Intersecting transcriptomic profiling and long noncoding RNAs in the jejunal mucosa of pre-weaning calves fed different diets

R. Weikard¹, F. Hadlich¹, H.M. Hammon¹, D. Frieten², C. Gerbert³, C. Koch³, G. Dusef² & C. Kühn¹,4

¹ Leibniz Institute for Farm Animal Biology (FBN), 18196 Dummerstorf, Germany
² University of Applied Sciences, 55411 Bingen, Germany
³ Educational and Research Centre for Animal Husbandry, Hofgut Neumühle, 67728 Münchweiler, Germany
⁴ University Rostock, Faculty of Agricultural and Environmental Sciences, 18059 Rostock, Germany
weikard@fbn-dummerstorf.de (Corresponding author)

Introduction

Long noncoding RNAs (lncRNAs) emerged as important regulatory components of mechanisms involved in biological processes regulating developmental, metabolic and immunological states of cells and tissues (e.g., Mercer et al. 2009, Rinn & Chang 2012, Chen et al. 2017). LncRNAs were defined as transcripts with more than 200 nucleotides in length that have no or low protein-coding potential (Guttman et al. 2009). They are transcribed similar to mRNAs, but their poor sequence conservation and low tissue-specific expression impede an accurate identification (Derrien et al. 2012, Iyer et al. 2015), and functional annotation and characterization of lncRNAs is a challenging task. Thus, lncRNAs are rarely annotated in the reference genomes of domesticated and farm animals and are just beginning to be explored (Weikard et al. 2016).

The holistic approach of whole transcriptome analysis by RNA sequencing (RNA-seq) allows to discover previously unknown transcripts (including lncRNAs), to identify complex transcript catalogues specific for cells and tissues and to unravel novel regulatory mechanisms on transcriptional level. We performed RNA-seq analysis to monitor changes in the jejunal mucosa transcriptome in response to different milk feeding regimen in pre-weaning calves. We hypothesized that lncRNAs play a regulatory role in mediating the effects of different early life milk diets in the gastrointestinal transcriptome of calves in the pre-weaning period because lncRNAs are known to act as regulatory integral in the modulation of immunological, metabolic, cellular and developmental processes. Thus, our analysis of the RNA-seq dataset had a focus on the identification of lncRNAs present in the jejunal mucosa of pre-weaning calves and aimed to functionally annotate lncRNAs differentially expressed between calf groups experienced a different diet in early life.

Material and methods

Animals and sampling

The experimental design of the study has been described in detail by Frieten et al. (2017). In brief, 12 male German Holstein calves were reared from birth until day 80 of age. Starting from day four, one calf group (RES, n=6) had restricted access to milk replacer diet (6 L/d, milk replacer: 125 g powder per L), whereas the second calf group (AL, n=6) received milk...
replacer ad libitum (max. 25 L/d; AL) for the first 8 weeks of age. The amount of milk replacer was stepped down linearly during week 9 - 10 in both groups, and milk was fed in amounts of 2 L/d until the end of the trial. Hay and concentrate were offered ad libitum to all calves starting from 10 days of age, water was freely available. At slaughtering (day 80) jejunum mucosa samples were collected and immediately snap-frozen in liquid nitrogen.

Transcriptome sequencing, transcript assembly and differential expression analysis

Stranded, indexed libraries (TruSeq mRNA library, Illumina) were sequenced in a 2x 80 cycle protocol on the Illumina HiSeq2500 sequencer platform. After quality control and adapter trimming, reads were subjected to genomic alignment and assembled using the bovine reference genome assembly UMD3.1 (ftp://ftp.ensembl.org/.../pub/release-83/fasta/bos_taurus/dna/) and the Ensembl annotation 83 (ftp://ftp.ensembl.org/.../pub/release-83/gtf/bos_taurus/) following the pipeline published by Trapnell et al. (2012). The final new transcript annotation file served for transcript quantification via Cuffdiff2 and edgeR (McCarthy et al. 2012). Differential expression analysis results were corrected for multiple testing (Benjamini & Hochberg, 1995) and considered as significant at q < 0.1.

Analysis of unknown transcripts and prediction of IncRNAs

The RNAseq dataset representing the transcripts not previously annotated in the bovine transcriptome (class code “u” according to the Cufflinks package, Trapnell, 2010) served as input for the identification of putative IncRNAs. Four different, independent bioinformatic IncRNA prediction tools were applied, each of which extracts specific sequence-based features and includes different filtering steps: PLEK (https://sourceforge.net/projects/plek/files/), CNCI (https://github.com/www-bioinfo-org/CNCI), FEELnc (https://github.com/tderrien/FEELnc) and PLAR (http://www.weizmann.ac.il/Biological_Regulation/IgorUlitsky/pipeline-IncRNA-annotation-rna-seq-data-plar). After performing IncRNA prediction using these four algorithms separately, the intersection between all four IncRNA prediction tools, and combinations consisting of three of them were determined.

Co-expression and pathway analysis

Weighted gene co-expression network analysis (WGCNA) implemented in the R package (Langfelder & Horvath 2008) was performed to construct a weighted correlation network with protein-coding genes expressed in the calf jejunum transcriptome. Network eigengene modules were correlated with differentially expressed IncRNAs. lncRNAs were considered to be significantly co-expressed with respective gene modules at p < 0.05. To identify IncRNA-related biological pathways and to infer potential functional roles for differentially expressed lncRNAs, we performed enrichment, pathway and network analyses using the Ingenuity analysis package (IPA, https://www.qiagenbioinformatics.com/products/ingenuity-pathway-analysis/). Protein-coding genes from eigengene modules related to significantly co-expressed gene module-lncRNA pairs were used as input.

Results and discussion
Prediction of lncRNAs

The classification of unknown transcripts (14,689 class code “u” transcripts) was dependent on the lncRNA prediction tool used (Table 1). PLEK was able to classify 98% of all transcripts included in the input dataset, which deviates clearly from the other three tools (21%, 32% and 33% for PLAR, FEELnc and CNCI) due to the fact that monoexonic transcripts were not filtered by PLEK. It is noticeable that FEELnc and CNCI classified a similar number of lncRNAs and novel mRNAs, whereas PLAR resulted in the lowest number transcripts classified from the dataset (Table 1).

Table 1. Analysis and classification of unknown transcripts.

<table>
<thead>
<tr>
<th>Bioinformatic tool</th>
<th>Total number of predicted transcripts</th>
<th>Predicted lncRNAs</th>
<th>%(^2)</th>
<th>Predicted novel mRNAs</th>
<th>%(^2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FEELnc</td>
<td>4,860</td>
<td>3,494</td>
<td>75</td>
<td>1,186</td>
<td>25</td>
</tr>
<tr>
<td>CNCI</td>
<td>4,784</td>
<td>3,626</td>
<td>76</td>
<td>1,158</td>
<td>24</td>
</tr>
<tr>
<td>PLAR</td>
<td>3,168</td>
<td>2,575</td>
<td>81</td>
<td>593</td>
<td>19</td>
</tr>
<tr>
<td>PLEK*</td>
<td>14,328</td>
<td>10,449</td>
<td>73</td>
<td>3,879</td>
<td>27</td>
</tr>
</tbody>
</table>

1 Classified by the respective tool from the input dataset comprising 14,665 class code “u” transcripts
2 LncRNAs relative to total number of transcripts classified by the respective bioinformatic tool
* Singletons included

The intersections between all four bioinformatic tools and all trio combinations were determined to evaluate the concordance between the lncRNA prediction tools and to extract lncRNAs with a high reliability. The intersection between all four tools showed a total of 1,055 lncRNA transcripts concordantly predicted by all bioinformatic tools (Figure 1).

Figure 1. Intersection of lncRNA prediction by PLEK, CNCI, FEELnc and PLAR.

The intersection between the trio combinations of lncRNA prediction tools revealed that the alignment-free pipelines, CNCI, FEELnc and PLEK, shared the highest number (1,812) of concordantly predicted lncRNA transcripts. The intersection of novel mRNA transcripts revealed that only 48 transcripts were concordantly predicted by all four bioinformatic methods, and again the trio combination of CNCI, FEELnc and PLEK, showed the highest congruence (456). Follow-up analyses of lncRNAs were performed using the intersection
dataset of the trio combination of alignment-free pipelines, CNCI, FEELnc and PLEK (lncRNA concordance dataset).

**Differentially expressed lncRNAs and co-expression analysis**

Ten differentially expressed loci were present in the lncRNA concordance dataset. All of them were downregulated in RES calves compared to AL calves, suggesting a potential functional relevance in the modulation of regulatory processes in the intestine associated with adaptation of the calf groups to the different diet.

WGCNA revealed that a total of 26 gene network modules were significantly correlated with at least one of the lncRNAs included in the analysis. The lncRNA-gene module relationships showed that almost all lncRNAs were co-expressed with several gene modules. Canonical pathway analysis of most tightly co-expressed gene network module datasets (p ≤ 0.01) with IPA highlighted a gene module comprising genes, which were predominantly enriched in “Remodeling of epithelial adherens junctions”, “ Tight junction signaling” and “Integrin signaling”. Another gene network module comprises genes, which were highly significant enriched in canonical pathways “Oxidative phosphorylation”, “Mitochondrial dysfunction”, “ NRF2-mediated oxidative stress response” and “TCA cycle II suggesting a potential regulatory function in metabolic processes related to energy metabolism and maintaining cellular homeostasis with a focus on mitochondrial processes.

**References**


