Initial RNA-seq analysis on the milk transcriptional profiling of two sheep breeds

A. Suárez-Vega1, B. Gutiérrez-Gil1, C. Esteban2 and J.J. Arranz1.

1University of León, 2Supercomputing Center of Castile and León, León, Spain.

ABSTRACT: In this study, massively parallel sequencing of transcripts (RNA-Seq) was used to make a preliminary characterization of the milk transcriptome of two dairy sheep breeds, Churra and Assaf, at day 120 of lactation. A total of 230 million paired-end reads were obtained from RNA sequencing. The 91% of these reads were mapped to a unique location in the ovine genome. Genes encoding caseins and whey protein showed the highest expression in this stage of lactation. Eight genes were identified as differentially expressed between the two dairy sheep breeds under study. A relationship with milk composition and lactation maintenance was established for some of the differentially expressed genes found.

Keywords: Sheep; Milk producition; RNA-seq

Introduction

Massively parallel DNA sequencing, widely referred as “next-generation sequencing” (NGS), is changing the biological research view by enabling the comprehensive and relatively inexpensive analysis of genomes and transcriptomes (Metzker (2010)). RNA-seq technologies have emerged proving a unique opportunity to quantify transcripts and identify differential regulation in a single experiment (Wang et al. (2009)).

Sheep production can lead many potential outputs (milk, meat, skin, fiber). Sheep’s milk is commonly used to produce many cultured dairy products. Dairy sheep industry is mainly concentrated in Europe and the countries on or near the Mediterranean Sea. The milk transcriptome profile has been analyze in different mammalian species (Shu et al. (2012), Wickramasinghe et al. (2012), Lemay et al. (2013)) in order to better characterize the genes involved in lactation, although an specific study on the ovine milk transcriptome has yet not been reported.

The aim of this research was performing a preliminary analysis of the milk somatic cell transcriptome profile in two dairy sheep breeds (Spanish Churra and Assaf) at day 120 of lactation. Churra is an autochthonous dairy breed reared in the northwest of Spain and Assaf is a specialized dairy breed, with a higher milk production (De la Fuente et al. (2006)). The stage of lactation analyzed in this study is interesting from the point of view of the different lactation length in these two dairy sheep breeds, normalized to 120 and 150 lactation days in Churra and Assaf, respectively. For Assaf ewes, day 120 of lactation corresponds to a transition stage from peak to final lactation while that time-point is related to nearly final lactation in Spanish Churra sheep.

Materials and Methods

Sample collection, RNA isolation and sequenc- ing. The milk samples were obtained from three Churra sheep and three Assaf sheep at the 120 lactation day. For each animal we collected 50 ml of fresh milk, from both mammary glands. Somatic cells were separated as described by Wickramasinghe et al. (2012) with some modifications. Cells were pelleted by centrifugation at 1800 rpm in 50 mL sterile tubes for 10 min. at 4°C in the presence of a final concentration of 0.5 mM EDTA. The cell pellet was washed in PBS twice, pH 7.2 with 0.5 mM EDTA and suspended in 500 μl of Trizol. RNA extraction continued following the Trizol protocol (Invitrogen, Carlsbad, CA, USA).

The integrity of the RNA was assessed using Agilent 2100 Bioanalyzer device (Agilent technologies, Santa Clara, CA, USA). The RNA integrity value (RIN) of the samples ranged between 8.2 and 8.9. Paired-end libraries with fragments of 300 bp were prepared using the True-Seq RNA-Seq sample preparation Kit v2 (Illumina, San Diego, CA, USA). The fragments were sequenced on an Illumina Hi-Seq 2000 sequencer (Fasteris SA, Plan-les-Ouates, Switzerland).

Mapping and assembly. The quality of the raw RNA-seq reads was assessed by using FastQC (www.bioinformatics.babraham.ac.uk/projects/fastqc). The quality control of the studied samples was later performed using the option HEADCROP of Trimmomatic (www.usadellab.org/cms/?page=trimmomatic). In this step we aimed to remove the first 10 bps of the paired-end reads in order to eliminate the bias induced by the Illumina hexamer priming in the nucleotide composition at the beginning of the reads. The last bases of the reads which a ‘per base sequence quality’ lower than 20 were also trimmed.

After the quality control, the paired-end reads were mapped against the ovine reference genome (OARv3.1) using TopHat v.2.0.10 (Trapnell et al. (2009)) with Bowtie2 (v2.2.0) applying the default settings. In addition, we used the –G option followed by the ENSEMBL transcriptome annotation, which allows the program to do a first alignment of the reads to the virtual sheep transcriptome. We also set the distance between both pairs to 50 bp (inner-mean distance) and the standard deviation to 150 bp. SAMtools (Li et al. (2009)) was used to sort the mapped...
Expression quantification and differential expression analysis. The estimation of the number of reads mapping to each transcript and the differential expression analysis was done with the Cufflinks package (Cufflinks, Cuffmerge, Cuffcompare, Cuffdiff and CummeRbund) following the analysis pipeline described by Trapnell et al. (2012). The parameter FPKM (Fragments Per Kilobase of transcript per Million fragments mapped) was used to quantify transcripts expression. FPKM was calculated based on the mapped transcript fragments, transcript length and sequencing depth. Differential expression analysis of the two groups of milk samples was performed using Cuffdiff. Cuffdiff was run using the multi-read correction (-u option) and the -b option which improves the accuracy of transcript abundance estimates by running a new bias detection. The CummeRbund Bioconductor R package was employed to analyze Cuffdiff outputs.

Results and Discussion

Global analysis of gene expression in day 120 of lactation. A total of 230 million paired-end reads were obtained from RNA sequencing. Three replicates were analyzed for each breed with reads ranging from 34 to 42 million paired-end reads per sample. Approximately, the 91% of the sequences (95 million paired reads in Assaf and 87 million paired-end reads in Churra) were mapped to a unique location in the ovine genome.

From the results obtained with cuffcompare (Table 1), it can be observed that more transcripts were identified for Churra than for Assaf (52,690 versus 47,510 reads). The distribution of the percentage across the different cuffcompare classes for the two breeds is very similar; the highest difference can be noticed for the unknown/intergenic transcripts which are slightly increased for the Churra breed (7% higher).

Table 1. Classification of the transcripts identified in the milk samples in relation to Ensembl annotated genes in the sheep genome.

<table>
<thead>
<tr>
<th>Class</th>
<th>N° transcripts</th>
<th>% Assaf</th>
<th>N° transcripts</th>
<th>% Churra</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>25024</td>
<td>52.67</td>
<td>25046</td>
<td>47.53</td>
</tr>
<tr>
<td>2</td>
<td>14</td>
<td>0.03</td>
<td>22</td>
<td>0.04</td>
</tr>
<tr>
<td>3</td>
<td>15294</td>
<td>32.19</td>
<td>16510</td>
<td>31.33</td>
</tr>
<tr>
<td>4</td>
<td>717</td>
<td>1.51</td>
<td>783</td>
<td>1.49</td>
</tr>
</tbody>
</table>

Figure 1 shows the six most highly expressed genes in the milk samples of Churra and Assaf breeds. Genes with the highest value of FPKM at day 120 of lactation are LALBA (α-lactoalbumin), ENSOARG00000005099 (LGB, β-lactoglobulin), CSN1S1 (α-S1-casein), CSN1S2 (α-S2-casein), CSN3 (κ-casein), CSN2 (β-casein). These highly expressed genes encode caseins and whey proteins, principal components of milk. From the Figure 1 we can observe that the caseins, with the exception of CSN1S1 are slightly more expressed in Churra than in Assaf, while the whey proteins (α-lactoglobulin and β-lactoglobulin) are lightly more expressed in Assaf than in Churra. These results agree with the differences observed in cheese yield between the two breeds, as the Churra breed is reported to have more protein and fat contents in milk compared to Assaf sheep (De la Fuente et al. (2006)).

Analysis of differentially expressed genes. The differential expression analysis of annotated genes performed with Cuffdiff reveal a total of eight differentially expressed genes (DE) between the two breeds, Churra and Assaf (Figure 2). Five of these genes showed the highest expression in Assaf sheep (GPRC5C (G-protein coupled receptor family C group 5 member C), FGL2 (Fibrinogen-
Like 2), LIPG (Endothelial Lipase), TGM3 (Protein-glutamine gamma-glutamyltransferase E), ENSOARG00000016020 and three showed the highest expression in Churra sheep (MGP (Matrix Gla Protein), PDK4 (Pyruvate Dehydrogenase Kinase, Isozyme 4), SLC02B1 (solute carrier organic anion transporter family, member 2B1) (Figure 2). As can be observed in the heatmap plot the most highly expressed DE gene is MGP in Churra (1378 FPKM). The fold change of these DE genes ranged from 2.47 to 3.65. The gene with the highest fold change between two conditions is FGL2.

![Figure 2. Heatmap plot of the differentially expressed (DE) genes between the two sheep breeds under study.](image)

A deeper analysis of the eight DE genes was performed. For some of these DE genes, a relationship with lactation maintenance and milk composition was found. This relation is discussed below.

FGL2 is a gene related to innate immunity; a study of the effects of increasing milking frequency in bovine mammary gland (Connor et al. (2008)) showed that under stress conditions, like increasing milking frequency, the FGL2 is down-regulated.

The GPRC5C gene encodes a membrane protein receptor. This gene product is included among the milk fat globule proteins where appears to be involved in signal transduction and lipid droplet transport (D’Alessandro et al. (2011)).

Other of the genes identified as DE in this work, LIPG, is involved in the lipid metabolism of the mammary gland. The LIPG protein influences milk fatty acid composition, showing a large negative correlation with CLAt10c12, CLAc11t9 and other trans-fatty acids, and large positive correlations with C16, SCFA and SFA (Mach et al. (2013)). The DE observed for LIPG between these two breeds could suggest differences in the milk fatty acid composition between Churra and Assaf, which have not been studied up to date.

The MGP gene encodes the Matrix-Gla protein, which acts as an inhibitor of calcification (Luo et al. (1997)). The absence of this protein makes the expression of osteopontin, another protein implicated in the modulation/inhibition of calcification, to be upregulated (Speer et al. (2002)). Osteopontin is a protein present in tissues elaborating fluids, such as lactating mammary gland, where it has been shown to be a regulator of lactation (Sheehy et al. (2009)). Levels of osteopontin are high during the initiation of lactation and the decrease, but remain high during involution (Rittling and Novick (1997)). Due to the similar functions of these two proteins, Matrix-Gla could act in a similar way to osteopontin during lactation.

**Conclusion**

The undertaken RNA-seq experiment has allowed presenting herein a preliminary study on the gene expression profiling in sheep’s milk somatic cells. RNA-seq has shown to be a powerful tool to analyze the ovine milk transcriptome.

Genes encoding caseins and whey proteins show the highest expression at day 120 of lactation in sheep’s milk. In general, from these six genes, whey proteins were slightly more abundant in Assaf and caseins in Churra.

The differential expression analysis revealed a total of eight DE genes between the two dairy breeds, Churra and Assaf. Some of these genes display known functions related to lactation and milk composition.

Further analysis on functional enrichment and covering other stages of lactation should be carried out in order to obtain a global expression profiling of the milk somatic cells in sheep.

**Acknowledgment**

Note: This work has been funded by a competitive grant from Spanish Ministry of Economy and Competitiveness (Mineco) project: AGL2012-34437. BGG is supported by Ramón y Cajal Program from Mineco.

**Literature Cited**


