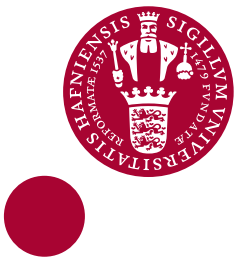

Copy number variations and genome-wide associations reveal putative genes and metabolic pathways involved with feed conversion ratio in beef cattle

Satana, Miguel Henrique de Almedia ; Junior, Gerson Antônio Oliveira ; Cesar, Aline Silva Mello; Freua, Mateus Castelani; Gomes, Rodrigo da Costra ; Silva, Saulo da Luz e ; Leme, Paulo Roberto; Fukumasu, Heidge; Carvalho, Minos Esperandio; Ventura, Ricardo Vierira; Coutinho, Luiz Lehmann; Kadarmideen, Haja; Ferraz, José Bento Sterman

Published in:
Journal of Applied Genetics

DOI:
10.1007/s13353-016-0344-7

Publication date:
2016

J Appl Genetics
DOI 10.1007/s13353-016-0344-7

ANIMAL GENETICS • ORIGINAL PAPER

Copy number variations and genome-wide associations reveal putative genes and metabolic pathways involved with the feed conversion ratio in beef cattle

Miguel Henrique de Almeida Santana[1,2] · **Gerson Antônio Oliveira Junior**[3] · **Aline Silva Mello Cesar**[3] · **Mateus Castelani Freua**[2] · **Rodrigo da Costa Gomes**[4] · **Saulo da Luz e Silva**[2] · **Paulo Roberto Leme**[2] · **Heidge Fukumasu**[2] · **Minos Esperândio Carvalho**[2] · **Ricardo Vieira Ventura**[2,5] · **Luiz Lehmann Coutinho**[6] · **Haja N. Kadarmideen**[1] · **José Bento Sterman Ferraz**[2]

Received: 29 September 2015 / Revised: 20 January 2016 / Accepted: 2 March 2016

Abstract The use of genome-wide association results combined with other genomic approaches may uncover genes and metabolic pathways related to complex traits. In this study, the phenotypic and genotypic data of 1475 Nellore (*Bos indicus*) cattle and 941,033 single nucleotide polymorphisms (SNPs) were used for genome-wide association study (GWAS) and copy number variations (CNVs) analysis in order to identify candidate genes and putative pathways involved with the feed conversion ratio (FCR). The GWAS was based on the Bayes B approach analyzing genomic windows with multiple regression models to estimate the proportion of genetic variance explained by each window. The CNVs were detected with PennCNV software using the log R ratio and B allele frequency data. CNV regions (CNVRs) were identified with CNVRuler and a linear regression was used to associate CNVRs and the FCR. Functional annotation of associated genomic regions was performed with the Database for Annotation, Visualization and Integrated Discovery (DAVID) and the metabolic pathways were obtained from the Kyoto Encyclopedia of Genes and Genomes (KEGG). We showed five genomic windows distributed over chromosomes 4, 6, 7, 8, and 24 that explain 12 % of the total genetic variance for FCR, and detected 12 CNVRs (chromosomes 1, 5, 7, 10, and 12) significantly associated [false discovery rate (FDR) < 0.05] with the FCR. Significant genomic regions (GWAS and CNV) harbor candidate genes involved in pathways related to energetic, lipid, and protein metabolism. The metabolic pathways found in this study are related to processes directly connected to feed efficiency in beef cattle. It was observed that, even though different genomic regions and genes were found between the two approaches (GWAS and CNV), the metabolic processes covered were related to each other. Therefore, a combination of the approaches complement each other and lead to a better understanding of the FCR.

Keywords CNV · Feed efficiency · Genomics · GWAS · Nellore cattle · SNPs

Communicated by: Maciej Szydlowski

✉ Miguel Henrique de Almeida Santana

[1] Faculty of Health and Medical Sciences, University of Copenhagen, Grønnegårdsvej 7, 1870 Frederiksberg, Denmark

[2] Faculdade de Zootecnia e Engenharia de Alimentos, University of São Paulo, Duque de Caxias Norte, 225, 13635-900 Pirassununga, Brazil

[3] Department of Animal Science, Iowa State University, Ames, IA 50011, USA

[4] Empresa Brasileira de Pesquisa Agropecuária, CNPGC/EMBRAPA, BR 262 km 4, 79002-970 Campo Grande, Brazil

[5] University of Guelph, 50 Stone Road East, Guelph, Ontario N1G 2W1, Canada

[6] Escola Superior de Agricultura Luiz de Queiroz, University of São Paulo, 13418-900 Piracicaba, Brazil

Introduction

Feeding cattle is a major cost in beef production and it directly affects the overall profitability of the meat industry. Several

Published online: 21 March 2016

strategies have been proposed to reduce this cost, focusing mainly on the improvement of feed efficiency. The feed conversion ratio (FCR) is a feed efficiency trait that measures the animal's capacity to convert feed consumed into the desired output (e.g., meat deposition or gained mass). The FCR is not a direct measurement, and it is computed as a function of the feed consumed, body weight gain, and duration of the trial (Arthur et al. 2001).

The discovery of DNA variants, such as single nucleotide polymorphisms (SNPs) and copy number variations (CNVs), which may be associated with the genetic variation of desired economic traits, has played an important role in livestock genetics (Kijas et al. 2011). In recent years, high-throughput technologies have led to the discovery of markers associated with economic traits by genome-wide association studies (GWAS), which allowed the identification of subsets of markers that explain an important portion of the variation of these traits (Barendse et al. 2007; Moore et al. 2009; Rolf et al. 2012). The use of genomic information can be a strategy for the improvement of interesting phenotypes such as the FCR by increasing the prediction accuracy of young animal candidates for genetic selection (Hayes et al. 2007), and, thus, accelerating genetic gain by reducing the generation interval. The association between markers and important phenotypes can be improved by using other genomic approaches, such as genomic prediction, to select animals (Bishop and Woolliams 2014; Kadarmideen 2014).

Additionally, several studies have reported the viability of using the information from SNPs to identify quantitative trait loci (QTL) and candidate genes associated with phenotypes of interest. Moreover, pathway analyses from GWAS results have been applied to aggregate information about genes and the physiology involved with important diseases and economical traits (Carbonetto and Stephens 2013), and to understand the molecular and physiological mechanisms involved in feed efficiency in beef cattle (Bolormaa et al. 2011; Snelling et al. 2011; Lu et al. 2013). Thus, the combination of GWAS with other approaches can be interesting to better explain these genes and pathways related to complex traits. There are alternative frameworks, such as CNVs, that can be useful to explain the variability and unveil the molecular architecture of complex traits (Hou et al. 2012b; Tamari et al. 2013; Bickhart and Liu 2014). The objective of this study was to identify candidate genes and putative pathways involved with the FCR in Nellore cattle (*Bos indicus*) from GWAS and CNV results.

Materials and methods

Ethical statement

No statement by the local ethics committee was required because the data used were from other experiments. DNA samples taken from each of the tests had been approved by the respective ethics committees of each study (Gomes et al. 2013; Santana et al. 2013; Alexandre et al. 2015).

Animals and phenotype

The phenotypic data of 1475 Nellore (*B. indicus*) young bulls and steers were used from 16 different studies focusing on feed efficiency conducted in Brazil from 2007 to 2013. The number of animals evaluated per test ranged from 45 to 120, and the animals had an average age of 574 ± 95 days and live weight of 381 ± 45 kg at the beginning of the experiments. These animals were from different breeding programs.

The tests were conducted in feedlots equipped with three different types of installation: two automated systems (GrowSafe and Calan Gates) and an individual pen system. Before testing, a period of adaptation to diets and facilities was conducted for no less than 21 days. Individual feed intake was measured daily for 70 to 90 days, with an average of 84 days. In addition, the feed was periodically analyzed for its chemical composition in order to adjust the dry matter intake (DMI). The diet was offered twice daily as total mixed ration. More details about the tests, diets, and managements appear in Gomes et al. (2013), Santana et al. (2013), and Alexandre et al. (2015).

During the experimental period, the animals were weighed regularly every 21 days to obtain the individual body weight (BW). These data were used to calculate the average daily gain (ADG), which was estimated as the slope of the linear regression of BW by individual experimental days. Feed efficiency was evaluated by the FCR, which was estimated by dividing the DMI by the ADG. Phenotypes (ADG, DMI, and FCR) were tested for normality (Shapiro–Wilk, $P < 0.05$) and the data that exceeded three standard deviations above or below the mean were considered outliers and excluded from further analyses.

Genotypes, imputation, and informativeness

The genotypic data from 3776 Nellore cattle were used in this study and these animals were genotyped with four different commercial products according to manufacturer: Illumina BovineSNP50® version 2 BeadChip (54,609 SNPs), Illumina BovineHD® Genotyping BeadChip (777,962 SNPs), GGP Indicus Neogen HD® (84,379 SNPs), and Affymetrix Axiom® Genome-Wide BOS 1 Array (648,874 SNPs). The genotypes were tested to ensure that they were determined correctly and clustering provided by the manufacturer was correct. First, only samples with genotype calls greater than 0.70 and call rate over 90 % were retained. Additionally, errors of duplicate samples were tested by calculating the proportion of alleles identical by state (IBS) of 10,000 SNPs randomly sampled, and all possible pairs of samples with IBS over 95 % were deleted.

In order to combine the genotypes of animals from these four different genotyping platforms and, thus, increase the sample size and marker density for association tests, a super-dense panel (SDP) was created and imputation was implemented. A total of 2604 animals were genotyped with Illumina BovineHD. Initially, the SDP (1,261,128 SNPs) was made by combining the genotypes of 279 (Affymetrix BOS 1) animals also genotyped using Illumina BovineHD (subset of 2604). Afterwards, imputation was performed for all animals with both genotypic and phenotypic data (1475 young bulls and steers). The imputation accuracy was tested by cross-validation for each SNP panel to the SDP in a parallel investigation and, for all scenarios, the concordance rate was higher than 97.5 % between the imputed and true genotypes. Imputation procedures were performed using FImpute 2.2 software (Sargolzaei et al. 2014).

After imputation, the genotypic data were evaluated for their informativeness and only SNPs in autosomal chromosomes (Chr) with minor allele frequency above 2 % and deviation from the Hardy–Weinberg equilibrium (χ^2-test, 1 *df*, $P > 1 \times 10^{-5}$) were kept. All quality control procedures were performed in R (R Development Core Team 2008). The final dataset had 1475 samples and 941,033 SNPs for the association test.

Genome-wide association study

Genome-wide associations for the FCR were made using Bayes B analyzing all SNP data simultaneously with different genetic variance for each SNP (Meuwissen et al. 2001; Habier et al. 2011). The prior genetic and residual variances were estimated using Bayes C (Garrick and Fernando 2013; Cesar et al. 2014), with π being 0.9999. The model equation was:

$$\boldsymbol{y} = \boldsymbol{Xb} + \sum_{j=1}^{k} \boldsymbol{a}_j \beta_j \delta_j + \boldsymbol{e}$$

where:

- **y** The vector of the phenotypic values
- **X** The incidence matrix for fixed effects
- **B** The vector of fixed effects defined above
- k The number of SNP loci (941,033)
- $\boldsymbol{a}_j$ The column vector representing the SNP covariate at locus j coded as the number of B alleles

β_j was the random substitution effect for locus j, which, conditional on σ^2_β, was assumed to be normally distributed $N(0, \sigma^2_\beta)$ when $\delta_j = 1$ but $\beta_j = 0$ when $\delta_j = 0$, with δ_j being a random 0/1 variable indicating the absence (with probability π) or presence (with probability $1 - \pi$) of locus j in the model, and e is the vector of the random residual effects assumed to be normally distributed $N(0, \sigma^2_e)$.

The variance σ^2_β (or σ^2_e) was *a priori* assumed to follow a scaled inverse Chi-square distribution with $\nu\beta = 4$ (or $\nu e = 10$) degrees of freedom and scale parameter S^2_β (or S^2_e), and these parameters for markers were derived as a function of the known genetic variance of the population. The genetic variance was based on the number of SNPs assumed to have non-zero effects based on parameter π being 0.9999 and the average SNP allele frequency. The posterior distributions of SNP effects was performed in GenSel software (Fernando and Garrick 2008) using Markov chain Monte Carlo (MCMC). The Bayesian multiple regression models with $\pi = 0.9999$ and about 80–105 SNP markers are fitted simultaneously in each MCMC iteration. Inference of associations in these multiple regression models is based on genomic windows rather than on single markers (Onteru et al. 2011), and these genomic windows were constructed based on the UMD3.1 bovine assembly.

In this study, 2527 windows (1 MB) were performed across the autosomal chromosomes. Samples of the proportion of genetic variance explained by each window were obtained by dividing the variance of the window by the variance of the whole genome in that iteration. The window was computed by multiplying the number of alleles that represent the SNP covariates for each consecutive SNP in a window by their sampled substitution effects in that iteration.

CNV detections and association analyses

The CNV analyses were performed in 2253 animals genotyped in Illumina BovineHD. The CNVs were identified by PennCNV (Wang et al. 2007), which uses the luminosity measure of log R ratio (LRR) and B allele frequency (BAF) in the predictions. The LRR is predicted from the ratio of the expected normalized intensity of a sample and observed normalized intensity, while the BAF is calculated from the difference between the expected position of the cluster group and the actual value (Winchester et al. 2009). A PennCNV perl script was used to eliminate calls from low-quality samples based on the standard deviation of LRR (>0.30), the default for BAF drift (>0.01), and waviness factor (>0.05).

The CNV regions (CNVRs) were determined by merging overlapping CNVs identified in two or more samples (Redon et al. 2006; Hou et al. 2012a). They were inferred by CNVRuler (Kim et al. 2012). Regions of very low density of overlapping (recurrence parameter <0.1) were not used in the analyses for a more robust definition of the beginning and end of regions. CNVRs with less than 5 % allele frequency were also not included in the association analyses. Linear regression was used to determine associations between CNVRs and the FCR. The false discovery rate (FDR) method was used for multiple comparison correction, where an FDR <0.05 was considered to have a potentially significant association.

Quantitative trait loci, candidate genes, and enrichment analysis

The top five SNP windows with the highest posterior mean proportion of genetic variance and the significant CNVR were considered the most important regions associated with FCR and were declared as promising QTL. Gene annotations were obtained from the Ensembl Genes 81 Database using BioMart software (Kinsella et al. 2011). Functional annotation was performed using the Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.7 (Huang et al. 2009) and pathway analyses were based on data available in the Kyoto Encyclopedia of Genes and Genomes (KEGG).

Results and discussion

There was no evidence ($P=0.38$, $P=0.47$, and $P=0.55$ for the FCR, DMI, and ADG, respectively) that the phenotypic data were not normally distributed according to the Shapiro–Wilk test. The top five associated genomic windows identified herein explained 6.92 %, 3.02 %, 0.83 %, 0.63 %, and 0.60 % of the genetic variance for the FCR, respectively, accounting for 12 % of the total genetic variance (Table 1).

These genomic regions are distributed over five different Chr: 4, 6, 7, 8, and 24 (Fig. 1) and harbor 135 genes based on the *Bos taurus* genome assembly UMD3.1. Of those 135 genes, 117 are annotated and used *in silico* for the functional analysis; the functional annotation tool uses gene-GO term enrichment analysis to look at the internal relationships among hundreds of other terms (Table 2).

The FCR is a measure of animal efficiency and relates the conversion of consumed feed into the desired output, such as muscle growth (protein accretion). Therefore, we identify QTL regions associated with the FCR that harbor important genes involved with pathways related to muscle development (myogenesis) and enhanced adipogenesis. The main pathways identified in GWAS were the JAK-STAT signaling pathway, cytokine–cytokine receptor interaction, ribosome and toll-like receptor signaling pathway. The cytokines is a class of endogenous feeding-regulatory substances (Plata-Salamán 2001) which activate the JAK-STAT pathway, positively or negatively regulating the differentiation of myoblasts (Jang and Baik 2013). The protein accretion and muscle growth is directly related to ribosome metabolism in that the protein synthesis is regulated by amino acids, which actively influence the protein synthesis by affecting the number of ribosomes, as reported by Wannemacher et al. (1971). On the other hand, toll-like receptors (TLRs) are associated with adipogenesis and insulin resistance (Senn 2006; Yan et al. 2010).

According to Animal QTLdb, the QTL region Chr4 overlaps with four (5271, 14675, 18437, and 18438) QTL previously reported for the FCR and two (18436 and 18435) for the ADG (Sherman et al. 2008, 2009, 2010). In this QTL region, two potential candidate genes were identified: Carnitine O-octanoyltransferase (*CROT*) and Insulin-like growth factor 2 mRNA binding protein 3 (*IGF2BP3*). These genes are associated with the generation of precursor metabolites and energy, fatty acid metabolism, carnitine metabolic process, and regulation of cytokine biosynthesis, respectively (Table 2). The relationship between IGFs proteins and feed efficiency is well known. These proteins have been considered as promising physiological markers for feed efficiency traits since the 1990s, mainly IGF1 and IGF2 (Stick et al. 1998; Arthur et al. 2004; Moore et al. 2005; Herd and Arthur 2009). An SNP in IGF2 was previously associated with the FCR in *Bos taurus* cattle (Sherman et al. 2008).

Chr6 at 81 Mb was previously described as a suggestive region of the signature of artificial selection for production purposes (Pérez O'Brien et al. 2014). In this region, the gene *TECRL* was identified (Table 2) and is associated with lipid production in the body, i.e., lipid metabolic process (GO:0006629) and oxidation–reduction process (GO:0055114). Another gene harbored in the region Chr6 at 81 Mb is *EPHA5* (Table 2), which is related to the regulation of insulin secretion involved in the cellular response to glucose stimulus (GO:0061178), which is also important to the lipid metabolism. In EphA5 knockout mice, behavioral changes were observed due to altered concentrations of serotonin (5-HT) and the metabolite 5-HIAA in the hypothalamus, resulting in an increase of body weight (Mamiya et al. 2008).

In the Chr7 QTL region, several candidate genes were identified as *COMP*, *CRLF1*, *UBA52*, *INSL3*, *PGPEP1*, *MPV17L2*, *IFI30*, *PIK3R2*, *JAK3*, *MAP1S*, *COLGALT1*, and *SLC27A1*, as well as one QTL previously described for the FCR and one for the DMI (Nkrumah et al. 2007; Lu et al. 2013). These genes are involved in biological processes such as lipid metabolism, apoptotic process (GO:0006915), cellular protein metabolic process (GO:0044267), carbohydrate metabolic process (GO:0005975), glycogen biosynthetic process (GO:0005978), fibroblast growth factor receptor signaling pathway (GO:0008543), protein transport (GO:0015031), and cellular response to insulin stimulus (GO:0032869). These biological processes are highly associated with muscle (myogenesis) and adipose (adipogenesis) development; these are mechanisms related to animal feed efficiency. Chr8 QTL harbors two miRNAs (bta-mir-873 and bta-mir-876) and three QTL for the ADG (Santana et al. 2014a) and Chr24 QTL harbors seven genes *DOK6*, *5S_rRNA*, *CCDC102B*, *TMX3*, *HIGD1D*, *DSEL*, and *CDH19*, which were not directly associated with molecular or biological processes related to feed efficiency.

A total of 139,089 CNVs were identified by PennCNV using 2253 HD genotypes, of which about 32 % (44,558) were non-redundant, unique CNVs. CNVs were not detected in approximately 28 % (622) of the animals. Gurgul et al.

Table 1 Genomic variance explained by quantitative trait loci (QTL) regions associated with the feed conversion ratio (FCR) in Nellore cattle by Bayes B

#SNP/window	Start position	End position	Chromosome	Variance (%)	Cumulative variance(%)
389	81011906	81997627	6	6.92	6.92
373	14004499	14998872	8	3.02	9.94
249	32000405	32995535	4	0.83	10.77
335	5004457	5999368	7	0.63	11.40
557	9007160	9999594	24	0.60	12.00

(2015) reported 33 % unique CNVs in 849 Holstein animals and 48 % in animals without detection. Several CNVRs (2667) were determined in autosomal chromosomes after CNVRuler analyses, representing 9 % of those chromosomes. The pattern of the different types of CNVRs was specific for each segment, with 1111 loss, 938 gain, and 617 mixed regions. Hou et al. (2012a), working with BovineHD SNP chip in 147 Holstein animals, reported 443 CNVRs and also more loss than gain regions, with a loss to gain ratio of 1.7. Likewise, Wu et al. (2015) found more losses than gain and mixed regions in Simmental cattle.

The detected CNVRs were used in the association analyses, resulting in 16 non-fixed regions, where 12 (chromosomes 1, 5, 7, 10, and 12) had significant association (FDR < 0.05) with the FCR (Table 3). All of them were defined as mixed CNVRs, meaning that the boundary of the regions was constituted by both "gain" and "loss" of CNVs.

Compared to SNPs, CNVs can be defined as a segment of DNA that displays copy number differences by comparison with reference genomes (Redon et al. 2006; Scherer et al. 2007; Liu et al. 2010). In other words, a CNV is the variation on the number of copies of a particular genomic region or gene from one individual to another, including changing gene structure and dosage, alternating gene regulation, and exposing recessive alleles (Zhang et al. 2009; Clop et al. 2012). The PennCNV algorithm is the most common in CNVs detection from SNP arrays (Kadri et al. 2012; Hou et al. 2012a; Xu et al. 2014b). It is also considered to be a software with relatively low false-positive rates, supporting the viability of these calls (Dellinger et al. 2010). Hou et al. (2012a) reported a high correlation between PennCNV estimations of copy number and qPCR copy number estimates.

The main goal of CNV detection is to identify their association with interested traits. Seroussi et al. (2010) were the first to work with complex traits in livestock, finding a significant association between CNVRs and protein and fat production, and herd life in Holstein cattle. Xu et al. (2014a) reported 33 candidate CNVRs using BovineSNP50 arrays in an Angus population. The authors reported one deletion in Chr7 associated with resistance to gastrointestinal nematodes. In another study, Xu et al. (2014a) conducted CNV analyses on a Holstein population and reported 34 CNVs on 22 chromosomes with significant association ($P < 0.05$ and FDR correction) with milk production traits.

Searching for genes in BioMart software (Kinsella et al. 2011) on these 12 significant regions resulted in 51 genes spread over four Chr (5, 7, 10, and 12). The metabolic pathways found in KEGG related with these genes were endocytosis (*FLJ20531*, *IKZF4*, *RNF126*, *ZC3H10*, *ZNF496*, *ZNF672*), glycerophospholipid metabolism (*DGKA*, *PPAP2C*), insulin signaling pathway (*SHC2*), histidine metabolism (*AMDHD1*, *HAL*), olfactory transduction (*OR10A7*, *OR10P1*, *OR11L1*, *OR14C36*, *OR2AJ1*, *OR2AK2*, *OR2AP1*, *OR2B11*, *OR2G2*, *OR2G3*, *OR2L13*, *OR2M4*, *OR2T1*, *OR2T6*, *OR6C1*, *OR6C2*, *OR6C75*, *OR6C76*, *OR6F1*, *OR6J1*), oxidative phosphorylation (*ATP5D*, *NDUFS7*), and

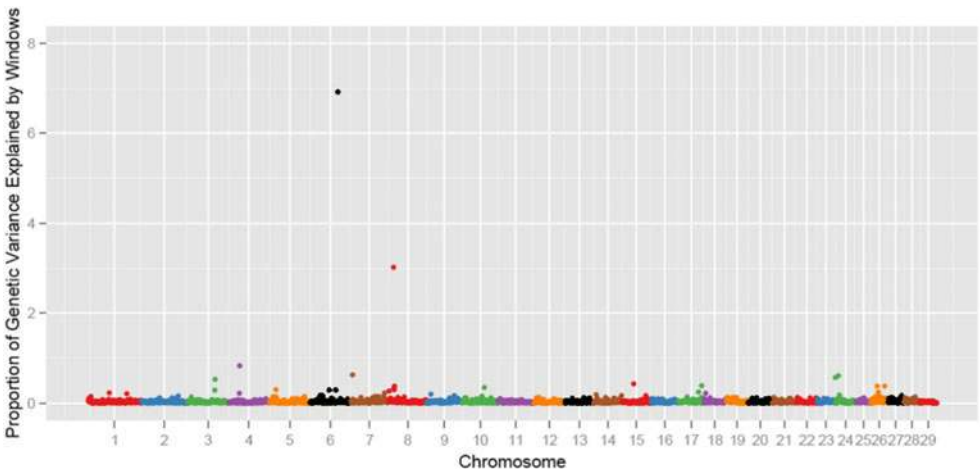

Fig. 1 Manhattan plot of the genome-wide association study (GWAS) for the feed conversion ratio (FCR) in Nellore cattle by Bayes B

Table 2 Symbol, name, and gene ontology of candidate genes identified from the genome-wide association study (GWAS) for the feed conversion ratio (FCR) in Nellore cattle

Candidate gene symbol	Name	Gene ontology/biological process
COLGALT1	Collagen beta(1-O)galactosyltransferase 1	Extracellular matrix organization (GO:0030198)
COMP	Cartilage oligomeric matrix protein	Skeletal system development (GO:0001501); growth plate cartilage development (GO:0003417); apoptotic process (GO:0006915); cell adhesion (GO:0007155); organ morphogenesis (GO:0009887); extracellular matrix organization (GO:0030198); negative regulation of apoptotic process (GO:0043066)
CRLF1	Cytokine receptor-like factor 1	Positive regulation of cell proliferation (GO:0008284); positive regulation of tyrosine phosphorylation of Stat3 protein (GO:0042517); negative regulation of neuron apoptotic process (GO:0043524)
CROT	Carnitine O-octanoyltransferase	Generation of precursor metabolites and energy GO:0006091; fatty acid metabolic process (GO:0006631); carnitine metabolic process (GO:0006635); fatty acid beta-oxidation (GO:0009437)
EPHA5	EPH receptor A5	Regulation of insulin secretion involved in cellular response to glucose stimulus (GO:0061178)
IFI30	Interferon, gamma-inducible protein 30	Cytokine-mediated signaling pathway (GO:0019221); negative regulation of fibroblast proliferation (GO:0048147); protein stabilization (GO:0050821); oxidation–reduction process (GO:0055114)
IGF2BP3	Insulin-like growth factor 2 mRNA binding protein 3	Regulation of cytokine biosynthetic process (GO:0042035); anatomical structure morphogenesis (GO:0009653)
INSL3	Insulin-like 3	Inositol biosynthetic process (GO:0006021); lipid metabolic process (GO:0006629); biological process (GO:0008150); phospholipid biosynthetic process (GO:0008654)
JAK3	Janus kinase 3	JAK-STAT cascade involved in growth hormone signaling pathway (GO:0060397)
MAP1S	Microtubule-associated protein 1S	Deoxyribonuclease activity (GO:0004536); protein binding (GO:0005515); microtubule binding (GO:0008017); microtubule binding (GO:0008017); tubulin binding (GO:0015631); beta-tubulin binding (GO:0048487); actin filament binding (GO:0051015); actin filament binding (GO:0051015)
MPV17L2	MPV17 mitochondrial membrane protein-like 2	Mitochondrial ribosome assembly (GO:0061668); positive regulation of mitochondrial translation (GO:0070131)
PGPEP1	Pyroglutamyl-peptidase I	Proteolysis (GO:0006508)
PIK3R2	Phosphoinositide-3-kinase, regulatory subunit 2	Cellular glucose homeostasis (GO:0001678); phospholipid metabolic process (GO:0006644); phosphatidylinositol biosynthetic process (GO:0006661); insulin receptor signaling pathway (GO:0008286); insulin receptor signaling pathway (GO:0008286); fibroblast growth factor receptor signaling pathway (GO:0008543); protein transport (GO:0015031); cellular response to insulin stimulus (GO:0032869); response to endoplasmic reticulum stress (GO:0034976); phosphatidylinositol-3-phosphate biosynthetic process (GO:0036092)
SLC27A1	Solute carrier family 27 (fatty acid transporter), member 1	Medium-chain fatty acid transport (GO:0001579); long-chain atty acid metabolic process (GO:0001676); phosphatidylethanolamine biosynthetic process (GO:0006646); phosphatidic acid biosynthetic process (GO:0006654); phosphatidylglycerol biosynthetic process (GO:0006655); phosphatidylcholine biosynthetic process (GO:0006656); phosphatidylserine biosynthetic process (GO:0006659); long-chain fatty acid transport (GO:0015909); response to insulin (GO:0032868); adiponectin-activated signaling pathway (GO:0033211); cellular lipid metabolic process

Table 2 (continued)

Candidate gene symbol	Name	Gene ontology/biological process
		(GO:0044255); positive regulation of protein serine/threonine kinase activity (GO:0071902)
TECRL	Trans-2,3-enoyl-CoA reductase-like	Lipid metabolic process (GO:0006629); oxidation–reduction process (GO:0055114)
UBA52	Ubiquitin A-52 residue ribosomal protein fusion product 1	G1/S transition of mitotic cell cycle (GO:0000082); G2/M transition of mitotic cell cycle (GO:0000086); activation of MAPK activity (GO:0000187); protein polyubiquitination (GO:0000209); mitotic cell cycle (GO:0000278); toll-like receptor signaling pathway (GO:0002224); MyD88-dependent toll-like receptor signaling pathway (GO:0002755); carbohydrate metabolic process (GO:0005975); glycogen biosynthetic process (GO:0005978); glucose metabolic process (GO:0006006); cellular protein metabolic process (GO:0044267)

retinol metabolism (*RDH5*). Retinol metabolism was related to the rump fat thickness in another GWAS in Nellore cattle. In that study, the authors indicated that this metabolic pathway participates in cell proliferation and differentiation, increased fat deposition, and release of growth hormone in the pituitary (Santana et al. 2015). Lipid and protein metabolism, which are related to body composition and protein turnover, are widely known and discussed as important bases for the physiology of feed efficiency (Richardson and Herd 2004; Moore et al. 2009; Herd and Arthur 2009). Glycerophospholipid and histidine pathways found in this study reinforce the physiological linkage of lipid and protein metabolism with feed efficiency. Likewise, the endocytosis pathway was previously associated with the DMI in Angus cattle (Rolf et al. 2012), and differences in the gene expression profile of Yorkshire pigs related to this molecular mechanism were also observed (Lkhagvadorj et al. 2010). Interestingly, all of the genes in this study for the endocytosis pathway are genes that are encoding zinc finger-type proteins. These proteins have been identified as candidate genes for the DMI in Nellore cattle (Santana et al. 2014b) and Holstein–Friesian dairy cows (Veerkamp et al. 2012).

Energy metabolism was also observed in our CNVR analysis by the enriched oxidative phosphorylation and insulin pathways. The importance of energy metabolism for feed efficiency is well known. Processes such as oxidative phosphorylation are essential for cellular energetic efficiency and, therefore, to the energy required by the animal for basal metabolic maintenance. The relation between energy maintenance and feed efficiency has been widely discussed (Castro Bulle et al. 2007; Hoque et al. 2009). In *Bos taurus* cattle, differential gene expression of the oxidative phosphorylation pathway with feed efficiency was demonstrated and the authors suggested that there is a relationship between cellular energetic efficiency and residual feed intake (Kelly et al. 2011).

Another interesting molecular process that was found in this study was an insulin pathway, one of the main hormones involved in energy and lipid metabolism, which plays an important role in controlling feed intake and efficiency in beef cattle (Richardson and Herd 2004; Rolf et al. 2012; Karisa et al. 2014). Several studies have shown the importance of insulin in feed intake. Do et al. (2014) found an association between insulin, olfactory response, and feed efficiency in pigs. The authors suggested that insulin secretion could be an intermediate stimulus to the olfactory pathway influencing feed efficiency. In fact, it has been shown that insulin levels may modulate the response of odor perception by smell (Palouzier-Paulignan et al. 2012), and this perception may indicate the importance of genes that affect smell and taste in cattle intake (Veerkamp et al. 2012). Moreover, the genes linked to the olfactory pathway may participate in nutrient uptake by acting as chemical sensors in the gut (Veerkamp et al. 2012). In this study, most of the genes found in the CNVR

Table 3 Copy number variation regions (CNVRs) associated with the FCR

CNVR	Chromosome	Start	End	Size	FDR
1	5	58386640	58441130	54491	<0.01*
2	12	74844575	74942860	98286	<0.01*
3	5	117356476	117639815	283340	<0.01*
4	5	59421039	59627471	206433	<0.01*
5	12	75016673	75141024	124352	<0.01*
6	7	44437375	44444959	7585	<0.01*
7	12	72748544	72882991	134448	<0.01*
8	10	22709100	22952910	243811	0.01*
9	1	93730576	93819471	88896	0.01*
10	12	73657824	73666963	9140	0.01*
11	12	72518489	72739627	221139	0.02*
12	7	42736530	43353211	616682	0.04*

FDR False discovery rate

*FDR < 0.05

analysis were linked to the olfactory pathway. Two other studies have also found the potential relationship of the olfactory pathway with feed intake in cattle (Veerkamp et al. 2012; Lindholm-Perry et al. 2015).

Combining different methods of analysis (e.g., GWAS with CNV) can identify more effectively the genomic regions and pathways associated with complex traits (Ritchie et al. 2015), such as feed efficiency. This genome-wide study identified QTL and CNV regions that were associated with the FCR. Our results showed that the two approaches found distinct genomic regions, but the identified genes are part of the same biological processes, mainly lipid, protein, and energetic metabolism. These findings contribute to the knowledge of the genetic basis of feed efficiency. The benefits of using more than one type of approach has been demonstrated, especially when it is possible to integrate multi-omics data in a systems biology approach (Ritchie et al. 2015). This integration can be critical for a full understanding of complex traits from the physiological point of view. More biologically interesting and promising results could be obtained using a multi-stage multi-omic approach (Kadarmideen 2014), which includes genome, epigenome, transcriptome, proteome, and metabolome in the same integrated analysis.

Acknowledgments This research was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP, 2012/02039-9, 2013/20571-2, 2014/14121-7) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, 442345/2014-3). The contributions of Núcleo de Criadores de Nelore do Norte do Paraná, Luciano Borges (Rancho da Matinha), and Eduardo Penteado Cardoso (Faz. Mundo Novo) are gratefully acknowledged.

References

Alexandre PA, Kogelman LJA, Santana MHA et al 2015. Liver transcriptomic networks reveal main biological processes associated with feed efficiency in beef cattle. BMC Genomics 16:1073 doi:10.1186/s12864-015-2292-8

Arthur PF, Archer JA, Johnston DJ, Herd RM, Richardson EC, Parnell PF (2001) Genetic and phenotypic variance and covariance components for feed intake, feed efficiency, and other postweaning traits in Angus cattle. J Anim Sci 79:2805–2811

Arthur PF, Archer JA, Herd RM (2004) Feed intake and efficiency in beef cattle: overview of recent Australian research and challenges for the future. Aust J Exp Agric 44:361–369. doi:10.1071/EA02162

Barendse W, Reverter A, Bunch RJ, Harrison BE, Barris W, Thomas MB (2007) A validated whole-genome association study of efficient food conversion in cattle. Genetics 176:1893–1905. doi:10.1534/genetics.107.072637

Bickhart DM, Liu GE (2014) The challenges and importance of structural variation detection in livestock. Front Genet 5:37. doi:10.3389/fgene.2014.00037

Bishop SC, Woolliams JA (2014) Genomics and disease resistance studies in livestock. Livest Sci 166:190–198. doi:10.1016/j.livsci.2014.04.034

Bolormaa S, Hayes BJ, Savin K et al (2011) Genome-wide association studies for feedlot and growth traits in cattle. J Anim Sci 89:1684–1697. doi:10.2527/jas.2010-3079

Carbonetto P, Stephens M (2013) Integrated enrichment analysis of variants and pathways in genome-wide association studies indicates central role for IL-2 signaling genes in type 1 diabetes, and cytokine signaling genes in Crohn's disease. PLoS Genet 9:e1003770. doi:10.1371/journal.pgen.1003770

Castro Bulle FCP, Paulino PV, Sanches AC, Sainz RD (2007) Growth, carcass quality, and protein and energy metabolism in beef cattle with different growth potentials and residual feed intakes. J Anim Sci 85:928–936. doi:10.2527/jas.2006-373

Cesar ASM, Regitano LCA, Mourão GB et al (2014) Genome-wide association study for intramuscular fat deposition and composition in Nellore cattle. BMC Genet 15:39. doi:10.1186/1471-2156-15-39

Clop A, Vidal O, Amills M (2012) Copy number variation in the genomes of domestic animals. Anim Genet 43:503–517. doi:10.1111/j.1365-2052.2012.02317.x

Dellinger AE, Saw SM, Goh LK, Seielstad M, Young TL, Li YJ (2010) Comparative analyses of seven algorithms for copy number variant identification from single nucleotide polymorphism arrays. Nucleic Acids Res 38:1–14. doi:10.1093/nar/gkq040

Do DN, Strathe AB, Ostersen T, Pant SD, Kadarmideen HN (2014) Genome-wide association and pathway analysis of feed efficiency in pigs reveal candidate genes and pathways for residual feed intake. Front Genet 5:307. doi:10.3389/fgene.2014.00307

Fernando RL, Garrick DJ (2008) GenSel—User manual for a portfolio of genomic selection related analyses

Garrick DJ, Fernando RL (2013) Implementing a QTL detection study (GWAS) using genomic prediction methodology. Methods Mol Biol 1019:275–298. doi:10.1007/978-1-62703-447-0_11

Gomes RC, Silva SL, Carvalho ME et al (2013) Protein synthesis and degradation gene SNPs related to feed intake, feed efficiency, growth, and ultrasound carcass traits in Nellore cattle. Genet Mol Res 12:2923–2936

Gurgul A, Jasielczuk I, Szmatoła T et al (2015) Genome-wide characteristics of copy number variation in Polish Holstein and Polish Red cattle using SNP genotyping assay. Genetica 143(2):145–155. doi:10.1007/s10709-015-9822-9

Habier D, Fernando RL, Kizilkaya K, Garrick DJ (2011) Extension of the bayesian alphabet for genomic selection. BMC Bioinformatics 12:186. doi:10.1186/1471-2105-12-186

Hayes BJ, Chamberlain AJ, McPartlan H, Macleod I, Sethuraman L, Goddard ME (2007) Accuracy of marker-assisted selection with single markers and marker haplotypes in cattle. Genet Res 89:215–220. doi:10.1017/S0016672307008865

Herd RM, Arthur PF (2009) Physiological basis for residual feed intake. J Anim Sci 87:E64–E71. doi:10.2527/jas.2008-1345

Hoque MA, Hosono M, Oikawa T, Suzuki K (2009) Genetic parameters for measures of energetic efficiency of bulls and their relationships with carcass traits of field progeny in Japanese Black cattle. J Anim Sci 87:99–106. doi:10.2527/jas.2007-0766

Hou Y, Bickhart DM, Chung H et al (2012a) Analysis of copy number variations in Holstein cows identify potential mechanisms contributing to differences in residual feed intake. Funct Integr Genomics 12:717–723. doi:10.1007/s10142-012-0295-y

Hou Y, Bickhart DM, Hvinden ML et al (2012b) Fine mapping of copy number variations on two cattle genome assemblies using high density SNP array. BMC Genomics 13:376. doi:10.1186/1471-2164-13-376

Huang DW, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4:44–57. doi:10.1038/nprot.2008.211

Jang Y-N, Baik EJ (2013) JAK-STAT pathway and myogenic differentiation. JAK-STAT 2:e23282. doi:10.4161/jkst.23282

Kadarmideen HN (2014) Genomics to systems biology in animal and veterinary sciences: Progress, lessons and opportunities. Livest Sci 166:232–248. doi:10.1016/j.livsci.2014.04.028

Kadri NK, Koks PD, Meuwissen TH (2012) Prediction of a deletion copy number variant by a dense SNP panel. Genet Sel Evol 44:7. doi:10.1186/1297-9686-44-7

Karisa B, Moore S, Plastow G (2014) Analysis of biological networks and biological pathways associated with residual feed intake in beef cattle. Anim Sci J 85:374–387. doi:10.1111/asj.12159

Kelly AK, Waters SM, McGee M, Fonseca RG, Carberry C, Kenny DA (2011) mRNA expression of genes regulating oxidative phosphorylation in the muscle of beef cattle divergently ranked on residual feed intake. Physiol Genomics 43:12–23. doi:10.1152/physiolgenomics.00213.2009

Kijas JW, Barendse W, Barris W et al (2011) Analysis of copy number variants in the cattle genome. Gene 482:73–77. doi:10.1016/j.gene.2011.04.011

Kim J-H, Hu H-J, Yim S-H, Bae JS, Kim S-Y, Chung Y-J (2012) CNVRuler: a copy number variation-based case–control association analysis tool. Bioinformatics 28:1790–1792. doi:10.1093/bioinformatics/bts239

Kinsella RJ, Kähäri A, Haider S et al (2011) Ensembl BioMarts: a hub for data retrieval across taxonomic space. Database (Oxford) 2011: bar030. doi: 10.1093/database/bar030

Lindholm-Perry AK, Kern RJ, Kuehn LA et al (2015) Differences in transcript abundance of genes on BTA15 located within a region associated with gain in beef steers. Gene 572:42–48. doi:10.1016/j.gene.2015.06.076

Liu GE, Hou Y, Zhu B et al (2010) Analysis of copy number variations among diverse cattle breeds. Genome Res 20:693–703. doi:10.1101/gr.105403.110

Lkhagvadorj S, Qu L, Cai W et al (2010) Gene expression profiling of the short-term adaptive response to acute caloric restriction in liver and adipose tissues of pigs differing in feed efficiency. Am J Physiol Regul Integr Comp Physiol 298:R494–R507. doi:10.1152/ajpregu.00632.2009

Lu D, Miller S, Sargolzaei M et al (2013) Genome-wide association analyses for growth and feed efficiency traits in beef cattle. J Anim Sci 91:3612–3633. doi:10.2527/jas.2012-5716

Mamiya PC, Hennesy Z, Zhou R, Wagner GC (2008) Changes in attack behavior and activity in EphA5 knockout mice. Brain Res 1205:91–99. doi:10.1016/j.brainres.2008.02.047

Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829

Moore KL, Johnston DJ, Graser H-U, Herd R (2005) Genetic and phenotypic relationships between insulin-like growth factor-I (IGF-I) and net feed intake, fat, and growth traits in Angus beef cattle. Aust J Agric Res 56:211–218. doi:10.1071/AR04248

Moore SS, Mujibi FD, Sherman EL (2009) Molecular basis for residual feed intake in beef cattle. J Anim Sci 87:E41–E47. doi:10.2527/jas.2008-1418

Nkrumah JD, Sherman EL, Li C et al (2007) Primary genome scan to identify putative quantitative trait loci for feedlot growth rate, feed intake, and feed efficiency of beef cattle. J Anim Sci 85:3170–3181. doi:10.2527/jas.2007-0234

Onteru SK, Fan B, Nikkilä MT, Garrick DJ, Stalder KJ, Rothschild MF (2011) Whole-genome association analyses for lifetime reproductive traits in the pig. J Anim Sci 89:988–995. doi:10.2527/jas.2010-3236

Palouzier-Paulignan B, Lacroix M-C, Aimé P et al (2012) Olfaction under metabolic influences. Chem Senses 37:769–797. doi:10.1093/chemse/bjs059

Pérez O'Brien AM, Utsunomiya YT, Mészáros G et al (2014) Assessing signatures of selection through variation in linkage disequilibrium between taurine and indicine cattle. Genet Sel Evol 46:19. doi:10.1186/1297-9686-46-19

Plata-Salamán CR (2001) Cytokines and feeding. Int J Obes Relat Metab Disord 25(Suppl 5):S48–S52. doi:10.1038/sj.ijo.0801911

R Development Core Team (2008) R: A language and environment for statistical computing

Redon R, Ishikawa S, Fitch KR et al (2006) Global variation in copy number in the human genome. Nature 444:444–454. doi:10.1038/nature05329

Richardson EC, Herd RM (2004) Biological basis for variation in residual feed intake in beef cattle. 2. Synthesis of results following divergent selection. Aust J Exp Agric 44:431–440. doi:10.1071/EA02221

Ritchie MD, Holzinger ER, Li R, Pendergrass SA, Kim D (2015) Methods of integrating data to uncover genotype–phenotype interactions. Nat Rev Genet 16:85–97. doi:10.1038/nrg3868

Rolf MM, Taylor JF, Schnabel RD et al (2012) Genome-wide association analysis for feed efficiency in Angus cattle. Anim Genet 43:367–374. doi:10.1111/j.1365-2052.2011.02273.x

Santana MHA, Rossi Junior P, Almeida RD, Schuntzemberger AMDS (2013) Blood cell and metabolic profile of Nellore bulls and their correlations with residual feed intake and feed conversion ratio. Rev Bras Saúde e Produção Anim 14:527–537. doi:10.1590/S1519-99402013000300018

Santana MHA, Utsunomiya YT, Neves HHR et al (2014a) Genome-wide association study for feedlot average daily gain in Nellore cattle (Bos indicus). J Anim Breed Genet 131:210–216. doi:10.1111/jbg.12084

Santana MHA, Utsunomiya YT, Neves HHR et al (2014b) Genome-wide association analysis of feed intake and residual feed intake in Nellore cattle. BMC Genet 15:21. doi:10.1186/1471-2156-15-21

Santana MHA, Ventura RV, Utsunomiya YT et al (2015) A genomewide association mapping study using ultrasound-scanned information identifies potential genomic regions and candidate genes affecting carcass traits in Nellore cattle. J Anim Breed Genet 132:420–427. doi:10.1111/jbg.12167

Sargolzaei M, Chesnais JP, Schenkel FS (2014) A new approach for efficient genotype imputation using information from relatives. BMC Genomics 15:478. doi:10.1186/1471-2164-15-478

Scherer SW, Lee C, Birney E et al (2007) Challenges and standards in integrating surveys of structural variation. Nat Genet 39:S7–S15. doi:10.1038/ng2093

Senn JJ (2006) Toll-like receptor-2 is essential for the development of palmitate-induced insulin resistance in myotubes. J Biol Chem 281:26865–26875. doi:10.1074/jbc.M513304200

Seroussi E, Glick G, Shirak A et al (2010) Analysis of copy loss and gain variations in Holstein cattle autosomes using BeadChip SNPs. BMC Genomics 11:673. doi:10.1186/1471-2164-11-673

Sherman EL, Nkrumah JD, Murdoch BM et al (2008) Polymorphisms and haplotypes in the bovine neuropeptide Y, growth hormone receptor, ghrelin, insulin-like growth factor 2, and uncoupling proteins 2 and 3 genes and their associations with measures of growth, performance, feed efficiency, and carcass merit in beef cattle. J Anim Sci 86:1–16. doi:10.2527/jas.2006-799

Sherman EL, Nkrumah JD, Li C, Bartusiak R, Murdoch B, Moore SS (2009) Fine mapping quantitative trait loci for feed intake and feed efficiency in beef cattle. J Anim Sci 87:37–45. doi:10.2527/jas.2008-0876

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. doi:10.2527/jas.2008-1759

Snelling WM, Allan MF, Keele JW et al (2011) Partial-genome evaluation of postweaning feed intake and efficiency of crossbred beef cattle. J Anim Sci 89:1731–1741. doi:10.2527/jas.2010-3526

Stick DA, Davis ME, Loerch SC, Simmen RC (1998) Relationship between blood serum insulin-like growth factor I concentration and

postweaning feed efficiency of crossbred cattle at three levels of dietary intake. J Anim Sci 76:498–505

Tamari M, Tanaka S, Hirota T (2013) Genome-wide association studies of allergic diseases. Allergol Int 62:21–28. doi:10.2332/allergolint.13-RAI-0539

Veerkamp RF, Coffey MP, Berry DP et al (2012) Genome-wide associations for feed utilisation complex in primiparous Holstein-Friesian dairy cows from experimental research herds in four European countries. Animal 6:1738–1749. doi:10.1017/S1751731112001152

Wang K, Li M, Hadley D et al (2007) PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. Genome Res 17:1665–1674. doi:10.1101/gr.6861907

Wannemacher RW Jr, Wannemacher CF, Yatvin MB (1971) Amino acid regulation of synthesis of ribonucleic acid and protein in the liver of rats. Biochem J 124:385–392

Winchester L, Yau C, Ragoussis J (2009) Comparing CNV detection methods for SNP arrays. Brief Funct Genomic Proteomic 8:353–366. doi:10.1093/bfgp/elp017

Wu Y, Fan H, Jing S et al (2015) A genome-wide scan for copy number variations using high-density single nucleotide polymorphism array in Simmental cattle. Anim Genet 46(3):289–298. doi:10.1111/age.12288

Xu L, Cole JB, Bickhart DM et al (2014a) Genome wide CNV analysis reveals additional variants associated with milk production traits in Holsteins. BMC Genomics 15:683. doi:10.1186/1471-2164-15-683

Xu L, Hou Y, Bickhart DM et al (2014b) A genome-wide survey reveals a deletion polymorphism associated with resistance to gastrointestinal nematodes in Angus cattle. Funct Integr Genomics 14:333–339. doi:10.1007/s10142-014-0371-6

Yan X, Zhu MJ, Xu W et al (2010) Up-regulation of Toll-like receptor 4/nuclear factor-kappaB signaling is associated with enhanced adipogenesis and insulin resistance in fetal skeletal muscle of obese sheep at late gestation. Endocrinology 151:380–387. doi:10.1210/en.2009-0849

Zhang F, Gu W, Hurles ME, Lupski JR (2009) Copy number variation in human health, disease, and evolution. Annu Rev Genomics Hum Genet 10:451–481. doi:10.1146/annurev.genom.9.081307.164217