COMPUTATIONAL ANALYSIS OF BOVINE MIRNA'S ACROSS DIFFERENT TISSUES AND SPECIES

by

Naga Sridhar Betrapally

A Thesis
Submitted to the
Graduate Faculty
of
George Mason University
in Partial Fulfillment of
The Requirements for the Degree
of
Master of Science
Bioinformatics and Computational Biology

Committee:

_____	Dr. Donald Seto, Thesis Director
_____	Dr. Iosif Vaisman, Committee Member
_____	Dr. Jason Kinser, Committee Member
_____	Dr. Huzefa Rangwala, Committee Member
_____	Dr. James D. Willett, Director, School of Systems Biology
_____	Dr. Timothy L. Born, Associate Dean for Academic and Student Affairs, College of Science
_____	Dr. Vikas Chandhoke, Dean, College of Science
Date: 8/24/2012	Fall Semester 2012 George Mason University Fairfax, VA

Computational Analysis of Bovine miRNAs across Different Tissues and Species

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science at George Mason University

by

Naga Sridhar Betrapally
Bachelor of Engineering
Visvesvaraya Technological University, 2008

Director: Donald Seto, Professor
Department of Bioinformatics and Computational Biology

Fall Semester 2012
George Mason University
Fairfax, VA

ACKNOWLEDGEMENTS

I would like to thank my Parents and Friends for their constant support during the research. I would also like to thank Drs. Sonstegard, McDaneld, Smith and Alexander for providing the data to carry out the research, Dr. Nobuta for carrying out the Flicker analysis on the dataset, Drs. Rangwala, Seto, Kinser and Vaisman for their invaluable help during the research and in preparation of the thesis document. Finally, I would like to thank the Johnson Center, Fenwick Library for providing a clean, quiet, and well-equipped repository where I have spent many hours carrying out my research.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF EQUATIONS

ABSTRACT

COMPUTATIONAL ANALYSIS OF BOVINE MIRNAS ACROSS DIFFERENT TISSUES AND SPECIES

Naga Sridhar Betrapally

George Mason University, 2012

Thesis Director: Dr. Donald Seto

MicroRNAs (miRNAs) are small non-coding RNA molecules of approximately 22 nucleotides in length. They are present as genome-encoded stem-loop precursors that recognize target mRNAs by base pairing, which then regulates their expression. Due to their influence in the expression of hundreds of genes, they play a role in regulation of gene expression for numerous biological processes such as in animal development, apoptosis, fat metabolism and hematopoietic differentiation. Initial studies showed that most miRNAs are conserved among related species. However, recent studies have shown that newly identified miRNAs tend to be species specific. miRNAs are known to have differential expression patterns during development and across tissues but there is not much known about relative abundance and specificity of expression patterns among tissues for most bovine miRNAs. Profiling of bovine miRNAs and evaluation of their expression patterns were carried out in this study for a total of 64 different tissues from

bovine Calf and Fetus. The study was carried out to identify tissue specificity and tissue class specificity. Analysis based on this principle revealed tissue specificity for a certain class of miRNAs. Further analysis and deep sequencing of the data shall help us identify the functional role of miRNAs in these tissues.

1. INTRODUCTION

miRNAs are described as ~22-nt RNA molecules that originate from fold-back precursors and can regulate the expression of genes [1]. They are an emerging class of gene regulators that are endogenously produced as small non-protein coding RNAs. They negatively regulate gene expression at the post transcriptional level by homologous interactions with the 3' UTR (untranslated region) and more rarely with the coding region of the target mRNA [2,3]. Numerous biological processes in animal development, apoptosis, fat metabolism and hematopoietic differentiation have been reported to be regulated by miRNAs [4-8]. In addition, miRNAs can increase protein translation by binding to complementary promoter sequences, extending the important function of miRNA to protein expression [8-10].

miRNAs are transcribed as long transcripts which are called primary miRNA or pri-miRNA [11]. pri-miRNAs are characterized by the presence of loop and stem structures with imperfect complimentary region. pri-miRNAs on being transcribed undergo their first maturation reaction in the nucleus. In the first step of processing the reaction, the pri-miRNA is processed to a 60-70 nucleotides intermediate precursor called pre-miRNA [12]. The pre-miRNA is exported to the cytoplasm for further processing [13]. In the cytoplasm, the loop structure is cleaved to yield a double stranded miRNA with two 5' phosphates and 3' two nucleotide overhang structure. Upon further

processing and the destruction of the complimentary strand, the mature miRNA is incorporated into RISC (RNA Induced Silencing Complex) to form an active RISC complex, and directs either translational repression or site-specific cleavage of the mRNA target (Figure 1).

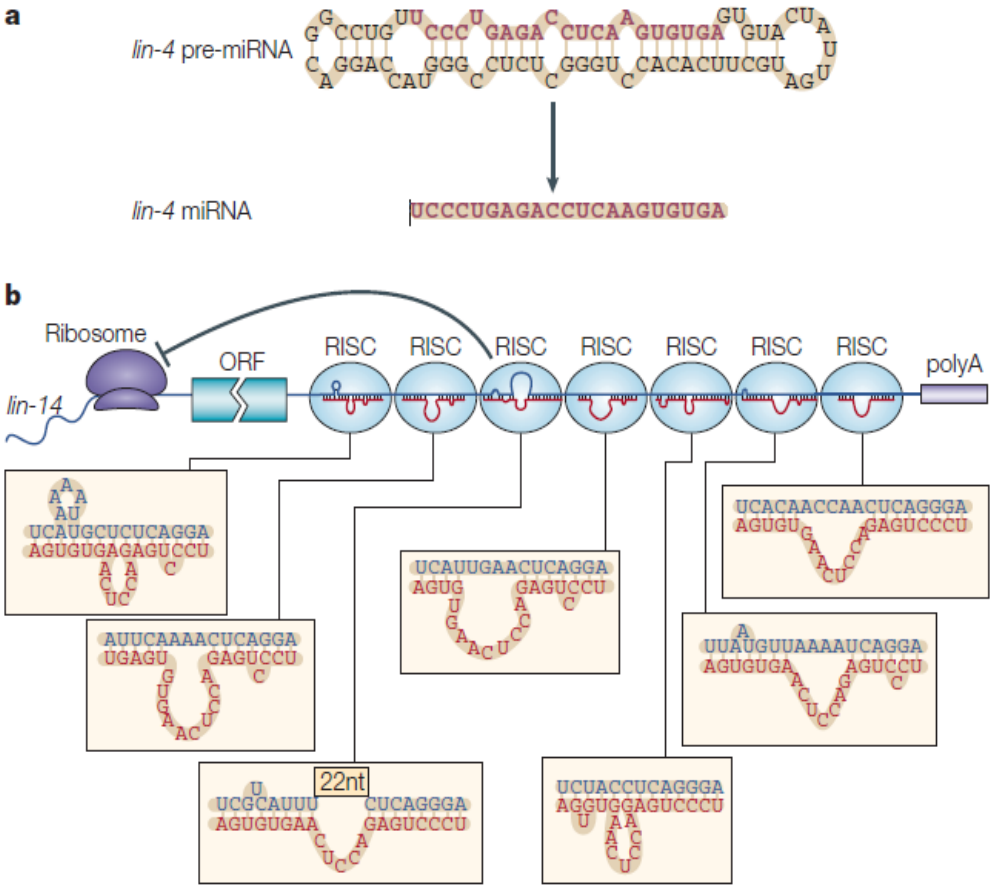

Figure 1. An example of miRNA structure and its processing steps. (a) The precursor structure and mature microRNA (miRNA) sequence of lin-4. (b) Processing of pri-miRNA to pre-miRNA (blue). Maturation of the pre-miRNA occurs in the nucleus to form a mature miRNA [14].

In order to identify miRNAs, the following three criteria need to be fulfilled. The mature miRNA should be expressed as a distinct transcript of ~22 nucleotides. Mature miRNA should originate from a precursor with a characteristic secondary structure, such as a hairpin or fold-back, which does not contain large internal loops or bulges. Mature miRNA should occupy the stem part of the hairpin and mature miRNA should be processed by Dicer, as determined by an increase in accumulation of the precursor in Dicer-deficient mutants [15].

Reported here is the analysis of tissues from bovine calf and fetus. A total of 32 different tissues were obtained for Calf and 32 different tissues were obtained for Fetus. Table 1 shows the different tissues used for the study. The tissues were grouped into their tissue classes based on Brenda Tissue Ontology [15]. The sequences obtained from both the species were run on an Illumina Genome Analyzer [35] and the obtained miRNA sequences were then used for carrying out clustering analysis and data exploration study. The motivation to carry out this analysis was to identify relative abundance and specificity of expression patterns among tissues for most bovine miRNAs. Elucidation of the expression patterns of different miRNAs among different tissues will help in understanding the roles of miRNAs in gene expression.

Table 1. Grouping of tissues in different tissue classes based on Brenda Tissue Ontology.

Tissue class	Tissues classified in the tissue class
Viscus	Abomasum, Cecum, Duodenum, Ileum, Jejunun, Large Intestine, Omasum, Reticulum, Rumen, Stomach Omasum
Reproductive System	Oviduct, Uterus intercaruncular, Vas deferens
Muscular System	Biceps fem, Infra Spin, Longisimus Dorsi Muscle, Semitend, Sternomandibuilaris Skeletal, Tongue Muscle
Nervous System	Frontal Cortex, Basal Ganglia, Cerebellum, Hippocampus, Medulla, Midbrain, Pons, Spinal Cord
Integument	Nasal Epithelium, Outer tongue surface, Oral mucosa
Glands	Adrenal, Adrenal Cortex, All pituitary, Anterior pituitary
Embryonic Structure	Umbilical Cord, Placenta
Connective Tissue	Marbling, SubQ, White fat, Bone, Paracardial

2. MATERIALS AND METHODS

In order to study the miRNAs data from the bovine Calf and Fetus, it was necessary to identify tools which help in the analysis process. It was important to identify tools which shall allow for aligning the reads to the genome, carry out cluster analysis and identify their statistical significance.

2.1 Flicker 3.0

Flicker is an in house tool developed by Illumina [16]. It has been developed in order to carry out analysis of small RNA data. Flicker first aligns the reads to the 3' adapter sequences and aligns them using the Smith-Waterman algorithm [17]. Once it has aligned the reads to the adapter sequences, it trims the reads in order to remove the adapter sequences. These reads are then aligned to different elements such as mitochondrial RNA, rRNA and primers by using the BLAST principle [18]. After filtering the reads for these contaminant elements, the reads are then aligned to the genome sequences, hairpin loop sequences, and mature miRNAs. These alignments are then stored in a BAM alignment file which contains all the alignment results and chromosomal co-ordinates [19]. This alignment file can then be visualized using Integrative Genomics Viewer [20] in order to visually inspect the alignments of the reads to the bovine genome.

2.2 Hits Normalized Abundance

Hits normalized abundance was used to measure the abundance ratios of miRNA expression in this study. Next Generation Sequencing has various popular standardized technologies that have developed in a short period of time offering cost effective solutions for sequencing large number of sequences [21,22]. In order to sequence for small RNAs including mRNA and small RNA transcriptome profiling, Illumina has developed a new technology called Sequence by Synthesis (SBS) [23]. This method uses reversible nucleotide terminators to sequence short DNA fragments (signatures). The sequence length is about ~33-35 bases and is long enough to capture the full length sequence of the small RNA molecules such as miRNA. The total number of signatures sequenced by SBS differs from library to library. In order to compare the expression level of a particular signature across the libraries, abundance levels which are the total number of hits to the gene must be normalized. For example, abundance levels of a particular miRNA in the library Liver will be different from the abundance levels of the same miRNA in the library Kidney. Thus in order to measure all of the abundance levels on the same common scale, the values have to be normalized, i.e., adjusting the different abundance levels from libraries to the common scale in order to be comparable to other libraries. This calculation is performed for each signature and for each library, based on the total abundance of the libraries. In order to normalize them, a round number that is close to the total number of sequences is used. This number is called total raw abundance. The abundance levels are considered as a raw value. Each raw value is divided by the total number of sequences, i.e., total raw abundance and multiplied by a normalization factor. The normalization factor is a value that is close to the adjusted total

raw abundance and is rounded off to a multiple of a million. These normalized values allow us to accurately compare the expression levels of signatures across libraries. These normalized values are referred to as Hits Normalized Abundance [23]. The hits normalized abundance is computed as follows:

Equation 1. Hits Normalized Abundance

$$Hits\ Normalized\ abundance = \frac{x}{y}\ X\ N$$

where x = abundance values of each miRNA in the library,

y = total number of sequences in the library,

N = Normalization factor which is a multiple of a million.

2.3 Clustering analysis using PermutMatrix

Cluster analysis groups data based on information found in the data that describes the objects and their relationships. The goal is that the objects within a group be similar to one another and different from the objects in other groups. Hierarchical clustering is a clustering approach that produces subclusters which is a set of nested clusters that are organized as a tree. The trees are formed by starting with each point as a singleton cluster and then repeatedly merging the two closest clusters based on the shortest distance between them until a single, all-encompassing tree is formed [37]. PermutMatrix is a tool which can be used to implement hierarchical clustering [24]. It provides for a graphical environment and has been designed to graphically explore gene expression data. PermutMatrix approaches to simultaneously display the clustering tree, colored representation of the data matrix and supplements them with several optimal linear reordering methods such as reorganization of the leaves of a clustering tree,

unidimensional scaling and seriation [8,24,25]. The reorganization of the leaves of a clustering tree by PermutMatrix has been used in the clustering analysis of the dataset in order to sort the miRNA hits normalized abundance values in the descending order. The use of this characteristic is explained in more detail in 3.2.

2.4 Detection of differentially abundant features

In order to identify the differential expression based on hits normalized abundance among different tissues, it is necessary to use a statistical method to quantify the significance of the results. Metastats is a statistical method which is designed to identify differentially abundant features in a dataset [26]. It takes in an input file containing hits abundance values for 2 sets of data. It provides an output file containing the mean, variance, standard error and p-value as an output. Here, the mean is defined as the central tendency of a collection of numbers taken as the sum of the numbers divided by the size of the collection.

The variance of a random variable is the squared deviation of that variable from its mean. If a variable, X has the mean μ; the variance of X is given by:

Equation 2. Variance

$$Variance(X) = (X - \mu)^2$$

The standard error is the standard deviation of the sample from the mean and is given by:

Equation 3. Standard error

$$Standard\ error = \frac{s}{\sqrt{n}}$$

where, s is the sample standard deviation and n is the size of the sample.

In statistics, p-value for any hypothesis test is the α level at which we would be indifferent between accepting or rejecting the null hypothesis given the sample data at

hand. That is, the p-value is the α level at which the given value of the test statistic (such as t) is on the borderline between the acceptance and rejection regions. The p-value is important because it tells us exactly how significant our results are without performing repeated significance tests at different α levels. The p-value results obtained here are then used to quantify the results which in turn shall help in identifying the significance of the miRNA expression in the dataset.

2.5 Brenda Tissue Ontology

Brenda Tissue Ontology represents a comprehensive structured encyclopedia of tissue terms [15]. Brenda Tissue Ontology contains more than 4600 different anatomical structures, tissues, cell types and cell lines, classified under generic categories corresponding to the rules and formats of the Gene Ontology Consortium [34]. The parent-child relationship permits the depiction of the hierarchical structure of the ontology which contains terms at various levels of detail such as tissues falling under particular tissue classes [15]. Figure 2 shows an example of the parent-child relationship for an animal. The animal is categorized as the parent and the tissue classes under its tree are the child nodes. Further exploration of the child node will give us the details of the sub nodes of the child node which shall give information regarding the tissues under the tissue classes. This information was used to define the tissue classes for the bovine species under this study.

Figure 2. Brenda Tissue Ontology for an animal and its child terms [15].

3 RESULTS AND DISCUSSION

3.1 Data analysis

For this study, the 32 different tissue samples of bovine Calf and 32 different tissue samples of Fetus were sequenced by using the Illumina Genome Analyzer [35].A total of 790 (786,318,569) million sequence reads were sequenced for all the tissues. 120 (120,123,642) million reads from the total sequence reads aligned to the miRBase [27-30] database which is 15.2% reads that aligned to the miRBase. Table 2 shows an example of different lengths and the total counts of sequence reads for the hypothalamus tissue obtained from the bovine Calf. Table 2 shows read lengths ranging from 12 to 26. For these reads, we observe that the length of the majority of the reads range from 18 to 22 as compared to other lengths. These results were observed to be consistent for all the tissues indicating that our dataset contained a lot of miRNAs since for a read to be categorized as a miRNA, it should range from 18 to 22 nucleotides [1] and thus a very good dataset was obtained for the analysis.

3.2 Clustering analysis

Clustering analysis of all the tissues was carried out using PermutMatrix. A hierarchical clustering was carried out for all the tissues by using squared Pearson distance [36] for the measure of the distance between data points and complete linkage for the measure of dissimilarity between data points. The squared Pearson measures the

similarity in shape between two profiles and can also capture inverse relationships [36]. The formula for the squared Pearson distance is given by:

Equation 4. Squared Pearson Distance

$$d = 1 - 2r$$

where d is the squared Pearson distance and r is the Pearson correlation.

The Pearson correlation is given by:

$$r = Z(x).Z(y)/n$$

The Pearson correlation is the dot product of the z-scores of the vectors x and y. The z-score of x is constructed by subtracting from x its mean and dividing by its standard deviation [36].

A tab delimited file with hits normalized abundance values were submitted to PermutMatrix. The rows were sorted to order the miRNAs in descending order. The hierarchical clustering results are as shown in Figure 3. Figure 3 displays the forming of a hierarchical cluster for the tissues for a subset of 25 miRNAs showing high hits normalized abundance values. The hierarchical clustering tree data was analyzed using Brenda Tissue Ontology as a reference. The tissue systems were found to have conformity to Brenda Tissue Ontology. For example, the tissues that falls under the connective tissue system, Marbling, Sub-cutaneous and Paracardial all fall under a single tree. There was however an exception for the embryonic tissue system, where Umbilical cord and Placentome don't fall under a single tree. Since we observe that most of the

tissue systems cluster in accordance to the Brenda Tissue Ontology, we conclude that we had a good clustering result for the dataset.

Table 2. Table showing different lengths and respective counts of sequence reads obtained for Hypothalamus tissue of bovine calf.

Length	Count
12	8909
13	14469
14	21251
15	36120
16	33846
17	130451
18	376532
19	45728
20	311231
21	390484
22	652661
23	353479
24	126212
25	32965
26	17249

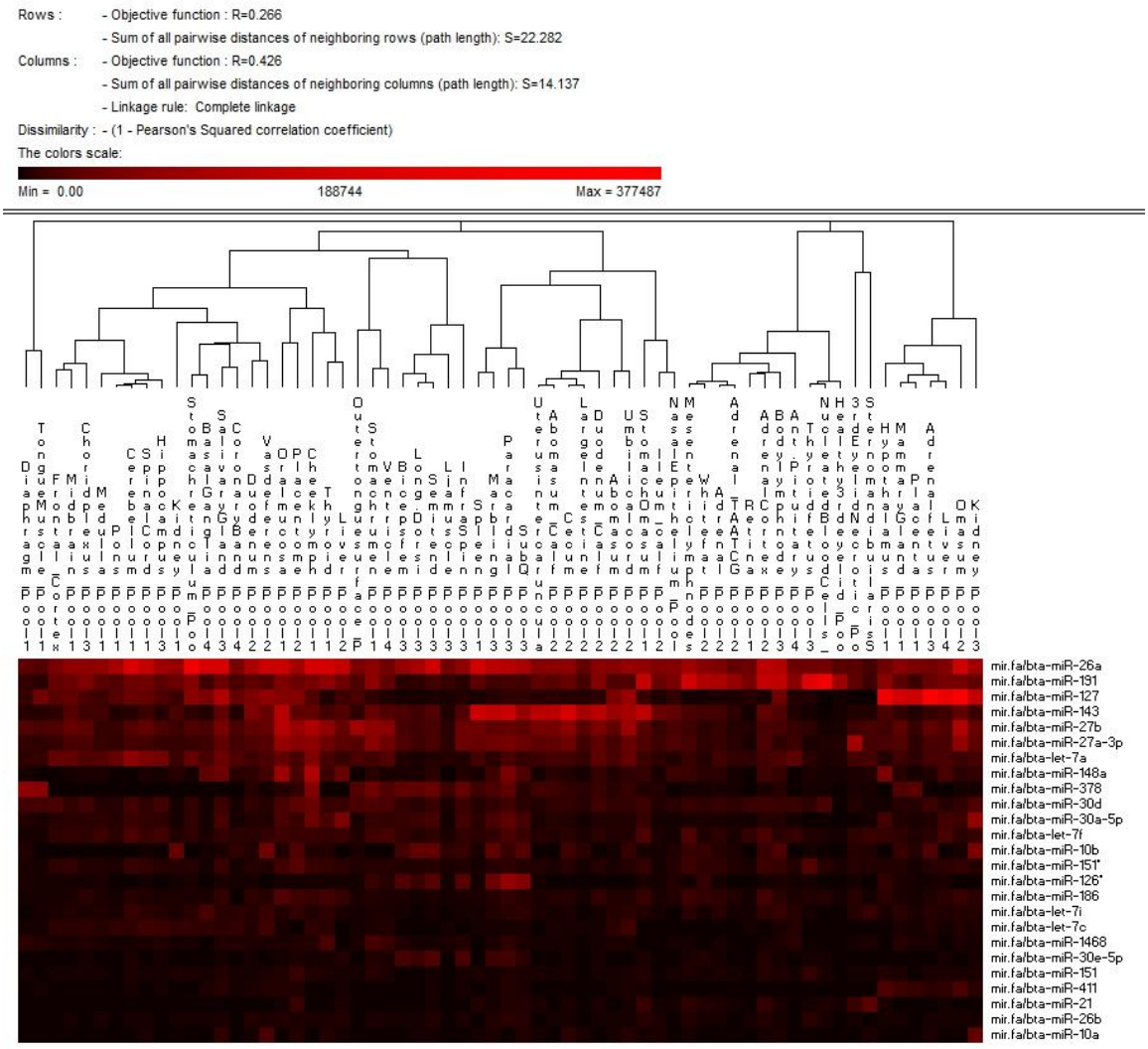

Figure 3. : A hierarchical clustering of the 64 tissues from bovine Calf and Fetus. The color scale varies from black to bright red where black represents low HNA value and bright red represents high HNA value.

3.3 Calf vs. Fetus

Comparison study of tissues present in both Calf and Fetus was carried out. Three different tissues, Abomasum, Kidney and Liver miRNA datasets were available for the analysis. In order to study these tissues, a one against one comparison analysis was done using a bar plot for the top 50 well expressing miRNAs. The miRNAs were first sorted to arrange them in descending manner and the top 50 well expressing miRNAs were used as a subset. When this comparison was carried out, we could see the difference in expression between the developmental stage of a tissue and a fully developed tissue. Figure 4 shows us the differential expression in Kidney of bovine Calf and Fetus. We observe that only four miRNAs express better in Calf as compared to Fetus which has 21 better expressing miRNAs (Table 3). This shows that the miRNAs in Kidney of Fetus are more active and play an important role in developmental stage. The same expression pattern was seen for Abomasum of bovine Calf and Fetus (Figure 5). We observe that 4 miRNAs express better in Abomasum of Calf and 5 miRNAs express better in Abomasum of Fetus (Table 4). However, for the Liver of Calf and Fetus expression results were observed to be different. We could observe that there were 10 miRNAs express better in Calf and only 8 miRNAs expressed better in Fetus (Table 5). Thus we observe variable expression patterns in these tissues and this could be attributed to the regulatory role of the miRNAs in the tissues at different stages of development. In order to identify whether these results occurred randomly or whether they are statistically significant, p-values for the tissue pairs were calculated and plotted. For example, the tissue pair of Kidney in Calf and Fetus was studied for their p-values in order to identify

their significance. The hits normalized abundance values for these pairs of tissues were saved in a text file and submitted to Metastats to calculate their p-values. In statistics, the p-value is the probability of obtaining a test statistic at least as extreme as the one that was actually observed, assuming that the null hypothesis is true. In statistical testing, a p-value of 0.05 or less is said to be statistically significant [39]. p-value is used here in order to ascertain whether the obtained results here have occurred randomly or to find out if they have a statistical significance. If they have values lower than 0.05, they show us that the miRNAs are variably expressed in the different tissue pairs. From the obtained results, it could be seen that the p-values were significantly lesser and shows us that the variable expression for the tissue pairs was not randomly obtained. Figure 7 shows us the low p-values obtained for the different tissue pairs for miRNAs. For miR-92b, miR-26, miR-99b and miR-30f all the three tissue pairs showed a low p-value score indicating that there is variability in expression for all the three tissue pairs. This shows that these miRNAs play a regulatory role and variably express in developmental stages and fully developed stage (Table 6).

Figure 4. Kidney expression in Calf and Fetus. x-axis displays the different miRNAs expressed in both Calf and Fetus kidney, y-axis displays the different hits normalized abundance values. Blue bars represent kidney of Calf and red bars represent kidney of Fetus.

Table 3. Tables showing differentially expressing miRNAs in Calf and Fetus for the tissue Kidney.

miRNA's Expression in Calf
miR-191
miR-143
miR-125b
miR-7c

miRNA's Expression in Fetus:
miR-127
miR-26a
miR-30a-5p
miR-10b
miR-27b
miR-10a
miR-27a-3p
miR-30d
miR-99b
miR-92
miR-92a
let-7f
miR-148a
miR-660
miR-30f
miR-411
miR-204
miR-186
let-7i
miR-410
miR-30c

Figure 5. Abomasum expression in Calf and Fetus. x-axis displays the different miRNAs expressed in both Calf and Fetus abomasum, y-axis displays the different hits normalized abundance values. Blue bars represent abomasum of Calf and red bars represent abomasum.

Table 4. Tables showing differentially expressing miRNAs in Calf and Fetus for the tissue Abomasum.

miR high Expression in Calf
miR-143
miR-27a-3p
miR-378
miR-24-3p

miR high Expression in Fetus
miR-127
miR-145
miR-92
miR-92a
miR-7c

Table 5. Tables showing differentially expressing miRNAs in Calf and Fetus for the tissue Liver.

miRNA highly expressed in Calf:
miR-26a
miR-27b
miR-27a-3p
miR-30a-5p
miR-143
miR-186
miR-30d
miR-192
miR-26b
miR-30c-5p

miRNA highly expressed in Fetus:
miR-127
miR-7a
miR-486
miR-92
miR-92a
miR-204
miR-411
miR-25

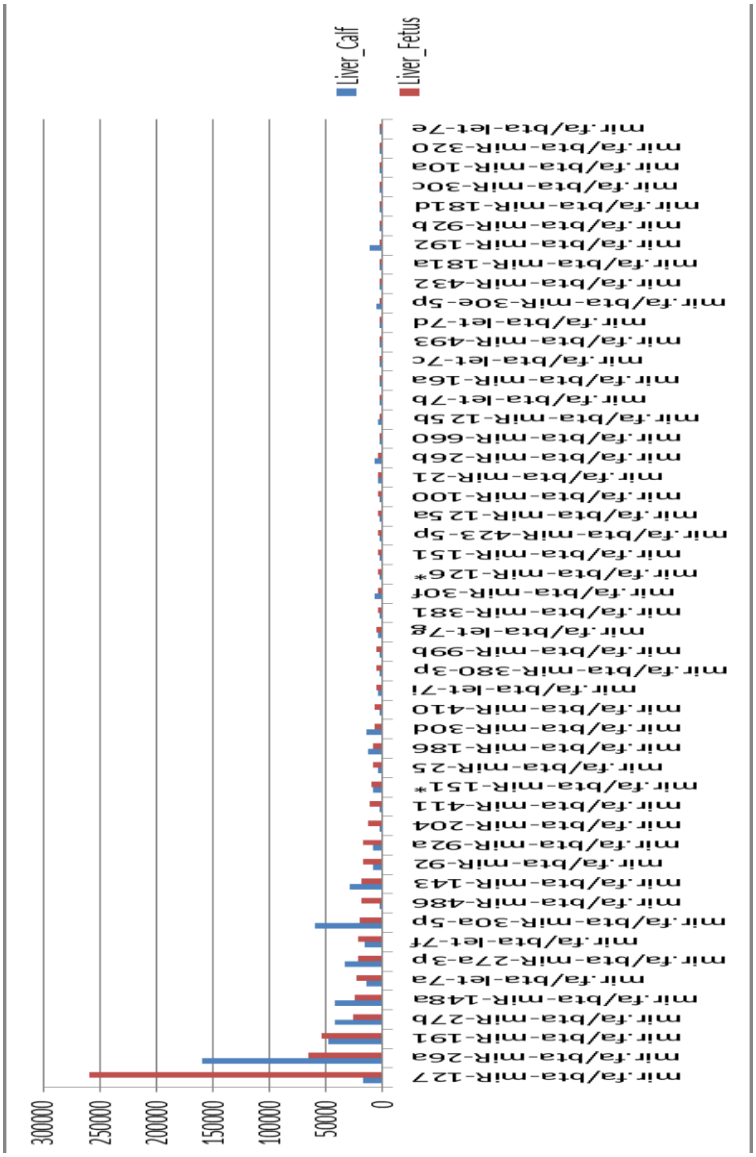

Figure 6. Liver expression in Calf and Fetus. x-axis displays the different miRNAs expressed in both Calf and Fetus kidney, y-axis displays the different hits normalized abundance values. Blue bars represent Liver of Calf and red bars represent Liver of Fetus.

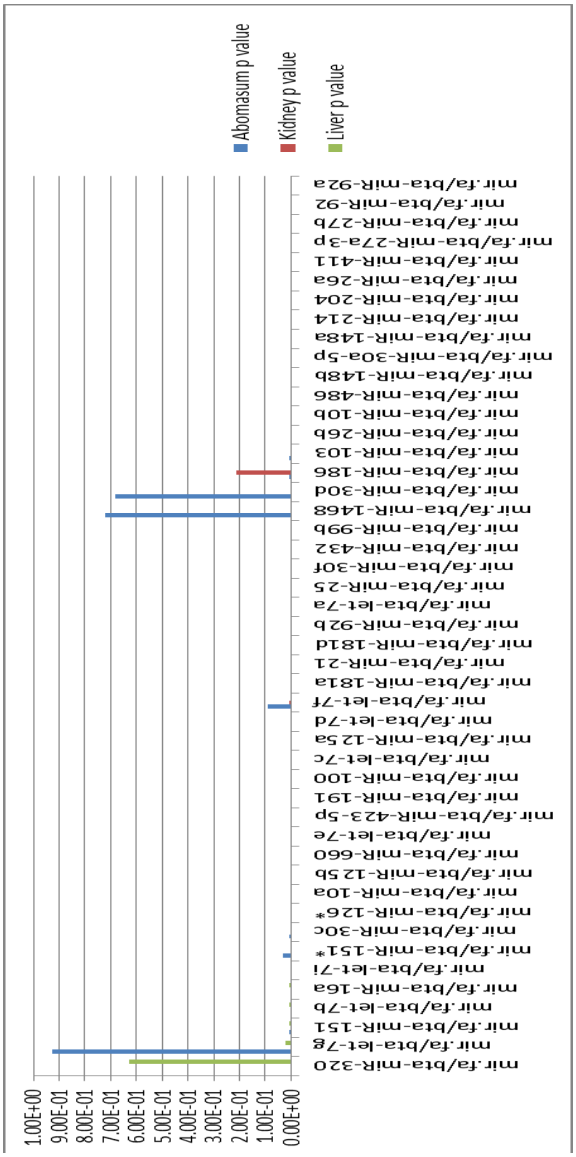

Figure 7. A plot showing the p-values of different tissue pairs for different miRNAs. x-axis represents the miRNAs and the y-axis represents the p-value.

Table 6. Table showing a subset of miRNAs with significant p-values for the tissue pairs.

Name of microRNA	Abomasum p-value	Kidney p-value	Liver p-value
mir.fa/bta-miR-92b	6.56E-32	4.16E-11	9.56E-110
mir.fa/bta-let-7a	6.70E-20	0	9.98E-166
mir.fa/bta-miR-25	4.62E-10	9.18E-144	3.18E-248
mir.fa/bta-miR-30f	5.25E-18	7.91E-296	2.64E-255
mir.fa/bta-miR-432	0	0	2.25E-260
mir.fa/bta-miR-99b	1.02E-20	2.19E-156	9.00E-303

3.4 Well-expressed miRNAs

Well-expressed miRNAs are those miRNAs which have a high hits normalized abundance value across all the tissues. Another criterion to define them as well-expressing miRNAs was that the lowest expressing miRNA should have a hits normalized abundance value of atleast 10% of the highest expressing miRNA for that particular tissue. On filtering based on this criterion, 15 different miRNAs were found to satisfy the criterion. These 15 different miRNAs are as shown in Table 7. Also, these 15 miRNAs are the most conserved miRNAs among the total miRNAs that hit the miRBase

Table 7. Table shows well-expressing miRNAs across all tissues. Table shows the miRNAs and the tissues which have the highest expression for the miRNA and the lowest expression for the same miRNA.

No.	miRNA	Highest Expression Tissue	Lowest Expression Tissue
1	miR-26a	Cerebellum (210885.5)	Sternomandibuilaris Skeletal(20347)
2	miR-27b	Oral Mucosa(128262.5)	Nucleated Blood Cells(2171.5)
3	miR-27a-3p	3rd eyelid necrotic (104943.5)	Retina (3737.66)
4	let-7a	Cerebellum (78931)	Long. Dorsi (3983)
5	miR-30d	Cheek lymph (42568.41)	Healthy 3rd Eyelid (1148)
6	miR-151*	Placentome(30826.5)	Oral Mucosa (1544.5)
7	miR-186	Thyroid_Pool1(19238.67)	Coronary Band(1459.5)
8	let-7f	Mesenteric lymph node(24770)	Oral Mucosa (1544.5)
9	miR-30f	Stomach Rumen (10449)	Healthy 3rd Eyelid (586)
10	miR-423-5p	Nucleated Blood Cells(5111)	SubQ(299.67)
11	let-7g	Cerebellum(12267.75)	Nasal Epithelium(575.83)
12	let-7d	Umbilical Cord(3419)	Healthy 3rd Eyelid(253)
13	miR-26b	Paracardial (8652)	Diaphragm(792.33)
14	miR-23a	Spleen(4632)	Retina (184)
15	miR-361	Oral Mucosa(1544.5)	Spleen(142.5)

database. All the other miRNAs did not show the same conservation across tissues. Cerebellum has the highest hits normalized abundance value for 3 different miRNAs. This shows that Cerebellum has a lot of regulatory control handled by the miRNAs. Brain being one of the complex organs of the body and being involved in decision making process, a further study into why there is high expression of miRNAs in the Cerebellum shall help in understanding the functional role of the miRNAs. For miR-27a-3p, we see that it expresses highest in an infected necrotic eye and expresses the lowest in the retina. This again displays the regulatory role of the miRNAs in the tissues and the role they play in functional regulation.

In order to identify the top occurring tissues which have the highest hits normalized abundance values in well-expressed miRNAs, the top 10 highest hits normalized values for each well-expressing miRNA were selected and their occurrences were counted. Based on this count, there were 28 tissues which had high expression values for the well-expressed miRNAs. The results are as shown in Table 8. Cheek lymph occurs in 11 different well-expressing miRNAs, Placentome occurs in 10 different well-expressing miRNAs, Cerebellum occurs in 8 different well-expressing miRNAs. These results show the regulatory role by miRNAs in these tissues since there is a high expression of miRNAs in these tissues [6]. Thus due to the higher the expression of these miRNAs in these tissues, they clearly play a regulatory role [2,6,38]. The occurrence of 10 different well-expressing miRNAs in Placentome indicates the regulatory role played by miRNAs in the developmental stage [6,38]. To understand the occurrence of 8 different well-expressing miRNAs in Cerebellum needs deep-sequencing in order to

identify the role played by these miRNAs and to detect if the tissue specificity of these miRNAs contributes in the decision making process of the brain. The results from the Table 8 also show different gland tissues: cheek lymph, mesenteric lymph node, salivary gland and thyroid. Glands function by secreting hormones into the body which in turn affects the cell metabolism. Presence of high expressing miRNAs in the gland tissues shows the regulatory role performed by these miRNAs. A deep sequencing study can help to validate these findings.

Table 8. Top occuring tissues in well-expressing miRNAs with the count of their occurrences as the highest occurring tissue for different miRNAs.

No.	Top Occuring Tissues in Well Expressing miRNAs	No. of occurrences as highest expressing tissue
1	Cheek lymph	11
2	Placentome	10
3	Cerebellum	8
4	Mesenteric lymph node	7
5	Salivary Gland	7
6	Omasum	7
7	Thyroid - Pool1	7
8	Ventricle	7
9	Adrenal Cortex	6
10	Stomach rumen	6
11	Oral Mucosa	6
12	Paracardial	5
13	Umbilical Cord	5
14	Stomach Omasum	4
15	Thyroid fetus	3
16	Body lymph node	3
17	Liver - Pool 2	3
18	Hippocampus	3
19	Spinal Cord	3
20	Duodenum	3
21	Outer tongue surface	3
22	Frontal Cortex	3
23	Liver - Pool 4	2
24	Kidney Pool3	2
25	Nucleated Blood Cells	2
26	Spleen	2
27	Semitend	2
28	Stomach Reticulum	2

4 SUMMARY AND CONCLUSION

The hierarchical clustering analysis on the 64 different tissues of bovine Calf and Fetus was carried out to see if certain tissue classes grouped together and showed conformity in accordance to the Brenda Tissue Ontology. It was seen that most of the tissue classes showed conformity to the Brenda Tissue Ontology except for one tissue class, embryonic tissue. These results show us that the dataset followed a Brenda Tissue Ontology. On confirmation of these results, the different tissue classes were grouped according to Brenda Tissue Ontology and studied to see tissue specificity of miRNAs. Even though we observed many miRNAs being highly conserved across different tissues, there are certain miRNAs which showed tissue specificity. Studying the tissues in its different developmental stages showed higher expression in early stages of development in comparison to fully developed tissue. These results hence show the relative abundance and tissue specificity of miRNAs. Further analysis by deep sequencing will allow for understanding of their regulatory role in physiology or function modification in these tissues.

5 FUTURE DIRECTIONS

A pipeline for prediction of novel miRNAs is being implemented. In order to identify novel miRNAs, miRDeep a computational prediction tool is being used which allows carrying out the prediction by using the bovine genome as a reference [31,32]. The alignment files of all the tissues generated using Flicker are submitted to miRDeep and possible novel miRNAs are identified. These predictions shall help us in identifying microRNA genes. Identification and validation of these novel microRNA genes shall allow for better understanding of the regulatory role of miRNAs played in bovine Calf and Fetus.

REFERENCES

REFERENCES

1. **Berezikov, E., Cuppen, E., Plasterk, R.H.** (2006). Approaches to microRNA discovery. Nat Genet **38**.

2. **Ambros, V.** (2001). microRNAs: tiny regulators with great potential. Cell **107(7):**823-6.

3. **Ambros, V., Bartel, B., Bartel, D.P., Burge, C.B., Carrington, J.C., Chen, X., Dreyfuss, G., Eddy, S.R., Griffiths-Jones, S., Mashall, M., Matzke, M., Ruvkun, G., Tuschl, T.** (2003). A uniform system for microRNA annotation. RNA **9(3):**277-9.

4. **Song, L., Tuan, R.S. (**2006). MicroRNAs and cell differentiation in mammalian development. Birth Defects Res C Embryo Today **78(2):**140-9.

5. **Lee, C.T., Risom, T., Strauss, W.M.** (2006). MicroRNAs in mammalian development. Birth Defects Res C Embryo Today **78(2):**129-39.

6. **Wienholds, E., Plasterk, R.H.** (2005). MicroRNA function in animal development. FEBS Lett **579(26):**5911-22.

7. **Alvarez-Garcia, I., Miska, E.A.** (2005). MicroRNA functions in animal development and human disease. Development **132(21):**4653-62.

8. **Jin, W., Grant, J.R., Stothard, P., Moore, S.S., Guan, L.L.** (2009). Characterization of bovine miRNAs by sequencing and bioinformatics analysis. BMC Mol Biol **16**; 10:90.

9. **Place, R.F., Li, L.C., Pookot, D., Noonan, E.J., Dahiya, R.** (2008). MicroRNA-373 induces expression of genes with complementary promoter sequences. Proc Natl Acad Sci USA **105(5):**1608-13.

10. **Vasudevan, S., Tong, Y., Steitz, J.A.** (2007). Switching from repression to activation: microRNAs can up-regulate translation. Science **318(5858):**1931-4.

11. **Lee, Y., Jeon, K., Lee, J.T., Kim, S., Kim, V.N.** (2002). MicroRNA maturation: stepwise processing and subcellular localization. EMBO J **21(17):**4663-70.

12. **Lee, Y., Kim, M., Han, J., Yeom, K.H., Lee, S., Baek, S.H., Kim, V.N.** (2004). MicroRNA genes are transcribed by RNA polymerase II. EMBO J **23(20):**4051-60.

13. **Yi, R., Qin, Y., Macara, I.G., Cullen, B.R.** (2003). Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. Genes Dev **17(24):** 3011-6.

14. **He, L., Hannon, G.J.** (2004). MicroRNAs: small RNAs with a big role in gene regulation. Nat Rev Genet **5(7):** 522-31.

15. **Gremse, M., Chang, A., Schomburg, I., Grote, A., Scheer, M., Schomburg, D.** (2011). The BRENDA Tissue Ontology (BTO): the first all-integrating ontology of all organisms for enzyme sources. Nucleic Acids Res **39:**D507-13.

16. http://www.illumina.com/support/documentation.ilmn. (Accessed on 8/23/12).

17. **Smith, T.F., Waterman, M.S.** (1981). Identification of common molecular subsequences. J Mol Biol **147(1):**195-7.

18. **Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J**. (1990). Basic local alignment search tool. J Mol Biol **215(3):**403-10.

19. **Li, H., Handsaker, Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R**. (2009). The Sequence alignment/map (SAM) format and SAMtools. Bioinformatics **25**, 2078-9.

20. **James, T.R., Helga, T., Wendy, W., Mitchell, G., Eric, S.L., Gad, G., Jill, P.M.** (2011). Integrative Genomics Viewer. Nature Biotechnology **29**, 24-26.

21. **Lu, C., Tej, S.S., Luo, S., Haudenschild, C.D, Meyers, B.C., Green, P.J.** (2005). Elucidation of the small RNA component of the transcriptome. Science **309:**1567-1569.

22. **Mallory, A.C., Vaucheret, H.** (2006). Functionf of microRNAs and related small RNAs in plants. Nat Genet **38** Suppl: S31-S36.

23. **Nobuta, K., McCormick, K., Nakaon, M., and Meyers, B.** (2010). Bioinformatics analysis of small RNAs in plants using next generation sequencing technologies. Methods Mol Biol **592:**79-106.

24. **Caraux, G., Pinloche, S.** (2005). PermutMatrix: a graphical environment to arrange gene expression profiles in optimal linear order. Bioinformatics **21(7):**1280-1281.

25. **Eisen, M.B., Spellman, P.T., Brown, P.O., Botstein, D.** (1998). Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci USA **95(25):**14863-8.

26. **White, J., Nagarajan, N., Pop, M**. (2009). Statistical Methods for Detecting Differentially Abundant Features in Clinical Metagenomic Samples. PLoS Comp. Bio **5(4).**

27. **Kozomara, A., Griffiths-Jones, S.** (2011). miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic Acids Res **39:**D152-57.

28. **Griffiths-Jones, S., Saini, H.K., van Dongen, S., Enright, A.J.** (2008). miRBase: tools for microRNA genomics. Nucleic Acids Res **36:**D154-58.

29. **Griffiths-Jones, S., Grocock, R.J., van Dongen, S., Bateman, A., Enright, A.J.** (2006). miRBase: microRNA sequences, targets and gene nomenclature. Nucleic Acids Res **34:**D140-44.

30. **Griffiths-Jones, S.** (2004). The microRNA Registry. Nucleic Acids Res **32:**D109-111.

31. **Li, Y., Zhang, Z., Liu, F., Vongsangnak, W., Jing, Q., and Shen, B.** (2012). Performance comparison and evaluation of software tools for microRNA deep-sequencing data analysis. Nucleic Acids Res **40(10):**4298-305.

32. **Friedlander, M., Chen, W., Adamidi, C., Maaskola, J., Einspanier, R., Knespel, S., and Rajewsky, N.** (2008). Discovering microRNAs from deep sequencing data using miRDeep. Nat Biotechnol **26(4):**407-15.

33. http://www.illumina.com/documents/products/techspotlights/techspotlight_sequencing.pdf (Accessed on 8/23/12).

34. **Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., Harris, M.A., Hill, D.P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J.C., Richardson, J.E., Ringwald, M., Rubin, G.M., Sherlock, G.** (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet **25(1):**25-9.

35. http://www.illumina.com/technology/small_rna.ilmn.(Accessed on 8/23/12).

36. **Motulsky, H.** (1995). Intuitive Biostatistics: A Nonmathematical Guide to Statistical Thinking. Oxford University Press.

37. **Tan, P., Kumar, V., Steinbach, M.** (2006). Introduction to Data Mining. Pearson Education, Inc.

38. **Lim, L.P., Lau, N.C., Garrett-Engel, P., Grimson, A., Schelter, J.M., Castle, J., Bartel, D.P., Linsley, P.S. and Johnson, J.M.** (2005). Microarray analysis shows that some microRNAs downregulate large number of target mRNAs. Nature. **463:**769-773.

39. **Rosner, B.** (2011). Fundamentals of Biostatistics. Pacific Grove.

CURRICULUM VITAE

Naga Sridhar Betrapally graduated from Visvesvaraya Technological University, Bangalore, India, in 2008. He received his Bachelor of Engineering from Visvesvaraya Technological University in 2008. He was employed as a Graduate Research Assistant in George Mason University for one and a half years. He shall receive his Master of Science in Bioinformatics and Computational Biology from George Mason University in 2012.