Comparative Analysis Of Bovine Milk And Mammary Gland Transcriptome Using RNA-Seq

J.F. Medrano[*], G. Rincon[*], A. Islas-Trejo[*]

Introduction

Gene expression profiling of mammary tissue is a powerful tool that has been used to catalog and characterize the roll of genes expressed during lactation, and has the potential of discovering the nature of some of the unique functional properties of milk. The transcriptome has been widely analyzed using microarrays; however, high throughput next-generation (Next-Gen) sequencing techologies have emerged providing a unique opportunity to quantify and annotate transcriptomes (Marioni et al. (2008), Mortazavi et al. (2008)). RNA-Seq is the first sequencing method that allows very accurate quantification of the expression levels of the entire transcriptome with a high dynamic range (Wang et al. (2009), Wilhelm and Landry (2009)). This technology generates sequences on an unprecedented scale at a fraction of the costs required for traditional sequencing, allowing the application of sequencing approaches to biological questions that would not have been economically or logistically practical before.

Numerous studies have examined gene expression in the bovine mammary gland by performing mammary biopsies; however, such techniques are invasive, disturb the normal lactation process, are labor intensive and are costly. These factors limit more dynamic studies of the mammary transcriptome. An alternative sampling procedure has been proposed by isolating mRNA directly from somatic cells that are naturally released into milk during lactation.

In the present study, a comparative transcriptome analysis from milk and mammary tissue was performed using RNA-Seq technology. Considering the high level of caseins and whey proteins expressed in the mammary gland, we decided to include a normalized mRNA library from mammary tissue in the analysis in order to compare the global gene expression pattern in normalized vs. standard RNA-Seq mammary tissue.

Material and methods

Animals and RNA-Seq sample preparation. A mammary tissue biopsy was obtained from a second lactation Holstein cow at 170 days in milk from the University of California Dairy Research Facility. Somatic cell count was performed before samples were collected. Biopsy was performed using a rotating stainless steel cannula described by Farr et al. (1996). The University of California Davis Institutional Animal Care and Use Committee (IACUC) approved all animal procedures. Milk samples were collected before the biopsy was performed; the samples were kept on ice and processed immediately for RNA extraction.

[*] Department of Animal Science, University of California, Davis, California 95616, USA

Mammary tissue was suspended in RNA-later (Ambion Inc, Austin,TX), cut into small pieces and immediately put in liquid nitrogen for future extraction of RNA. 50 ml of fresh milk was obtained from each quarter, and somatic cells were separated as described by Boutinaud et al. (2002) with some modifications. Cells were pelleted by centrifugation at 2000g 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, pH 7.2 with 0.5 *mM* EDTA and resuspended in 800 µl of Trizol. RNA extraction continued following the Trizol protocol (Invitrogen, Carlsbad, CA). mRNA was isolated and purified using RNA-Seq sample preparation Kit (Illumina, CA). Then mRNA was fragmented to approximately 200bp and first and second strand cDNA was synthesized. End repair and adapter ligation was performed on DNA fragments. Trimmer–Direct cDNA normalization methodology (Evrogen) was used to develop the mammary gland normalized cDNA library. The fragments were sequenced on an Illumina system GA sequencer (Illumina, CA).

Transcriptome analysis. Sequence reads were assembled and analyzed using CLC Bio genome work bench software. NCBI assembly Btau_4.0 was used to map all the sequence reads to the bovine genome. Gene expression was estimated by calculating read density as 'reads per kilobase of exon model per million mapped reads' (RPKM) for 27,368 annotated genes. SVS7 from Helixtree version 6.3.1 (Golden Helix Inc., Bozeman, Montana, USA) was used to map RPKM values for the expressed genes in the genome and to perform the "in silico" normalization analysis. Gene Ontology analysis was performed to establish the function of differentially expressed genes among samples. Baggerley's test on expression proportions was used to test for RPKM differences between samples (Baggerly et al 2003).

Results and discussion

In this study the mammary gland and milk transcriptome from the same cow were analyzed using RNA-Seq technology. A total of 43.6 million reads were assembled to the consensus Btau_4.0 genome assembly. RPKM values higher that 0.2 were used to establish the genes that were expressed in each sample. Table 1 shows the total number of genes expressed in each sample and the number of unique genes expressed in milk, mammary tissue and normalized mammary cDNA library.

Table 1: Number of genes expressed in milk and mammary gland transcriptome

Sample	Normalized cDNA library	Mammary Gland	Milk
Total Genes	9,007	12,573	13,737
Unique genes	313	570	1,572

The milk transcriptome showed the highest number of expressed genes followed by mammary gland. The normalized cDNA library showed the lowest number of expressed genes, representing a difference of 3,027 genes that were lost with normalization but were observed in the mammary gland transcriptome. Normalization decreased the level of expression of genes by 40% on average, but some genes were missing during this process and some other genes were selectively affected when compared to the whole transcriptome

normalization. This was the case for the expression of the beta-lactoglobulin gene (LGB) that showed a reduction of 90% in the normalized library. We found that by performing an "in silico" normalization of the transcriptome, highly expressed genes can be removed from the analysis at desired RPKM levels of expression as shown in Figure 1, without the risk of non-specifically loosing transcripts which is the case in in-vitro normalized libraries. The high dynamic range of expression levels detected in RNA-Seq allowed us to accurately examine all the genes that were expressed in the mammary gland at very high and low levels.

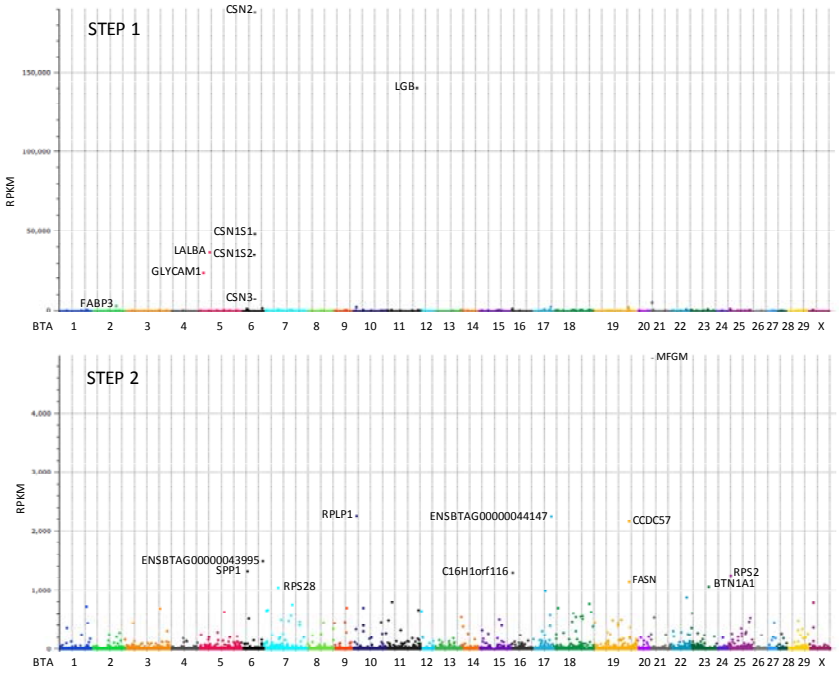

Figure 1: "In silico" normalization of mammary gland transcriptome mapped to the Btau_4.0 assembly. Genes are represented as colored dots; each chromosome is shown in a different color and represented as BTA1 to BTA X. **STEP 1**: High RPKM values, note that mostly caseins and LGB are highly expressed. **STEP2**: Genes with intermediate RPKM values after removing caseins, LGB, GLYCAM1, FABP3 and LALBA genes from the analysis. (This process can be performed repeatedly to improve the detection of genes that are expressed at low RPKM)

Genes that are shared between milk and mammary gland transcriptomes, as well as genes that were uniquely expressed in each tissue are summarized in Figure 2. Most of the genes expressed in the mammary gland transcriptome were also present in milk. Only 570 genes showed a specific pattern of expression in the mammary gland transcriptome. Gene Ontology (GO) analysis revealed that most of these genes were related to structural processes occuring in the mammary gland. A total of 1,572 genes were uniquely expressed in milk and those genes were related to immunity, organ development and behavior as shown in Figure 2. Baggerley's test on expression proportions revealed that only 1,984 genes from 11,672 showed significant differences in expression pattern between milk and mammary tissue.

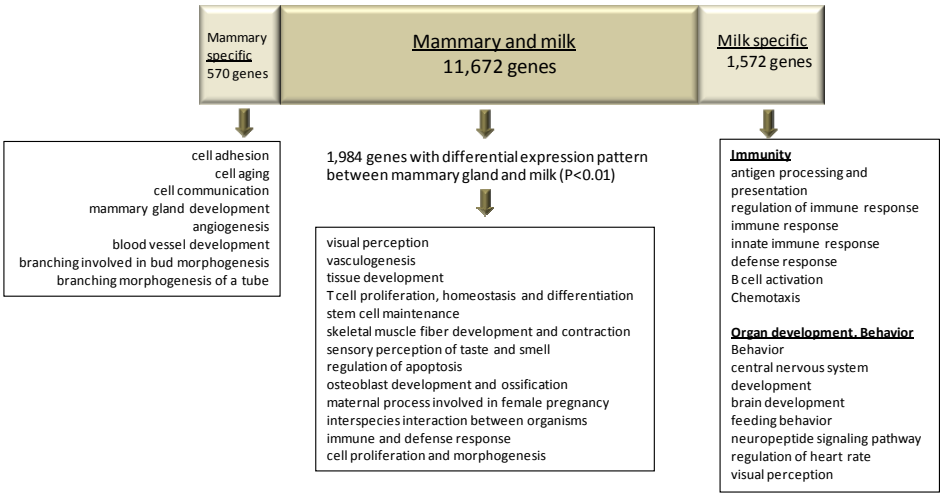

Figure 2: Summary of expressed genes in milk and mammary tissues. GO description is shown in black boxes. A total of 11,672 genes were observed in the milk and mammary gland transcriptomes. Baggerley's test showed that 1,984 genes had a differential pattern of expression.

Conclusion

RNA-Seq is a powerful tool to analyze transcriptomes with a high dynamic range allowing the quantification of genes that are expressed at extreme RPKM levels. With this technology there is no need to create a normalized cDNA library to study gene expression profiles, even in tissues like mammary gland that have a handful of genes expressed at extremely high levels. An "in-silico" normalization can be effectively performed to analyze expression at different levels in the transcriptome. According to our findings, milk is highly representative of the mammary gland transcriptome and can be used as an alternative sample to study mammary gland expression eliminating the need to perform a tissue biopsy.

Our analysis showed that milk also expresses a unique set of genes related to immunity, organ development and behavior that will require further investigation to ascertain their functionality in this tissue.

References

Baggerly, K.A., Deng, L., Morris, J.S., *et al.* (2003). *Bioinformatics*, 19:1477-1483.

Boutinaud, M., Rulquin, H., Keisler, D.H., *et al.* (2002). *J. Anim. Sci.*, 80:1258-1269.

Farr, V.C., Stelwagen, K., Cate, L.R. *et al.* (1996). *J. Dairy Sci.*, 79:543-549.

Marioni, J.C., Mason, C.E., Mane, S.M. *et al.* (2008). *Genome Res.*, 18:1509-1517.

Mortazavi, A., Williams, B.A., McCue, K. *et al.* (2008). *Nat. Methods*, 5:621-628.

Wang, Z., Gerstein, M. and Snyder, M. (2009). *Nature Reviews*, 10:57-63.

Wilhelm, B.T. and Landry, J.R. (2009). *Methods*, 48:249-257.