EXOSOMAL SMALL RNA SURVEY OF BOVINE MILK: DISCOVERY OF SMALL RNA SIGNATURES ACROSS BACTERIAL INFECTIONS OF THE BOVINE MAMMARY GLANDS

by

Kshama Aswath

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

Committee:

_____	Dr. Tad Sonstegard, Dissertation Director
_____	Dr. Donald Seto, Committee Chair
_____	Dr. Jason Kinser, Committee Member
_____	Dr. Lakshmi Kumar Matukumalli, Committee Member
_____	Dr. James Willet, Director, School of Systems Biology
_____	Dr. Donna M. Fox, Associate Dean, Office of Student Affairs & Special Programs, College of Science
_____	Dr. Peggy Agouris, Dean, College of Science
Date: _____	Spring Semester 2015 George Mason University Fairfax, VA

Exosomal Small RNA Survey of Bovine Milk: Discovery of Small RNA Signatures across Bacterial Infections of the Bovine Mammary Glands

A Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at George Mason University

by

Kshama Aswath
Master of Arts
Wayne State University, 2001
Master of Science
Bangalore University, 1998
Bachelor of Science
Bangalore University, 1996

Director: Tad Sonstegard, Scientist
Animal Genetics Improvement Laboratory, Agriculture Research Services, USDA

Spring Semester 2015
George Mason University
Fairfax, VA

Dedication

This is dedicated to my dad, Prof. G. Aswathanarayana, my mom, Prof. M. R. Lakshmi Devi, my loving husband Jayaram and kids, Aditya and Amogha. I would not have reached this stage in my academic pursuit, if not, for the constant support and encouragement that my parents showered on me and for teaching me the value of education, guiding me with their wisdom and tough love and mostly, never losing faith in my abilities. I dedicate this to my loving husband who is my best friend, confidant and the toughest critic I have ever known, who believed in me and supported my academic goals, since the day we met and without whose constant support, sacrifice and faith this would not have been possible. My lovely kids have been the strength behind my determination to keep going through those challenging and frustrating times, who showered on me their loving hugs and their awesome jokes and who always were very understanding of the long hours and late nights, times when I was attending late evening classes, studying for exams and preparing the dissertation. I am so blessed and thankful for the support and love I have from my family in being where I am today in my educational pursuit.

Acknowledgements

My heartfelt thanks goes to Dr. Steven Schroeder, who provided computational resources and spent countless hours striking insightful discussions, providing exhaustive feedback about approaches in computational analysis, molding a bioinformatician out of a wet-lab biologist. Dr. Schroeder's constant guidance and support in this research was not only a guiding light but also a lesson that taught me to approach a computational challenge without getting intimidated. My love for bioinformatics is forever indebted to his training and guidance.

I extend my sincere thanks to Dr. Tad Sonstegard, my advisor, for setting high standards and supporting my research goals, by not only providing laboratory resources for the experimental work but, whose thought provoking questions kept me on my toes constantly. His guidance in helping me write a better dissertation has been immense. I am blessed and thankful to have him as my advisor!

My respectful thanks goes to my committee chair person, Dr. Donald Seto and my committee members Dr. Jason Kinser, Dr. Lakshmi Kumar Matukumalli, for their constant encouragement and support through all these years of graduate school and for all the time and effort they put in reviewing and guiding me through the process of dissertation preparation.

A special thanks to the bioinformatics administrative staff at George Mason University, School of Systems Biology, Ms. Diane St. Germain, for making sure things are taken care of at the administrative end of graduate process and Mr. Chris Ryan, for his server support throughout the graduate school years.

Last but not the least, I would extend my sincere thanks to my collaborators Mr. Jiajee Sun, Dr. Reinhardt's group at Ames, Iowa and Dr. David Lynn at TEAGSC, Dublin, Ireland for making me a part of this study and Dr. David Rosenkranz at Johannes-Guttenberg University, for providing timely guidance with proTRAC software.

Table of Contents

Page
List of Tables viii
List of Figures ix
Abstract xi
Chapter 1 : Mastitis 1
1.1 Stages of mastitis in dairy cattle 2
1.2 Pathogens causing mastitis in dairy cattle 2
1.2.1 Modes of entry of the pathogens 4
1.2.2 Infection response in milk producing cows 5
1.3 Current available detection methods and limitations 5
1.4 Summary 7
Chapter 2 : Significance of the proposed study and goals 8
2.1 Goal 1: To build a small RNA profile in bovine milk exosomes 9
2.2 Goal 2: To assess differential expression of miRNA genes 10
2.3 Goal 3: To create a comprehensive list of functional bovine miRNAs 11
2.4 Goal 4: To explore possible piRNAs in bovine milk exosomes 11
2.5 Summary 12
Chapter 3 : Introduction to miRNAs 13
3.1 Biogenesis of miRNA 14
3.2 Functions of miRNA 16
3.3 miRNAs expression across various tissues 17
3.4 Exosomes: Significance in research 17
3.5 Summary 19
Chapter 4 : Data collection - experimental procedures 21
4.1 Induction of Infections in Cows 21
4.1 Induction of Infections in Cows 22
4.2 Isolation of exosomes from milk samples 24

4.3 Total RNA extraction by Trizol method 24
4.4 Small RNA library preparation 29
4.5 Next Generation sequencing setup 31
4.6 Results 32
4.7 Summary 32
Chapter 5 : Data analysis - bioinformatics procedures 33
5.1 Quality check on the raw FASTQ data 33
5.2 Assessment of sequence read lengths across raw data 35
5.3 Adaptor clipping 36
5.4 Preliminary survey of non-coding RNAs matching Rfam/cmscan screening 37
5.5 Mapping reads to the *Bos taurus* UMD3.1.73 using Novoalign 38
5.6 Counting reads using HTSeq 39
5.7 Differential expression analysis using DESeq2, an R package 41
5.8 Summary 42
Chapter 6 : Results 43
6.1 Small RNA survey using Rfam / cmscan 43
6.2 *S. uberis* analysis and read statistics 44
6.3 *S. uberis*: differential expression results 46
6.4 Summary: *S. uberis* differential expression 48
6.5 Comparison of miRNA types across each sample to the differentially expressed genes: *S. uberis* samples 49
6.6 Clustering miRNA genes against *S. uberis* infections 52
6.7 *S. aureus* analysis and read statistics 53
6.8 *S. aureus* differential expression of miRNA genes in infections over the control samples: results 54
6.9 Summary: *S. aureus* differential expression 57
6.10 Comparison of *S. aureus* miRNA types across differentially expressed miRNA genes under *S. aureus* infections. 58
6.11 Top 15 highly expressed miRNA genes in *S. aureus* infections 59
6.12 Clustering miRNA genes under *S. aureus* infections 60
6.13 *S. uberis* and *S. aureus* combined data: results 61
6.14 Cross comparison of differentially expressed genes across individual infections and combined infections 63

6.15 Summary: Combined *S. uberis* and *S. aureus* differential expression of miRNA genes 67
Chapter 7 : Exploration of possible piRNAs profiles across *S. aureus* and *S. uberis* infections 69
7.1 Significance of piRNA in this study 70
7.2 Bioinformatics analysis of piRNAs in bovine milk exosomes 71
7.3 Method 71
7.4 Results 73
7.5 Validation of piRNA reads in the data set 75
7.6 Summary 77
Chapter 8 : Conclusion and future direction of research 78
Appendix 82
A.1 Introduction to piRNAs 82
A.2 Argonaut proteins and piRNAs 83
A.3 Clusters 83
A.4 Other features of piRNAs 84
A.5 piRNA Biogenesis 85
A.6 Ping-Pong Model 86
A.7 Somatic pathway 87
A.8 Biological functions of piRNA and associated proteins 88
References 90

List of Tables

Title **Page**

Table 4.1.Validation of Total RNA extraction using Agilent Bioanalyzer. Electropherograms of each sample indicate the lengths of RNA in the samples. 26

Table 6.1. Number of unique Rfam matches to each different types of RNA species across control and *S. aureus* (represented as _ac) and *S. uberis* (represented as _ub) infections. 44

Table 6.2. List of up-regulated miRNAs, across *S. uberis* infections, from the DESeq2 analysis .Corresponding miRBASE accession numbers are indicated. (www.mirbase.org). 48

Table 6.3. List of down-regulated miRNAs, across *S. uberis* infections, from the DESeq2 analysis. Corresponding miRBASE accession numbers are provided (www.mirbase.org). 48

Table 6.4. List of 12 up-regulated miRNAs, across all *S. aureus* data provided by Ensembl searches against miRBASE database (www.mirbase.org) 56

Table 6.5. List of down-regulated miRNAs across all *S. aureus* data provided by Ensembl searches against miRBASE database (www.mirbase.org) 56

Table 6.6. List of eight miRNAs up-regulated across combined samples of *S. aureus* and *S. uberis* searched against Ensembl, with miRBASE IDs (www.mirbase.org). 62

Table 6.7. List of miRNAs down-regulated across combined *S. aureus* and *S. uberis* samples, searched against Ensemble, with miRBASE IDs (www.mirbase.org). 63

Table 6.8. Ten differentially expressed genes unique to the combined data analysis matched to miRBASE (www.mirbase.org) on Ensembl (www.ensembl.org). 65

Table 6.9. Top 12 highly expressed genes across combined samples of *S. uberis* and *S. aureus*. 66

Table 7.1. Data on clusters of piRNA *S. aureus* and *S. uberis* samples from proTRAC output.Table indicates the sample names with the total number of clusters identified with read depths. 74

Table 7.2. Sequences matching published bovine piRNAs (as a subset or as entire sequence match) of *S. aureus* and *S. uberis* data, to the *Bos taurus*, published, ovary and testes piRNA sequences, from piRBANK. 76

List of Figures

Title **Page**

Figure 3.1. Schematic representation of the biogenesis of miRNA showing transcription of miRNA in the nucleus to its transport into the cytosol. Zhu, H., & Fan, G. -C. (2011). Extracellular/circulating microRNAs and their potential role in cardiovascular disease. American Journal of Cardiovascular Disease, 1(2), 138–149. 15

Figure 4.1. Experimental workflow. A general overview of the experimental steps from sample preparation to the sequencing setup on Illumina Genome Analyzer. 22

Figure 4.2. Agilent Bioanalyzer pico-gel analysis of total RNA concentration across 18 exosomal RNA samples. Samples from *S. uberis* infection is shown with each sample lane with a band in the size range between 25 and 200 nucleotides, indicating short RNA size populations..... 28

Figure 4.3. Each sample lane shows a band in the size range between 25 and 200 nucleotides, indicating short RNA size populations. Samples from *S. aureus* infection is shown with each sample lane with a band in the size range between 25 and 200 nucleotides, indicating short RNA size populations 29

Figure 5.1. Represents a schematic outline of bioinformatics workflow followed to perform differential expression and RNA database matching. 34

Figure 5.2. Overall size distribution of the raw FASTQ reads plotted against the number of reads (read depth) at each size intervals. High prevalence of 18nt length fragments indicated. 36

Figure 6.1. Cutadapt trimmed read lengths plotted against their depth for all *S. uberis* samples. A greater number of reads are seen in the length range of 18nt to 25nt, reflecting typical miRNA lengths. 46

Figure 6.2. DESeq2 dispersion plot showing log2Foldchange of differentially expressed genes (Red squares) with a p-value of <0.05 as significantly differentially expressed.... 47

Figure 6.3. Venn diagram showing a cross comparison of differentially expressed genes of *S. uberis* treatment to the individual samples. 50

Figure 6.4. DESeq2-Rlog transformed heat map with top 15 highly expressed miRNA genes in *S. uberis* infections. Y-axis shows Ensembl IDs of genes. The x-axis provides the sample names. High expression is indicated with darker color. 51

Figure 6.5. DESeq2 clustering of miRNA genes across *S. uberis* infected animals. Clustering shows no effect of the infections on the individual animal. Darker blue rectangles indicate stronger clustering..... 53

Figure 6.6. Read size distribution across depth for all *S. aureus* samples and controls. A greater number of reads are seen in the length range of 18nt to 25nt, reflecting typical miRNA lengths. 54

Figure 6.7. DESeq2 dispersion plot showing up and down-regulated miRNA genes across *S. aureus* infections. A log2Foldchange of >1.0 are up-regulated while genes with a log2Foldchange value of -1 are down-regulated across the red threshold line..... 55
Figure 6.8. Venn diagram showing comparison of miRNA types across differentially expressed miRNAs of individual conditions, under *S. aureus* treatment. (bioinformatics.psb.ugent.be/webtools/Venn) 58
Figure 6.9. DESeq2-Rlog transformed heat map indicates top 15 highly expressed genes across all samples, *S. aureus* infection. Corresponding Ensembl IDs of these genes are provided on the y-axis with sample names in the x-axis. Darker color indicates more expression. 60
Figure 6.10. DESeq2, RLog transformed normalized clustering across all miRNA genes in *S. aureus* treated and control samples. Clustering shows no effect on the infections on the animals as the animals seem to cluster to itself over the infections..... 61
Figure 6.11. DESeq2 dispersion plot of all Aureus and Uberis combined samples. A log2Foldchange of >1.0 represents up-regulated genes and a log2Foldchange with -1 represents genes that are down-regulated. These genes are indicated in red..... 62
Figure 6.12. Venn diagram of cross comparison of differentially expressed genes. Comparison against combined samples (green), to the *S. aureus* differentially expressed genes (blue) and *S. uberis* differentially expressed genes (pink)..... 64
Figure 6.13. Heat map of RLog transformed data of top 30 highly expressed genes. Top 11 of these are represented with darker shades of blue. 66
Figure 6.14. DESeq2-Heatmap visualization of clustering of combined samples of *S. uberis* and *S. aureus* samples. Ensemble IDs are indicated on the y-axis of the graph. Sample IDs on the x-axis represents the animals across conditions. Clustering shows no effect of the infection. 67
Figure 7.1. Schematic workflow of piRNA data analysis. Each step indicates the processing of data set and is explained in section 7.3..... 73

Abstract

EXOSOMAL SMALL RNA SURVEY OF BOVINE MILK: DISCOVERY OF SMALL RNA SIGNATURES ACROSS BACTERIAL INFECTIONS OF THE BOVINE MAMMARY GLANDS

Kshama Aswath, Ph.D.

George Mason University, 2015

DissertationDirector: Dr. Tad Sonstegard

Mastitis, an endemic disease, affects the mammary glands of lactating cows and contributes to an estimated annual loss of two billion dollars to the U.S. dairy economy (Hogeveen, Huijps, & Lam, 2011). Many bacterial pathogens, for example, *Streptococcus uberis*, *Staphylococcus aureus* and *Escherichia coli*, if left undetected and untreated at sub-clinical stages of infection, can cause an endemic disease like mastitis (Barkema, Green, Bradley, & Zadoks, 2009). Mastitis negatively affects the quality and quantity of milk produced from a lactating animal (De Vliegher, Fox, Piepers, McDougall, & Barkema, 2012) and extensive use of antibiotics to treat this disease can have possible downstream implications on consumer safety, which, in turn pose a threat to human health. Recent research has shown the potential of circulating miRNAs as non-invasive biomarkers in disease diagnosis and therapeutics (Zen & Zhang, 2012). miRNAs present

in the milk are shown to be packaged into exosomes, a type of extracellular micro vesicle secreted by various cells (Valadi *et al.*, 2007).

The current study is a part of a larger comprehensive study of the functional genomics and proteomics of mastitis in dairy cattle, which was initiated at the National Animal Disease Center (NADC) in Ames, Iowa. Briefly, three groups of scientists (USDA-ARS scientists at NADC and Beltsville and TEAGASC scientists from Grange, Ireland) planned to survey the mammary-derived proteome and transcriptome (mRNA and microRNA) pre- and post-infection with *S. uberis*. To complement these studies of intracellular expression changes due to infection, the extracellular component of milk in the form of exosomes was also surveyed for changes in protein and RNA content due to infection with either *S. uberis* or *S. aureus*. This latter study of exosomes was the focus of this dissertation. Specifically, we profiled lactation related bovine miRNAs contained in milk exosomes prior to and two days post-infection with *S. uberis* and *S. aureus*, across nine Holstein-Friesian cattle.

Exosomal pellets collected pre- and post-infection were subjected to total RNA isolation, size fractionation, small RNA library construction and next-generation sequencing using the Illumina Genome Analyzer IIIx. A suite of bioinformatics tools relevant to small RNA expression analyses were implemented to identify transcript expressed and find differential expression patterns between infected and non-infected mammary glands. A total of 75.76 million reads were generated from nine libraries across the *S. aureus* infection challenge experiment and 309 different types of miRNAs were identified. For the *S. uberis* infection challenge, a total of 59.52 million reads were

generated from six libraries to reveal 109 unique miRNAs. Differential expression analysis found 16 miRNAs changed during *S. uberis* infection, while 35 miRNAs were differentially expressed during the first 48 hours of *S. aureus* infection. This novel information of total RNA contained in milk exosomes, provides a first look at the comprehensive list of miRNAs and other small RNA transcripts found across healthy and infected (bacteria) bovine mammary gland. Differences in miRNA profiles between healthy and infected milk could provide a key target to the development of molecular diagnostic markers that can be used to more accurately phenotype early (sub-clinical) detection of an endemic disease like mastitis.

Chapter 1 : Mastitis

Mastitis, an inflammatory disease of the udder tissue (Barkema *et al.,* 2009), is one of the major endemic diseases of the dairy industry that has a negative effect on the quantity and the quality of milk production in cattle (Hogeveen *et al.,* 2011). Safety of milk is of utmost concern to the food industry and to the consumer, as milk is the main component for a variety of food products and serves as a major protein source, in addition to being nutritious for child development (Reinhardt, Sacco, Nonnecke, & Lippolis, 2013).

According to the data presented by Lombard *et al.,* the U.S. dairy industry demographics includes approximately 60,000 herds, comprising 9.2M cattle consisting mainly of the Holstein breed, with herd averages of approximately 122 cows producing an estimated 97,000kg of milk per lactation cycle and a bulk tank somatic cell count (BTSCC) limit of 750,000 cells per ml (National mastitis council annual meeting proceedings, 2013). Although the estimated losses to the dairy industry worldwide due to mastitis ranges from $91 to $150 per cow, in the U.S. this rises to approximately $200 per cow, accounting to an overall loss of about an estimated two billion dollars annually to the U.S. dairy economy (Hogeveen *et al.,* 2011). Infections of mammary glands that lead to mastitis could have many repercussions, ranging from reduced milk production to a lower quality of milk (Bradley, 2002) in terms of taste and texture to the cow being

culled from the herd due to damage to the mammary gland. Hogveen *et al.,* (2011) highlighted the relatively huge economic losses to the dairy industry from mastitis. They attribute the loss to the dairy economy not only to the decreased production of milk in the dairy cattle but also to a huge cost associated with implementing preventive measures, efficacious treatment of the disease, and loss of producing cows form the herd.

1.1 Stages of mastitis in dairy cattle

A disease like mastitis can manifest in two different stages. The clinical stage of mastitis exhibits a range of visual symptoms including inflamed udder, abnormally watery milk, high fever, swollen quarters and loss of appetite. One of the preventive measures is to remove the cows from the milking string once they reach this stage of mastitis.

In sub-clinical stages, even though there is an underlying infection, the symptoms are silent, for example, a change in the composition of milk at the level of Somatic Cell Counts (SCC), including a high bacterial presence in the milk and a decrease in the overall milk production in the affected cows (Deb *et al.,* 2013).These cows may still be in the milking string if the SCC has not risen high enough and can lead to clinical mastitis, if not detected in time.

1.2 Pathogens causing mastitis in dairy cattle

Many microorganisms can infect the mammary gland and lead to mastitis. This includes: different types of bacteria, mycoplasma, yeasts and algae. All have been

implicated in causing bovine mastitis (Bradley, 2002). Typically, mastitis-causing pathogens are classified into two major categories being either 'contagious' or 'environmental' pathogens. However this classification is not as clear as it seems, as many environmental pathogens like *E. coli* and *S. uberis* could persist in the udder and could be spread among cows through contaminated milking equipments (Bradley, 2002).

Pathogens are generally adapted to survive within the host organism, specifically in the affected mammary glands. These types of pathogens are capable of inducing sub-clinical mastitis, resulting in an increase in the SCC, specifically leukocytes (predominantly neutrophils) and epithelial cells. These types of infections can typically spread from cow to cow either through aerosols, body fluids, or through contaminated milking equipments (Bradley, 2002). Some of the examples of such contagious pathogens are *S. aureus, S. dysgalactiae* and *S. agalactiae* (Barkema *et al.,* 2009).

Environmental pathogens, on the other hand, are an opportunistic invader, that are not typically adapted to survive within a host, but can invade, multiply and be easily treated. In most cases, infections caused by environmental pathogens, like the Streptococci species (*S. uberis and S. dysgalactiae*) and coliform species (*Klebsiella* mastitis and *Escherichia coli*) also can lead to clinical manifestation of mastitis (Gey, Werckenthin, Poppert, & Straubinger, 2013). A study has shown that the environmental pathogens, like *E. coli* and *S. uberis* are known to be implicated as the common pathogens in clinical cases, with *E. coli* (at 27%) and *S. uberis* (at 23%), whereas the contagious pathogens were implicated only in about 18% of the clinical cases (Bradley, 2002).

Infections of certain pathogens like the *E. coli* that cause acute clinical mastitis can be treated easily by the use of antibiotics. Infections caused by *Klebsiella. mastitis* and *S. aureus* can prove to be problematic and are a concern to the dairy industry due to the potential emergence of antibiotic resistance (De Vliegher *et al.,* 2012).

1.2.1 Modes of entry of the pathogens

There can be different modes of entry for bacteria into the bovine udder. Bacterial pathogens that can potentially cause mammary gland infections could enter the udder from the surrounding environment or could be already internally present in an animal (Bradley, 2002). However, the primary mode of entry of the microorganisms into the udder is mainly through the teat canals. In the cases of mycoplasmal infections, the main mode of entry is believed to be through aerosol transmission from an infected cow, leading to another udder infection (Hogeveen *et al.,* 2011).

Many bacterial agents like the *S. aureus* can be transferred from any contaminated milking equipments used between cows. Although new infections caused by environmental pathogens occur during dry periods, these pathogens can also be introduced during the actual process of milking (Sordillo & Streicher, 2002).

1.2.2 Infection response in milk producing cows

A natural layer of defense in combating udder infections is by way of anatomical and physical barriers. Studies have shown that the mammary glands in lactating cows can shut off the sphincters in their teats preventing the bacterial entry into the udder. Keratin, a waxy substance derived from the teat canal lining, is known to partially occlude the lumen and thereby prevent bacterial entry (Viguier, Arora, Gilmartin, Welbeck, & O'Kennedy, 2009).

Inflammation of the udder is seen as a physiological response to the infection. Inflammation is characterized by the influx of many components like Polymorphonuclear Leukocytes (PMNL), serum components, and fluids; all of which are released by the mammary gland to combat the bacterial pathogens (Sordillo & Streicher, 2002).

Studies have shown the presence of somatic cells in the milk of healthy cows, as shedding of epithelial cells. This is a normal process, which is seen in many lactating animals including cows. A base level estimation of somatic cells in the milk of uninfected cows is about 10^4 cells/ml to 10^5 cells/ml (Leitner *et al.,* 2012). Shoshani *et al.*, demonstrated that approximately 50% of the normal milk composition is epithelial cells, shed by the cows and the other estimated 50% constitutes leukocytes.

1.3 Current available detection methods and limitations

Standardized tests like the California Mastitis Test (CMT), Wisconsin Mastitis Test (WMT), and an electronic test offered by the Dairy Herd Improvement Organization

(DHIA) are some of the few methods available currently to detect bacterial infections (Deb *et al.,* 2013). These tests are based on the somatic cell counts, which are due predominantly to polymorphonuclear neutrophil (PMN) counts (Dohoo & Meek, 1982).

According to the DHIA standard, normal cows need to exhibit a lesser than 10,000 SCC/ml of milk to be considered Grade-A milk, or the herd average (in bulk tank) of less than 750,000 SCC/ml in the United States (Shittu, Abdullahi, Jibril, Mohammed, & Fasina, 2012). Bulk Tank SCC (BTSCC) include 400,000 SCC/ml for the European Union, Australia, New Zealand, and Canada, where as in Brazil, a limit of 1,000,000 SCC/ml is established (Dohoo & Meek, 1982). Now there is pressure on the United States to reduce the BTSCC levels to less than 400,000 SCC/ml for the export of milk into the global market, which is a production challenge. SCC counts of above 200,000 SCC/ml might indicate varying extent of sub-clinical mastitis cows in the herd. Although there is a slight change in the European and North American assessments of the somatic cell counts, a threshold of counts greater than 300,000 SCC/ml per cow is considered to be an infection (Shittu *et al.,* 2012).

The other types of detection methods can be based on many parameters for example enzyme levels, pH etc. Tests based on enzymatic assays measure N-acetyl-b-D-glucosaminidase (NAG-ase), lactate dehydrogenase (LDH) levels and can be very quick. However, the disadvantage is that it is a laboratory based test and cannot be performed at the "cow-side", as an immediate detection method. Identification of the causative pathogen by bacteriological cultures or by immune response (antibodies) tests such as milk ELISA and PCR, although are more specific, are time consuming. A user friendly

pH test at the cow-side is quick but is not as sensitive as the other tests (Viguier *et al.*, 2009).

1.4 Summary

A disease like mastitis poses many challenges to the dairy industry, because it impacts the profitability to the dairy industry not only by affecting the milk production but in the cost associated with treating the disease and implementing preventive measures. Hence, there is a need for the development of robust, quick detection methods to phenotype animals at the sub-clinical stages of bacterial infection, which can be administered at the cow-side and is cost effective for farmers. Such a test must provide quick, accurate detection of infection; thereby, reducing the impact of the disease on the dairy industry by helping identify resistant animals or allowing early veterinary intervention to reduce antibiotic use.

Chapter 2 : Significance of the proposed study and goals

A disease like mastitis needs to be addressed at many levels. Administering standards and developing innovative ways to identify early bacterial infections in milk producing cows are significant aspects of the U.S dairy industry in preventing huge economic losses. Although there are many detection tools available in identification of mastitis, there are many advantages and disadvantages associated with their implementation and analysis techniques (Viguier *et al*., 2009). There is a lack of robust identification tools that are based on innate molecular signatures within the cellular components, which could potentially indicate the infection at sub-clinical stages even when there is no visual indication of the disease.

Previous studies show the importance of breast milk and its components in innate and adaptive immunity of offspring (Zhou *et al*., 2012) and various studies by (Hu, Drescher, & Chen, 2012), (Valadi et al., 2007) and (Meckes, 2015) demonstrate the role of exosomal miRNAs in cell-cell communication and its potential as biomarkers in disease diagnosis. This dissertation focuses on addressing the molecular signatures in bovine milk exosomes, in an attempt to identify specific RNA transcripts as signatures of infection that can be developed in future work into biomarkers for animal phenotyping or on farm use as diagnostic management tools.

Apart from the immunological and nutritional qualities milk provides to an organism's sustenance, it also serves as an ideal bio-fluid for scientific research, due to its ease of availability through non-invasive collection. Because it is a raw commodity produced on farm daily, bovine milk is a great research resource for identification of molecular biomarkers to detect sub-clinical levels if bacterial infection prior to onset of a disease likes mastitis

This dissertation research is the first survey of RNA signatures in milk exosomes of cattle across patho-phyiological conditions caused by two representative bacterial strains. Each of the following sub-sections provides the goals that are addressed in this research to identify microRNA signatures indicative of bovine milk exosomes from infected mammary tissue, which could be used as a biomarker for early detection of mastitis, in the future.

2.1 Goal 1: To build a small RNA profile in bovine milk exosomes

The main aim of this dissertation is to build RNA profiles from bovine milk exosome samples, across different health conditions of the mammary gland, using artificially infected glands challenged with *S. uberis* and *S. aureus* bacterial strains and normal milk exosomal samples. Implementation of a suite of recent bioinformatics software packages and a thorough computational analyses are performed on the data

generated using the Illumina Genome Analyzer IIIx to identify differentially expressed microRNAs as signatures of infections

Previous studies have shown the importance of circulating miRNAs in the body fluids in humans as potential biomarkers in disease diagnosis (Zen & Zhang, 2012) and (Izumi *et al*., 2015). Presence of miRNAs in bovine milk exosomes and its uptake by human macrophages was demonstrated by Izumi *et al*., (2015). Exosomes are shown to be involved in cell-cell communication across various cell types (D'Incà & Pucillo, 2015) and are considered to potentially have a role in transporting immune related components between cells (Bobrie, Colombo, Raposo, & Théry, 2011)

This study focuses on providing a comprehensive list of bovine miRNAs, in normal milk and milk from artificial infections induced by two different bacterial strains, one representing *S. aureus* infection, which can be antibiotic resistant, and the other, *S. uberis* infection, which is a better pathogen for controlled time course studies. Completion of this goal contributes towards the existing knowledge of RNA biology in higher organisms like the dairy cows across patho-physiological conditions.

2.2 Goal 2: To assess differential expression of miRNA genes

Previous studies have shown changes in miRNA expression patterns in diseases like diabetes, heart conditions and many types of cancers across mouse and humans (Kadakkuzha, 2014). This research focused on extending that understanding to dairy cattle, to determine any differential expression based on the miRNA content in milk exosomes either derived from bacterial infections or healthy cows.

2.3 Goal 3: To create a comprehensive list of functional bovine miRNAs

This research aims at identification of all signatures small RNAs found in normal bovine milk exosomes or those induced by one or both types of artificial bacterial infections. A previous study demonstrates the role of immune related miRNAs in the breast milk and their effect on the off spring (Zhou *et al*., 2012). Studies have shown the role of exosomes in cell-cell communication (Valadi *et al*., 2007) and its effect on human pathophysiology (D'Incà & Pucillo, 2015), which can have huge implications ultimately on human consumption.

This study characterizes the profiles of bovine miRNAs in the exosomal samples across healthy and infected cows that are known and may have a functional role in immunity or other physiological pathways across other organisms.

2.4 Goal 4: To explore possible piRNAs in bovine milk exosomes

Further, this study explores the possibility of the presence of the PIWI associated RNAs, known as the piRNAs in the bovine milk exosomes. This study implements a novel software called proTRAC to explore any known piRNAs and to determine piRNA clusters in the bovine genome, using the piRNA cluster information from known human piRNA clusters from piRBANK, a database of piRNAs.

2.5 Summary

In conclusion, this research explores the molecular signatures comprising small non-coding RNAs like miRNAs and piRNAs found in bovine milk exosomes from normal and infected mammary tissue, utilizing the most current sequencing technology and a whole suite of novel bioinformatics methods. This study adds onto to the existing knowledge in RNA biology; thereby, helping develop more robust detection methods, in identification of infections at sub-clinical stages, thereby contributing towards addressing early diagnosis of a complex disease like mastitis. This research also lays foundation in building a biomarker based on the profiled miRNAs that can have a significant impact in early disease detection affecting animal well-being and health.

Chapter 3 : Introduction to miRNAs

miRNAs are short non-coding RNA that are approximately 18 to 22nt long. miRNAs are shown to be present in many organisms for example, humans, *C. elegans*, *D. melanogaster* and in plants. miRNA was characterized for the first time in the *C. elegans* genome, exhibiting complementarity to the 3' end of the target lin-4 and let-7 genes (Cortez *et al.*, 2011) and thereby silencing the genes. Studies have shown a conservation in the miRNA sequences and the expression of miRNAs in monotremes like platypus (Murchison *et al*., 2008). Non-coding RNAs are shown to be involved in many biological functions such as gene expression, chromatin modification and many different types of diseases in organisms (Kadakkuzha, 2014).

Understanding the miRNA signatures in various types of tissues and cellular entities of body fluids like exosomes is an important component upon which functional research can be built. The presence of these small RNA molecules, which are non-coding, have been shown to have regulatory roles in germ cells (Gigli & Maizon, 2013) and in physiological pathways including their dysregulation in diseases like cancer (Mulrane, McGee, Gallagher, & O'Connor, 2013).

Building a profile of molecular signatures like small RNAs across cellular components paves way for the comparative study of the entities across tissues and across organisms and helps us to understand better their biological significance.

3.1 Biogenesis of miRNA

A thorough understanding of the biogenesis of these molecules, from their origin of synthesis to their location of action is a very important aspect of a pathophysiological study. This section elucidates the biogenesis of these small RNA molecules and their various features.

Figure 3.1 illustrates the biogenesis pathway of miRNAs in eukaryotic organism. miRNAs are transcribed from the genes that encode the miRNAs which are present in many independent loci of the genome or, in some cases could be processed from the introns of other protein or RNA coding genes during splicing. Primary transcripts of miRNA are long and non-functional, and are present in the nucleus of the cell. This long primary miRNA (pri-miRNA) transcript is shuttled out of the cytoplasm and processed further to obtain small functional miRNAs of 22nt lengths.

The processing of pri-miRNA into a functional miRNA is a two-step process. Initially, RNA POL II produces the pri-miRNA from the genome. This pri-miRNA is then processed by a microprocessor complex called a Drosha-DGCR8 complex (Weber *et al*., 2010) in the nucleus to form a double-stranded pre-miRNA with secondary structures. The pre-miRNA is then exported into the cytoplasm in a controlled manner through a protein complex called Exportin. In the cytoplasm, the pre-miRNA is further processed by RNAse III Dicer to form a single guide strand of mature miRNA, and the other passenger strand is degraded. The mature miRNA functions via the miRNA RISC

complex (RNA-induced silencing complex) in binding to the target mRNA and initiating downstream regulation of the mRNA transcript (Zhu & Fan, 2011).

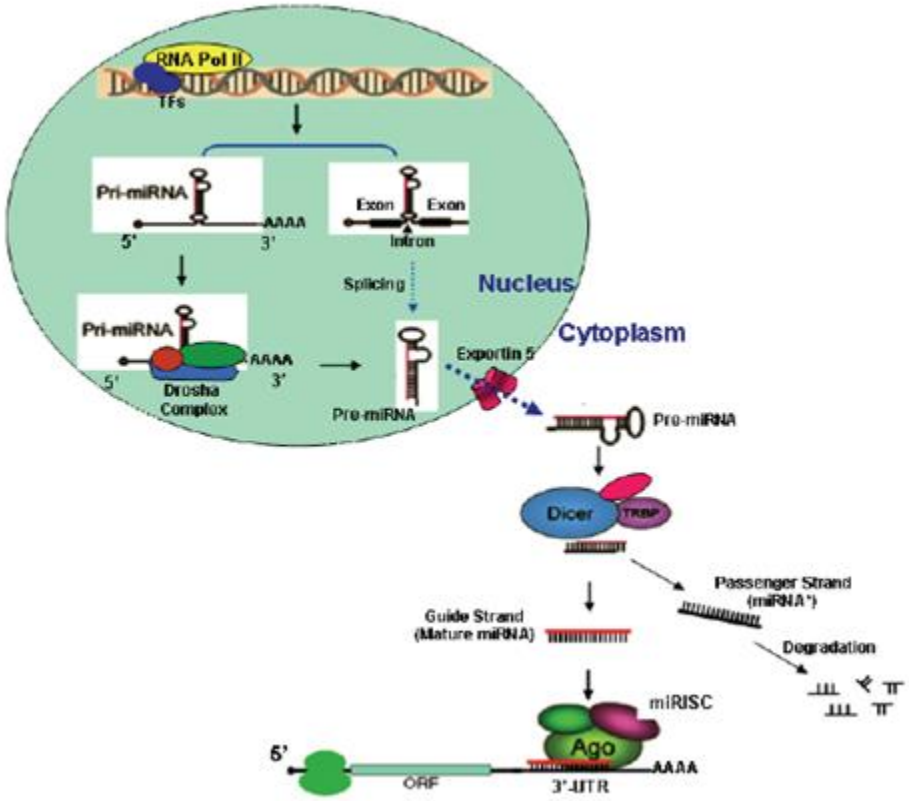


Figure 3.1. Schematic representation of the biogenesis of miRNA showing transcription of miRNA in the nucleus to its transport into the cytosol. Zhu, H., & Fan, G. -C. (2011). Extracellular/circulating microRNAs and their potential role in cardiovascular disease. American Journal of Cardiovascular Disease, 1(2), 138–149.

3.2 Functions of miRNA

miRNAs are known to play regulatory roles in the development of an organism, including protein secretion and gene silencing. The regulatory role of miRNAs ranges from down-regulation of gene expression at the post transcriptional level to post transcriptional degradation of mRNA by binding to them, and to increasing the translation of proteins. Studies show evidence of many immune-related miRNAs to be abundant in breast milk exosomes of humans (Gigli & Maizon, 2013). miRNAs are also shown to play a regulatory role in certain stages of embryonic development, and the concentration of certain types of miRNAs are shown to be high at certain disease states (Pigati *et al*., 2010). In certain types of cancer, selective miRNA genes are suppressed thereby decreasing the amount of miRNA, in order to keep the tumorigenesis going (Mulrane *et al*., 2013). This indicates strongly that the expression of certain types of miRNA could play a positive regulatory role in cancer growth in the body. Many types of cancers also have shown that an increase in specific miRNA contents leads to tumor formation. This dysregulation of miRNA indicates the complexity of the miRNA function and its involvement in disease manifestation (Wang, Wang, Lu, Song, & Cui, 2010).

3.3 miRNAs expression across various tissues

Literature survey shows that miRNA expression across tissues to be varied. There are many miRNAs that are universally expressed across all types of tissues and some that are very tissue specific (Guo *et al*., 2014). The regulatory role of miRNA depends on the tissue type (Pigati *et al*., 2010). Various miRNAs are either down or up-regulated depending on the disease or normal conditions of the tissues (Lawless et al., 2014). Certain types of miRNA are formed at the trigger of an injury (Pigati *et al*., 2010). miRNAs are characterized at various developmental stages of the organism and some unique types across mammary, ovaries and testes tissues (Wang *et al*., 2010). Many are present in various body fluids, highlighting their potential role as a peregrine signaling molecule (Cortez *et al*., 2011).

Characterizing the unique types of miRNAs in normal and mastitis affected milk samples in cattle might have a great potential leading to the development of a miRNA biomarker aiding in early detection of the disease, and could lead to a better understanding of the variations in the resistance of the cows to the disease.

3.4 Exosomes: Significance in research

Exosomes are shown to carry many molecular signatures like miRNAs, mRNAs, proteins and lipids (Valadi *et al*., 2007). Also the miRNAs in the human breast milk

exosomes are shown to be protected against the harsh pH conditions in the cell and are known to play a key role in imparting immunity to the offspring (Zhou *et al*., 2012).

Since the discovery of exosomes in 1987, these micro-vesicles, ranging from 30 to 150 nm, were ignored as they were considered to be the "garbage bins" of cells known to carry unwanted biomolecules out of the cells (Bobrie *et al*., 2011). However, the break-through in exosomal study came through in 1996 when, a study published by Raposo *et al.*, showed that exosomes secreted by EBV (Epsein-Barr Virus) actually induced antigen - specific MHC class II restricted T cell response (Raposo *et al*., 1996). In 1998, a study conducted by Lotvall demonstrated that the exosomes that carry many biomolecules actually have very functional and intact mRNA and miRNAs that play a significantly role in cell-cell communication(Valadi *et al*., 2007). Many extracellular components like exosomes have been shown to be secreted into body fluids like sweat, blood, tears, urine, and milk (Weber *et al*., 2010) and a recent study demonstrates the uptake of exosomes by the human macrophages (Izumi *et al*., 2015).

These previous studies provide a strong platform to explore the miRNA signatures in the exosomes of the bovine mammary glands infected with bacteria that cause mastitis, to better understand the molecular signatures in an endemic infection like mastitis.

Depending on the type of cell, the exosomes contents may vary (Valadi *et al*., 2007). Studies have shown exosomes carry biomolecules like various types of RNAs to proteins to chemical entities as morphogens and prions (Patel, 2014). Exosomes are shown to be secreted during various types of stress responses including inflammation,

coagulation, angiogenesis, and cell death (D'Incà & Pucillo, 2015). Certain specific cell types, like the MC/9 and HMC-1 cells, are known to secrete exosomes that contain about 1300 mRNAs and about 121 miRNAs (D'Incà & Pucillo, 2015)

Exosomes are shown to be of endocytic origin membrane-bound micro vesicles that are about 40nm-100nm in diameter. Many types of cells *in vitro* have shown to release exosomes into the extracellular environment (D'Incà & Pucillo, 2015). Although exosomal contents have shown to range from proteins to nucleic acids, mainly comprising of RNAs like mRNA and large volume of small RNAs, isolated exosomal cargo does not show any evidence of DNA (Valadi *et al*., 2007). A study conducted by Valadi *et al*. concluded that the presence of RNA could determine the role of exosomes as a vehicle aiding in cell-cell communication. Although the mRNAs seem to be functional, due to the lack of the protein synthesis machinery within the exosomes, they lack the ability to synthesize protein (Patel, 2014). One of the interesting reasons exosomes are studied is their ability to clear unwanted cargo such as proteins out of the cells, and hence have a huge potential as a diagnostic biomarker in various diseases (Jin *et al*., 2014).

3.5 Summary

Based on the previous studies on functional roles of exosomes in cell-cell communication (Bobrie *et al*., 2011) and significance of the inner composition of the exosomes (Valadi *et al*., 2007), there is enough evidence to explore the constituents of bovine milk exosomes during bacterial infections.

A main aim of this research is to study specifically the miRNA content of exosomes from the bovine milk samples and to understand the effects of bacterial infections at the level of the miRNA signatures in exosomes. This research also profiles any potential piRNAs in bovine milk exosomes, particularly in comparison to homologous human piRNA loci across both types of bacterial infections of the mammary glands. Understanding these molecular signatures could enhance the knowledge in RNA biology and help develop a molecular biomarker in detection of disease like mastitis.

Chapter 4 : Data collection - Experimental procedures

This is a comprehensive study conducted to understand the effect of mastitis on the protein, mRNA and miRNA contents of milk exosomes, in cattle induced with bacterial infections, initiated at National Animal Disease Center (NADC), Ames, Iowa.

A proteomics study was performed at NADC, Agricultural Research Services (ARS), USDA (Ames, Iowa). A transcriptomics study was performed at the level of mRNA at Teagasc, Animal and Bioscience Research Department, Grange (Dunsany, County Meath, Ireland) by Dr. David Lynn's group and this miRNA study was performed at ARS / USDA (Beltsville, MD).

Figure 4.1 shows a general outline of the experimental procedure performed to extract the small RNAs from the bovine milk exosomal samples. A traditional Trizol method was used in RNA extraction. Small RNA library was built using the Illumina TruSeq small RNA protocol according to the manufacturer's protocol. Sequencing reaction is set up on Genome Analyzer IIIx according to the Illumina guidelines. Each of the steps is further explained in detail in the following subsections.

4.1 Induction of Infections in Cows

Holstein-Friesian cattle, free of any history of mastitis disease, were sourced to USDA-NADC for this study. Cows were at 90 days-in-milk (DIM) from the calving dates of 3/26/11-4/25/11. Milk samples were collected aseptically from all four quarters and subjected to bacteriological examination weekly six times pre-trial and on the day of the trial (Eckersall, Young *et al.*, 2006; Petzl, Zerbe *et al.*, 2008) before infection.

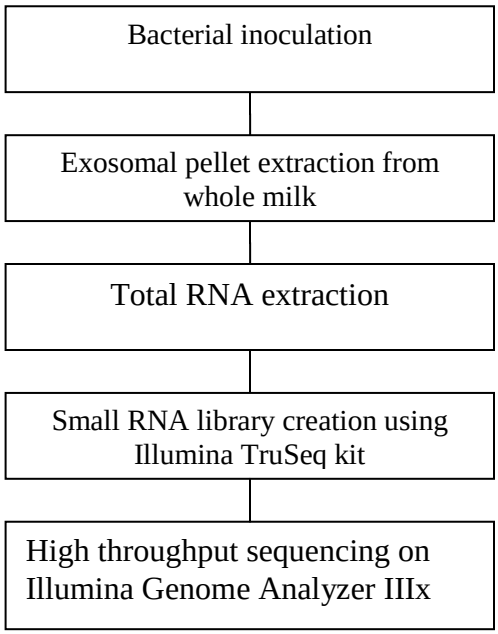


Figure 4.1. Experimental workflow. A general overview of the experimental steps from sample preparation to the sequencing setup on Illumina Genome Analyzer.

4.1 Induction of Infections in Cows

Holstein-Friesian cattle, free of any history of mastitis disease, was sourced to USDA-NADC for this study. Cows were at 90 days-in-milk (DIM) from the calving dates

of 3/26/11-4/25/11. Milk samples were collected aseptically from all four quarters and subjected to bacteriological examination weekly six times pre-trial and on the day of the trial (Eckersall, Young *et al.*, 2006; Petzl, Zerbe *et al.*, 2008) before infection. This served as the control samples (d0 or day 0 samples). The first 5 ml of milk was discarded and the following 10 ml was used in the bacteriological study and for Somatic Cell Count (SCC).

The health of the animal was documented daily three months before the pathogens were inoculated. Five Holstein cows (biological replicates) were infected with 400 cfu of *S. aureus* (Newbolt strain) and four Holstein cows were infected with 500 cfu of *S. uberis* (strain 0140) for 48 hrs, to induce sub-clinical infection.

Milk was used to assess the somatic cell counts (SCC) and the bacterial cell counts. Rectal temperatures were obtained twice a day to ensure that the animals did not have elevated body temperatures due to these induced infections. All of the animals remained normal during the course of the experiment. Infected milk samples were collected after 48hrs post infection and are referred as d2 or day2, in the protocols.

Control and Infected milk were subjected to bacteriological tests and to assessment of the number of SCC. Control milk SCC averaged at 157,000 ± 43,000 cells /ml with no detectable bacteria. After the *S. aureus* infections the SCC was at an average of 4,902,000 ± 533,000 cells/ml and bacterial counts of 8738 ± 3359 cfu/ml. The visual appearance of milk was normal even at this stage of infection.

4.2 Isolation of exosomes from milk samples

Exosomes were extracted at NADC (Ames, Iowa) from milk samples of nine Holstein-Friesian cattle before and after the infections. Collected milk was centrifuged at 10,000 g for 15 minutes. After the skim milk was cleared of any floating milk fat pellets, the supernatant was washed using 10 volumes of ice cold PBS (pH7) along with a protease inhibitor cocktail, for up to three times, until the supernatant was clear. This supernatant was used to extract the exosomes by centrifuging in a 50 Ti rotor at 150,000 g at 4 ºC for 60 minutes. Supernatant, which was the crude whey portion, was carefully separated from the bottom pellet. The loose pellet of exosomes layering the bottom tight casein pellet, was carefully transferred into another tube and re-suspended in PBS + PI solution (Reinhardt *et al.,* 2013).

Exosomal samples obtained from the above procedure were concentrated at 500 µg/sample and shipped to AGIL, ARS, USDA, for miRNA analysis.

4.3 Total RNA extraction by Trizol method

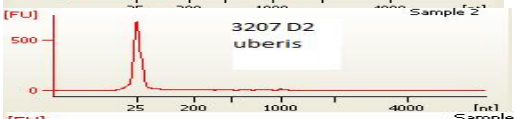
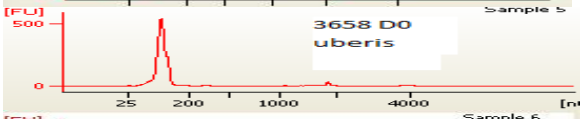
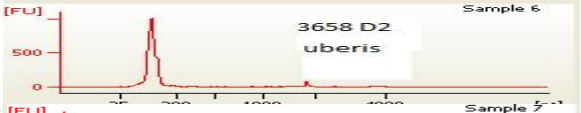
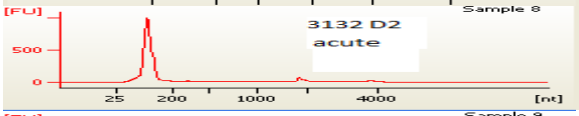
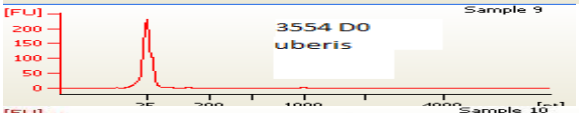
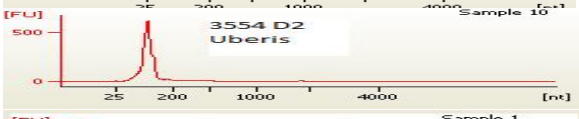
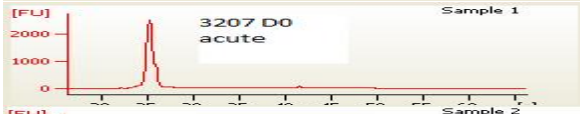
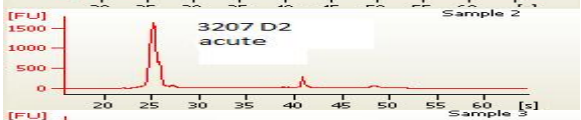
Total RNA extraction and further processing of the samples were carried on at Animal Genetics Improvement Laboratory, Beltsville, MD.

Exosomal pellets were thawed on ice and resuspended in 500 µl of Trizol solution from Ambion. After a vigorous vortex to break the pellet well, samples were incubated for 5 minutes at room temperature and 100 ul Chloroform is added at a 0.2 ml of Chloroform to 1 ml of Trizol volume. Samples were vigorously mixed by shaking for 15

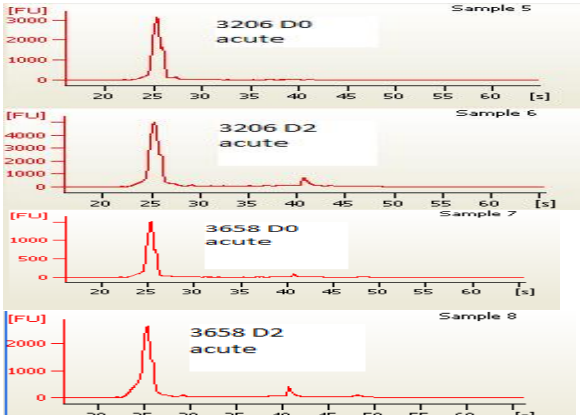
seconds each, and incubated further for 2-3 minutes at room temperature. Centrifugation at 12000 g was performed for 15 minutes at 4ºC to separate layers. The aqueous layer was transferred into a new 1.5 ml tube and 10 µg of RNAse-free glycogen from Illumina was added to the aqueous layer, along with 250 µl of 100 % isopropanol. This was mixed well by tilting the tubes by hand. After 10 minutes incubation at room temperature, the samples were centrifuged again, at 12000 g for 10 minutes at 4°C. Supernatant was removed and the gel like RNA pellet was retained for a wash with 500 µl of 75 % ethanol at a 1 ml/ml of ethanol to Trizol ratio. Samples were then briefly vortexed and centrifuged at 7500 g at 4ºC for 5 minutes. The pellet at the bottom was air dried and resuspended in 12 ul volume of ultrapure RNAse-free water provided by Illumina and incubated in heat block at 55°C for 15minutes and frozen at - 80°C until further use.

Total RNA extracted using the protocol elucidated in section 3.3, was further quantified using the Agilent Bioanalyzer using pico chips following the manufacturer's protocol. Table 4.1 provides the yield of total RNA in pg/µl, in column three, including the electropherograms in the last column. The purity of RNA species in the samples is clearly indicated based on size and the sharpness of the peak at the appropriate lengths of 25 to 200 nucleotides. Concentration of the extracted total RNA is indicated as pg/ul. Electropherograms indicate a population of RNA predominantly between 25 nt and 200 nt sizes.

Table 4.1.Validation of Total RNA extraction using Agilent Bioanalyzer. Electropherograms of each sample indicate the lengths of RNA in the samples.

Sample	Infection	Concentration	RIN	Electropherograms
3207 – D0	*S. uberis*	207 pg/ul	1	
3207-D2	*S. uberis*	534 pg/ul	1	
3318-D0	*S. uberis*	227 pg/ul	1	
3318-D2	*S. uberis*	3,966 pg/ul	1	
3658-D0	*S. uberis*	5,460 pg/ul	2.7	
3658-D2	*S. uberis*	10,167 pg/ul	2.7	
3132-D0	*S. aureus*	535 pg/ul	4.1	
3132-D2	*S. aureus*	11,365 pg/ul	2.7	
3554-D0	*S. uberis*	60 pg/ul	1	
3554-D2	*S. uberis*	6,786 pg/ul	2.6	
3207-D0	*S. aureus*	23,236 pg/ul	2.5	
3207-D2	*S. aureus*	16,597 pg/ul	N/A	
3504-D0	*S. aureus*	33,311 pg/ul	2.8	
3504-D2	*S. aureus*	70,212 pg/ul	-	

3206 – D0	*S. aureus*	31,577 pg/ul	2.6
3206-D2	*S. aureus*	66,174 pg/ul	2.8
3658-D0	*S. aureus*	14,439 pg/ul	2.7
3658-D2	*S. aureus*	35,553 pg/ul	-



Agilent Bioanalyzer provides an electrophoresis quality assessment of the samples. Figure 4.2 and Figure 4.3 is a pico-gel electrophoresis of the samples. Band sizes are in the range of 25 to 200 nt lengths which is the range of the RNA molecules extracted. A secondary band of higher lengths is also seen in some samples that could be a technical artifact or contamination.

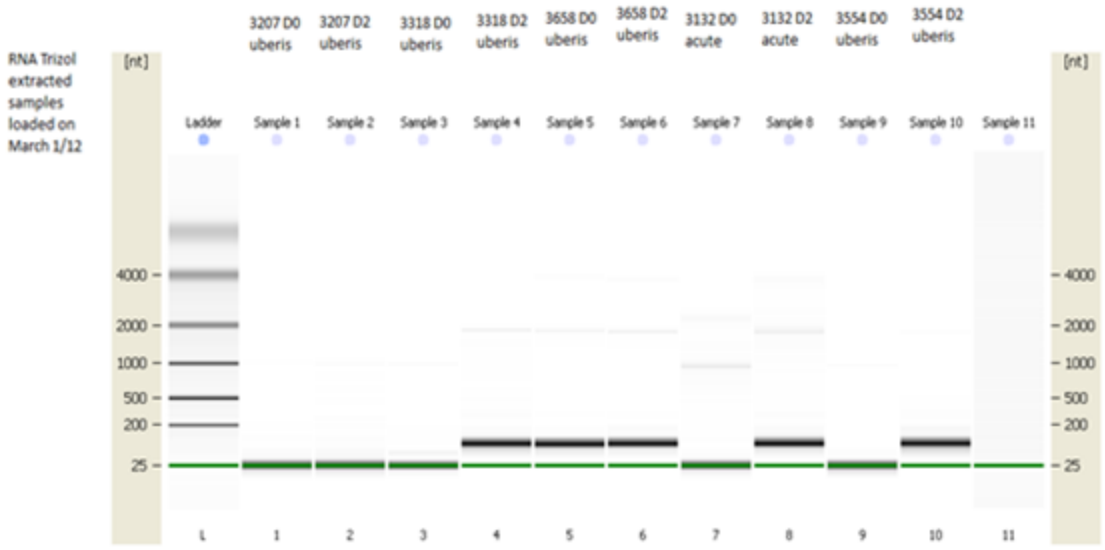


Figure 4.2. Agilent Bioanalyzer pico-gel analysis of total RNA concentration across exosomal RNA samples. Samples from *S. uberis* and *S. aureus* (indicated as acute) infections are shown with each sample lane with a band in the size range between 25 and 200 nucleotides, indicating the presence of short RNA size populations.

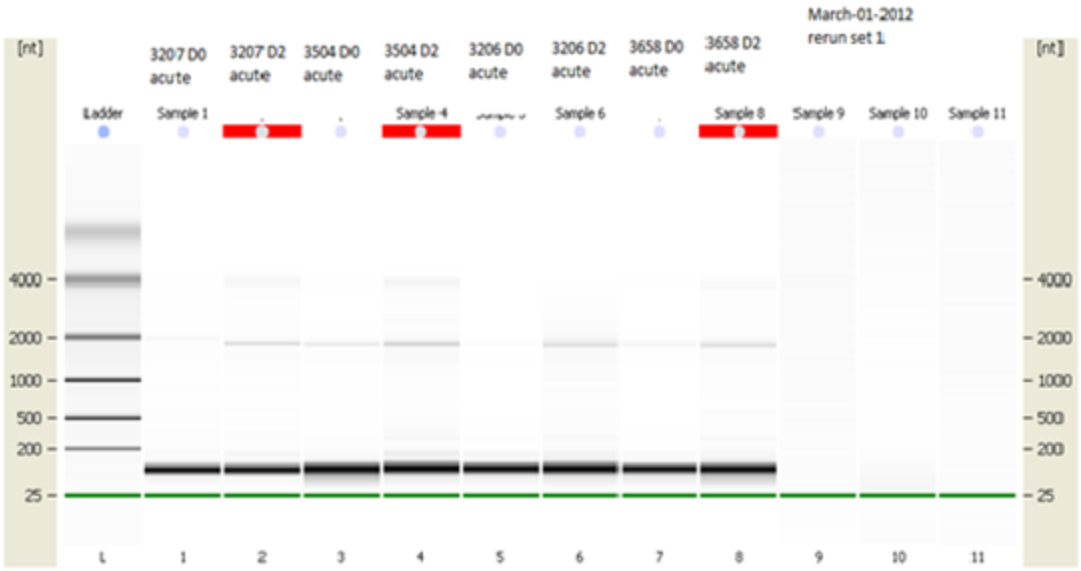


Figure 4.3. Agilent Bioanalyzer pico-gel analysis of total RNA concentration across exosomal RNA samples. Samples from *S. aureus* (indicated as acute) infections are shown with each sample lane with a band in the size range between 25 and 200 nucleotides, indicating the presence of short RNA size populations.

4.4 Small RNA library preparation

TruSeq small RNA library preparation kit, Catalog #RS-930-1012, part #15004197, Rev C. March 2011 (Illumina, San Diego, CA), was used to build the small RNA library. Total RNA extracted from the exosomes of infected and control milk samples were used as the source for library construction. All reagents used were provided in the kit except for the truncated ligase from New England Bio Labs (NEB) (Ipswich, MA) truncated ligase deletion mutant-2 from Epicenter and RNAse inhibitor.

5 µl of the total RNA extracted from all of the 18 exosomal pellets from milk samples was used in ligation with 3' adaptor. After incubation at 70°C for 2 minutes and immediate cooling on ice a mix of ligation buffer, RNAse inhibitor and T4 RNA ligase (truncated) were added and the samples were incubated at 28°C in a thermo cycler for an hour. Reaction is stopped by adding a Stop Solution (STP) and incubation was further continued for 15 minutes at 28°C.

5' adaptor was added to the reaction mix incubated for 70°C for 2 minutes and a combination of 10 mM Adenosine-Tri-Phosphate (ATP) and T4 ligase was added and further incubated for an hour at 28°C. Ligated samples were frozen up to a month at -80°C.

Adaptor-ligated total RNA was subjected to Reverse Transcriptase PCR (RT-PCR) using 12.5 mM Deoxyribonucleotide triphosphate (dNTP), RT primer (Illumina), 5X first strand buffer, Dithiothreitol (DTT) 100 mM and RNAse inhibitor and single strand synthesis II RNA transcriptase at 50°C for one hour. PCR amplification of the library was performed in MJ thermo cyclers using the reagents provided by Illumina by heating to 98°C, cooling to 60°C and then reheating to 72°C for 11 to 15 cycles.

PCR-amplified products were run on a 3% agarose gel in 1X TBE and 4 µl of ethidium bromide. 2 µl of ethidium bromide was pre-added to the running buffer. Custom ladder and High Resolution ladder (supplied with Illumina TruSeq small RNA kit) were used as standards for comparison of the band sizes. Band sizes between 140 bp to 160 bp were excised using a UV box. Approximately a band size of 22nt RNA fragment with both the 3' and 5' adaptors would add up to 147 nt in length. Bands are excised by tightly

packing the band-cuts on the gel to enhance the chances of acquiring as accurate as possible of the pool of miRNAs in the final concentration. Gel-purified band is subjected to further purification using QiaPrep gel purification kit and eluted using 11 µl of the Elution buffer (EB). Samples were stored at -20ºC until further use. Libraries were quantitated using Agilent nano and Qbit.

Libraries were quantified using Agilent and Qbit assays according to the manufacturer's protocol and 1 µl of the samples were used to construct small RNA libraries as explained in Section 4.5.

4.5 Next Generation sequencing setup

Small RNA libraries, with the exception of sample 3132-d0 (ac), *S. aureus* infected sample, were diluted to 1 nM concentration. Equal volumes of 1nM libraries that were to be multiplexed were combined together. 9µL of the multiplex pool was denatured at room temperature for 5minutes, using 0.5 µL nM , 2 N NaOH and 0.5 µL of Elution Buffer (Supplied by Qiagen, Valencia, CA) to keep the final concentration to about 1 nM. Further dilution was achieved by using 5 µl of the denatured library and adding 995 µl of pre-chilled Hybridization buffer supplied by Illumina. This is performed in order to have a final concentration for clustering at a ratio of denatured library (1 µL) for every 1pM of final concentration for clustering. Reaction mixture was vortexed and 120 µL of this dilution was aliquoted into a strip of PCR tubes compatible with the cBot. Samples were then loaded onto the cBot along with flow cell and reagents. Version 7 recipe designed for Single Read chemistry on the cBot was used for sequencing.

Samples were multiplexed by combining five unique indexed samples into a single reaction run. The index numbers indicates the type of indices used in the experiment. All the indices were provided by Illumina in the TruSeq kit catalog ("Illumina_sequencing_introduction.pdf").

4.6 Results

Total RNA was successfully extracted from the exosomes and concentrations of the RNA were assessed using Agilent Bioanalyzer. Results of extraction can be found in Table 1, section 3.3.1.

A total of 17 small RNA libraries were constructed successfully, not including sample 3132 control from the *S. aureus* infection, due to low yield. Libraries were sequenced successfully on the Illumina Genome Analyzer and raw data generated was bioinformatically analyzed as described in Chapter 5.

4.7 Summary

Data collection was performed using a whole host of experimental procedures from the induction of infection in the selected dairy cows, at the farm, to the preparation of the small RNA libraries in the laboratory and were sequenced on the Illumina Genome Analyzer IIIx. Sequencing based on the Illumina's sequence by synthesis procedure yielded short reads of 36 nucleotide length fragments. A 3 GB data output was further analyzed using a host of bioinformatics methods as elucidated in Chapter 5.

Chapter 5 : Data analysis - Bioinformatics procedures

Data analysis was performed using series of different programs on a Fedora platform ("https://getfedora.org/en/", RedHat, 2015), which provided a greater flexibility in analysis. Illumina Genome Analyzer IIIx generated about 3GB of single end, 36 base length sequence data. Bioinformatics analysis was performed using many packages and each step in elucidated in the following subsections of this chapter.

5.1 Quality check on the raw FASTQ data

Raw data output was in FASTQ format, which stores both the biological sequence information and its corresponding quality scores. For example:

```
@HWUSI-EAS1768:40:64YUEAAXX:1:37:8217:1035 1:N:0:ATCACG
GCATTGGTNGTTCAGTGGTAGAATTCTCGCCTGGAA
+
GGGGGBBC#A7??;;A>B?BGGGGGGGGGGDDGBG:
```

Quality statistics was computed on all the FASTQ files from the sequencer, to determine the quality of each individual base and across the 36 base length fragments using a FASTQC program on Galaxy server (Giardine *et al.,* 2005). According to the Phred quality scores of a range of two to 40 quality of each individual nucleotide was assessed in every single sequence. On an average the Phred scores across all samples fell in the range of 32 to 38, which is within the acceptable read quality for Illumina reads.

Not counting the last couple bases with lower scores, which is attributed to their frequency to be prone to errors, reads in this experiment passed the quality check and further analyzed.

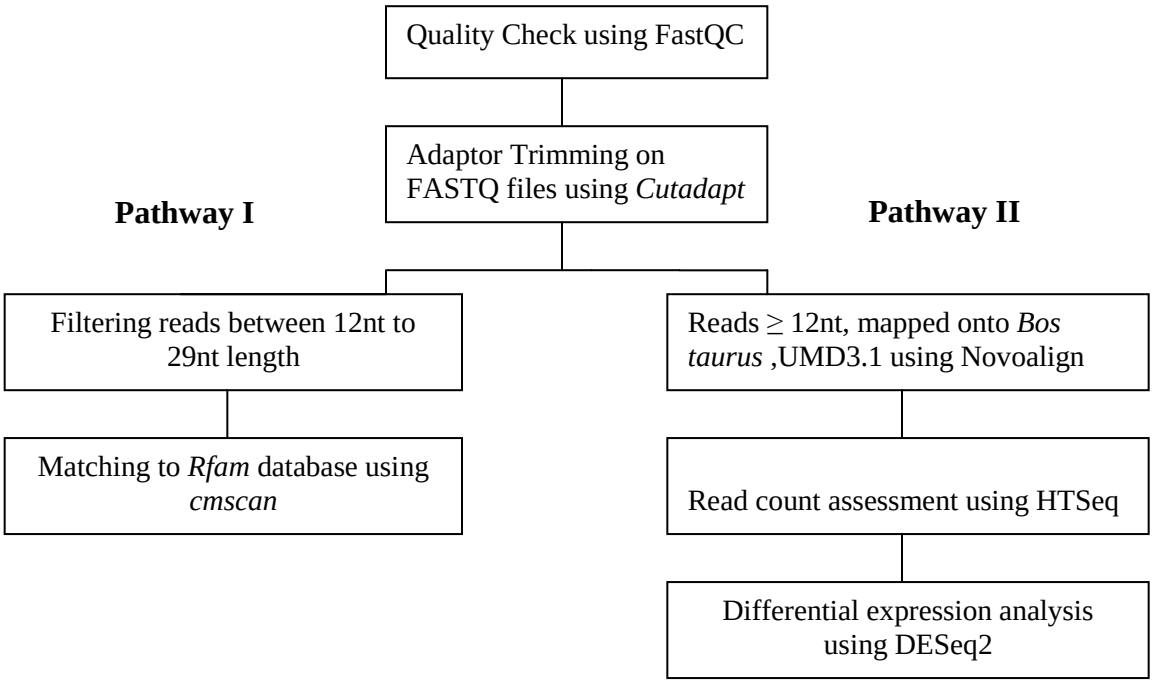


Figure 5.1. Represents a schematic outline of bioinformatics workflow followed to perform differential expression and RNA database matching.

Two different approaches were taken to perform data analysis of the reads as shown in Figure 5.1. Pathway I describes the steps taken in Rfam analysis of the reads that were Cutadapt trimmed and screened to pool reads between 12 nt to 29 nt in length.

Pathway II shows steps involved in the assessment of differential expression of the miRNA genes from the Cutadapt trimmed reads to mapping reads onto the bovine genome. These reads were then passed through HTSeq to obtain read depths and then differential expression is assessed using DESeq2, an R based package.

5.2 Assessment of sequence read lengths across raw data

FASTQ files were further sorted in ascending order of the lengths to get an overall size distribution and the depth (# of reads) before adaptor trimming. This provides an overall view of sequence sizes in the raw data as shown Figure 5.2 Read depth or the number of reads is plotted against the sequence read lengths to visualize the distribution of reads across various lengths. A high density of reads is seen in at 18nt ranges with a shorter peak around 24nt length.

Short fragments that can interfere as "noise" during the process of mapping are removed below a certain length, to help increase computational speed during the mapping process onto the reference genome. A general threshold of filtering the sequences based on size lengths is about 18 nt long. In this study the sizes below 12 nt were discarded.

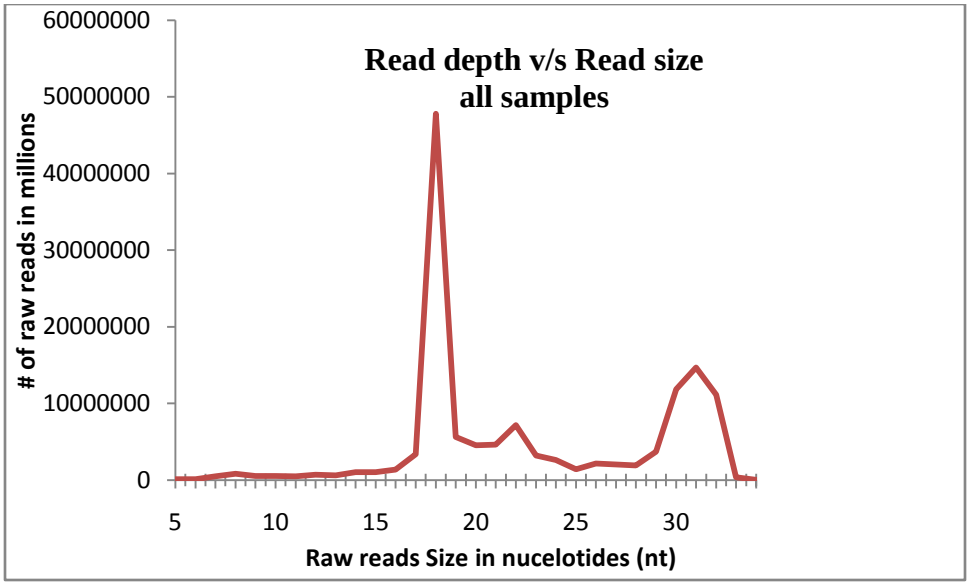


Figure 5.2. Overall size distribution of the raw FASTQ reads plotted against the number of reads (read depth) at each size intervals. High prevalence of 18nt length fragments indicated.

5.3 Adaptor clipping

Gel extraction of the desired size bands, corresponding to the miRNAs, as explained in Section 3.4, is a process to enrich for microRNAs. However, many a times artifacts like the small RNA primers and adaptors get left over in the pools and huge amounts of "primer_dimers" of Illumina sequencing primers and adaptor combinations remain, contaminating the read pools. A clean up process is initiated computationally by using programs that trim off the adaptor sequences and the adaptor only sequences from the reads.

Adaptor clipping was performed using the Cutadapt, v1.3(Martin, 2011), a software program which is specifically designed to trim adaptors from the NGS data.

The following adaptor sequence "TGGAATTCTCGGGTGCCAAGG" was used to add to the ends of the small RNA molecules. This adaptor sequence is trimmed off from the 3' end of the reads. A FASTQ sample file is needed to run the Cutadapt program along with the sequence of the adaptor to be trimmed. An example of the command is shown below.

```
>cutadapt -a AACCGGTT input.fastq > output.fastq
```

5.4 Preliminary survey of non-coding RNAs matching Rfam/cmscan screening

Raw FASTQ data was subjected to a preliminary analysis by the process of homology matching of sequences to Rfam, a database of non-coding RNAs. FASTQ reads were size trimmed to a length ≥ 12nt, in order to get rid of any spurious short read noise.

Rfam, a database of non-coding RNA sequences and annotations, hosting a vast number of RNA genes, including predicted RNA structures based on a probabilistic model known as *Covariance model* (CM) (Nawrocki et al., 2014), was used to match the read sequences to infer the different types of RNA population in the data set. In conjunction with *Infernal* (INFERence of RNA ALignment), a special instance of CM, characterization of RNA homologs with conserved secondary structures was performed. Rfam/cmscan program was implemented at USDA, courtesy of Dr. Steven Schroeder, to identify the various RNA signatures in unmapped, raw data from Illumina Genome Analyzer IIIx.

A FASTA file of the sample sequences, sized to have sequences >12nt is matched to the output of the cmscan, with the best e-value scores for sequences and the Rfam annotations are collected for different types of RNAs using awk scripts.

5.5 Mapping reads to the *Bos taurus* UMD3.1.73 using Novoalign

Bos taurus UMD3.1 is an assembly generated by the *Ensembl* at Center for Bioinformatics and Computational Biology (CBCB) at University of Maryland. UMD3.1 a third release in December 2009 was used as a reference genome in this study. Currently there are many freely available software packages in performing the mapping of the reads such as Bowtie (Langmead, Trapnell, Pop, Salzberg, & others, 2009), BWA (H. Li & Durbin, 2009) SOAP (R. Li, Li, Kristiansen, & Wang, 2008) and Novoalign (Novocraft, 2010). However for this study Novoalign from Novocraft was used in mapping reads onto the bovine genome.

Novoalign from Novocraft was used in following analysis for mapping the reads onto the bovine genome. Novoalign software program is based on Needleman-Wunsch algorithm that is freely available, with limited features, and also is available as a commercial copy with advanced features. Novoalign assigns Affine Gap Penalties, to find global optimum alignment and although is slower than Burrows-Wheeler based aligners like Bowtie and BWA, accuracy is increased, as its goal is to find the absolute best alignment for a short read to a reference genome sequence (http://www.novocraft.com/main/downloadpage.php, 2010).

In profiling small RNA signatures there is a preference in specificity over speed. Another important feature of Novoalign is an option (-m) that specifically looks for new miRNAs while mapping onto the genome. This feature was used in performing the mapping onto the bovine genome. Using the miRNA mode (-m) option to map, sets an additional score based on Needleman-Wunsch alignment to the opposite strand. As miRNA precursors form hairpin-like structures, this should provide a better score for the adjacent opposite strand alignment, thereby mapping potentially novel miRNAs as well.

Novoalign supports reads from various platforms including the Illumina Genome Analyzer that was used in this experiment and single-end reads ("Novoalign | Novocraft," 2010). Reads were aligned using Novoalign software onto the UMD3.1 *Bos Taurus* genome, using the "-m" option.

```
> Novoalign -o SAM -m -d /home/schroeder/Data/Genomes/Bos_taurus.UMD3.1.73 -F
ILM1.8 -f ../NoShort/3132_d0_acute_02_ca.fastq
```

5.6 Counting reads using HTSeq

Given a set of reads aligned to a reference genome, an estimate of how many reads mapped to the features of the reference genome, like the exons, is an interesting assessment that is a required step in NGS data analysis. HTSeq (Anders, Pyl, & Huber, 2015) a short form for High-Throughput Sequencing data, is a Python based package that estimates the number of reads mapping onto the gene features given a reference genome.

A gtf file of the bovine genome, with all known miRNA genes, was created from Ensembl and is referred to as the *Bos taurus_u_miRNA.gtf*. For the samples in the *S.*

uberis infections, the mapped reads were aligned to the *Bos taurus_u_miRNA.gtf* files and the regular bovine .gtf file encompassing all the exons that are not just the miRNA genes. Reads that mapped onto the miRNA genes, unambiguously or uniquely were counted using the HTSeq parameters.

```
> htseq-count -s no -a 10 input.SAM input_Reference.gtf -o
HTseq.SAM > HTSeq.txt
```

An option of "-s no " is chosen to consider read that are mapped regardless of the strandedness. Quality checks of 10 and above only are chosen to perform the counts. These reads are further subjected to differential expression analysis on DESeq2, an R-based package.

In order to prevent calling a read wrongly as highly expressed in situations when one read maps equally well to multiple features or if one of the gene reacts positively to a treatment uniquely mapped reads alone are accounted for in the analysis. A false positive can occur when the over expression of reads for multiple mapping condition is applied to both locations thereby falsely calling the other gene to be differentially expressed as well. Considering uniquely mapped reads alone for further analysis, although seem to under represent the output number of genes, it does not skew the data as the "fold change" between samples or conditions will be the same. Hence all the uniquely aligned SAM alignments are used for further processing through HTSeq to get the count data which can then be passed into DESeq2 to quantitate the differential expression.

5.7 Differential expression analysis using DESeq2, an R package

Differential expression analysis is performed on HTSeq count data using a R-BioConductor package, DESeq2 (Love, Huber, & Anders, 2014). There are many normalization methods statistically, that can be applied to normalize a NGS data set (Dillies *et al.*, 2012). However just counting the number of reads mapping to a gene is sub optimal for further analysis, as it does not consider the library composition bias when a larger gene expresses more reads than a small gene, and a direct estimation does not truly reflect the true (absolute) differential expression of the samples. Hence the general normalizations, like the Reads per kilobase per million (RPKM) in single end sequencing or Fragments per kilo base per million (FPKM) in case of paired end sequencing, are not ideal. Deseq2 package takes into account the library sizes and library composition biases and uses the Wald significance test and tries to fit negative generalized linear model (GLM) for each gene (Love *et al.*, 2014). DESeq2 differential analysis is performed on the HTSeq output for all samples across both infection and normal conditions.

In order to visualize and perform downstream analysis DESeq2 relies on transformation of the count data and one method is the use of logarithmic transformation of the raw counts. Two functions *rlog* and *variance stabilizing transformation* were used to transform the data for visualization. DESeq2 performs data quality control and assessment using clustering and visualization based on comparison of raw counts, *rlog* transformed and variance stabilized transformation. *rlog* transformation plots are shown through the entire analysis.

5.8 Summary

Data analysis is performed in a systematic manner, starting from a quality check of the overall data set using the FastQC to performing the differential expression using DESeq2. Analysis was performed at ARS, USDA (Beltsville, MD). 3GB of single end sequencing data was pre-processed to the desired size after adaptor clipping to retain reads above 12nt in length. Visualization tools in DESeq2 were implemented to view the differential expression of the miRNA genes. Results of the analysis are discussed in detail in Chapter 6.

Chapter 6 : Results

Small RNA extracted from the bovine milk exosomes were used to extract the RNA and individual small RNA libraries were built per sample. These libraries were further sequenced on the Illumina Genome analyzer and various bioinformatics suites were implemented in analysis. This Chapter discusses the results obtained from the bioinformatics analysis, across both types of infections across Rfam survey and the differential expression analysis.

6.1 Small RNA survey using Rfam / cmscan

Results of cmscan analysis, provided various RNA species from high abundance of different types of miRNAs to few unique types of tRNA, rRNA fragment matches. Many ncRNA, tmRNA, scRNAs, snoRNA and snRNAs were also identified. Rfam is a database of redundant RNA sequences provided an output of redundant matches with various e-values. Unique representations of each type of matches, with the lowest e-values were extracted and a table of various RNA unique matches was constructed.

Table 6.1 shows the total numbers of various types of unique small RNAs in the data set. Number of miRNA matches yielded unique types of miRNAs across each sample however the least diverse type of non-coding RNAs were the tRNA and rRNAs

which were seen predominantly the same across all samples. Rfam output indicates the presence of miRNAs in the data set and provides annotations to the various unique types of miRNAs.

Table 6.1. Number of unique Rfam matches to each different types of RNA species across control and *S. aureus* (represented as _ac) and *S. uberis* (represented as _ub) infections.

sample ID	Types of miRNA	Types of tRNA	Types of rRNA
3206_d0_ac01	188	2	5
3206r_d2_ac15	226	2	6
3132_d0_ac02	238	2	5
3207r_d0_ac03	200	2	5
3207_d2_ac07	335	2	5
3504_d0_ac17	170	2	5
3504_d2_ac05	276	2	5
3658_d0_ac13	100	2	5
3658_d2_ac14	145	2	5
3207_d0_ub16	210	2	5
3207_d2_ub04	255	2	5
3554_d0_ub08	136	2	5
3554_d2_ub09	196	2	5
3658_d0_ub10	162	2	5
3658_d2_ub11	234	2	5
3318r_d0_ub06	130	2	5
3318r_d2_ub12	188	2	5

6.2 *S. uberis* analysis and read statistics

RNA extracted from the exosomal pellets were used in the preparation of small RNA libraries that was sequenced on the Illumina Genome Analyzer IIIx as described in

the Chapter 3. Raw reads were analyzed using various bioinformatics tools as described in Chapter 5.

Illumina GA IIIx produced 63,956,365 total reads from small RNA libraries constructed under *S. uberis* infections out of which Cutadapt v1.3 trimming yielded 60,791,546 raw reads as output. Figure 6.1 indicates the overall size distribution of the raw reads according to the sizes in increasing order of lengths. Reads predominantly fall in the range of 18nt to 25nt lengths, validating the miRNA abundance based on the sizes.

After the size filtering to get reads ≥ 12 nt lengths and quality trimming, a total of 59,525,707 were passed through the Novoalign mapper using *Bos taurus UMD3.1*. A total of 96 % of reads were mapped onto the bovine genome whereas a total of 4% remained unmapped onto the bovine genome.

Mapped reads were used to assess the depth using a miRNA gtf file created from Ensemble for *Bos taurus* genome. Out of the 1153 different miRNA features that HTSeq identified from Ensemble for *Bos taurus* genome, an overall 1,358,455 reads mapped to 109 miRNA genes.

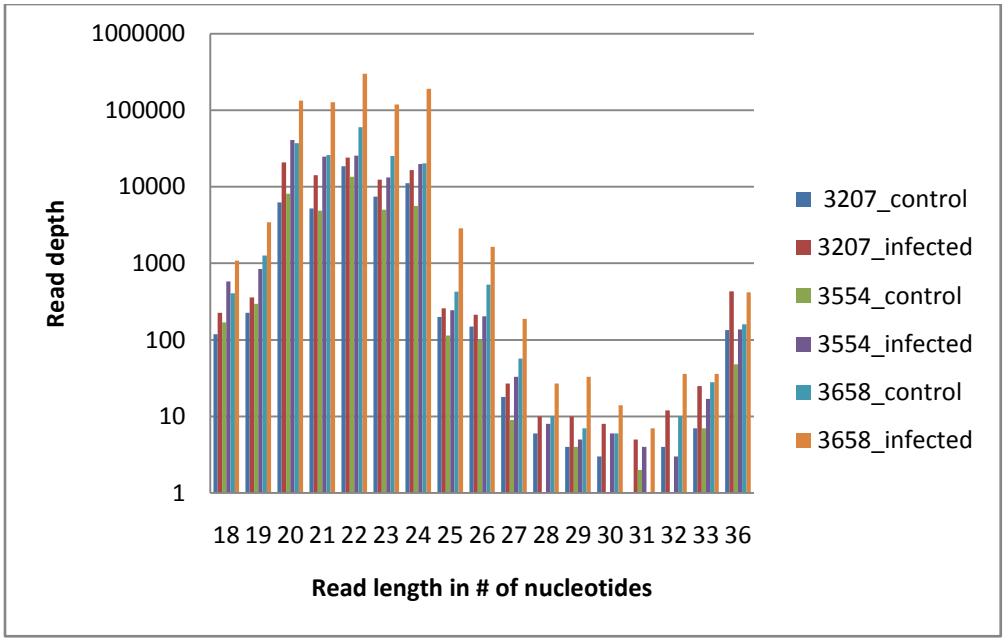


Figure 6.1. Cutadapt trimmed read lengths plotted against their depth for all *S. uberis* samples. A greater number of reads are seen in the length range of 18nt to 25nt, reflecting typical miRNA lengths.

6.3 *S. uberis*: differential expression results

An HTSeq count data was assessed for each genomic feature on the *Bos taurus* genome and mapping the sample reads. The count data from HTSeq was further processed into DESEq2 to determine differential expression of miRNA genes. A total of 6 miRNA genes were over expressed and 10 miRNA genes under expressed across *S. uberis* infections in comparison to the control samples, with a p-value of < 0.05.

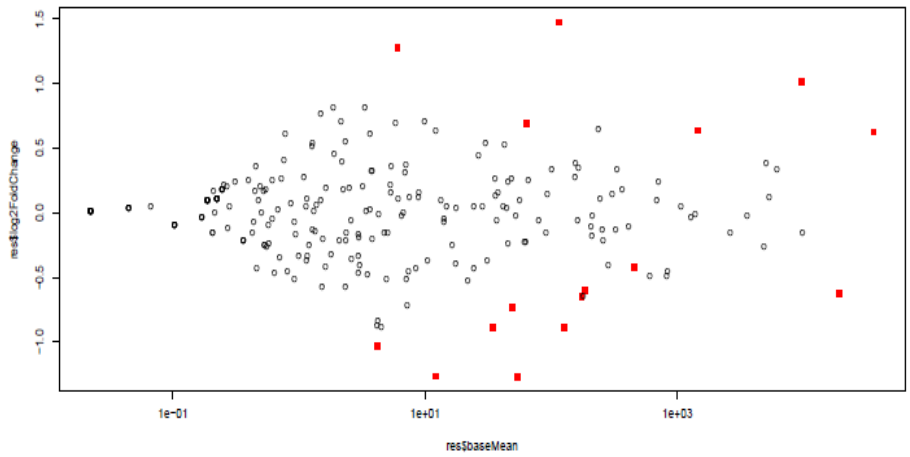


Figure 6.2. DESeq2 dispersion plot showing log2Foldchange of differentially expressed genes (Red squares) with a p-value of <0.05 as significantly differentially expressed..

DESeq2 uses the log2Fold change over the base mean values of the miRNA genes across samples to perform a dispersion plot, representing up-regulated and down-regulated genes. A dispersion plot with differentially expressed genes indicated in red, Figure 6.2.

A list of miRNAs with their miRBASE ids and accession numbers corresponding to the Ensembl IDs from HTSeq are listed as up-regulated and down-regulated miRNA genes. Table 6.2 provides a list of up-regulated genes with a log2Foldchange of >1.0 and p-value < 0.05, and Table 6.3 provides a list of genes that are down-regulated, with a log2Foldchange of -1.0 and p-value of < 0.05 across *S. uberis* infections in comparison to the controls.

Table 6.2. List of up-regulated miRNAs, across *S. uberis* infections, from the DESeq2 analysis .Corresponding miRBASE accession numbers are indicated. (www.mirbase.org).

Ensembl Gene ID	miRBase ID(s)	miRBase Accession(s)
ENSBTAG00000029804	bta-mir-30a	MI0005054
ENSBTAG00000029870	bta-mir-98	MI0005025
ENSBTAG00000029871	bta-mir-30e	MI0005018
ENSBTAG00000029982	bta-mir-142	MI0005011
ENSBTAG00000036410	bta-mir-141	MI0009742
ENSBTAG00000036423	bta-mir-146b	MI0009745

Table 6.3. List of down-regulated miRNAs, across *S. uberis* infections, from the DESeq2 analysis. Corresponding miRBASE accession numbers are provided (www.mirbase.org).

Ensembl Gene ID	miRBase ID(s)	miRBase Accession(s)
ENSBTAG00000029771	bta-mir-140	MI0005010
ENSBTAG00000029798	bta-mir-532	MI0005061
ENSBTAG00000029841	bta-mir-181a-2	MI0004757
ENSBTAG00000029920	bta-mir-25	MI0005067
ENSBTAG00000029945	bta-mir-148a	MI0004737
ENSBTAG00000030109	bta-mir-744	MI0009891
ENSBTAG00000036372	bta-mir-28	MI0009785
ENSBTAG00000036420	bta-mir-500	MI0009851
ENSBTAG00000037337	bta-mir-877	MI0009903
ENSBTAG00000043869	bta-mir-1468	MI0012207

6.4 Summary: *S. uberis* differential expression

Some of the known mammary gland specific miRNAs like the bta-miR-146, bta-miR-30e and bta-miR-142 are seen up-regulated in the bovine milk exosomal content under *S. uberis* infection as shown in Table 3. Bta-miR-98, a member of the bta-let-7

family, is also up-regulated under *S. uberis* infections indicated in Table 3. A previous study shows bta-miR-142 to be up-regulated specifically under *S. uberis* infections of the mammary monocytes (Lawless, Foroushani, Mccabe, O'Farrelly, & Lynn, 2013) which is concurrent to our findings. miR-142 is also shown to have a role in immune response in a study involving lupus patients (Carlsen *et al.*, 2013) where miR-142 expression is increased. Down regulation of certain types of miRNA could be associated with the genes they regulate and target prediction of those miRNA genes could reveal more information about their behavior, which is not the focus of this study. Except for miR-1468, most of the differentially expressed miRNAs are also found in other exosomal studies and listed in a exosomal miRNA database, Exocarta (Mathivanan, Fahner, Reid, & Simpson, 2012).

6.5 Comparison of miRNA types across each sample to the differentially expressed genes: *S. uberis* samples

A set of differentially expressed genes was compared to each sample, individually across control and infected conditions and a Venn diagram was created. Figure 6.3 shows the number of unique miRNA types across *S. uberis* samples under both conditions. The numbers indicate the total number of miRNA genes (red and yellow circles) individual samples compared to the number of differentially expressed genes (green circle) across all sample under *S. uberis* infection**.**

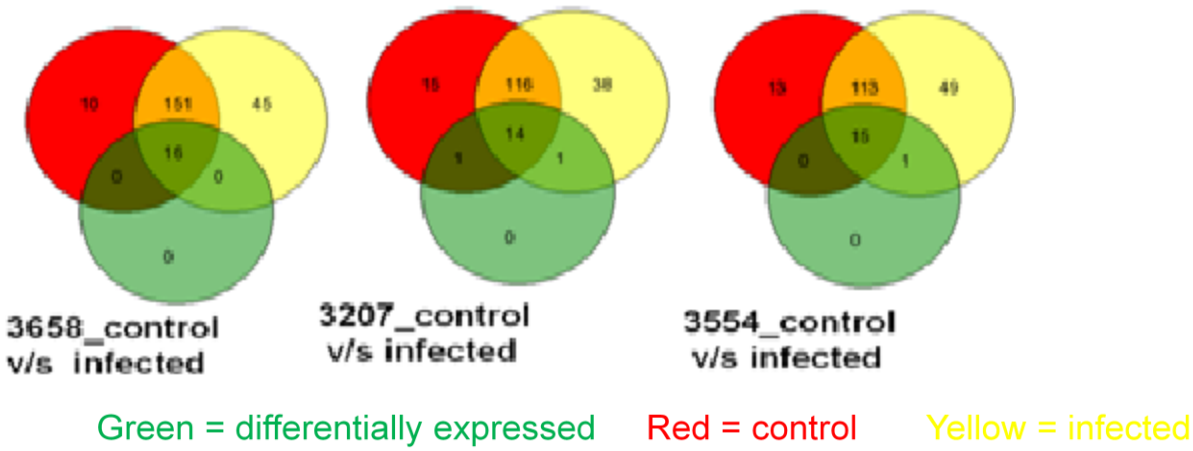


Figure 6.3. Venn diagram showing a cross comparison of differentially expressed genes of *S. uberis* treatment to the individual samples.

A total of 14 miRNA genes are seen differentially expressed across all three animals under *S. uberis* infections compared to the controls. bta-miR-142 was absent in 3207 and 3554 control samples. bta-miR-744 was absent in 3207 infected sample.

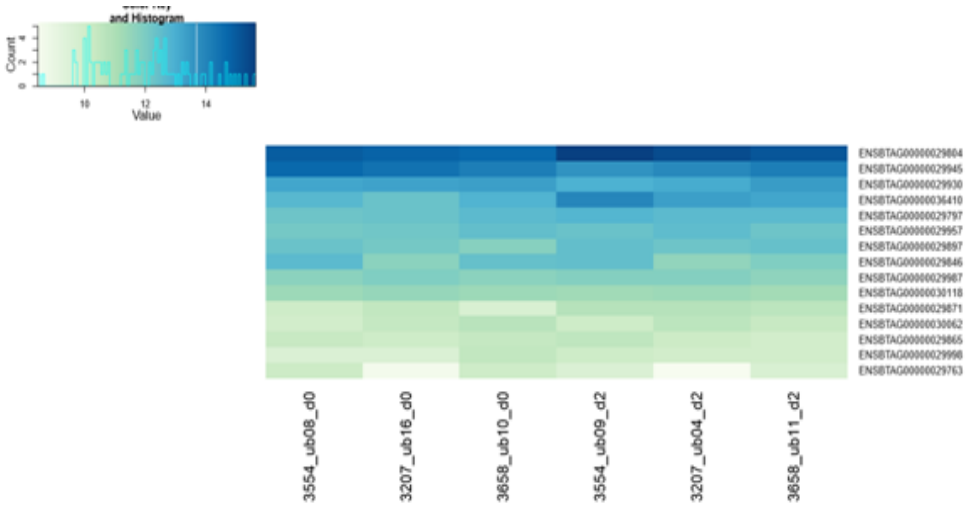


Figure 6.4. DESeq2-Rlog transformed heat map with top 15 highly expressed miRNA genes in *S. uberis* infections. Y-axis shows Ensembl IDs of genes. The x-axis provides the sample names. High expression is indicated with darker color.

DESeq2 was used to build a heat map of the top 15 highly expressed miRNA genes across all *S. uberis* infected samples as indicated in Figure 6.4. Each of the genes is indicated on y-axis with their corresponding Ensembl ID. The x-axis provides the sample names. High expression is indicated with darker color.

Their corresponding Ensemble IDs are provided along with the sample names. Common differentially-expressed miRNA types are compared to other published miRNAs from previous studies and it was seen that miRNAs, miR-30a, -141, -21 and -27b were present in the top 15 differentially expressed miRNAs, in comparison to porcine and human milk exosomal miRNAs (Chen *et al.,* 2014). Presence of immune related miRNA in human breast milk is elucidated by (Zhou *et al.,* 2012). It was demonstrated by Zhou *et al.,* that out of 87 well characterized pre-miRNAs, 59 immune related pre-miRNAs were present in human breast milk. In comparison to human breast

milk data, miR-30a, -148a, and -141 were found to be present in the bovine exosomes and be differentially expressed.

6.6 Clustering miRNA genes against *S. uberis* infections

Clustering of the data sets against conditions was performed to see if the bacterial infections had any effect on the sample's miRNA genes. Figure 6.5 shows the clustering of miRNA genes across *S. uberis* infections. All the control samples and infection samples does not show to be clustered to each other as a group. The darker blue rectangles indicate stronger clustering patterns and each individual animal, indicated by its sample ID, is shown to be highly clustered to itself than to the treatment or infection. Clustering did not seem to provide any conclusive effect of the infections on the miRNA content of the exosomes across samples that could be attributed to many factors as highlighted in the discussion section of this chapter.

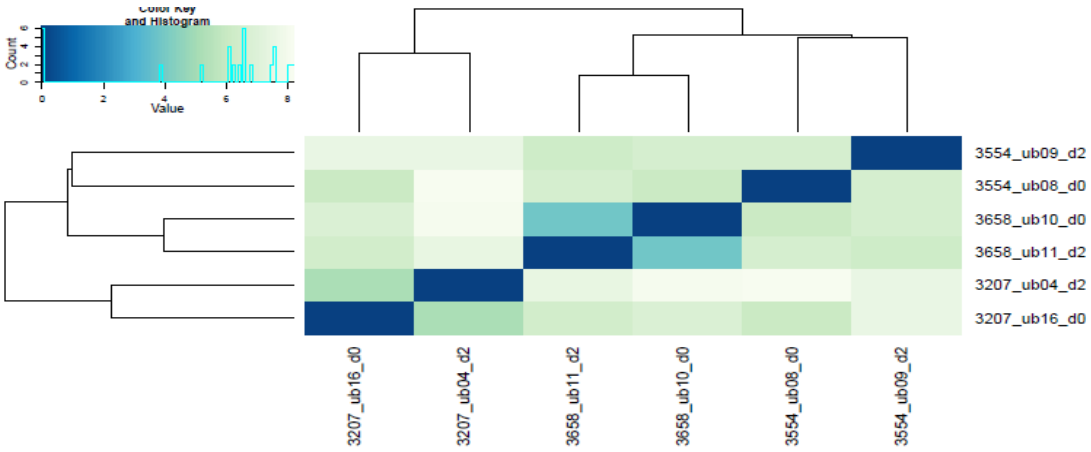


Figure 6.5. DESeq2 clustering of miRNA genes across *S. uberis* infected animals. Clustering shows no effect of the infections on the individual animal. Darker blue rectangles indicate stronger clustering

6.7 *S. aureus* analysis and read statistics

Mapping of *S. aureus* experiment reads onto the bovine genome, UMD3.1, provided a FASTQ data of 63,956,365 total reads out of which Cutadapt v1.3 trimmed raw reads, were 60,791,546. After the size filtering to get reads ≥12 nt lengths and quality trimming, a total of 59,525,707 reads were passed through the Novoalign mapping software using *Bos taurus UMD3.1*. An 83% mapping rate was achieved with 17% unmapped reads across 9 libraries. A total reads of 7,042,365 mapped across 309 miRNA genes from the Ensembl Bos_taurus_1.73.gtf out of 1153 total miRNA gene features.

Assessment of size distribution of the reads along with the depth reflected majority of the reads in the range of 18 nt to 25 nt lengths, validating the presence of miRNA reads sizes as shown in Figure 6.6.

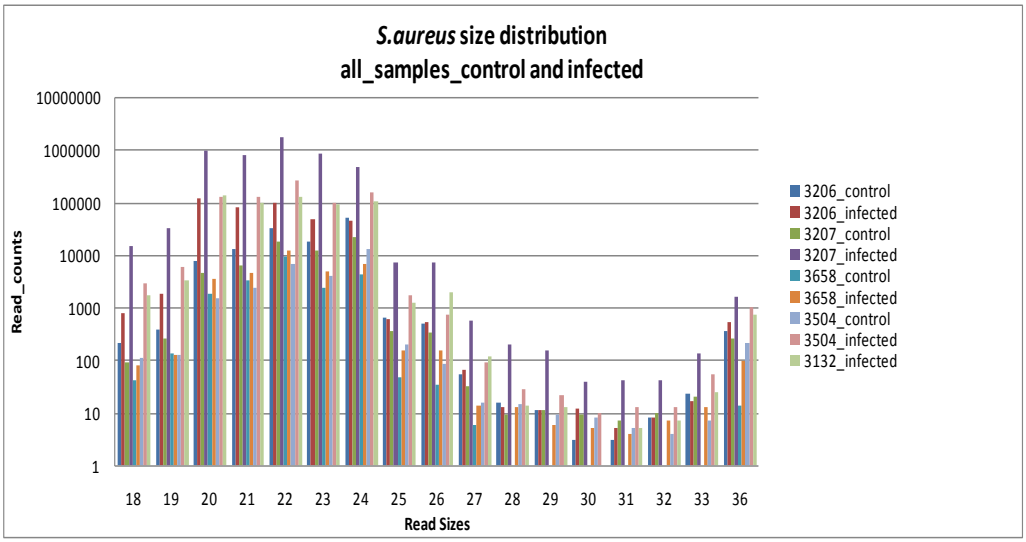


Figure 6.6. Read size distribution across depth for all *S. aureus* samples and controls. A greater number of reads are seen in the length range of 18nt to 25nt, reflecting typical miRNA lengths.

6.8 *S. aureus* differential expression of miRNA genes in infections over the control samples: results

Differential expression analysis on DESeq2, across HTSeq counts of *S. aureus* infected samples to the controls, provided a total of 35 miRNA genes that are differentially expressed across the samples with *S. aureus* infections.

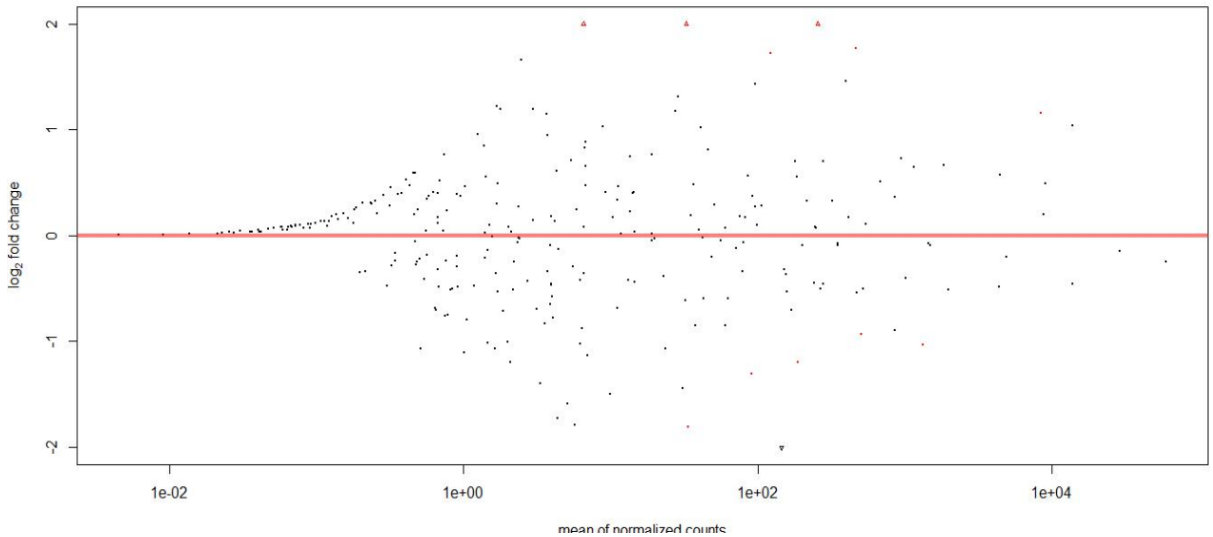


Figure 6.7. DESeq2 dispersion plot showing up and down-regulated miRNA genes across *S. aureus* infections. A log2Foldchange of >1.0 are up-regulated while genes with a log2Foldchange value of -1 are down-regulated across the red threshold line.

DESeq2 dispersion plot uses the log2Fold change over the base mean values of the miRNA genes. Figure 6.7 shows both up-regulated and down-regulated genes in red. Overall there are 23 miRNA genes that are down-regulated with a log2Foldchange of -negative values, and a p-value of < 0.05, and 12 miRNA genes up-regulated. List of these genes with the corresponding miRBASE and Ensembl IDs are provided in Table 6.4 and Table 6.5 retrieved from Ensembl biomart.

Table 6.4. List of 12 up-regulated miRNAs, across all *S. aureus* data provided by Ensembl searches against miRBASE database (www.mirbase.org)

Ensembl Gene ID	miRBase ID(s)	miRBase Accession(s
ENSBTAG00000029762	bta-mir-183	MI0009756
ENSBTAG00000029768	bta-mir-99a	MI0004751
ENSBTAG00000029918	bta-mir-99b	MI0005469
ENSBTAG00000029957	bta-mir-191	MI0005034
ENSBTAG00000029978	bta-mir-363	MI0005069
ENSBTAG00000029982	bta-mir-142	MI0005011
ENSBTAG00000030056	bta-mir-215	MI0005016
ENSBTAG00000036361	bta-mir-152	MI0009748
ENSBTAG00000036361	bta-mir-2957	MI0015944
ENSBTAG00000036367	bta-mir-374b	MI0009816
ENSBTAG00000036410	bta-mir-141	MI0009742
ENSBTAG00000036418	bta-mir-223	MI0009782
ENSBTAG00000036423	bta-mir-146b	MI0009745

Table 6.5. List of down-regulated miRNAs across all *S. aureus* data provided by Ensembl searches against miRBASE database (www.mirbase.org)

Ensembl Gene ID	miRBase ID(s)	miRBase Accession(s)
ENSBTAG00000029771	bta-mir-140	MI0005010
ENSBTAG00000029772	bta-let-7a-3	MI0005452
ENSBTAG00000029774	bta-let-7f-2	MI0004734
ENSBTAG00000029797	bta-mir-186	MI0005033
ENSBTAG00000029822	bta-mir-423	MI0005046
ENSBTAG00000029833	bta-mir-326	MI0009799
ENSBTAG00000029850	bta-mir-26b	MI0004745
ENSBTAG00000029865	bta-mir-151	MI0004738
ENSBTAG00000029870	bta-mir-98	MI0005025
ENSBTAG00000029871	bta-mir-30e	MI0005018
ENSBTAG00000029880	bta-let-7e	MI0005455
ENSBTAG00000029896	bta-mir-181b-2	MI0005013
ENSBTAG00000029914	bta-mir-365-2	MI0009812
ENSBTAG00000029935	bta-let-7f-1	MI0005062
ENSBTAG00000029967	bta-mir-103-1	MI0004736
ENSBTAG00000029989	bta-mir-23b	MI0005066
ENSBTAG00000030013	bta-mir-660	MI0005468
ENSBTAG00000037319	bta-mir-296	MI0009786
ENSBTAG00000038879	bta-mir-1307	MI0010474
ENSBTAG00000045223	bta-mir-2419	MI0011467
ENSBTAG00000045264	bta-mir-2465	MI0011525
ENSBTAG00000045721	bta-mir-2892	MI0013067
ENSBTAG00000046072	bta-mir-2284x	MI0014499

6.9 Summary: *S. aureus* differential expression

Some of the known mammary gland specific miRNAs, for example, the bta-miR-142 and bta-miR-223were seen up-regulated in the bovine milk exosomal content under *S. aureus* infections.bta-miR-142 is shown to be up-regulated specifically under *S. uberis* infections of the mammary monocytes (Lawless *et al.*, 2013) which is concurrent to our findings in both *S. uberis* and in *S. aureus* infections. miR-142 is also shown to have a role in immune response in a study involving lupus patients (Carlsen *et al.*, 2013) where miR-142 expression is increased.

Another important class of miRNA genes that are seen up-regulated are the miR-99a and miR-99b, which have been shown in previous study to aid in epithelial to mesenchymal transition during disease development and progression in murine mammary gland cells during and in proliferation and migration of breast cancer cells (Turcatel, Rubin, El-Hashash, & Warburton, 2012).

Another well known miRNA, miR-223 is up-regulated in the *S. aureus* infections. Previous studies have implicated its role as a potential biomarker in many types of cancers (Liu, Chen, Kuo, Cheng, & Lin, 2010).

6.10 Comparison of *S. aureus* miRNA types across differentially expressed miRNA genes under *S. aureus* infections.

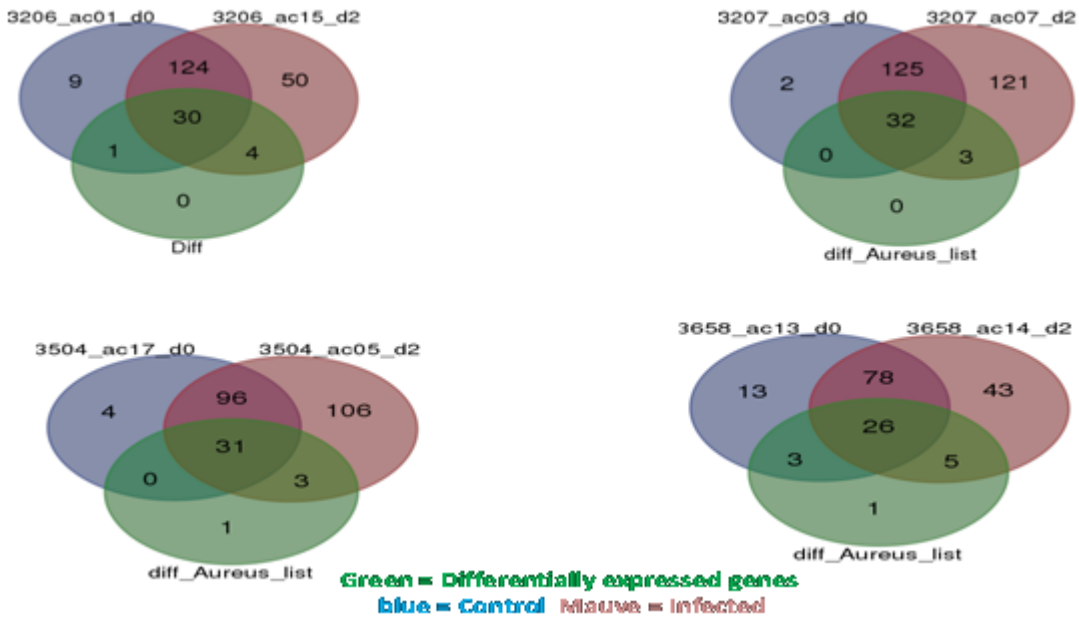


Figure 6.8. Venn diagram showing comparison of miRNA types across differentially expressed miRNAs of individual conditions, under *S. aureus* treatment. (bioinformatics.psb.ugent.be/webtools/Venn)

Figure 6.8 shows comparison of the differentially expressed genes from *S. aureus* analysis to the individual animal across each treatment condition reveals miRNA type that is prevalent in all samples. bta-mir-223 and bta-mir-142 were seen differentially expressed in all four animals across infection conditions. bta-miR-326 was seen in 3206_control and 3658_control conditions.

6.11 Top 15 highly expressed miRNA genes in *S. aureus* infections

Common differentially expressed miRNA types, miR-30a, -141, -21 and -27b were present in the top 15 differentially expressed miRNAs, in comparison to porcine and human milk exosomal miRNAs (Chen *et al.*, 2014). Presence of immune related miRNA in human breast milk is elucidated by (Zhou *et al.*, 2012). Out of 87 well characterized pre-miRNAs, 59 immune related pre-miRNAs were demonstrated to be present in human breast milk by Zhou *et al*. In comparison to human breast milk data, miR-30a and -141 were found to be present in the bovine exosomes and be differentially expressed across *S. aureus* treatment.

bta-mir-660 was one of the top 15 highly expressed in *S. aureus* infections and is also found in the exosomal content of Exocarta, a database of exosomal miRNAs (Mathivanan *et al.*, 2012). Figure 6.9 indicates the heat map of top 15 highly expressed genes. Darker blue color represents higher level of expression and the count analysis of is shown in the subset heat map.

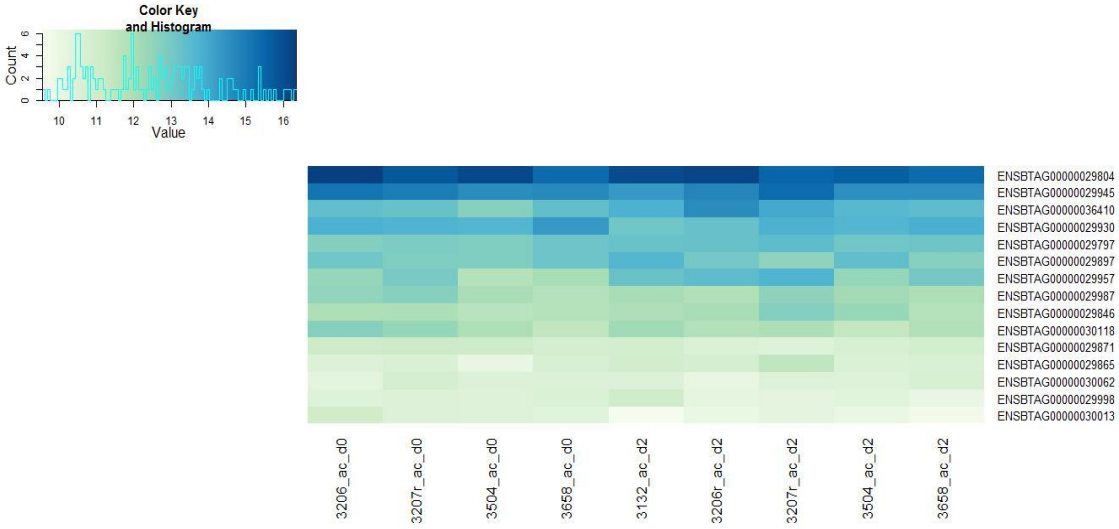


Figure 6.9. DESeq2-Rlog transformed heat map indicates top 15 highly expressed genes across all samples, *S. aureus* infection. Corresponding Ensembl IDs of these genes are provided on the y-axis with sample names in the x-axis. Darker color indicates more expression.

6.12 Clustering miRNA genes under *S. aureus* infections

DESeq2 was used to perform clustering of the miRNA genes across animals, to visualize if the bacterial infections had any effect on miRNA genes. A clustering heat map is shown in Figure 6.10 Clustering did not provide any conclusive effect of the infections on the miRNA content of the exosomes. Control samples of three animals seem to cluster together but the clustering is not seen in their corresponding infection conditions.

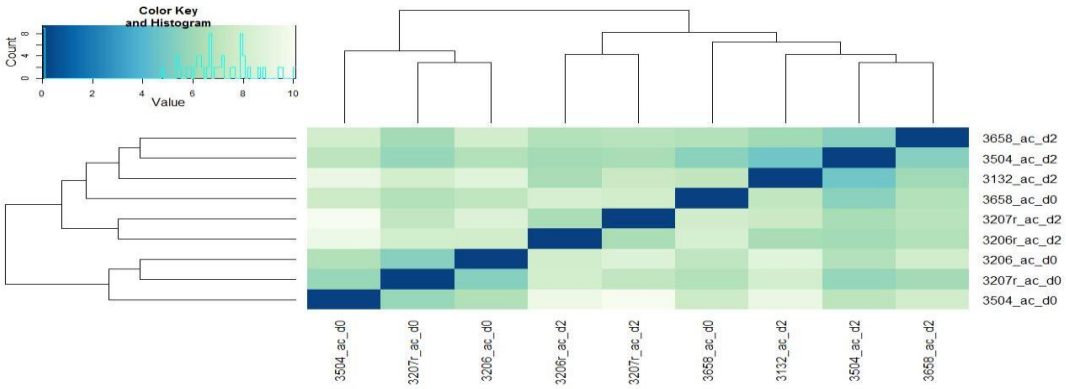


Figure 6.10. DESeq2, RLog transformed normalized clustering across all miRNA genes in *S. aureus* treated and control samples. Clustering shows no effect on the infections on the animals as the animals seem to cluster to itself over the infections.

6.13 *S. uberis* and *S. aureus* combined data: results

As clustering analysis from *S. uberis* and *S. aureus*, individually, did not provide any conclusive effect of the treatment on the miRNA genes, an alternative approach was implemented to see if the small samples size could be the issue. Hence all samples were combined to assess any effect of infection on miRNA genes from the large pooled samples.

All the steps initiated in the analysis of *S. uberis*, S. aureus were applied to the large sample pool in a similar way, and differentially expressed genes were analyzed. Figure 6.11 shows the dispersion plot generated by DESeq2, depicting the up-regulated and the down-regulated genes across all samples, shown in red.

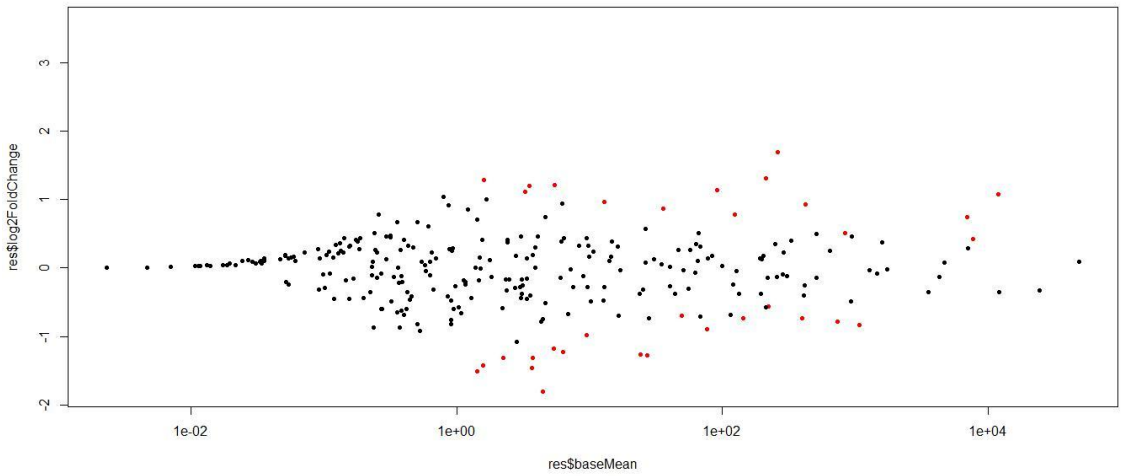


Figure 6.11. DESeq2 dispersion plot of all Aureus and Uberis combined samples. A log2Foldchange of >1.0 represents up-regulated genes and a log2Foldchange with -1 represents genes that are down-regulated. These genes are indicated in red.

Table 6.7 shows eight up-regulated miRNA genes with a p-value of <0.05 and >1 log2Foldchange. A total of 26 miRNA genes were down-regulated across the overall combined samples with a p-value of < 0.05 and log2Foldchange of < 1 and are listed in Table 6.7.

Table 6.6. List of eight miRNAs up-regulated across combined samples of *S. aureus* and *S. uberis* searched against Ensembl, with miRBASE IDs (www.mirbase.org).

Ensembl Gene ID	miRBase ID(s)	miRBase Accession(s)
ENSBTAG00000029762	bta-mir-183	MI0009756
ENSBTAG00000029830	bta-mir-146a	MI0009746
ENSBTAG00000029909	bta-mir-7-2	MI0010462
ENSBTAG00000029918	bta-mir-99b	MI0005469
ENSBTAG00000029924	bta-mir-34a	MI0005464
ENSBTAG00000036410	bta-mir-141	MI0009742
ENSBTAG00000036423	bta-mir-146b	MI0009745
ENSBTAG00000044594	bta-mir-502b	MI0009854

Table 6.7. List of miRNAs down-regulated across combined *S. aureus* and *S. uberis* samples, searched against Ensemble, with miRBASE IDs (www.mirbase.org).

Ensembl Gene ID	miRBase ID(s)	RBase Accession
ENSBTAG00000029762	bta-mir-183	MI0009756
ENSBTAG00000029768	bta-mir-99a	MI0004751
ENSBTAG00000029771	bta-mir-140	MI0005010
ENSBTAG00000029772	bta-let-7a-3	MI0005452
ENSBTAG00000029797	bta-mir-186	MI0005033
ENSBTAG00000029798	bta-mir-532	MI0005061
ENSBTAG00000029822	bta-mir-423	MI0005046
ENSBTAG00000029841	bta-mir-181a-2	MI0004757
ENSBTAG00000029847	bta-mir-18a	MI0004740
ENSBTAG00000029850	bta-mir-26b	MI0004745
ENSBTAG00000029870	bta-mir-98	MI0005025
ENSBTAG00000029918	bta-mir-99b	MI0005469
ENSBTAG00000029924	bta-mir-34a	MI0005464
ENSBTAG00000029955	bta-mir-193a	MI0005014
ENSBTAG00000029957	bta-mir-191	MI0005034
ENSBTAG00000029967	bta-mir-103-1	MI0004736
ENSBTAG00000030013	bta-mir-660	MI0005468
ENSBTAG00000036367	bta-mir-374b	MI0009816
ENSBTAG00000036410	bta-mir-141	MI0009742
ENSBTAG00000036423	bta-mir-146b	MI0009745
ENSBTAG00000037319	bta-mir-296	MI0009786
ENSBTAG00000037337	bta-mir-877	MI0009903
ENSBTAG00000037358	bta-mir-885	MI0009904
ENSBTAG00000045223	bta-mir-2419	MI0011467
ENSBTAG00000046072	bta-mir-2284x	MI0014499

6.14 Cross comparison of differentially expressed genes across individual infections and combined infections

Further, a Venn diagram is created to visualize the comparison of the differentially expressed genes from individual *S. aureus* (blue) and *S. uberis* (Pink) infections to the differentially expressed genes across combined samples (green) Figure 6.12. This comparison provided four commonly differentially expressed genes across all combinations.

Two genes, bta-mir-30e and bta-mir-142, are differentially expressed across both aureus and uberis infections (individual analysis). bta-mir-30e is one of the most commonly seen miRNA is bovine skeletal muscles (Coutinho *et al.,* 2006) bta-mir-142 is

seen in adipose tissues and mammary glands. However is not seen being differentially expressed in the combined analysis but bta-miR -30 is one of the highly expressed miRNA in the combined analysis.

A Venn diagram of differentially expressed genes across this combination is created as shown in Figure 6.12. The numbers in the Venn diagram indicate the number of differentially expressed genes across different types of conditions.

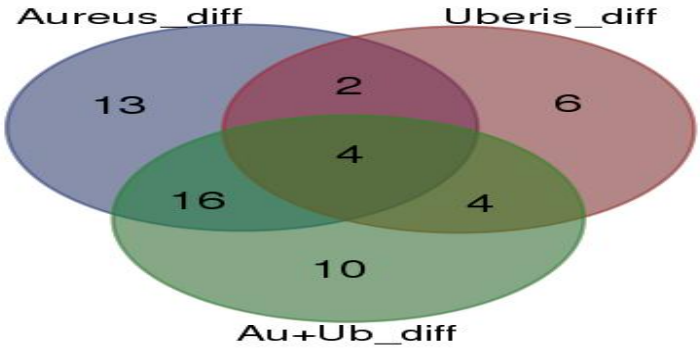


Figure 6.12. Venn diagram of cross comparison of differentially expressed genes. Comparison against combined samples (green), to the *S. aureus* differentially expressed genes (blue) and *S. uberis* differentially expressed genes (pink).

Comparison of the miRNAs from "combined" large dataset of differentially expressed genes to the differentially expressed genes of individual treatments provided 10 genes that are not seen in individually differentially expressed gene list. Table 6.8. Ten differentially expressed genes unique to the combined data analysis matched to miRBASE (www.mirbase.org) on Ensembl (www.ensembl.org).Table 6.8 lists the ten differentially expressed genes unique to the combined samples.

Table 6.8. Ten differentially expressed genes unique to the combined data analysis matched to miRBASE (www.mirbase.org) on Ensembl (www.ensembl.org).

Ensembl Gene ID	miRBase ID(s)	miRBase Accession(s)
ENSBTAG00000029812	bta-mir-324	MI0009798
ENSBTAG00000029830	bta-mir-146a	MI0009746
ENSBTAG00000029909	bta-mir-7-2	MI0010462
ENSBTAG00000029924	bta-mir-34a	MI0005464
ENSBTAG00000029955	bta-mir-193a	MI0005014
ENSBTAG00000037358	bta-mir-885	MI0009904
ENSBTAG00000044594	bta-mir-502b	MI0009854
ENSBTAG00000046770		
ENSBTAG00000047838	bta-mir-2887-1	MI0013060
ENSBTAG00000047838	bta-mir-2887-2	MI0013061

Combined analysis of data sets from *S. aureus* and *S. uberis*, provided differential expression of a gene, with Ensembl ID, ENSBTAG00000046770 with no miRBASE annotation characterized as novel miRNA

DESeq2 is used to generate a heat map of miRNA genes across combined samples to visualize the top 30 highly expressed miRNA genes as shown in Figure 6.13. Higher the expression of the genes darker the color indicated in the Figure 6.13.

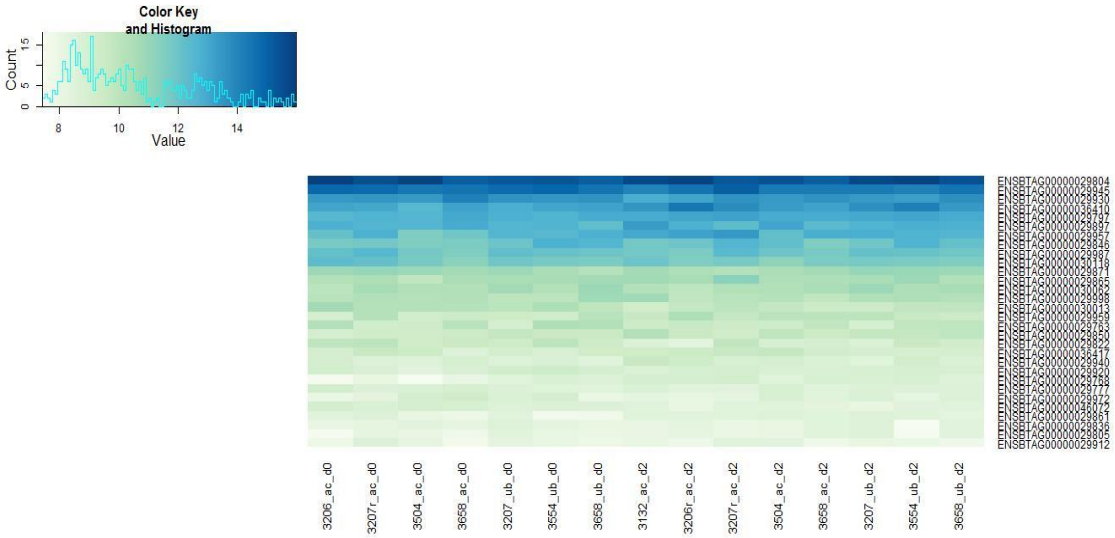


Figure 6.13. Heat map of RLog transformed data of top 30 highly expressed genes. Top 11 of these are represented with darker shades of blue.

Top 12 of the highly expressed genes are indicated in the y-axis as the Ensembl IDs and x-axis with the samples. These Top 11 highly expressed genes are listed out in Table 6.9.

Table 6.9. Top 12 highly expressed genes across combined samples of *S. uberis* and *S. aureus*.

Ensembl Gene ID	miRBase ID(s)	miRBase Accession(s)
ENSBTAG00000029797	bta-mir-186	MI0005033
ENSBTAG00000029804	bta-mir-30a	MI0005054
ENSBTAG00000029846	bta-mir-27b	MI0004760
ENSBTAG00000029871	bta-mir-30e	MI0005018
ENSBTAG00000029897	bta-mir-21	MI0004742
ENSBTAG00000029930	bta-mir-22	MI0005041
ENSBTAG00000029930	bta-mir-3600	MI0015943
ENSBTAG00000029945	bta-mir-148a	MI0004737
ENSBTAG00000029957	bta-mir-191	MI0005034
ENSBTAG00000029987	bta-mir-30d	MI0004747
ENSBTAG00000030118	bta-mir-182	MI0009755
ENSBTAG00000036410	bta-mir-141	MI0009742

Clustering of the samples across conditions was performed to assess the effect of infection on the animals. A visualization of the clustering of miRNA genes across the combined *S. uberis* and *S. aureus* samples was performed using DESeq2 as shown in Figure 6.14 across all miRNA genes of combined samples.

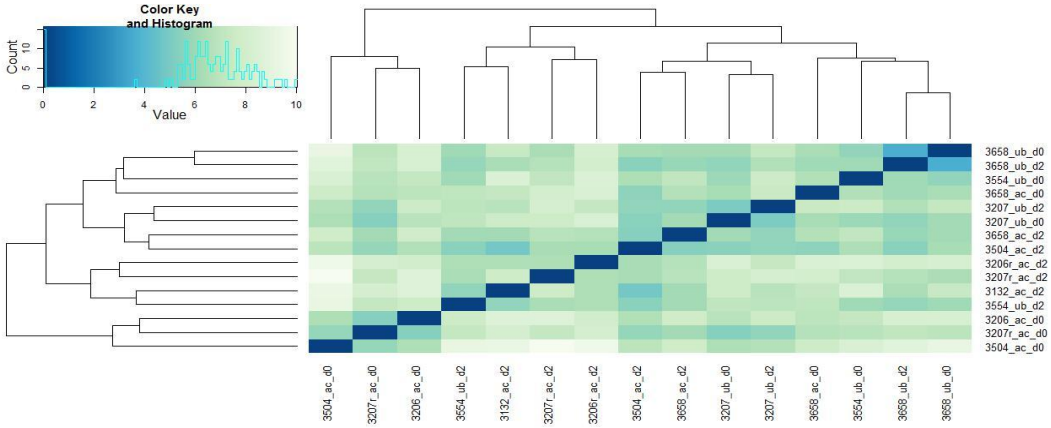


Figure 6.14. DESeq2-Heatmap visualization of clustering of combined samples of *S. uberis* and *S. aureus* samples. Ensemble IDs are indicated on the y-axis of the graph. Sample IDs on the x-axis represents the animals across conditions. Clustering shows no effect of the infection. .

6.15 Summary: Combined *S. uberis* and *S. aureus* differential expression of miRNA genes

Combined analysis of *S. uberis* and *S. aureus* provided 10 differentially expressed miRNAs that are neither expressed differentially in either individual. This could be an

additive effect due to the pooling of samples thereby creating a larger sample size for analysis.

Among these 10 differentially expressed miRNA from the combined data, bta-mir-885, bta-mir-502b, bta-mir-2887-1 and bta-mir-193a were not seen in the Exocarta (Mathivanan et al., 2012) set of known exosomal miRNAs. Ensembl ID ENSBTAG00000046770 matched to a novel miRNA without a miRBASE name. Clustering of the reads based on the treatment provided no conclusive evidence of the effect of infections on the samples indicating the possibility that the genetics of the animals was perhaps overtaking the effect of infection. In other words the animals were seen to behave the same regardless of the effect of infection.

Another interesting small RNA molecule explored was the piRNA. Details of the piRNA study are elucidated in Chapter 7.

Chapter 7 : Exploration of possible piRNAs profiles across *S. aureus* and *S. uberis* infections

piRNAs are another category of small non coding RNAs that are known to play a significant role in embryogenesis (Brennecke *et al.,* 2008) and in regulating the developmental pathway of the embryos (Samji, 2009) In exploring the diverse exosomal small RNA populations, profiling piRNA signatures could potentially open up venues for further research in understanding functional significance of these molecules in micro-vesicles like exosomes and their role in a non germ line cellular location.Introduction to the piRNA sequence and cluster features along with biogenesis is elucidated in Appendix of this dissertation.

Recent study has shown the role of piRNAs in regulation of many pathways, including their role in dysregulation of certain types of cancer (Iyengar *et al.,* 2014). piRNAs are also characterized in many body fluids like blood and gastric juice and is represented as an potentially efficient biomarker in disease identification (Bahn *et al.,* 2015). Maternal piRNAs are demonstrated to serve as epigenetic vectors in defining the biogenesis of piRNAs in the offspring (oocytes) in *D. melanogaster* (Le Thomas *et al.,* 2014).

A specific type of histone protein, H3K9me3, is enriched in the piRNA clusters which is essential for the piRNA biogenesis (Le Thomas *et al.,* 2014). Maternally inherited piRNAs are shown to serve as epigenetic vectors that are necessary to maintain

high levels of H3K9me3 in the offspring (Le Thomas *et al.*, 2014). It has been shown that piRNA knockouts in *D. melanogaster* oocytes leads to these proteins not to be anchored and hence demonstrates the significance of role of piRNAs in the epigenetic pathways (Le Thomas *et al*., 2014). A recent study built the small RNA profiles in saliva in healthy women (Bahn *et al*., 2015).

This study explores the possibility of the presence of piRNAs in bovine milk exosomes as a part of the small RNA profiling effort.

7.1 Significance of piRNA in this study

One of the significant roles of piRNAs are their epigenetic role (Le Thomas *et al.*, 2014) and their association with PIWI proteins helps protect germ line cells. However the presence of any piRNAs in milk would open up new investigations of their roles in body fluids and potentially their effect on the offspring. It have been shown that the maternally inherited piRNAs are shown to serve as epigenetic vectors that are necessary to maintain high levels of H3K9me3 in the offspring (Le Thomas *et al.*, 2014) in oocytes. Profiling any piRNA molecules in milk is a novel discovery of bovine piRNAs in the milk exosomes.

7.2 Bioinformatics analysis of piRNAs in bovine milk exosomes

piRNA analysis was performed based on the proTRAC - *probabilistic TRACking and Analysis of clusters*, a software created by Dr.Rosenkranz and Zischler (Rosenkranz & Zischler, 2012).

proTRAC statistically determines the minimum density for a region to be tracked as a cluster based on the binomial distribution which works on the assumption that a cluster is called a piRNA cluster when they significantly deviate from a uniform distribution. This software further charecterizes reads to the clusters based on the density of reads, the features of the reads with T in position 1 and A in position 10 and the length of the clusters across the genome.

7.3 Method

Novoalign mapped SAM files for each sample, was matched to the Novoalign SAM files of known human piRNAs from piRBANK (Rosenkranz & Zischler, 2012). SAM files were converted to BED files and intersecting BED files were used to determine the matching co-ordinates. These files were converted to FASTA files and processed through piRMAPPER (Rosenkranz & Zischler, 2012), another PERL based package and finally cluster coordinates were determined using proTRAC. Figure 7.1 shows a schematic representation of the steps involved in the piRNA analysis.

piRMAPPER is specifically designed to map piRNA sequences to the genomes using a specialized algorithm that requires a perfect 5' seed match (default:18nt) and

optionally allows non-template 3' ends as well as internal mismatches in the part of the sequences that follows the seed region. Default parameters were used in the mapping of the sample Novoalign data onto the piRMAPPER.

proTRAC are provides the options to normalize the hits for each sequence by the number of sequence reads, number of genomic hits and the number of total mapped reads per million mapped reads.

Advantages of proTRAC is that it considers piRNA structural features in cluster assignment, for example the 1U/10A and the strand bias of reads.

Many available softwares incorporates heuristic methods depending on the need for specificity or sensitivity in analysis. Many methods used by Girard *et al.,* and Lau *et al.,*. are based on high specificity, which are very successful in pulling out sequences with high abundance and and requiring minimum number of single copy piRNA loci. However, the disadvantage of this method proves to be the elimination of recently duplicated clusters with very few piRNA reads (Rosenkranz & Zischler, 2012).

On the other hand, methodology implemented by Lakshmi and Agarwal (2007) emphasise the high sensitivity compromising specificity leading to high mapping of sequences to regions that may not be piRNA clusters like the transposable elements (Rosenkranz & Zischler, 2012).

In summary, proTRAC tries to take into account the high abundance, sequence features like the 1U/10A, strand bias as well as the frequency of piRNA loci through the length of the genome including those loci with low abundance of reads (Rosenkranz & Zischler, 2012).

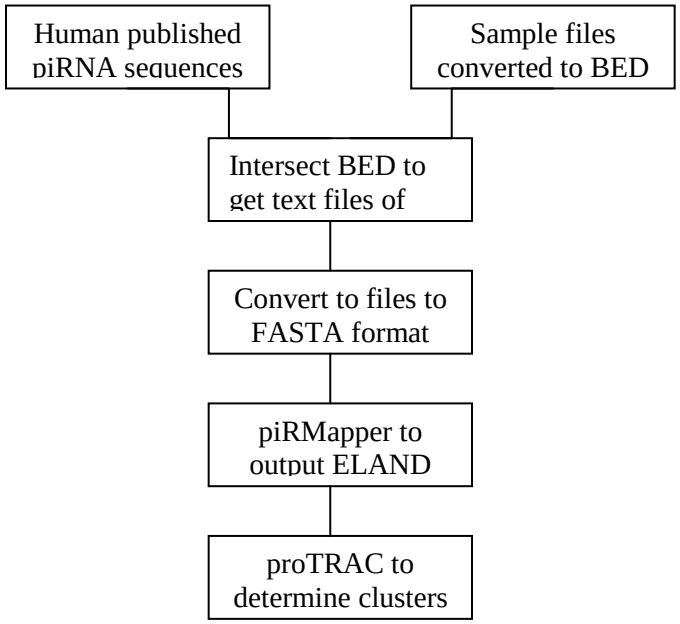


Figure 7.1. Schematic workflow of piRNA data analysis. Each step indicates the processing of data set and is explained in section 7.3

7.4 Results

proTRAC considers the dataset in the FASTA format and a reference genome file in order to perform the cluster analysis. The following command implements the software to run on a Unix system.

```
> perl proTRAC_2.0.5.pl -map piRNAs.fasta.map -genome genome.fasta –nr –nh
```

Output of proTRAC provides clusters and individual FASTA files for each cluster. Table 7.1 represents the cluster data of piRNAs as determined by proTRAC. Table 9 shows the sample ID, chromosome, coordinate and the direction of the clusters. *S. aureus* and *S. uberis* provides the following two clusters that are common between the two groups.

Chr 22: Cluster 7769250-7840680 Chr 3: Cluster 101145523-101169564

Table 7.1. Data on clusters of piRNA *S. aureus* and *S. uberis* samples from proTRAC output.Table indicates the sample names with the total number of clusters identified with read depths.

sample name	# of piRNA clusters	# reads / clusters	Unique sequences / clusters	direction	bi	plus	minus
3658_d2_uberis_11	24	1122	87	2bi/13+/9-	2	13	9
3658_d0_uberis_10	4	856	27	0/2+/2-	0	2	2
3554_d2_uberis_09	2	109	8	1bi/0+/1-	1	0	1
3554_d0_uberis_08	13	1578	66	2bi/7+/4-	2	7	4
3318r_d2_uberis_12	30	32300	98	6bi/15+/9-	6	15	9
3318r2_d0_uberis_06	0	0	0	0	0	0	0
3207_d2_uberis_04	1	198	26	0bi/1+/0-	0	1	0
3207_d0_uberis_16	5	19	8	1bi/2+/2-	1	2	2
3206r_d2_acute_15	9	20194	54	1bi/0+/8-	1	0	8
3206_d0_acute_01	8	10799	55	2bi/3+/3-	2	3	3
3658_d2_acute_14	6	88	14	3bi/2+/1-	3	2	1
3658_d0_acute_13	0	0	0	0	0	0	0
3504_d2_acute_ca_05	0	0	0	0	0	0	0
3504_d0_acute_ca_17	33	7192	92	4bi/15+/14-	4	15	14
3207_d2_acute_ca_07	0	0	0	0	0	0	0
3207r_d0_acute_ca_03	10	538	68	2bi/4+/4-	2	4	4
3132_d2_acute_02	4	746	23	1+/3-	0	1	3

7.5 Validation of piRNA reads in the data set

Use of proTRAC in profiling piRNA is a new method implemented in this research that identifies the possible piRNA clusters on the genomes. A recent study published on March 26, 2015 in Cell Reports (Ketting *et al.*,) charecterized bovine piRNA in mammalian oocytes and early embryos, and implemented proTRAC for charecterzing piRNAs. Based on the published piRNA sequences from the paper, a simple sequence match was implemented to check for any piRNAs in our data set.

Data from across all *S. aureus* and *S. uberis* reads are matched either completely or as a subset of the sequence from the published bovine piRNA sequences. *as* shown in Table 7.2. Total unique reads across all clusters in the dataset, to the number of reads matching testes or ovary is indicated in the columns.

Table 7.2. Sequences matching published bovine piRNAs (as a subset or as entire sequence match) of *S. aureus* and *S. uberis* data, to the *Bos taurus*, published, ovary and testes piRNA sequences, from piRBANK.

proTRAC_clusters	Total unique piRNAs across all clusters in data set	Reads matching the testes piRNA either as a whole or as a subset	# matching the Ovary piRNA either as a whole or as a subset
3207r_d0_aureus	68	47	43
3207r_d2_aureus	no data	n/a	n/a
3206r_d2_aureus	54	33	36
3206_d0_aureus	55	31	34
3658_d2_aureus	14	6	6
3658_d0_aureus	no data	n/a	n/a
3132_d2_aureus	24	5	5
3504_d2_aureus_ca	no data	n/a	n/a
3504_d0_aureus_ca	92	61	59
3658_d2_uberis_11	87	54	55
3658_d0_uberis_10	27	23	22
3554_d2_uberis_09	9	5	4
3554_d0_uberis_08	67	57	53
3318r_d2_uberis_12	99	61	61
3318r2_d0_uberis_06	no data	n/a	n/a
3207_d2_uberis_04	26	0	0
3207_d0_uberis_16	8	1	3

7.6 Summary

proTRAC outputs a few samples with no clusters indicating that there are no possible piRNA cluster locations. This seems to be an artifact of the software rather than a biological fact. There could be many reasons that this situation could happen for example, there could be a low abundance of reads that fits the proTRAC defined features of a read to be considered a piRNA and hence could not assign it to any location and hence discarded or the "noise" in the exosomal content, for example a high abundance of short reads less than the length requirements of a piRNAs, that could skew the prediction. Another possible reason for seeing low number of reads matching piRNAs clusters could be due to the fact that sequences could be degraded mostly as it is the exosome. Although previous studies have demonstrated the integrity of miRNAs in milk exosomes (Zhou *et al.*, 2012), there is not much information on the stability of piRNAs in the exosomes. Functional role of miRNAs and the abundance of immune related miRNAs in exosomes is demonstrated by previous studies (Zhou *et al.*, 2012) however there is no kown functional roles attributed from previous studies for the presence of exosomal piRNAs.

Although the results from matching the reads to the published data cannot be conclusive, it does indicate that the presence of populations of reads that match the known piRNAs.

Further experimental proof is needed to backup the validation of the presence of piRNAs in exosomes and their possible functional role across infection conditions needs to be further evaluated in a body fluid like milk.

Chapter 8 : Conclusion and future direction of research

This study focused on profiling small RNA signatures across *S. aureus* and *S. uberis* infections of the bovine milk exosomes of the dairy cattle as an effort to develop molecular biomarkers in detection of a disease like mastitis, at sub-clinical stages. Previous studies have been promising in demonstrating the potential of exosomal miRNAs in disease detection and therapeutics (Hu et al., 2012) and (Cortez *et al*., 2011). Previous studies show the importance of breast milk and its components in innate and adaptive immunity of offspring (Zhou *et al*., 2012) and various studies by (Hu et al., 2012), (Valadi et al., 2007)., and (Meckes Jr. & Raab-Traub, 2011) demonstrate the role of exosomal miRNAs in cell-cell communication and its potential as biomarkers in disease diagnosis.

This is a comprehensive study conducted to understand the effect of mastitis on the protein, mRNA and miRNA contents of milk exosomes, in cattle induced with bacterial infections, initiated at National Animal Disease Center (NADC), Ames, Iowa.

Proteomic study was performed at NADC, Agricultural Research Services (ARS), USDA (Ames, Iowa). A transcriptomics study was performed at the level of mRNA at Teagasc, Animal and Bioscience Research Department, Grange (Dunsany, County Meath, Ireland) by Dr. David Lynn's group and this dissertation, focused on profiling lactation related bovine milk miRNAs, under two days post-infection with *S. uberis* and

S. aureus bacterial inoculations, across nine Holstein-Friesian cattle. was performed at ARS / USDA (Beltsville, MD).

Exosomal miRNA study proved to be successful in profiling miRNAs in the control and the *S. aureus* and *S. uberis* infections of the bovine milk exosomes of the dairy cattle. Majority of the reads sequenced across the Next Generation Illumina Platform mapped onto the bovine genome, providing a 96% mapping across *S. uberis*, six libraries and 83% mapping across *S. aureus*, nine libraries. A read size analysis represented a high population of reads with sizes ranging from 18nt to 24nt lengths indicating high abundance of the miRNA read ranges. Read depth of the *S. aureus* infections were considerably higher than *S. uberis* infections indicating possible higher amounts of cell shedding in response to the *S. aureus* infections in general.

Although clustering of the samples across treatments did not yield a conclusive understanding of the effect of treatment on miRNA genes, we could successfully identify some of the key miRNA types that could potentially be used as biomarkers in identification of the disease at a 48hr period inoculation.

Some of the known mammary gland specific miRNAs like the bta-miR-146, bta-miR-30e and bta-miR-142 and bta-miR-98 a member of the bta-let-7 family were shown to be up-regulated in the bovine milk exosomal content under *S. uberis* infection. Previous study demonstrates bta-miR-142 to be up-regulated specifically under *S. uberis* infections of the mammary monocytes (Lawless *et al.*, 2013) which is concurrent to our findings. miR-142 is also shown to have a role in immune response in a study involving lupus patients (Carlsen *et al.*, 2013) where miR-142 expression is increased. Down

regulation of certain types of miRNA genes seen in this study could be associated with the genes they regulate. Target prediction of those miRNA genes could reveal more information about their behavior, which is not the focus of this study.

Some of the known mammary gland specific miRNAs like the bta-miR-142 and bta-miR-223 were seen up-regulated in the bovine milk exosomal content under *S. aureus* infections. These two miRNAs were differentially expressed across all animals in the *S. aureus* study as candidate genes across infection condition.

Another important class of miRNA genes that are seen up-regulated are the miR-99a and miR-99b, which have been shown in previous study to aid in epithelial to mesenchymal transition during disease development and progression in murine mammary gland cells during and in proliferation and migration of breast cancer cells (Turcatel *et al.,* 2012). Apart from the above mentioned miRNA genes, another well known miRNA; miR-223 is up-regulated in the *S. aureus* infections. Previous studies have implicated its role in many types of cancer, implicating its role as a potential biomarker (Liu *et al.*, 2010).

Many of the miRNAs profiled in this study matches the information from Exocarta (Mathivanan *et al.,* 2012), a database on known miRNAs in micro vesicles like exosomes, except for miR-1468, which is differentially regulated in *S. uberis* infections.

This study also was successful in identification of piRNA clusters on the bovine genome, homologous to the human piRNA locations. piRNAs in the bovine milk exosomes are identified in comparison to the bovine ovary and testes data published in a recent study (Roovers *et al*., 2015). One of the significant roles of piRNAs are their

epigenetic role (Le Thomas *et al*., 2014) and their association with PIWI proteins helps protect germ line cells. It have been shown that the maternally inherited piRNAs are shown to serve as epigenetic vectors that are necessary to maintain high levels of H3K9me3 in the offspring (Le Thomas *et al*., 2014) in oocytes.

Discovery of piRNA molecules in bovine milk exosomes opens up many novel directions of research pertaining to the understanding of the functional role of piRNA molecules in somatic body fluid like milk and its potential role as a biomarker in disease detection.

This knowledge of the small RNAs in the bovine milk exosomes not only adds to the understanding of the RNA biology of various cellular components in higher organism like a cow, but also could prove valuable in extending this understanding to human breast milk exosomes across disease conditions. Currently there is a study of such type across normal breast milk (Zhou *et al.,* 2012) in humans, however this study does not address infections of the mammary tissue.

This is a novel study that provides a comparative profile of miRNAs across two types of infections which can be further extended to study mastitis across other causative agents and in turn its effect on the offspring's' innate and adaptive immunity.

Target prediction for the profiled miRNAs across both experimental conditions could provide useful in the future in better understanding the roles of these miRNA genes and also contributes towards the development of a molecular biomarker in identification of the disease like mastitis.

Appendix

A.1 Introduction to piRNAs

piRNAs are characterized as short non- protein coding RNAs which typically range from 26-33 nucleotides in length, longer than the miRNAs. piRNAs are found in abundance in germ line cells of mammals (Haifan Li *et al.*, 2009). A recent study has characterized piRNAs in the somatic follicle cells (Hannon *et al.*, 2009) that line the nurse cells in *D. melanogaster*. So far no piRNAs have been discovered in plants, although many other types of small RNAs like the siRNA and miRNA are reported in Arabidopsis (Matzke *et al.*, 2002) and in Petunia (Westhead *et al.*, 2009). piRNAs are enriched with repeats and hence previously categorized as rasiRNAs or repeat associated small interfering RNAs. One of the unique features of the piRNAs is the presence of a 5' Uridine which is found in most vertebrates and invertebrates (Marucci G *et al.*, 2009). piRNAs have also known to have a 3' monophosphate and a 5' modification of a 2-0-methylation (Joshua-Tor *et al.*, 2007). Unlike the short RNAs, the piRNAs have no known secondary structures and the individual sequences are not conserved (Lin and Thompson, 2009).

A.2 Argonaut proteins and piRNAs

piRNAs are known to be associated with certain specific class of Argonaut family of proteins called the PIWI proteins (Meister *et al.,* 2008), hence the name piRNAs. The Argonauts proteins mainly encompass the AGO clad and the PIWI clad which aid in the transportation of piRNAs to the target mRNA, thereby bringing about gene silencing (Höck & Meister, 2008). These proteins are known to be highly conserved at the level of amino acid, and many mammals express multiple forms isoforms of the same protein. (Hock and Meister, 2008).

All Argonaut proteins are characterized by the presence of a PAZ domain and the PIWI domain. The PAZ domain is known to have an RNaseIII activity while the PIWI domain is shown to have an RNAse H activity (Hock and Meister, 2008). The piRNA associates with the PIWI clad of Argonaut proteins are hence are called PIWI-associated RNAs (or piRNAs). Similarly in mouse, the piRNAs associate with MIWI proteins, while in zebra fish, there are two types of Argonauts associated with piRNAs, ZIWI and ZILI. In *C. elegans,* the previously known 21U-RNAs were discovered to be piRNAs. PRG1 and PRG2 are known to be the Argonauts of *C. elegans* and specifically PRG1 is known to associate with the piRNAs (Miska *et al.,* 2008).

A.3 Clusters

Another unique feature of piRNAs that they are derived from the genomic regions on the chromosome, that tends to group together in clusters that could be strand-specific,

depending on its orientation on the strand. Each cluster is likely to represent a long single stranded-RNA precursor (Yamanaka, Siomi, & Siomi, 2014). Sometimes they are seen to originate from two non-overlapping divergent precursors.

Although the individual piRNA sequences are not conserved they are transcribed from genomic clusters whose locations are highly conserved (Hetch *et al.,* 2008). piRNAs map onto the genome in clusters of 20 to 90 kb either in the sense strand or anti-sense strand. piRNA clusters are not uniformly seen mapping onto the genome and they tend to map to various regions like the intergenic regions, exonic and intronic regions of the genome (Yamanaka *et al.,* 2014). Although the clusters are not very unique, they are seen to be evolutionarily conserved. Although most piRNAs tend to cluster they could range as a single or few per cluster to large numbers.

A.4 Other features of piRNAs

piRNAs are known to exhibit certain preference of nucleotides at specific sites depending on the species. For example, many vertebrate and invertebrate piRNAs show a huge tendency in having a uracil-rich 5' region. Vertebrates prefer to have adenine at the 10th position (Cora *et al.,* 2014). Also, in many organisms like the *C. elegans*, zebra fish, *Drosophila melanogaster* it is seen that along with the 5' monophosphate there also exists a 3' modification which is highlighted a possible role of the piRNA in the epigenetic mode of regulation of piRNA (Brennecke *et al.,* 2008). One such example is shown in the *D. melanogaster* where the PIWI associated piRNA is influenced toward

epigenetic role in gene regulation (Brennecke *et al*., 2008). Evolutionary conservation of clusters of piRNAs could hold tremendous potential in understanding the significance of these small molecules and its role in a biological system.

A.5 piRNA Biogenesis

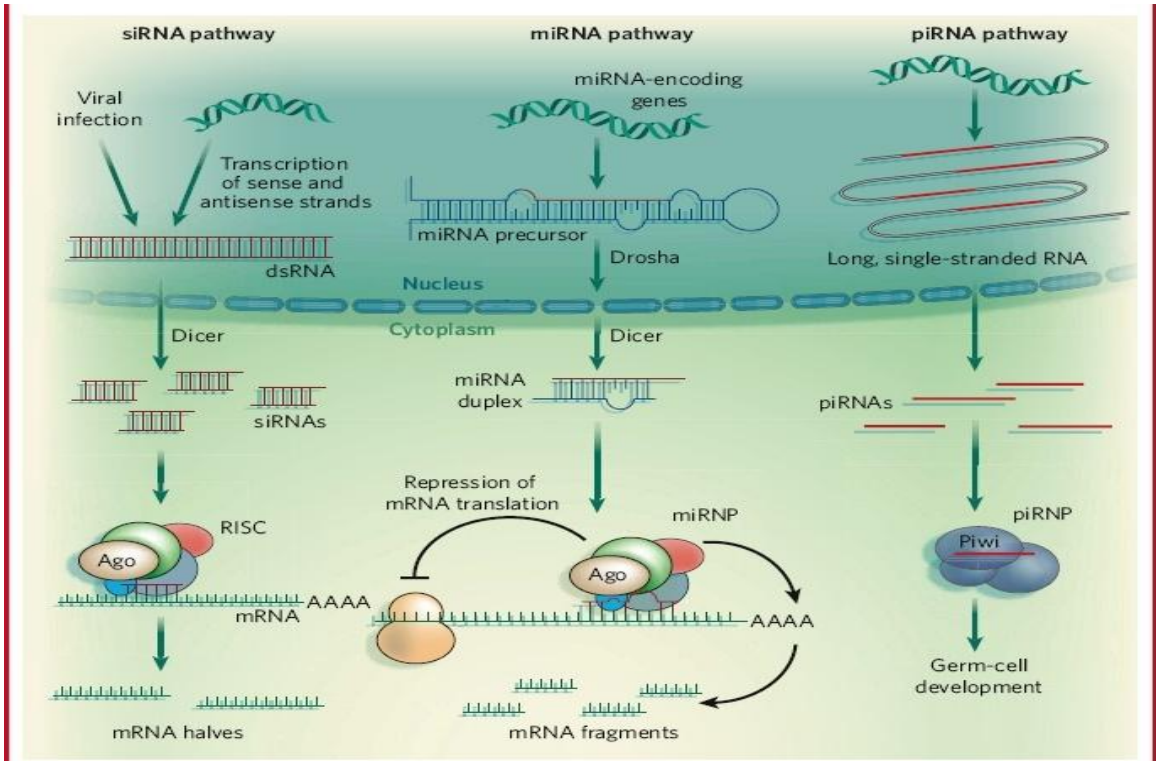


Figure 1. piRNA biogenesis. A comparative biogenesis of small RNA pathways showing various proteins involved in processing of the precursor molecules across nucleus and cytosol. Taken from Siomi *et al.,* Nat Rev Mol Cell Biol 2009 Feb;10(2):126-39.

Figure 1. shows the piRNA biogenesis in contrast to the miRNA biogenesis pathway and siRNA pathways. piRNAs are known to have a dicer independent mechanism of biogenesis unlike its other short RNA counterparts like the siRNAs and the

miRNAs, which needs the dicer component. Also, the piRNAs are known to have single-stranded precursors when compared to the siRNAs and miRNAs which need the double-stranded precursors. Although many questions are still left unanswered in understanding exact mechanism of piRNA biogenesis, there are some evidence to backup certain features but more needs to be done.

A.6 Ping-Pong Model

One of the proposed models for piRNA biogenesis is the ping-pong model (Klattenhoff & Theurkauf, 2007) as shown in Figure 2.This model is elucidated based on certain structural characteristic features of the piRNAs and bioinformatics analysis. One such structural feature is the complementarity of the piRNAs that associate with AGO3 proteins, to the piRNAs that associate with the AUB or the PIWI proteins. piRNA associated with the AUB and the PIWI are antisense to the transposon regions of the genome in the *Drosophila melanogaster* (Brennecke *et al.,* 2007).

One speculation about how the mature piRNAs are cleaved from the precursor is that perhaps the AUB /PIWI associated piRNAs are cleaved into small transcripts due to the RNAseH activity associated with the folds of the Argonauts (Weick & Miska, 2014). Also certain types of PIWIs are shown to have Ago2 mediated slicer activity and that feature perhaps could help cleave the piRNA precursors into mature piRNAs (Klattenhoff & Theurkauf, 2007).

According to the germ line piRNA pathway, the piRNAs that preferentially bind to the AGO3 are antisense in nature and bind to the sense strand of the transposon regions

of the genome and cleave them into short transcripts (Zamore, 2010). These short transcripts or the secondary piRNAs, that are the sense strands, bind to the AUB/PIWI, which further binds to the antisense strand of the transposon and cleaves it to release the antisense short transcripts and the antisense short transcripts that bind to the AGO3 and so on. This is the cyclic mechanism that is referred to as ping-pong model (Thomson and Lin, 2009).

A.7 Somatic pathway

A study reported an increase in the piRNAs associated with the AGO3/AUB but nothing substantial was observed in the PIWI/AGO3 (Brennecke *et al*., 2008). piRNA pathway was shown to be much simpler than the ping-pong pathway, in the somatic cells (Li *et al., 2008* and Malone *et al., 2009*). The study was performed in the follicle cells of the drosophila ovary. Mostly, the transposon are found more in the germ line cells but one specific transposable element called GYPSY evades the germ line and plays a role in the follicle cells. The group-detected PIWI proteins in the follicle cells as against AGO3 and AUB, however previously PIWI was known to reside in the nucleus (Brennecke *et al*., 2008). This suggested that the PIWI had to have a way of shuttling between the cytoplasm and the nucleus.

(Brennecke *et al.,* 2008) through the RNAi studies, showed the significant need for three specific kinds of proteins that are involved in the shuttling of PIWI between the nucleus and cytoplasm. It is shown that the Yb and the Armitage proteins are necessary for the PIWI to move from the cytoplasm to nucleus. Mutants of Armitage and Yb have

been shown to retain PIWI in the cytoplasm. On the other hand, a lack of another kind of protein, Zucchini, holds the PIWI in the cytoplasm part of the nuclear envelope. Also in the absence of piRNA, the PIWI is retained in the cytoplasm (Brenneck *et al., 2008*). These data suggests that the piRNA guides the PIWI proteins to its destination.

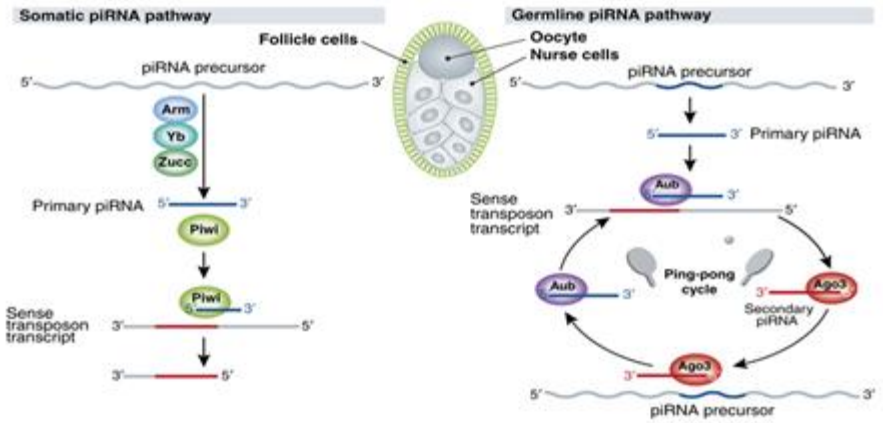


Figure 2: Comparative biogenesis of somatic and germ line piRNA pathway. Germ line piRNA pathway involves Ago3 and Aub dependent piRNA amplification cycle, whereas in the somatic cells, pathway occurs through a PIWI-dependent, Aub-Ago3 independent pathway.

A.8 Biological functions of piRNA and associated proteins

piRNAs and their associated proteins play a significant role in many functions of the cell system. PIWI proteins are known to play an epigenetic role (Aravin *et al.*, 2008). piRNAs along with the associated PIWI proteins, play a role in regulating transposon activity (Hartig, Tomari, & Forstemann, 2007). PIWI proteins are also known to be

involved in DNA elimination and DNA integrity. PIWI proteins associated with the piRNAs play a role in spermatogenesis, in germ cells (Brennecke *et al.*, 2008).

References

Adapter and quality trimming of Illumina data | ARK Genomics. (2015) Retrieved March 22, 2015, from http://www.ark-genomics.org/events-online-training-eu-training-course/adapter-and-quality-trimming-illumina-data

Anders, S., & Huber, W. (2010).Differential expression analysis for sequence count data.*Genome Biol*, *11*(10), R106.

Anders, S., McCarthy, D. J., Chen, Y., Okoniewski, M., Smyth, G. K., Huber, W., & Robinson, M. D. (2013). Count-based differential expression analysis of RNA sequencing data using R and BioConductor. *Nature Protocols*, *8*(9), 1765–1786. http://doi.org/10.1038/nprot.2013.099

Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*, *31*(2), 166–169. http://doi.org/10.1093/bioinformatics/btu638

Aravin, A. A., Sachidanandam, R., Bourc'his, D., Schaefer, C., Pezic, D., Toth, K. F., Hannon, G. J. (2008). A piRNA Pathway Primed by Individual Transposons Is Linked to De Novo DNA Methylation in Mice. *Molecular Cell*, *31*(6), 785–799. http://doi.org/10.1016/j.molcel.2008.09.003

Aravin, A. A., Sachidanandam, R., Girard, A., Fejes-Toth, K., & Hannon, G. J. (2007).Developmentally Regulated piRNA Clusters Implicate MILI in Transposon Control. *Science*, *316*(5825), 744–747. http://doi.org/10.1126/science.1142612

Assis, R., & Kondrashov, A. S. (2009). Rapid repetitive element-mediated expansion of piRNA clusters in mammalian evolution. *Proceedings of the National Academy of Sciences*, *106*(17), 7079–7082.

Bahn, J. H., Zhang, Q., Li, F., Chan, T.-M., Lin, X., Kim, Y., Xiao, X. (2015). The Landscape of MicroRNA, Piwi-Interacting RNA, and Circular RNA in Human Saliva. *Clinical Chemistry*, *61*(1), 221–230. http://doi.org/10.1373/clinchem.2014.230433

Barkema, H. W., Green, M. J., Bradley, A. J., & Zadoks, R. N. (2009).Invited review: The role of contagious disease in udder health. *Journal of Dairy Science*, *92*(10), 4717–4729. http://doi.org/10.3168/jds.2009-2347

Betel, D., Sheridan, R., Marks, D. S., & Sander, C. (2007).Computational analysis of mouse piRNA sequence and biogenesis.*PLoS Computational Biology*, *3*(11), e222.

Blog, R: piClust: A density based piRNA clustering algorithm | RNA-Seq Blog. Retrieved from http://www.rna-seqblog.com/piclust-a-density-based-pirna-clustering-algorithm/

Bobrie, A., Colombo, M., Raposo, G., & Théry, C. (2011). Exosome Secretion: Molecular Mechanisms and Roles in Immune Responses. *Traffic*, *12*(12), 1659–1668.

Bradley, A. J. (2002). Bovine mastitis: An Evolving Disease. *The Veterinary Journal*, *164*(2), 116–128. http://doi.org/10.1053/tvjl.2002.0724

Brennecke, J., Malone, C. D., Aravin, A. A., Sachidanandam, R., Stark, A., & Hannon, G. J. (2008). An Epigenetic Role for Maternally Inherited piRNAs in Transposon Silencing. *Science (New York, N.Y.)*, *322*(5906), 1387–1392. http://doi.org/10.1126/science.1165171

Buermans, H. P., Ariyurek, Y., van Ommen, G., den Dunnen, J. T., & AC't Hoen, P. (2010). New methods for next generation sequencing based microRNA expression profiling. *BMC Genomics*, *11*(1), 716.

Carlsen, A. L., Schetter, A. J., Nielsen, C. T., Lood, C., Knudsen, S., Voss, A., Heegaard, N. H. H. (2013). Circulating MicroRNA Expression Profiles Associated With Systemic Lupus Erythematosus. *Arthritis & Rheumatism*, *65*(5), 1324–1334. http://doi.org/10.1002/art.37890

Chen, T., Xi, Q.-Y., Ye, R.-S., Cheng, X., Qi, Q.-E., Wang, S.-B., others. (2014). Exploration of microRNAs in porcine milk exosomes.*BMC Genomics*, *15*(1), 100.

Chevillet, J. R., Kang, Q., Ruf, I. K., Briggs, H. A., Vojtech, L. N., Hughes, S. M., … Tewari, M. (2014). Quantitative and stoichiometric analysis of the microRNA content of exosomes. *Proceedings of the National Academy of Sciences*, *111*(41), 14888–14893. http://doi.org/10.1073/pnas.1408301111

Chitsaz, H., Salari, R., Sahinalp, S. C., & Backofen, R. (2009). A partition function algorithm for interacting nucleic acid strands. *Bioinformatics*, *25*(12), i365–i373.

Chuma, S., & Pillai, R. S. (2009). Retrotransposon Silencing by piRNAs: Ping-Pong Players Mark Their Sub-Cellular Boundaries. *PLoS Genetics*, *5*(12), e1000770. http://doi.org/10.1371/journal.pgen.1000770

Cora, E., Pandey, R. R., Xiol, J., Taylor, J., Sachidanandam, R., McCarthy, A. A., & Pillai, R. S. (2014). The MID-PIWI module of Piwi proteins specifies nucleotide-

and strand-biases of piRNAs. *RNA*, *20*(6), 773–781. http://doi.org/10.1261/rna.044701.114

Cortez, M. A., Bueso-Ramos, C., Ferdin, J., Lopez-Berestein, G., Sood, A. K., & Calin, G. A. (2011).MicroRNAs in body fluids—the mix of hormones and biomarkers. *Nature Reviews Clinical Oncology*, *8*(8), 467–477. http://doi.org/10.1038/nrclinonc.2011.76

Coutinho, L. L., Matukumalli, L. K., Sonstegard, T. S., Van Tassell, C. P., Gasbarre, L. C., Capuco, A. V., & Smith, T. P. L. (2006). Discovery and profiling of bovine microRNAs from immune-related and embryonic tissues. *Physiological Genomics*, *29*(1), 35–43. http://doi.org/10.1152/physiolgenomics.00081.2006

Deb, R., Kumar, A., Chakraborty, S., Verma, A. K., Tiwari, R., Dhama, K., Kumar, S. (2013). Trends in diagnosis and control of bovine mastitis: a review. *Pakistan Journal of Biological Sciences: PJBS*, *16*(23), 1653–61.

De Vliegher, S., Fox, L. K., Piepers, S., McDougall, S., & Barkema, H. W. (2012). Invited review: mastitis in dairy heifers: Nature of the disease, potential impact, prevention, and control. *Journal of Dairy Science*, *95*(3), 1025–1040. http://doi.org/10.3168/jds.2010-4074

Dillies, M.-A., Rau, A., Aubert, J., Hennequet-Antier, C., Jeanmougin, M., Servant, N., Jaffrézic, F. (2012). A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. *Briefings in Bioinformatics*, bbs046. http://doi.org/10.1093/bib/bbs046

D'Incà, F., & Pucillo, C. E. (2015). Exosomes: tiny clues for mast cell communication. *Molecular Innate Immunity*, *6*, 73. http://doi.org/10.3389/fimmu.2015.00073

Dohoo, I. R., & Meek, A. H. (1982). Somatic Cell Counts in Bovine Milk. *The Canadian Veterinary Journal*, *23*(4), 119–125.

Fedora (2013). Retrieved April 12, 2015, from https://getfedora.org/

Garmire, L. X., & Subramaniam, S. (2012). Evaluation of normalization methods in mammalian microRNA-Seq data.*RNA*, *18*(6), 1279–1288. http://doi.org/10.1261/rna.030916.111

Gerstl, M. P., Hackl, M., Graf, A. B., Borth, N., & Grillari, J. (2013).Prediction of transcribed PIWI-interacting RNAs from CHO RNAseq data.*Journal of Biotechnology*, *166*(1-2), 51–57. http://doi.org/10.1016/j.jbiotec.2013.04.010

Gey, A., Werckenthin, C., Poppert, S., & Straubinger, R. K. (2013). Identification of pathogens in mastitis milk samples with fluorescent in situ hybridization. *Journal of*

Veterinary Diagnostic Investigation, *25*(3), 386–394. http://doi.org/10.1177/1040638713486113

Giardine, B., Riemer, C., Hardison, R. C., Burhans, R., Elnitski, L., Shah, P., Nekrutenko, A. (2005). Galaxy: A platform for interactive large-scale genome analysis. *Genome Research*, *15*(10), 1451–1455. http://doi.org/10.1101/gr.4086505

Gigli, I., & Maizon, D. O. (2013).microRNAs and the mammary gland: A new understanding of gene expression. *Genetics and Molecular Biology*, *36*(4), 465–74. http://doi.org/10.1590/S1415-47572013005000040

Grivna, S. T. (2006). A novel class of small RNAs in mouse spermatogenic cells. *Genes & Development*, *20*(13), 1709–1714. http://doi.org/10.1101/gad.1434406

Grivna, S. T., Pyhtila, B., & Lin, H. (2006). MIWI associates with translational machinery and PIWI-interacting RNAs (piRNAs) in regulating spermatogenesis. *Proceedings of the National Academy of Sciences*, *103*(36), 13415–13420.

Gu, Z., Eleswarapu, S., & Jiang, H. (2007). Identification and characterization of microRNAs from the bovine adipose tissue and mammary gland.*FEBS Letters*, *581*(5), 981–988. http://doi.org/10.1016/j.febslet.2007.01.081

Hartig, J. V., Tomari, Y., & Forstemann, K. (2007).piRNAs--the ancient hunters of genome invaders. *Genes & Development*, *21*(14), 1707–1713. http://doi.org/10.1101/gad.1567007

Hartl, T. A., Smith, H. F., & Bosco, G. (2008). Chromosome Alignment and Transvection Are Antagonized by Condensin II. *Science*, *322*(5906), 1384–1387. http://doi.org/10.1126/science.1164216

Hata, T., Murakami, K., Nakatani, H., Yamamoto, Y., Matsuda, T., & Aoki, N. (2010).Isolation of bovine milk-derived microvesicles carrying mRNAs and microRNAs. *Biochemical and Biophysical Research Communications*, *396*(2), 528–533. http://doi.org/10.1016/j.bbrc.2010.04.135

Hatem, A., Bozdağ, D., Toland, A. E., & Çatalyürek, Ü. V. (2013).Benchmarking short sequence mapping tools.*BMC Bioinformatics*, *14*(1), 184.

Hettinga, K. A., van Valenberg, H. J. F., Lam, T. J. G. M., & van Hooijdonk, A. C. M. (2009). The influence of incubation on the formation of volatile bacterial metabolites in mastitis milk. *Journal of Dairy Science*, *92*(10), 4901–4905. http://doi.org/10.3168/jds.2008-1818

Höck, J., & Meister, G. (2008).The Argonaut protein family. *Genome Biol*, *9*(2), 210.

Hogeveen, H., Huijps, K., & Lam, T. (2011). Economic aspects of mastitis: New developments. *New Zealand Veterinary Journal*, *59*(1), 16–23. http://doi.org/10.1080/00480169.2011.547165

Huang, X., Yuan, T., Tschannen, M., Sun, Z., Jacob, H., Du, M., Wang, L. (2013). Characterization of human plasma-derived exosomal RNAs by deep sequencing.*BMC Genomics*, *14*(1), 319. http://doi.org/10.1186/1471-2164-14-319

Hu, G., Drescher, K. M., & Chen, X.-M. (2012). Exosomal miRNAs: Biological Properties and Therapeutic Potential. *Frontiers in Genetics*, *3*. http://doi.org/10.3389/fgene.2012.00056

illumina_sequencing_introduction.pdf. (2013).

Iyengar, B. R., Choudhary, A., Sarangdhar, M. A., Venkatesh, K. V., Gadgil, C. J., & Pillai, B. (2014). Non-coding RNA interact to regulate neuronal development and function. *Frontiers in Cellular Neuroscience*, *8*. http://doi.org/10.3389/fncel.2014.00047

Izumi, H., Tsuda, M., Sato, Y., Kosaka, N., Ochiya, T., Iwamoto, H., Takeda, Y. (2015). Bovine milk exosomes contain microRNA and mRNA and are taken up by human macrophages. *Journal of Dairy Science*. http://doi.org/10.3168/jds.2014-9076

Jin, P., Alisch, R. S., & Warren, S. T. (2004).RNA and microRNAs in fragile X mental retardation. *Nature Cell Biology*, *6*(11), 1048–1053.

Jin, W., Ibeagha-Awemu, E. M., Liang, G., Beaudoin, F., Zhao, X., & Guan, L. L. (2014). Transcriptome microRNA profiling of bovine mammary epithelial cells challenged with Escherichia coli or Staphylococcus aureus bacteria reveals pathogen directed microRNA expression profiles. *BMC Genomics*, *15*, 181. http://doi.org/10.1186/1471-2164-15-181

JOHANSSON, I., & KORKMAN, N. (1952). Heritability of the udder proportions in dairy cows. *Hereditas*, *38*(2), 131–151.

Jones, G. M., & Bailey, T. L. (2009).Understanding the basics of mastitis. Retrieved from https://vtechworks.lib.vt.edu/handle/10919/48392

J. Simpson, R., & Mathivanan, S. (2012). Extracellular Microvesicles: The Need for Internationally Recognised Nomenclature and Stringent Purification Criteria. *Journal of Proteomics & Bioinformatics*, *05*(02). http://doi.org/10.4172/jpb.10000e10

Júnior, L., Ferreira, J. E., Lange, C. C., Brito, M. A. V. P., Santos, F. R., Silva, M. A. S., … Souza, G. N. de. (2012). Relationship between total bacteria counts and somatic

cell counts from mammary quarters infected by mastitis pathogens.*Ciência Rural*, *42*(4), 691–696.

Kadakkuzha, B. (2014). Role of noncoding RNAs in diseases.*RNA & DISEASE*, *1*(1), 10–14800.

Klattenhoff, C., & Theurkauf, W. (2007).Biogenesis and germline functions of piRNAs. *Development*, *135*(1), 3–9. http://doi.org/10.1242/dev.006486

Kohno, K., Izumi, H., Uchiumi, T., Ashizuka, M., & Kuwano, M. (2003).The pleiotropic functions of the Y-box-binding protein, YB-1.*BioEssays*, *25*(7), 691–698. http://doi.org/10.1002/bies.10300

Kuramochi-Miyagawa, S. (2004). Mili, a mammalian member of piwi family gene, is essential for spermatogenesis. *Development*, *131*(4), 839–849. http://doi.org/10.1242/dev.00973

Kwon, C., Tak, H., Rho, M., Chang, H. R., Kim, Y. H., Kim, K. T., … Nam, S. (2014). Detection of PIWI and piRNAs in the mitochondria of mammalian cancer cells.*Biochemical and Biophysical Research Communications*, *446*(1), 218–223. http://doi.org/10.1016/j.bbrc.2014.02.112

Langmead, B., Trapnell, C., Pop, M., Salzberg, S. L., & others. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.*Genome Biol*, *10*(3), R25.

Lawless, N., Foroushani, A. B. K., Mccabe, M. S., O'Farrelly, C., & Lynn, D. J. (2013). Next generation sequencing reveals the expression of a unique miRNA profile in response to a gram-positive bacterial infection. *PloS One*, *8*(3), e57543. http://doi.org/10.1371/journal.pone.0057543

Lawless, N., Reinhardt, T. A., Bryan, K., Baker, M., Pesch, B., Zimmerman, D., Lynn, D. J. (2014). MicroRNA regulation of bovine monocyte inflammatory and metabolic networks in an in vivo infection model.*G3 (Bethesda, Md.)*, *4*(6), 957–71. http://doi.org/10.1534/g3.113.009936

Lawless, N., Vegh, P., O'Farrelly, C., & Lynn, D. J. (2014).The Role of microRNAs in Bovine Infection and Immunity. *Frontiers in Immunology*, *5*. http://doi.org/10.3389/fimmu.2014.00611

Leitner, G., Merin, U., Krifucks, O., Blum, S., Rivas, A. L., & Silanikove, N. (2012). Effects of intra-mammary bacterial infection with coagulase negative staphylococci and stage of lactation on shedding of epithelial cells and infiltration of leukocytes into milk: Comparison among cows, goats and sheep. *Veterinary Immunology and Immunopathology*, *147*(3-4), 202–210. http://doi.org/10.1016/j.vetimm.2012.04.019

Lerat, E. (2010). Identifying repeats and transposable elements in sequenced genomes: how to find your way through the dense forest of programs. *Heredity*, *104*(6), 520–533.

Le Thomas, A., Stuwe, E., Li, S., Du, J., Marinov, G., Rozhkov, N., Aravin, A. A. (2014). Transgenerationally inherited piRNAs trigger piRNA biogenesis by changing the chromatin of piRNA clusters and inducing precursor processing. *Genes & Development*, *28*(15), 1667–1680. http://doi.org/10.1101/gad.245514.114

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. http://doi.org/10.1093/bioinformatics/btp324

Lindow, M., & Gorodkin, J. (2007).Principles and Limitations of Computational MicroRNA Gene and Target Finding.*DNA and Cell Biology*, *26*(5), 339–351. http://doi.org/10.1089/dna.2006.0551

Li, R., Li, Y., Kristiansen, K., & Wang, J. (2008). SOAP: short oligonucleotide alignment program. *Bioinformatics*, *24*(5), 713–714. http://doi.org/10.1093/bioinformatics/btn025

Liu, T.-Y., Chen, S.-U., Kuo, S.-H., Cheng, A.-L., & Lin, C.-W. (2010). E2A-positive gastric MALT lymphoma has weaker plasmacytoid infiltrates and stronger expression of the memory B-cell-associated miR-223: possible correlation with stage and treatment response. *Modern Pathology*, *23*(11), 1507–1517. http://doi.org/10.1038/modpathol.2010.139

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.*Genome Biology*, *15*(12). http://doi.org/10.1186/s13059-014-0550-8

Luo, L., Ye, L., Liu, G., Shao, G., Zheng, R., Ren, Z., Li, F. (2010). Microarray-Based Approach Identifies Differentially Expressed MicroRNAs in Porcine Sexually Immature and Mature Testes. *PLoS ONE*, *5*(8), e11744. http://doi.org/10.1371/journal.pone.0011744

Malone, C. D., Brennecke, J., Dus, M., Stark, A., McCombie, W. R., Sachidanandam, R., & Hannon, G. J. (2009).Specialized piRNA Pathways Act in Germline and Somatic Tissues of the Drosophila Ovary. *Cell*, *137*(3), 522–535. http://doi.org/10.1016/j.cell.2009.03.040

Mardis, E. R. (2008). The impact of next-generation sequencing technology on genetics. *Trends in Genetics*, *24*(3), 133–141. http://doi.org/10.1016/j.tig.2007.12.007

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, *17*(1), pp–10.

Mathivanan, S., Fahner, C. J., Reid, G. E., & Simpson, R. J. (2012). ExoCarta 2012: database of exosomal proteins, RNA and lipids. *Nucleic Acids Research*, *40*(D1), D1241–D1244. http://doi.org/10.1093/nar/gkr828

Meckes, D. G. (2015). Exosomal Communication Goes Viral. *Journal of Virology*, JVI.02470–14. http://doi.org/10.1128/JVI.02470-14

Meckes Jr., D. G., & Raab-Traub, N. (2011). Microvesicles and Viral Infection. *Journal of Virology*, *85*(24), 1–2. http://doi.org/Article

Mette, M. F. (2002). Short RNAs Can Identify New Candidate Transposable Element Families in Arabidopsis. *PLANT PHYSIOLOGY*, *130*(1), 6–9. http://doi.org/10.1104/pp.007047

Meyer, S. U., Pfaffl, M. W., & Ulbrich, S. E. (2010). Normalization strategies for microRNA profiling experiments: a "normal" way to a hidden layer of complexity? *Biotechnology Letters*, *32*(12), 1777–1788. http://doi.org/10.1007/s10529-010-0380-z

Milking, milk production hygiene and udder health. (2015) Retrieved April 9, 2015, from http://www.fao.org/docrep/004/t0218e/t0218e04.htm

Milk Quality Around the World. (2009). Retrieved April 8, 2015, from http://www.slideshare.net/QualityMilk/milk-quality-around-the-world

Murchison, E. P., Kheradpour, P., Sachidanandam, R., Smith, C., Hodges, E., Xuan, Z., Hannon, G. J. (2008) Conservation of small RNA pathways in platypus. *Genome Research*, *18*(6), 995–1004. http://doi.org/10.1101/gr.073056.107

Naeem, A., Zhong, K., Moisá, S. J., Drackley, J. K., Moyes, K. M., & Loor, J. J. (2012). Bioinformatics analysis of microRNA and putative target genes in bovine mammary tissue infected with Streptococcus uberis. *Journal of Dairy Science*, *95*(11), 6397–6408. http://doi.org/10.3168/jds.2011-5173

Nawrocki, E. P., Burge, S. W., Bateman, A., Daub, J., Eberhardt, R. Y., Eddy, S. R., … Finn, R. D. (2014). Rfam 12.0: updates to the RNA families database. *Nucleic Acids Research*, gku1063. http://doi.org/10.1093/nar/gku1063

Nelson, C. D., Reinhardt, T. A., Beitz, D. C., & Lippolis, J. D. (2010).In Vivo Activation of the Intracrine Vitamin D Pathway in Innate Immune Cells and Mammary Tissue during a Bacterial Infection.*PLoS ONE*, *5*(11), e15469. http://doi.org/10.1371/journal.pone.0015469

Novoalign | Novocraft. (2010). Retrieved from http://www.novocraft.com/documentation/Novoalign-2/

Novocraft.pdf. (2010).

Ogawa, Y., Taketomi, Y., Murakami, M., Tsujimoto, M., & Yanoshita, R. (2013).Small RNA transcriptomes of two types of exosomes in human whole saliva determined by next generation sequencing. *Biological and Pharmaceutical Bulletin*, *36*(1), 66–75.

Ogorevc, J., Kunej, T., Razpet, A., & Dovc, P. (2009).Database of cattle candidate genes and genetic markers for milk production and mastitis.*Animal Genetics*, *40*(6), 832–851. http://doi.org/10.1111/j.1365-2052.2009.01921.x

Patel, T. (2014). Extracellular vesicle noncoding RNA: New players in the diagnosis and pathogenesis of cholangiocarcinoma. *Hepatology*, *60*(3), 782–784. http://doi.org/10.1002/hep.27185

Pezic, D., Manakov, S. A., Sachidanandam, R., & Aravin, A. A. (2014).piRNA pathway targets active LINE1 elements to establish the repressive H3K9me3 mark in germ cells. *Genes & Development*, *28*(13), 1410–1428. http://doi.org/10.1101/gad.240895.114

Pigati, L., Yaddanapudi, S. C. S., Iyengar, R., Kim, D.-J., Hearn, S. A., Danforth, D., Duelli, D. M. (2010). Selective Release of MicroRNA Species from Normal and Malignant Mammary Epithelial Cells.*PLoS ONE*, *5*(10). http://doi.org/10.1371/journal.pone.0013515

proTRAC_documentation.pdf. (2015).

Raposo, G., Nijman, H. W., Stoorvogel, W., Liejendekker, R., Harding, C. V., Melief, C. J., & Geuze, H. J. (1996). B lymphocytes secrete antigen-presenting vesicles. *The Journal of Experimental Medicine*, *183*(3), 1161–1172. http://doi.org/10.1084/jem.183.3.1161

Recipes — cutadapt 1.9.dev0 documentation. (2009). Retrieved March 21, 2015, from http://cutadapt.readthedocs.org/en/latest/recipes.html

Reinhardt, T. A., Sacco, R. E., Nonnecke, B. J., & Lippolis, J. D. (2013). Bovine milk proteome: Quantitative changes in normal milk exosomes, milk fat globule membranes and whey proteomes resulting from Staphylococcus aureus mastitis. *Journal of Proteomics*, *82*, 141–154. http://doi.org/10.1016/j.jprot.2013.02.013

Resistance to mastitis.(n.d.). Retrieved April 9, 2015, from http://ansci.illinois.edu/static/ansc438/ mastitis/resistance.html

Riedmann, L. T., & Schwentner, R. (2010).miRNA, siRNA, piRNA and Argonauts: News in small matters. *RNA Biology*, *7*(2), 133–139. http://doi.org/10.4161/rna.7.2.11288

Robert, V. J. P. (2005). Chromatin and RNAi factors protect the C. elegans germline against repetitive sequences. *Genes & Development*, *19*(7), 782–787. http://doi.org/10.1101/gad.332305

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010).edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1), 139–140. http://doi.org/10.1093/bioinformatics/btp616

Roovers, E. F., Rosenkranz, D., Mahdipour, M., Han, C.-T., He, N., Chuva de Sousa Lopes, S. M., Ketting, R. F. (2015). Piwi Proteins and piRNAs in Mammalian Oocytes and Early Embryos. *Cell Reports*, *10*(12), 2069–2082. http://doi.org/10.1016/j.celrep.2015.02.062

Rosenkranz, D., & Zischler, H. (2012).proTRAC-a software for probabilistic piRNA cluster detection, visualization and analysis. *BMC Bioinformatics*, *13*(1), 5.

Rozhkov, N. V., Aravin, A. A., Zelentsova, E. S., Schostak, N. G., Sachidanandam, R., McCombie, W. R., … Evgen'ev, M. B. (2010). Small RNA-based silencing strategies for transposons in the process of invading Drosophila species.*RNA*, *16*(8), 1634–1645. http://doi.org/10.1261/rna.2217810

Ruffalo, M., LaFramboise, T., & Koyutürk, M. (2011). Comparative analysis of algorithms for next-generation sequencing read alignment. *Bioinformatics*, *27*(20), 2790–2796.

Sai Lakshmi, S., & Agrawal, S. (2007).piRNABank: a web resource on classified and clustered Piwi-interacting RNAs. *Nucleic Acids Research*, *36*(Database), D173–D177. http://doi.org/10.1093/nar/gkm696

Samji, T. (2009). PIWI, piRNAs, and Germline Stem Cells: What's the link? *The Yale Journal of Biology and Medicine*, *82*(3), 121–124.

Sapetschnig, A., & Miska, E. A. (2014).Getting a Grip on piRNA Cluster Transcription.*Cell*, *157*(6), 1253–1254. http://doi.org/10.1016/j.cell.2014.05.022

Schibler, L., Roig, A., Mahe, M.-F., Laurent, P., Hayes, H., Rodolphe, F., & Cribiu, E. P. (2006). High-resolution comparative mapping among man, cattle and mouse suggests a role for repeat sequences in mammalian genome evolution. *BMC Genomics*, *7*(1), 194.

Senti, K.-A., & Brennecke, J. (2010). The piRNA pathway: a fly's perspective on the guardian of the genome. *Trends in Genetics*, *26*(12), 499–509. http://doi.org/10.1016/j.tig.2010.08.007

Shittu, A., Abdullahi, J., Jibril, A., Mohammed, A. A., & Fasina, F. O. (2012).Sub-clinical mastitis and associated risk factors on lactating cows in the Savannah Region of Nigeria.*BMC Veterinary Research*, *8*(1), 134.

Siomi, M. C., Mannen, T., & Siomi, H. (2010). How does the Royal Family of Tudor rule the PIWI-interacting RNA pathway? *Genes & Development*, *24*(7), 636–646. http://doi.org/10.1101/gad.1899210

Sordillo, L. M., & Streicher, K. L. (2002).Mammary gland immunity and mastitis susceptibility. *Journal of Mammary Gland Biology and Neoplasia*, *7*(2), 135–146.

Sun, J., Zhang, B., Lan, X., Zhang, C., Lei, C., & Chen, H. (2014). Comparative Transcriptome Analysis Reveals Significant Differences in MicroRNA Expression and Their Target Genes between Adipose and Muscular Tissues in Cattle. *PLoS ONE*, *9*(7), e102142. http://doi.org/10.1371/journal.pone.0102142

Szakmary, A., Reedy, M., Qi, H., & Lin, H. (2009). The Yb protein defines a novel organelle and regulates male germline stem cell self-renewal in Drosophila melanogaster. *The Journal of Cell Biology*, *185*(4), 613–627. http://doi.org/10.1083/jcb.200903034

Tedder, P., Zubko, E., Westhead, D. R., & Meyer, P. (2009). Small RNA analysis in Petunia hybrida identifies unusual tissue-specific expression patterns of conserved miRNAs and of a 24mer RNA. *RNA*, *15*(6), 1012–1020. http://doi.org/10.1261/rna.1517209

The Galaxy Project: Online bioinformatics analysis for everyone. (2009). Retrieved April 12, 2015, from http://galaxyproject.org/

Thomson, T., & Lin, H. (2009). The Biogenesis and Function of PIWI Proteins and piRNAs: Progress and Prospect. *Annual Review of Cell and Developmental Biology*, *25*(1), 355–376. http://doi.org/10.1146/annurev.cellbio.24.110707.175327

Tijsterman, M., & Plasterk, R. H. (2004). Dicers at RISC: the mechanism of RNAi. *Cell*, *117*(1), 1–3.

truseq_smallrna_sampleprep_guide_15004197_e.pdf. (2012-a).

truseq_smallrna_sampleprep_guide_15004197_e.pdf. (2012.-b).

Turcatel, G., Rubin, N., El-Hashash, A., & Warburton, D. (2012).MIR-99a and MIR-99b Modulate TGF-β Induced Epithelial to Mesenchymal Plasticity in Normal Murine Mammary Gland Cells.*PLoS ONE*, *7*(1), e31032. http://doi.org/10.1371/journal.pone.0031032

Twine, N. A., Janitz, K., Wilkins, M. R., & Janitz, M. (2011). Whole Transcriptome Sequencing Reveals Gene Expression and Splicing Differences in Brain Regions Affected by Alzheimer's Disease. *PLoS ONE*, *6*(1), e16266. http://doi.org/10.1371/journal.pone.0016266

User guide — cutadapt 1.8.1 documentation. (2009). Retrieved April 12, 2015, from http://cutadapt.readthedocs.org/en/latest/guide.html#trimming-reads

Valadi, H., Ekstrom, K., Bossios, A., Sjostrand, M., Lee, J. J., & Lotvall, J. O. (2007). Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. *Nat Cell Biol*, *9*(6), 654–659. http://doi.org/10.1038/ncb1596

Viguier, C., Arora, S., Gilmartin, N., Welbeck, K., & O'Kennedy, R. (2009). mastitis detection: current trends and future perspectives. *Trends in Biotechnology*, *27*(8), 486–493. http://doi.org/10.1016/j.tibtech.2009.05.004

Wang, D., Wang, J., Lu, M., Song, F., & Cui, Q. (2010). Inferring the human microRNA functional similarity and functional network based on microRNA-associated diseases. *Bioinformatics*, *26*(13), 1644–1650. http://doi.org/10.1093/bioinformatics/btq241

Wang, L., Feng, Z., Wang, X., Wang, X., & Zhang, X. (2010). DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics*, *26*(1), 136–138.

Wang, M., Moisá, S., Khan, M. J., Wang, J., Bu, D., & Loor, J. J. (2012). MicroRNA expression patterns in the bovine mammary gland are affected by stage of lactation. *Journal of Dairy Science*, *95*(11), 6529–6535. http://doi.org/10.3168/jds.2012-5748

Warden, C. D., Yuan, Y.-C., & Wu, X. (2013).Optimal calculation of RNA-seq fold-change values. *International Journal of Computational Bioinformatics and In Silico Modeling*, *2*, 285–292.

Weber, J. A., Baxter, D. H., Zhang, S., Huang, D. Y., Huang, K. H., Lee, M. J., Wang, K. (2010). The MicroRNA Spectrum in 12 Body Fluids. *Clinical Chemistry*, *56*(11), 1733–1741. http://doi.org/10.1373/clinchem.2010.147405

Weick, E.-M., & Miska, E. A. (2014).piRNAs: from biogenesis to function. *Development*, *141*(18), 3458–3471. http://doi.org/10.1242/dev.094037

Xu, M., You, Y., Hunsicker, P., Hori, T., Small, C., Griswold, M. D., & Hecht, N. B. (2008). Mice Deficient for a Small Cluster of Piwi-Interacting RNAs Implicate Piwi-Interacting RNAs in Transposon Control. *Biology of Reproduction*, *79*(1), 51–57. http://doi.org/10.1095/biolreprod.108.068072

Yamanaka, S., Siomi, M. C., & Siomi, H. (2014).piRNA clusters and open chromatin structure. *Mobile DNA*, *5*, 22. http://doi.org/10.1186/1759-8753-5-22

Ying, S.-Y., Chang, D. C., & Lin, S.-L. (2008). The MicroRNA (miRNA): Overview of the RNA Genes that Modulate Gene Function. *Molecular Biotechnology*, *38*(3), 257–268. http://doi.org/10.1007/s12033-007-9013-8

Zamore, P. D. (2010). Somatic piRNA biogenesis. *The EMBO Journal*, *29*(19), 3219–3221. http://doi.org/10.1038/emboj.2010.232

Zen, K., & Zhang, C.-Y. (2012). Circulating MicroRNAs: a novel class of biomarkers to diagnose and monitor human cancers: Circulating MicroRNAs. *Medicinal Research Reviews*, *32*(2), 326–348. http://doi.org/10.1002/med.20215

Zhang, D., Xiong, H., Shan, J., Xia, X., & Trudeau, V. L. (2008). Functional insight into Maelstrom in the germline piRNA pathway: a unique domain homologous to the DnaQ-H 3'–5' exonuclease, its lineage-specific expansion/loss and evolutionarily active site switch. *Biology Direct*, *3*(1), 48. http://doi.org/10.1186/1745-6150-3-48

Zhou, Q., Li, M., Wang, X., Li, Q., Wang, T., Zhu, Q., Li, X. (2012). Immune-related microRNAs are abundant in breast milk exosomes. *International Journal of Biological Sciences*, *8*(1), 118.

Biography

Kshama Aswath graduated with a Master's degree in Biology in 2001 from Wayne State University, Detroit, MI and has a Masters degree from NMKRV College for Women, Bangalore University, Karnataka, India, in 1998. She has held positions as a research associate at Johns Hopkins and National Cancer Institute and recently as a team lead in a bio-risk management project, at National Institutes for Health, Bethesda, MD. Ms. Kshama Aswath looks forward to a career in the genomics research utilizing her molecular biology and bioinformatics skills.