EXPLORING THE REGULATORY POTENTIAL OF LONG NON-CODING RNA IN BOVINE FEED EFFICIENCY THROUGH COEXPRESSION IN LIVER AND MUSCLE

P.A. Alexandre[1,2], A. Reverter[2], L.R. Porto-Neto[2], R.B. Berezin[1], J.B.S. Ferraz[1] and H. Fukumasu[1]

[1]College of Animal Sciences and Food Engineering, USP, Pirassununga, SP, Brazil
[2]CSIRO Agriculture & Food, St. Lucia, Brisbane, QLD, 4067 Australia

SUMMARY

Long non-coding RNA (lncRNA) control gene expression through a variety of processes, thus, can be central regulators of important traits in livestock production, such as feed efficiency (FE). With the aim to identify regulatory lncRNA in liver and muscle of bulls divergent for FE, we used RNAseq data to obtain a list of novel lncRNA in bovine. Then, lncRNA in liver and muscle of high and low FE bulls were tested for differential expression (DE) and their potential to regulate gene expression in those tissues assessed through a regulatory impact factor algorithm. A total of 102 lncRNA were DE in both tissues and 126 were considered regulators of gene expression. We were able to highlight at least three lncRNA with the potential to be further explored in our ongoing study. The variety of mechanisms by which lncRNA influences gene expression make these molecules of particular interest to be validated as central regulators of FE.

INTRODUCTION

The flow of information from DNA to protein synthesis comprises several steps in which gene expression can be controlled, and lncRNA (long non-coding RNAs) are involved in several of them. Structurally, lncRNA have low or no potential for protein-coding and low conservation among species. In contrast to protein-coding genes, lncRNA are more tissue-specific and have a lower level of expression, although they also present multiple exons, polyA tail, 5 ‘cap and may contain CpG islands in their promoter regions (Weikard *et al.* 2016).

Long non-coding RNAs are involved in the regulation of gene expression either by recruiting chromatin-modifying complexes, regulating DNA methylation levels, modulating allele-specific expression or functioning as guides, precursors or sequester of miRNAs. They can also act as sequester of transcription factors or as scaffolds of multiprotein complexes. They act in the alternative splicing, in enzymes activity and can either be exported and act in other cells. In gene networks, they act as regulators, being expressed in coordinated pattern with other genes. This variety of lncRNA mechanisms of action confer these molecules great potential to regulate important traits for animal production, such as feed efficiency (FE), a proxy for both increased productivity and reduced environmental impact.

In cattle, a catalogue of lncRNA in multiple tissues was published by Koufariotis *et al*. (2015), but there is little overlap between those and lncRNA found by other studies (Weikard *et al.* 2016). This makes the tissue/cell/time/condition-specific character of the lncRNA evident and highlights the need to study these molecules in the context of FE. The objective of this study was to identify lncRNAs with regulatory potential for FE in beef cattle that could be explored as biomarkers of this trait.

MATERIALS AND METHODS

Tissue Sampling and Bioinformatics Analyses. We used RNAseq data (Illumina HiSeq 2500, 100bp, paired-end) from our database which comprises 86 samples from liver, skeletal muscle, adrenal gland, hypothalamus and pituitary of high and low FE Nellore (*Bos indicus*) bulls. Refer to

Alexandre *et al*. (2019) for details of the experimental resource and bioinformatics pipeline protocol for sequence alignments. In brief, Cufflinks pipeline (Ghosh and Chan 2016) was used to generate an annotation file from which new transcripts were selected and filtered to exclude those smaller than 200bp. Then, transcripts presenting open reading frame greater than 300bp between a START and a STOP codon were excluded. Transcripts were tested for similarity to the UniProt/SwissProt database using BLAST+ and discarded if E-value $<10^{-6}$. A final check was performed to consider as novel lncRNA only transcripts presenting more than 1 exon. Cuffcompare was used to compare novel transcripts to the NCBI annotation file containing predicted genes by the GNOMON method. Novel lncRNA were also compared to the bovine NONCODE database using BLAST+ and were considered significant when E-value<10-6.

Differential Expression and Gene Network Inference. Using data from liver and muscle RNAseq, read counts were generated by HTseq and expression values were estimated as fragments per kilobase of transcript per million mapped reads (FPKM). Genes with average expression lower than 0.2 FPKM across all samples in the two tissues were discarded. Differential expression (DE) analysis was performed as described in Alexandre *et al.* (2019). Genes were considered DE when the difference between the expression in high FE and low FE groups were greater than ±2.57 SD from the mean, corresponding to a t-test $P < 0.01$. In order to identify lncRNA with regulatory potential, the regulatory impact factor (RIF) metric was calculated considering as targets the totality of 14,466 mRNAs expressed in liver and muscle (Alexandre *et al*. 2019), and the DE lncRNA as potential regulators. Those scores deviating ±2.57 SD from the mean (corresponding to $P < 0.01$) were considered significant. For gene network inference, we input all the mRNAs plus the DE lncRNA significant according to RIF into the PCIT algorithm in order to identify significant connections that would establish an edge in the reconstruction of the network. Only edges involving at least one DE lncRNA were considered for the network visualisation and analysis. We used Cytoscape (Shannon *et al*. 2003) to visualise the network and BINGO plugin (Maere *et al.* 2005) to perform functional enrichment, using the whole genome as background, hypergeometric test and Benjamini–Hochberg FDR correction ($P < 0.05$).

RESULTS AND DISCUSSION

We were able to initially identify 174,793 novel transcripts of which 174,431 (99.8%) presented length greater than 200bp, a criterion used to distinguish small ncRNA from lncRNA. The filter that most excluded transcripts was ORFs greater than 300nt, maintaining only 8.5% (14,904) of the transcripts initially identified. Then, 5,865 transcripts were excluded because they presented high similarity with known proteins and 12 were excluded because they presented coding potential. From the resulting 9,027 transcripts, only 4,603 presented more than one exon and were considered as novel lncRNA.

Among the 4,603 transcripts identified as lncRNA, the most represented class was u - intergenic transcripts (Figure 1a), which is in accordance with the literature. However, when we compared the lncRNA identified with predicted genes in the bovine genome annotation, we found that a large portion of them are, actually, new isoforms (class j) or overlap exon of a predicted gene (class o). A few of them (6%) overlap perfectly with predicted genes (Figure 1b).

A total of 2,677 (58%) lncRNA presented high similarity with NONCODEv5, suggesting that a large part of those have already been identified previously. While that percentage indicates a certain level of efficiency in the identification of lncRNA, it also reflects a deficit in the annotation of these molecules, especially considering differences between the subspecies *taurus* and *indicus*, as most of the experiments for characterization of lncRNA were performed in taurine breeds. Given the regulatory role of this class of RNAs and their apparent lack of conservation among species, it is expected that important differences exist between zebu and taurine cattle, especially regarding adaptive responses.

Differential expression analysis was performed using all the 2,688 expressed lncRNA and resulted in 58 DE lncRNA in liver and 46 in skeletal muscle (Figure 2), being 42 up-regulated and 60 down-regulated in high FE (P<0.01), as two DE transcripts in both tissues were concordant. Thirty-eight of the DE lncRNA were isoforms of known/predicted genes and one, TCONS00029353, draws attention for being DE in muscle and a non-coding isoform of TSC22D1, a gene stimulated by transforming growth factor beta (TGFB) that regulates the transcription of multiple genes. In fact, TGFB1 was recently pointed as a central regulator of FE in muscle (Alexandre *et al.* 2019), making that lncRNA of particular interest. Two transcripts completely overlapped predicted genes LOC101905381 and LOC104970784 which are indeed described as a non-coding RNA, but there is no literature discussing their potential function.

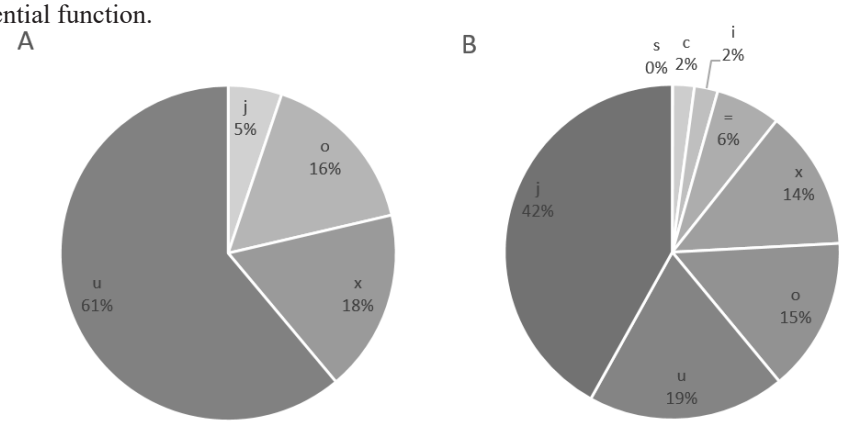

Figure 1. Distribution of the lncRNA identified according to their classification. A) Only annotation of known genes and B) Annotation including genes predicted by Gnomon (NCBI): = - overlap with known gene; c - contained in a known gene; i – overlap with known intron; j - new isoform; o - generic overlap with known exon; s - contains intron that overlaps known intron on the opposite string; u - intergenic transcript; x – overlap with known gene on opposite string

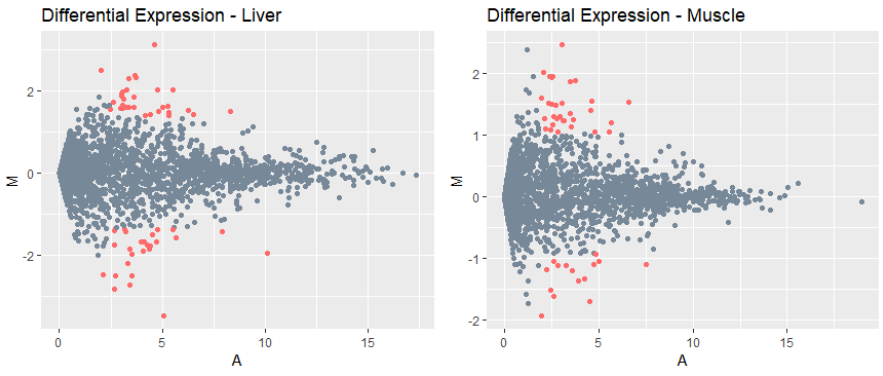

Figure 2. MA plots of 2,688 lncRNA expressed in liver and muscle between high and low feed efficiency (FE). Differential expression (M, y-axis) is plotted against average expression (A, x-axis). Pink dots represent statistically significant DE lncRNA (P<0.01)

A total of 8 lncRNA were considered regulators and a co-expression network containing their directly co-expressed mRNAs was created (Figure 3). Of these, two lncRNA are co-expressed with

most of the genes in the network and the functional enrichment of their co-expressed genes showed that these transcripts are related to biological functions already previously associated with FE in these tissues (Alexandre *et al*. 2019), supporting their potential as regulators of FE. TCONS00030954 is an intergenic transcript up-regulated in the liver of high FE animals and their co-expressed genes are related to translation (adjusted $P=3.8x10^{-7}$) and gene expression (adjusted $P=4.9x10^{-5}$). Moreover, it is co-expressed with NR2F6, also recently pointed as a central regulator of FE in liver. TCONS00089983 overlaps PTMA gene in the opposite string and is also up-regulated in the muscle of high FE bulls. Its co-expressed genes are enriched mainly for protein metabolic process (adjusted $P=1.1x10^{-2}$).

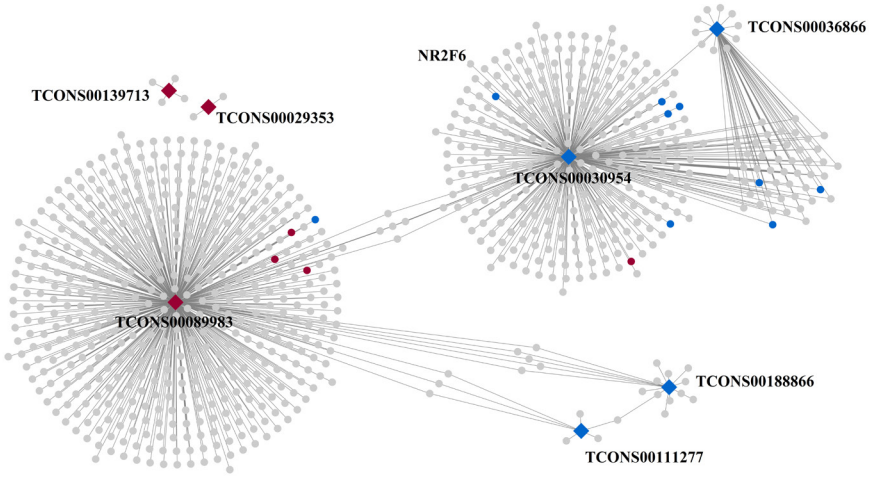

Figure 3. Gene co-expression network of eight lncRNA that are differentially expressed and with regulatory potential, and their 878 directly co-expressed mRNA. Ellipses represent mRNA, diamonds represent lncRNA; blue and red nodes represent differential expression in liver and muscle, respectively. Only coexpression correlations above 0.7 are represented

CONCLUSIONS

These are preliminary results of an ongoing study but at this point, we can already indicate that lncRNA have great potential as regulators of FE in beef cattle, in particular, three lncRNA that will be further explored regarding their functional effect as regulators of gene expression and their potential as biomarkers for selection purposes.

ACKNOWLEDGEMENTS

To São Paulo Research Foundation (FAPESP) Proc. 2015/22276-3 and 2014/07566-2.

REFERENCES

Alexandre P.A., Naval-Sanchez M., Porto-Neto L.R., Ferraz J.B.S., Reverter A. and Fukumasu H. (2019) *Front Genet*. **10**:3.

Ghosh S. and Chan C.K.K. (2016) *Methods Mol Biol*. **1374**:339.

Koufariotis L.T., Chen Y.P., Chamberlain A., Jagt C.V. and Hayes B.J. (2015) *PLoS ONE*. **10**:10.

Maere S., Heymans K. and Kuiper M. (2005) *Bioinformatics*. **21**:16.

Shannon P., Markiel A., Ozier O., Baliga N.S., Wang J.T., Ramage D. and Ideker T. (2003) *Genome Res.* **13**:11.

Weikard R., Demasius W. and Kuehn C. (2016) *Animal Genet*. **48**:1.