/Met	NP/NP	ATC/ATG	-
17	11830386	<i>POU4F2</i>	Leu/Phe	NP/NP	CTC/TTC	rs475409192
17	11830478	<i>POU4F2</i>	Ser/Ile	PNeu/NP	AGC/ATC	-
17	11830492	<i>POU4F2</i>	His/Gln	PP/PNeu	CAC/CAG	-
19	31767321	<i>MYOCD</i>	Ile/Ser	NP/PNeu	ATT/AGT	-
21	58170306	<i>CHGA</i>	Thr/Asn	PNeu/PNeu	ACC/AAC	rs109042853
21	58170321	<i>CHGA</i>	Pro/Leu	NP/NP	CCG/CTG	rs520684424
21	58172131	<i>CHGA</i>	Ser/Phe	PNeu/NP	TCT/TTT	rs109145556
21	58172484	<i>CHGA</i>	Pro/Ala	NP/NP	CCC/GCC	rs517883631
22	15195074	<i>CCK</i>	Met/Val	NP/NP	ATG/GTG	rs42891945
24	5619434	<i>CBLN2</i>	Val/Met	NP/NP	GTG/ATG	rs42038885
26	6855119	<i>DKK1</i>	Thr/Ser	PNeu/PNeu	ACT/AGT	rs210866503
X	20142496	<i>VGLL1</i>	Cys/Phe	NP/NP	TGT/TTT	-
X	20142501	<i>VGLL1</i>	Glu/Lys	PN/PP	GAA/AAA	-

NP: nonpolar; PP: polar positive; PN: polar negative; PNeu: polar neutral.

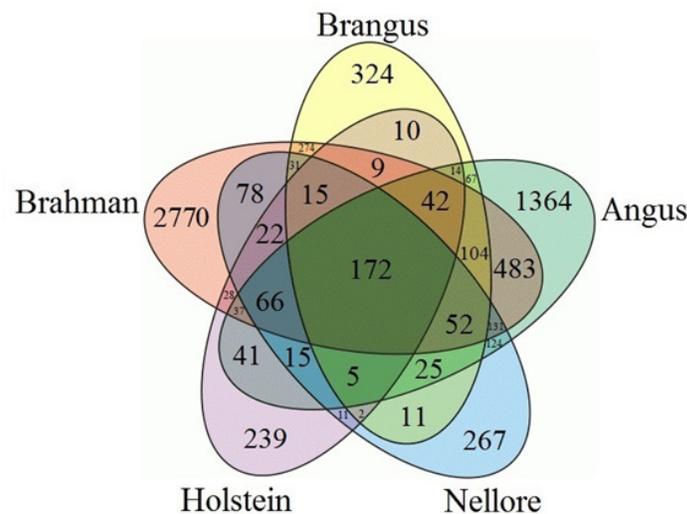
In the Brahman heifer samples, SNP discovery revealed 4316 variants within 55 genes. The average number of SNPs per gene was  $78.4 \pm 14.8$ . Genes with the lowest numbers of SNPs were *CCK* and *GATA1*, both with one SNP. The *MMD2* gene contained 467 SNPs. Genes in which no SNPs were detected included *CDX2*, *DPPA4*, *E2F3*, *FOXB2*, *FOXE1*, *MOS*, and *NHLHI*.

Samples from Nellore bulls revealed 1027 SNPs within 31 genes. The average number of SNPs per gene was  $33.1 \pm 7.5$ . The gene with the fewest SNPs was *PENK*, which possessed one SNP. The maximum number of SNPs was 213 in the *NRXN3* gene. The genes in which no SNPs were detected were: *CARTPT*, *CBLN1*, *CBLN2*, *CCK*, *CDX2*, *CHGA*, *CHGB*, *DACH2*, *DKK1*, *DPPA4*, *FOXB2*, *FOXE1*, *GATA1*, *IL22RA1*, *INHA*, *ITIH1*, *LHX4*, *LYSB*, *MEPE*, *MOS*, *NHLHI*, *PCSKIN*, *POU4F2*, *PROPI*, *SCG2*, *SIX6*, *SST*, *STMN1*, *TAC1*, *TP63*, and *VGLL1*. SNP discovery analysis within RNA from fibroblasts of Holstein calves revealed 728 SNPs within 31 genes, with an average of  $23.5 \pm 5.9$  SNPs per gene. Four genes contained one SNP

and included *INHA*, *SCG3*, *SST*, and *STMN1*. The gene *NRXN3* harbored the largest number of SNPs at 171. The genes in which no SNPs were detected included: *CARTPT*, *CBLN1*, *CBLN2*, *CBLN4*, *CCK*, *CDX2*, *CHGA*, *CHGB*, *DACH2*, *DKK1*, *DPPA4*, *FOXB2*, *GATA1*, *IL22RA1*, *ITIH1*, *LHX4*, *LYSB*, *MEPE*, *MOS*, *NHLH1*, *NTS*, *PCSKIN*, *PENK*, *PITX2*, *POU4F2*, *PROPI*, *RHEBL1*, *SCG2*, *SIX6*, *TAC1*, and *VGLL1*.

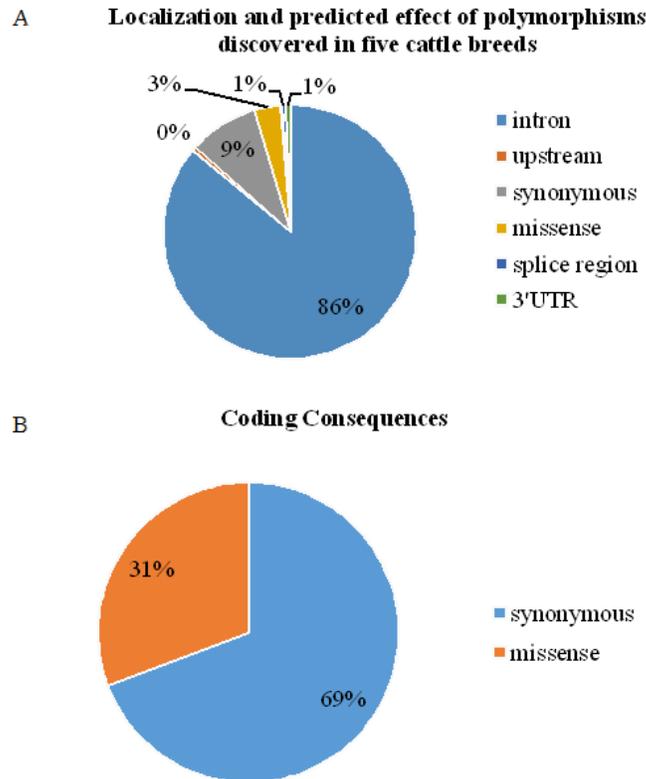
The studies involving Brangus and Brahman cattle were the only ones designed to specifically study differences between pre and post-pubertal animals (Fortes et al., 2012; Cánovas et al., 2014; Fortes et al., 2016). These are the only breeds from which RNA was extracted from tissues associated with reproduction. It is possible that the tissues from which the RNA was extracted influenced the number of detected SNPs, as the expression of some candidate genes is tissue specific.

After SNP discovery was executed in the five breed groups, a comparative analysis was performed. Overall, 172 SNPs were concordant in the five breed groups (Figure 3). These SNPs only existed within 13 genes: *CPNE5*, *FAM19A4*, *FOXN4*, *KLF1*, *LOC777593*, *MGC157266*, *NEBL*, *NRXN3*, *PEPT-1*, *PPP3CA*, *SCG5*, *TSG101* and *TSHR*. *NRXN3* was the gene with the largest number of SNPs in the four-breed comparison. In Brahman cattle, this gene had the second highest number of polymorphism, at 425. This gene is located on chromosome 10, and the length of this gene is 626,613 bp spanning eight exons and encoding a member of a family of proteins that function in the nervous system as receptors (Südhof, 2008). Furthermore, the number of mutations could depend on the gene length, and the *NRXN3* gene was the longest gene among those studied.



**Figure 3.** Venn diagram comparing SNPs present in Brangus, Angus, Brahman, Nellore, and Holstein cattle.

A study of 172 SNPs with VEP, and presented in Figure 4, revealed that 138 SNPs were novel (80.2%) and 34 SNPs (19.8%) were known. This analysis also revealed that these SNP existed in 13 genes. These results were surprising, as 77% of the SNPs were described as being in an intron despite the data being derived from RNA-Seq.



**Figure 4.** Descriptive statistics from the Variant Effect Predict (VEP) tool of Ensembl using the 172 SNPs that were consistent across breeds.

The *CPNE5* gene belongs to the copine gene family, and encodes a trafficking protein that guides ligands through the membrane based on their binding properties with phospholipids (Damer et al., 2007). This gene has also been shown to mediate intracellular processes by regulating various signaling pathways (Tomsig et al., 2003), as well as influencing obesity in humans (Wang et al., 2015). Other genes studied that are also associated with obesity include *NRXN3* (Heard-Costa et al., 2009) and *PEPT-1* (Zanni et al., 2015). Studies in humans have shown that obesity influences the age of puberty in girls. If girls have a relatively high body mass index, they are more likely to experience early menses (Kaplowitz, 2008).

*FOXN4* encodes a member of the protein superfamily forkhead box (Fox), which is evolutionarily conserved and acts as a transcriptional regulator influencing a broad spectrum of biological processes (Myatt and Lam, 2007). Transcription factors of this superfamily act as important regulators in development, homeostasis, and reproduction (Thackray, 2014).

Synthesis of hormones in the thyroid gland requires activation by thyroid stimulating hormone (TSH) and its receptor (TSHR) (Rocha et al., 2007; Bonger et al., 2009). Mutinati et al. (2010) detected TSHR in luteal cells in cattle using immunohistochemistry, which suggests this protein may function in the ovary. As a result, in our study, this hormone could have influenced multiple tissues.

The *PPP3CA* gene is associated with metabolic signaling via MAPK (San Giovanni and Lee, 2013). In addition, it has been reported that blockade of this gene's action leads to infertility in male rats (Miyata et al., 2015). Dias et al. (2015) reported a significant association between a SNP in this gene and the probability of an early pregnancy; therefore, polymorphisms in this gene are considered strong predictors of puberty in Nellore cattle.

Among the 13 genes that contained SNPs in the five breed groups in the present study, *LOC777593*, *MGC157266*, *NEBL*, *KLF1*, and *TSG101* were associated with puberty in a gene expression study by Cánovas et al. (2014). Similarly, *SCG5* was the only gene associated with puberty in a study by DeAtley (2012). Nonetheless, the results of our study may be useful for the design of a focused SNP panel for use in genotype-to-reproductive phenotype association analyses in a multi-breed setting. The focused (i.e. functional) SNP panel is an approach that can reduce costs associated with genotyping (Habier et al., 2009). One strategy of selecting SNPs to design a focused or low-density panel, is the use of biologically relevant polymorphisms (Fortes et al., 2014). Because our study used candidate genes that had already been determined to be associated with puberty in *B. indicus*-influenced cattle, the SNPs discovered should be useful to predict fertility performance in purebred and/or composites of Angus, Brahman, and Nellore cattle. We studied Holstein cattle in order to gain insights into the breadth of existence of these SNPs across the many breeds of cattle. The SNPs detected in this study should be located within an exon, 3'UTR, or 5'UTR, as we used reads from RNA-Seq, which have the potential to be causative mutations. Designing a focused SNP panel with causative mutations may be advantageous over SNPs selected to be of equal distance across the chromosomes on the current SNP-chip platforms, as these types of mutations are not dependent on linkage disequilibrium, and may be useful for selection over generations, and across breeds (Habier et al., 2009; Fortes et al., 2014).

## Transcript and gene discovery

Transcript discovery analysis revealed 160 novel transcripts within 42 candidate genes (Table 5). The genes that did not reveal new transcripts were *CBLN1*, *CDX2*, *DKK1*, *DPPA4*, *ELF5*, *FABP4*, *FOXB2*, *FOXE1*, *GATA1*, *INHA*, *KLF1*, *LYSB*, *NHLH1*, *NTS*, *PCSKIN*, *PENK*, *PEPT-1*, *POU4F2*, *SYCE1*, and *VAX2*. There was no change in the loci of the SNPs observed in the analysis using the VEP tool. Therefore, additional studies are needed to verify the mapping of SNPs identified from RNA-Seq and thereby to improve the annotation of the bovine reference genome.

In this analysis, some novel genes were discovered within the same region as five of the candidate genes described previously (Table 6). The mutations that affect our candidate genes may have also influenced the genes on the other DNA strand. It is unknown whether these genes influence puberty in cattle. More specifically, we need to delineate whether any transcriptional promoters exist on both strands or if there is RNA from both strands that could be transcribed in a manner that is influenced by SNPs. For genes that overlap at a locus, transcript synthesis cannot be regulated with the same polymerase because the enzyme only tracks sequences in one direction; however, the genes can share CpG islands, which can influence expression (Nakayama et al., 2007). Overlapping genes play a role in the regulation of gene expression at the level of transcription, mRNA processing, splicing and stability, or translation (Boi et al., 2004); therefore, there is the potential for interaction between these overlapping genes.

**Table 5.** Transcript discovery within candidate genes associated with puberty in *B. indicus*-influenced cattle.

Gene	Chr	Transcripts	Known transcripts	Unknown transcripts	Longest transcript	New 5' sequence	New 3' sequence
<i>SST</i>	1	2	1	1	604	No	No
<i>TP63</i>	1	2	1	1	3123	No	No
<i>STMN1</i>	2	2	1	1	982	No	No
<i>IL22RA1</i>	2	4	1	3	2775	No	No
<i>SCG2</i>	2	7	1	6	2362	No	No
<i>OVGP1</i>	3	4	1	3	2101	Yes	No
<i>TAC1</i>	4	8	3	5	1243	No	No
<i>RHEBL1</i>	5	4	1	3	637	Yes	No
<i>ADH6</i>	6	2	1	1	1525	Yes	No
<i>MEPE</i>	6	2	1	1	1572	No	No
<i>TECRL</i>	6	2	1	1	2197	No	Yes
<i>PITX2</i>	6	11	1	10	2164	Yes	No
<i>PPP3CA</i>	6	8	2	6	6721	Yes	No
<i>LOC777593</i>	7	3	1	2	637	No	No
<i>PROPI</i>	7	3	1	2	1190	No	No
<i>MGC157266</i>	8	3	1	2	2296	No	No
<i>NRXN3</i>	10	4	3	1	1497	No	No
<i>C10H11ORF46</i>	10	3	1	2	1017	Yes	No
<i>SLX6</i>	10	4	1	3	1478	No	No
<i>TSHR</i>	10	3	1	2	2473	Yes	No
<i>SCG3</i>	10	6	1	5	2820	No	No
<i>SCG5</i>	10	23	1	22	1018	No	No
<i>POMC</i>	11	10	1	9	1190	No	No
<i>CBLN4</i>	13	4	1	3	4589	No	Yes
<i>CHGB</i>	13	4	1	3	2433	Yes	No
<i>NEBL</i>	13	15	8	7	12,017	No	Yes
<i>MOS</i>	14	2	1	1	999	No	Yes
<i>LHX4</i>	16	5	1	4	1173	Yes	No
<i>FOXN4</i>	17	2	1	1	2384	Yes	No
<i>MYOCD</i>	19	9	1	8	8509	Yes	No
<i>CARTPT</i>	20	3	1	2	827	No	No
<i>CHGA</i>	21	2	1	1	1922	No	No
<i>ITIH1</i>	22	2	1	1	2910	No	No
<i>CCK</i>	22	3	1	2	804	No	No
<i>FAM19A4</i>	22	8	2	6	4567	Yes	No
<i>CPNE5</i>	23	6	1	5	1784	No	No
<i>E2F3</i>	23	7	1	6	1511	No	No
<i>CBLN2</i>	24	12	1	11	4486	Yes	No
<i>MMD2</i>	25	3	1	2	1039	No	No
<i>TSG101</i>	29	2	1	1	1434	No	No
<i>DACH2</i>	X	2	1	1	6086	No	Yes
<i>VGLL1</i>	X	4	1	3	1272	No	No

**Table 6.** Gene discovery and overlapping candidate genes that are associated with puberty in *B. indicus*-influenced cattle.

Gene	Chr	Unknown	Length	Start	End	Strand	Reads	Spliced reads
<i>LHX4</i>	16	No	47,307	62,868,725	62,916,031	+	214	128
Gene 3295	16	Yes	67,388	62,891,230	62,958,617	-	2,396	17
<i>FAM19A4</i>	22	No	198,805	32,709,906	32,908,710	+	1,127	123
Gene 4287	22	Yes	14,933	32,886,396	32,901,328	-	596	56
<i>MEPE</i>	6	No	11,386	38,279,966	38,291,351	-	80	3
Gene 4697	6	Yes	49,190	38,289,787	38,338,976	+	54	22
<i>LOC777593</i>	7	No	95,196	61,639,308	61,734,503	-	25	25
Gene 3471	7	Yes	27,048	61,682,521	61,709,568	+	53	44
Gene 1878	8	Yes	44,329	53,302,486	53,346,814	-	41	4
<i>FOXB2</i>	8	No	2,022	53,346,672	53,348,693	+	13	0

In conclusion, we were able to use RNA-Seq data to identify SNPs in Brangus, Brahman, Angus, Nellore, and Holstein cattle. These SNPs were located in genes associated with puberty. Our comparative analysis enhances understanding of the sequence polymorphisms

among these breeds. Furthermore, description of alternative splice variants within RNA, and the detection of overlapping genes, are important for improving annotation of the bovine genome and understanding how these loci may influence puberty in heifers.

### Conflicts of interest

The authors declare no conflict of interest.

### ACKNOWLEDGMENTS

The authors thank the state funding agency Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP, Grant #2011/21241-0 and #2009/16118-5) for providing financial support for Nellore research, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) for the fellowship granted to the first author, and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) for the fellowship granted for the sandwich period to the first author.

We also thank the Colorado State University John E. Rouse Endowments in Animal Breeding and Genetics Research for their support of graduate students and faculty. We also thank Dr. K.R. Stenmark of the University of Colorado, Denver for sharing data resources from Holstein calves and Dr. K. Cammack (South Dakota State University), R.R. Cockrum (Virginia Tech University), and K. Austin (University of Wyoming) for their assistance obtaining Angus tissue samples. The authors also acknowledge the University of Queensland for sharing data and supporting Dr. M. R. S. Fortes with a postdoctoral fellowship.

### REFERENCES

- Bohmanova J, Sargolzaei M and Schenkel FS (2010). Characteristics of linkage disequilibrium in North American Holsteins. *BMC Genomics* 11: 421. <http://dx.doi.org/10.1186/1471-2164-11-421>
- Boi S, Solda G and Tenchini ML (2004). Shedding light on the dark side of the genome: overlapping genes in higher eukaryotes. *Curr. Genomics* 5: 509-524. <http://dx.doi.org/10.2174/1389202043349020>
- Bonger KM, van den Berg RJBHN, Knijnenburg AD, Heitman LH, et al. (2009). Discovery of selective luteinizing hormone receptor agonists using the bivalent ligand method. *ChemMedChem* 4: 1189-1195. <http://dx.doi.org/10.1002/cmdc.200900058>
- Brumatti RC, Ferraz JBS, Eler JP and Formigoni EIB (2011). Desenvolvimento de índice de seleção em gado corte sob o enfoque de um modelo bioeconômico. *Arch. Zootec.* 60: 205-213. <http://dx.doi.org/10.4321/S0004-05922011000200005>
- Cánovas A, Rincon G, Islas-Trejo A, Wickramasinghe S, et al. (2010). SNP discovery in the bovine milk transcriptome using RNA-Seq technology. *Mamm. Genome* 21: 592-598. <http://dx.doi.org/10.1007/s00335-010-9297-z>
- Cánovas A, Reverter A, DeAtley KL, Ashley RL, et al. (2014). Multi-tissue omics analyses reveal molecular regulatory networks for puberty in composite beef cattle. *PLoS One* 9: e102551. <http://dx.doi.org/10.1371/journal.pone.0102551>
- Cánovas A, Cockrum R, Brown D, Riddle S, et al. (2016). Functional genomics of high altitude disease in angus cattle: leveraging-OMICS and systems biology to better understanding of the function and role of key contributing genes. Available at [[http://www.isag.us/Docs/Proceedings/ISAG\\_Proceedings\\_2016.pdf](http://www.isag.us/Docs/Proceedings/ISAG_Proceedings_2016.pdf)] Accessed 20 February 2016.
- Cunningham F, Amode MR, Barrell D, Beal K, et al. (2015). Ensembl 2015. *Nucleic Acids Res.* 43: D662-D669. <http://dx.doi.org/10.1093/nar/gku1010>
- Damer CK, Bayeva M, Kim PS, Ho LK, et al. (2007). Copine A is required for cytokinesis, contractile vacuole function, and development in *Dictyostelium*. *Eukaryot. Cell* 6: 430-442. <http://dx.doi.org/10.1128/EC.00322-06>
- Day ML and Nogueira GP (2013). Management of age at puberty in beef heifers to optimize efficiency of beef production. *Anim. Front.* 3: 6-11. <http://dx.doi.org/10.2527/af.2013-0027>
- DeAtley KL (2012). Neuropeptidome of hypothalamus and pituitary gland of pre- and post-pubertal Brangus heifers. Doctoral thesis, New Mexico State University, Las Cruces.

- Dias MM, Souza FRP, Takada L, Feitosa FLB, et al. (2015). Study of lipid metabolism-related genes as candidate genes of sexual precocity in Nellore cattle. *Genet. Mol. Res.* 14: 234-243. <http://dx.doi.org/10.4238/2015.January.16.7>
- Elsik CG, Tellam RL, Worley KC, Gibbs RA, et al.; Bovine Genome Sequencing and Analysis Consortium (2009). The genome sequence of taurine cattle: a window to ruminant biology and evolution. *Science* 324: 522-528. <http://dx.doi.org/10.1126/science.1169588>
- Fortes MRS, Reverter A, Hawken RJ, Bolormaa S, et al. (2012). Candidate genes associated with testicular development, sperm quality, and hormone levels of inhibin, luteinizing hormone, and insulin-like growth factor 1 in Brahman bulls. *Biol. Reprod.* 87: 58. <http://dx.doi.org/10.1095/biolreprod.112.101089>
- Fortes MRS, Porto-Neto LR, DeAtley KL, Reverter A, et al. (2014). Genetic markers in transcription factors of differentially expressed genes associated with post-partum anoestrus predict pregnancy outcome in an independent population of beef cattle. In: Proceedings of the 10th World Congress of Genetics Applied to Livestock Production, Vancouver.
- Fortes MRS, Nguyen LT, Weller MMDC, Cánovas A, et al. (2016). Transcriptome analyses identify five transcription factors differentially expressed in the hypothalamus of post- versus prepubertal Brahman heifers. *J. Anim. Sci.* 94: 3693-3702. <http://dx.doi.org/10.2527/jas.2016-0471>
- Habier D, Fernando RL and Dekkers JCM (2009). Genomic selection using low-density marker panels. *Genetics* 182: 343-353. <http://dx.doi.org/10.1534/genetics.108.100289>
- Heard-Costa NL, Zillikens MC, Monda KL, Johansson A, et al. (2009). NRXN3 is a novel locus for waist circumference: a genome-wide association study from the CHARGE Consortium. *PLoS Genet.* 5: e1000539. <http://dx.doi.org/10.1371/journal.pgen.1000539>
- Jin Y and Lee CG (2013). Single nucleotide polymorphisms associated with microRNA regulation. *Biomolecules* 3: 287-302. <http://dx.doi.org/10.3390/biom3020287>
- Kaplowitz PB (2008). Link between body fat and the timing of puberty. *Pediatrics* 121 (Suppl 3): S208-S217. <http://dx.doi.org/10.1542/peds.2007-1813F>
- Kuersten S and Goodwin EB (2003). The power of the 3' UTR: translational control and development. *Nat. Rev. Genet.* 4: 626-637. <http://dx.doi.org/10.1038/nrg1125>
- Li M, Riddle S, Zhang H, D'Alessandro A, et al. (2016). Metabolic reprogramming regulates the proliferative and inflammatory phenotypes of adventitial fibroblasts in pulmonary hypertension through the transcriptional co-repressor C-terminal binding protein-1. *Circulation* 134: 1105-1121. <http://dx.doi.org/10.1161/CIRCULATIONAHA.116.023171>
- Matoulkova E, Michalova E, Vojtesek B and Hrstka R (2012). The role of the 3' untranslated region in post-transcriptional regulation of protein expression in mammalian cells. *RNA Biol.* 9: 563-576. <http://dx.doi.org/10.4161/rna.20231>
- Miyata H, Satouh Y, Mashiko D, Muto M, et al. (2015). Sperm calcineurin inhibition prevents mouse fertility with implications for male contraceptive. *Science* 350: 442-445. <http://dx.doi.org/10.1126/science.aad0836>
- Mutinati M, Desantis S, Rizzo A, Zizza S, et al. (2010). Localization of thyrotropin receptor and thyroglobulin in the bovine corpus luteum. *Anim. Reprod. Sci.* 118: 1-6. <http://dx.doi.org/10.1016/j.anireprosci.2009.05.019>
- Myatt SS and Lam EWF (2007). The emerging roles of forkhead box (Fox) proteins in cancer. *Nat. Rev. Cancer* 7: 847-859. <http://dx.doi.org/10.1038/nrc2223>
- Nakayama T, Asai S, Takahashi Y, Maekawa O, et al. (2007). Overlapping of genes in the human genome. *Int. J. Biomed. Sci.* 3: 14-19.
- Nascimento AV, Matos MC, Seno LO, Romero AR, et al. (2016). Genome wide association study on early puberty in *Bos indicus*. *Genet. Mol. Res.* 15: <http://dx.doi.org/10.4238/gmr.15017548>.
- Rocha A, Gómez A, Galay-Burgos M, Zanuy S, et al. (2007). Molecular characterization and seasonal changes in gonadal expression of a thyrotropin receptor in the European sea bass. *Gen. Comp. Endocrinol.* 152: 89-101. <http://dx.doi.org/10.1016/j.ygcen.2007.03.001>
- Rodrigues HD, Kinder JE and Fitzpatrick LA (2002). Estradiol regulation of luteinizing hormone secretion in heifers of two breed types that reach puberty at different ages. *Biol. Reprod.* 66: 603-609. <http://dx.doi.org/10.1095/biolreprod66.3.603>
- SanGiovanni JP and Lee PH (2013). AMD-associated genes encoding stress-activated MAPK pathway constituents are identified by interval-based enrichment analysis. *PLoS One* 8: e71239. <http://dx.doi.org/10.1371/journal.pone.0071239>
- Snelling WM, Cushman RA, Keele JW, Maltecca C, et al. (2013). Breeding and Genetics Symposium: networks and pathways to guide genomic selection. *J. Anim. Sci.* 91: 537-552. <http://dx.doi.org/10.2527/jas.2012-5784>
- Südhof TC (2008). Neuroligins and neurexins link synaptic function to cognitive disease. *Nature* 455: 903-911. <http://dx.doi.org/10.1038/nature07456>
- Thackray VG (2014). Fox tales: regulation of gonadotropin gene expression by forkhead transcription factors. *Mol. Cell. Endocrinol.* 385: 62-70. <http://dx.doi.org/10.1016/j.mce.2013.09.034>

- Tomsig JL, Snyder SL and Creutz CE (2003). Identification of targets for calcium signaling through the copine family of proteins. Characterization of a coiled-coil copine-binding motif. *J. Biol. Chem.* 278: 10048-10054. <http://dx.doi.org/10.1074/jbc.M212632200>
- Vinsky M, Islam K, Chen L and Li C (2013). Communication: Association analyses of a single nucleotide polymorphism in the promoter of OLR1 with growth, feed efficiency, fat deposition, and carcass merit traits. *J. Anim. Sci.* 93: 193-197.
- Wang KS, Zuo L, Pan Y, Xie C, et al. (2015). Genetic variants in the CPNE5 gene are associated with alcohol dependence and obesity in Caucasian populations. *J. Psychiatr. Res.* 71: 1-7. <http://dx.doi.org/10.1016/j.jpsychires.2015.09.008>
- Yang Z and Nielsen R (2000). Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol. Biol. Evol.* 17: 32-43. <http://dx.doi.org/10.1093/oxfordjournals.molbev.a026236>
- Zanni E, Laudenzi C, Schifano E, Palleschi C, et al. (2015). Impact of a complex food microbiota on energy metabolism in the model organism *Caenorhabditis elegans*. *BioMed Res. Int.* 2015: 621709. <http://dx.doi.org/10.1155/2015/621709>
- Zhou S, Goldstein S, Place M, Bechner M, et al. (2015). A clone-free, single molecule map of the domestic cow (*Bos taurus*) genome. *BMC Genomics* 16: 644. <http://dx.doi.org/10.1186/s12864-015-1823-7>

## Supplementary material

**Table S1.** SNP mutations already annotated in dbSNP database, accessed at Ensembl (2016) within the region or around the 25 genes reported by Cánovas et al. (2014).

**Table S2.** SNP mutations already annotated in the dbSNP database, accessed at Ensembl (2016) within the region or around the top twenty transcription factors genes reported by Cánovas et al. (2014).

**Table S3.** SNP mutations already annotated in the dbSNP database, accessed at Ensembl (2016) within the region or around the genes related with neuropeptides observed by DeAtley (2012).

**Table S4.** SNP mutations already annotated in the dbSNP database (Ensembl, 2016) within the region or around the genes from a study based on SNPs from high-density chip associated with puberty in Nellore (Dias et al., 2015).