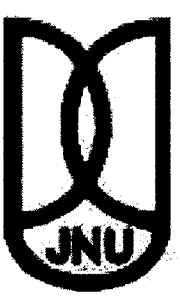
Deep sequencing based expression analysis for micro-RNA

Thesis submitted to Jawaharlal Nehru University in the partial fulfillment of the requirements for the award of

Master of Technology

Priyatama Pandey



Centre for Computational Biology and Bioinformatics
School of Computational and Integrative Sciences
Jawaharlal Nehru University
New Delhi - 110067, INDIA

May 2011

जवाहरलाल नेहरू विश्वविद्यालय
JAWAHARLAL NEHRU UNIVERSITY
संगणकीय एवं समेकित विज्ञान संस्थान
School of Computational and Integrative Sciences
नई दिल्ली- 110067
NEW DELHI- 110 067 (INDIA)

Hall No-6, Lecture Hall Complex, JNU
Tel. : (Direct) 26741517 (Off.) : 26704171
Fax : 011-26741586
Email : dean_sit@mail.jnu.ac.in

Declaration

I hereby declare that the work carried out in this thesis is entirely original. It has been carried out by me in the Center for Computational Biology and Bioinformatics, School of Computational and Integrative Sciences, Jawaharlal Nehru University, New Delhi. I further declare that it has not formed the basis for the award of any degree, diploma, membership or similar title of any University or Institution.

Priyatama Pandey

Priyatama Pandey
School of Computational & Integrative Sciences
Jawaharlal Nehru University, New Delhi

Supervisor

Rashi

Dr. Rashi Gupta
School of Computational & Integrative Sciences, Jawaharlal Nehru University
New Delhi

Supervisor

Alok Bhattacharya

Prof. Alok Bhattacharya
School of Life Sciences & School of Computational & Integrative Sciences
Jawaharlal Nehru University, New Delhi

Dean

Indira Ghosh

Prof. Indira Ghosh
School of Computational & Integrative Sciences,
Jawaharlal Nehru University
New Delhi

I would like to dedicate this thesis to my loving parents ...

Acknowledgment

With an overwhelming sense of pride and genuine obligation, I express my deepest regards to my supervisors Prof. Alok Bhattacharya and Dr. Rashi Gupta. They explore my research compatibility and provided their valuable guidance throughout my work. Prof. Alok made me to understand biological concepts, which was an unknown territory for me. Dr. Rashi Gupta explored and nurtured my statistical knowledge. Without their perspicuous comments, suggestions and erudite guidance, I would not be able to accomplish this work. Their patience to go through the draft meticulously has been incredible. Dr. Rashi has helped me a lot to improve my presentation and scientific writing skill through her comments and suggestions on this thesis.

I would also like to acknowledge Dean of our School, Prof. Indira Ghosh for motivating and providing us the best facilities for studies. The faculty of SC&IS have been very supportive and helped me whenever I approached them with a problem.

The love and support given by all dear seniors have been great source of strength. I would like to sincerely thank Mr. Sarbashis who helped me as adviser and friend. He was always there to clear my smallest doubts and answer my queries promptly. I am also highly thankful to Ms. Candida for her kind guidance and improving my knowledge on miRNAs. I would also like to thank my colleague Ravi for his kind help in learning LATEX and positive support.

The SC&IS office staff have been most cooperative and have been very helpful during past two years. I acknowledge the financial support from DBT.

A million thanks to my parents for being very supportive and giving freedom to pursue the M. Tech. in Computational and Systems Biology. They have always encouraged me and given opportunity to fulfill my dreams.

Above all I am very grateful to almighty God who brought me to this educational platform.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 The Central Dogma of Molecular Biology

In central dogma of genetics, first of all DNA is transcribed to RNA and then RNA is translated to protein. Protein is never translated back into RNA or DNA. DNA can not created from RNA and neither directly translated to protein. Hence it is unidirectional process in general as shown in (Figure 1.1).

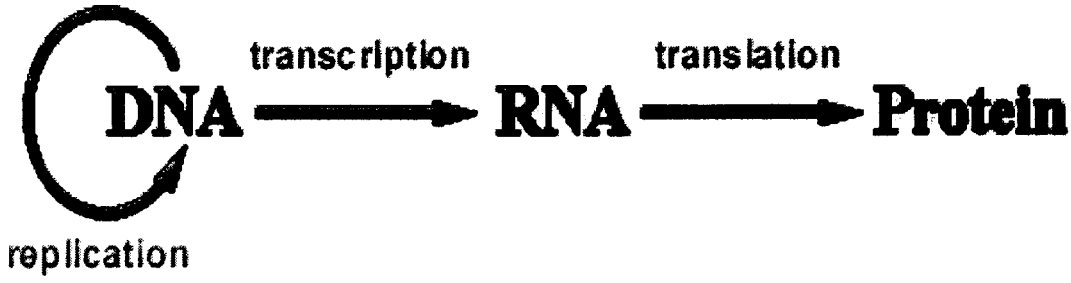


Figure 1.1: Central Dogma of life

The process of the transcription of DNA to RNA to protein is a central dogma. This dogma form a backbone of molecular biology and this is represented by four major stages

- The DNA replicates its information in a process with the help of enzyme called DNA polymerase. This enzyme works in collaboration with several others to copy the strand of DNA.
- In the process of transcription, a protein called RNA polymerase bind to the promoter region of the DNA and start copying the DNA message to make a messenger RNA (mRNA) molecule.
- The mRNA is processed i.e., exons are joint together and introns are removed by the process of splicing in the eukaryotic cells and mRNA migrates from the nucleus to cytoplasm.
- Messenger RNA carries information to ribosomes. Ribosomes read the information and use this information for the protein synthesis. This process is called translation.

  Proteins do not function for the production of protein, RNA or DNA. They are involved in almost all biological, structural or enzymatic activities. In human genome there are about 20,000-25,000 genes and almost 98% part of DNA is composed of non-coding regions.

1.2 Summary of current understanding of the role of RNA

RNA is single-stranded molecule in many biological roles and also has a short sequence of nucleotides. Each nucleotide consists of nitrogenous base, a ribose sugar and a phosphate group. RNA has the nucleobase uracil(U) instead of thymine(T), which is unmethylated form of thymine. DNA contain deoxyribose i.e., no hydroxyl group attached to the pentose ring in 2' position while RNA contains ribose. So RNA is less stable than DNA because of

the more inclined to hydrolysis. mRNA and rRNA are very important to translation of new proteins, the process can not occur without transfer RNA (tRNA).

There are excess of other reactions that happens after transcription. Such reactions that modify the transcripts are called post-transcriptional modifications. For example: poly(A) tail and 5' cap are added to eukaryotic pre-mRNA and introns are removed by the splicing process and only remaining exons are used as a template for protein synthesis after splicing.

Alternative splicing is a mechanism that select different combination of exons from the RNA and therefore it produces different proteins which greatly diversifies the transcriptome.

1.3 Standard non coding RNA

The ncRNA [1] is commonly the RNA that does not encode a protein but these RNAs contain information and take part in many functions. These non-coding RNAs involved in gene regulation, RNA processing and other roles. These RNAs determine most of our complex characteristics, play a significant role in disease and construct an unexplored world of genetic variation both within and between species.

In eukaryotes, non-coding RNA comes in several varieties:

1. **transfer RNA (tRNA) & ribosomal RNA (rRNA):**

   Ribosomal RNA, RNA component of ribosome, provides a mechanism for decoding mRNA into amino acid. It involved in peptide bond formation and interacts with tRNA during translation. tRNA carries the appropriate amino acids into the ribosome and transfers it to a growing polypeptide chain in translation. Most types of cells possess approximately 30 to 40 different tRNAs, with more than one tRNA corresponding to each amino acid. Hence these two RNAs play a critical role in the translation process.

2. **mRNA:**

Messenger RNA (mRNA) is the RNA that carries information from DNA to protein synthesis in the cell. Messenger RNA is essentially a copy of a section of DNA and serves as a template for the manufacture of one or more proteins. In eukaryotes, when RNA is first transcribed from DNA, it contain additional non-coding sequences that are interspersed within the coding sequence. This immature RNA molecule is referred to as precursor mRNA (pre-mRNA). The intervening non-coding sequences are called introns, and the segments of coding are known as exons. The introns are then removed by a process known as RNA splicing to produce the mature mRNA molecule.

3. **small ncRNA:**

small non-coding RNA has been classified in different subcategories like miRNAs, siRNAs, snoRNAs, piRNAs, snRNAs. SnRNAs are found in nucleus & play a crucial role in gene regulation by way of RNA splicing. miRNAs are approximately 20 to 26 nucleotides in length. They have been found in many species including flies, mice, and humans and inhibit gene regulation by repressing translation. miRNAs also play significant roles in cancer. Small interfering RNAs (siRNAs) are are only 21 to 25 base pairs in length, they also work to inhibit gene expression. Specifically, one strand of a double-stranded siRNA molecule can be incorporated into a complex called RISC. This RNA-containing complex can then inhibit transcription of an mRNA molecule. Small nucleolar RNAs (snoRNAs) were isolated from nucleolar extracts because of their abundance in this structure and guide chemical modification like methylation of other RNAs (See Figure 1.2)[2].

4. **large ncRNA:**

These non-protein coding transcripts are generally longer than 200 nucleotides. Long ncRNAs are located and transcribed within the intergenic sequences. It includes for example Xist, which direct the inactivation of a X-chromosome in female placental mammals. Long ncRNAs are noticed to be potential contributer in finding disease origin.

BOX 1 SMALL RNAS

Several types of small RNAs (sRNAs) have been identified (**Table 1**). Biogenesis of these small RNAs is for the most part dependent on the RNase III-type enzyme Dicer. Dicer processing results in small RNA products harboring 5′ phosphates and 3′ two-nucleotide overhangs on each strand. siRNAs, miRNAs, tasiRNAs, tncRNAs and scnRNAs are all dependent on Dicer for biogenesis. However, two classes of small RNAs, rasiRNAs and piRNAs, seem to be formed independently of Dicer processing.

Table 1 Classes of small RNAs

Class	Description	Size (nt)	Observed in	Mode of action	Biogenesis
siRNA	Small interfering RNA	~20–24	Mammals, At, Dm, Sp, Ce	Tend to have perfect complementarity to their mRNA target. Can be found dispersed throughout the entire message. Result in mRNA degradation	Dicer processing of long dsRNA[81] (Fig. 1a)
miRNA	MicroRNA	~20–24	Mammals, At, Dm, Sp, Ce	Mammalian miRNAs tend to contain mismatches, usually target untranslated regions of target mRNAs and result in translational suppression. Certain mammalian miRNAs can also facilitate mRNA degradation. Plant miRNAs tend to have perfect complementarity to their target and facilitate message degradation	Two-step processing of primary miRNA transcripts (pri-miRNA), first by the microprocessor complex[82,83] and then by Dicer[84–86] (Fig. 1b)
tasiRNA	*trans*-acting siRNA	~24	At, Dm, Sp	Act like siRNAs facilitating mRNA degradation[87]. However, tasiRNAs act in *trans*, targeting transcripts of genes other than the gene they are derived from	Dicer processing of RdRP-derived long dsRNA[87]
tncRNA	Tiny noncoding RNA	~20–21	Ce	Like tasiRNAs, tncRNAs also act in *trans*, resulting in message degradation[88,89]. May also facilitate translational suppression	tncRNAs are derived from noncoding sequences via Dicer processing in the nematode worm[88,89]
scnRNA	Small scan RNA	~28	Tt	Function in DNA elimination	Dicer-dependent small RNAs[90–92]
rasiRNA	Repeat-associated small interfering RNA	~24–29	At, Dm, Sp, Ce	Result in transcriptional silencing via chromatin remodeling (see Slicing and TGS)	Biogenesis is unclear, but rasiRNAs match repetitive sequence elements[13–22]
piRNA	Piwi-interacting RNA	~26–31	Mammals	Specifically expressed in the germ line, where they function in gametogenesis[53–57]. Mode of action is unclear	Biogenesis unclear, likely through processing of transcripts

At, *A. thaliana*; Dm, *D. melanogaster*; Sp, *S. pombe*; Ce, *C. elegans*; Tt, *T. thermophila*; RdRP, RNA-dependent RNA polymerase; dsRNA, double-stranded RNA.

Figure 1.2: Classes of small RNA

1.4 Methods for identification of RNA molecules

There are many methods for identifying RNA molecules like comparative sequence analysis method, sequencing methods and others.

1.4.1 Sequencing

Sequencing is the process by which we can measure and determine the primary sequence (or structure) of the molecule. RNA sequencing is the process of determining the order of nucleotides of the RNA fragment.

The most popular method for DNA sequencing is Dideoxy method or Sanger method. The dideoxy method gets its name because of the critical role played by these synthetic nucleotides (Dideoxynucleotides). Dideoxynucleotides are essentially the same as nucleotides except they contain a hydrogen group on the 3 carbon instead of a hydroxyl group (OH). When it added in the DNA strand, chain elongation stop because there is no 3'-OH for the next nucleotide to be attached. Hence this method is also called the chain termination method.

1.4.2 Modern approach for sequencing using Next Generation Sequencing

Recently, several sequencing platform has been developed that have very high throughput and low cost comparing to the traditional Sanger sequencing technology. Therefore, the high demand for low-cost sequencing has driven the development of high-throughput sequencing technologies [3]. Next-gen sequencing technology also known as deep sequencing approach, can sequence millions of short nucleotide sequences between 30 and 100 characters long. These sequences are generated by raw data in the form of noisy fluorescence intensity measurement. Next generation sequencing technology are able to convert these intensity measurement into discretized reads. So the process by which to identify a base sequence

from the fluorescence intensity trace data is called base calling.

Several different names are used to be these sequencing technology such as next-generation sequencing, massively parallel sequencing, second generation sequencing , ultra throughput sequencing. It can generate many hundreds of thousands or even millions of reads in a relatively short time.

While allowing powerful new applications like RNA-seq, the continually accelerating progress of technological change in the field of next generation sequencing also generate a glut of unused information.

1.4.3 Platforms for next generation sequencing

There are three platforms for massively parallel sequencing read production are use at present:

1. Roche/454 FLX
2. Applied Biosystems SOLiDTM System
3. Illumina/Solexa Genome Analyzer

Recently, another two massively parallel systems were announced:

- Helicos HeliscopeTM and
- Pacific Biosciences SMRT instruments

1. Roche/454 FLX Pyrosequencer

In Roche/454, the library fragments are mixed with beads whose surface carry oligonucleotides complimentary to the 454-specific adapter sequences on the fragment library, so each bead is associated with a single fragment. It involves Emulsion PCR [4] to

amplify the single stranded DNA from a fragment library and produces approximately one million copies of each DNA fragment on the surface of each bead. This strategy allows the 454 base-calling software to calibrate the light emitted by a single nucleotide incorporation. In the incorporation of a nucleotide during sequencing causes the release of pyrophosphate which initiates a series of downstream reactions that finally produce light by the firefly enzyme luciferase. The number of nucleotides incorporated is directly depend on the amount of light produced. The FLX instrument currently provides 100 flows of each nucleotide during an 8-h run, which produces an average read length of 250 nucleotides sequence. The resulting reads yield 100 Mb of quality data on average.

2. **Applied Biosystems SOLiDTM System**

This ligase-mediated sequencing approach of the Applied Biosystems SOLiD sequencer is based on the Emulsion PCR amplification similar to Roche. The DNA fragments for SOLiD sequencing are amplified on the surfaces of (1-m) small magnetic beads which provide enough signal during the sequencing reactions. Ligase-mediated sequencing begins by annealing a primer to the shared adapter sequences on each amplified fragment, and then DNA ligase is provided along with specific fluorescent-labeled 8mers, whose 4th and 5th bases are encoded by the attached fluorescent group. Each ligation step is followed by fluorescence detection, after which a regeneration step removes bases from the ligated 8mer including the fluorescent group and prepares the extended primer for next round of ligation. Since each fluorescent group on a ligated 8mer identifies a two-base combination, the resulting sequence reads can be screened for base-calling errors versus true polymorphisms versus single base deletions by aligning the individual reads to a known high-quality reference sequence.

3. Illumina Genome Analyzer

The Illumina sequencing-by-synthesis approach has following parts:

- Prepare genomic DNA sample
- Attach DNA to surface
- Bridge amplification
- Denature the double stranded molecules

The flow cell of the illumina system is an 8-channel sealed glass micro fabricated device that allow bridge amplification of DNA fragments on its surface and it uses DNA polymerase to produce multiple copies of a single DNA fragment, which is required to observed the signal intensity detection during sequencing. It is capable of sequencing different samples at the same time.

The Illumina system utilizes a sequencing-by-synthesis approach in which all four fluorescently labeled, 3 -OH blocked nucleotides are simultaneously added to the flow cell along with DNA polymerase. After the base incorporation step the fluorescent image of the flow cell is captured by CCD camera. An imaging step follows each base incorporation step. After each step, the 3' blocking group containing the fluorescent dye is chemically removed to prepare each strand for the next incorporation by DNA polymerase. It continue for specific number of steps and removing poor quality sequences and this series of steps permits the discrete reads the read length to be around 25 to 35 nucleotides and 120 nucleotides has come by the latest models. The base calling algorithm assign sequences and associates a quality score to each read. It remove the poor quality sequences of the illumina data in each run by the quality checking pipe line [5].

1.5 Applications of Next generation sequencing

NGS has application like in whole Genome Sequencing, Exome Sequencing, mRNA (transcriptome) Sequencing. RNA-sequencing to profile the mammalian transcriptome as well as whole human genome sequencing,single nucleotide polymorphism (SNP)-based association studies, Sequence capture, Resequencing, de novo sequencing, miRNA and Small RNAs sequencing, ChIP Sequencing, customized applications, and others. We explain some of its applications below:

1.5.1 RNA-seq

RNA-Seq is perhaps the most complex NGS application. Expression levels of specific genes, differential splicing, allele-specific expression of transcripts can be accurately determined by RNA-Seq experiments to address many biological-related issues. RNA-seq refers to the use of next generation sequencing technology to study of transcriptome. It generates millions of short reads of mRNA or cDNA. These short reads are mapped to the genome, resulting in a sequence of the read counts along with genomic positions. RNA-seq has higher throughput and low noise comparing to the other RNA measuring technologies such as qPCR and microarrays. RNA-seq [6] can measure the expression of millions of genes in a single experiment and it takes a few days only. It also generates digital results. Although microarray have been widely used to explore the gene expression but this experiment only generate analog results that leads to the unavoidable problems such as non-specific background noise and cross hybridization, which interrupt the transcriptome analysis. As a result data from microarray platform is sensitive to sample preparation, array platform and lab protocols. In contrast, RNA-Seq exhibits a high level of reproducibility and overcomes these limitations of microarrays.

1.5.2 Discovering and Identifying non coding RNAs

Non-coding RNAs system in different organism is very divers. So the Discovery of non-coding RNAs [7] is difficult to measure by computational methods alone. The read comes by NGS platforms is quantitative and pointing out the expression level of non-coding RNAs. This is detecting expression level changes and obtaining significant insights into their biology.

1.5.3 Illustration of DNA -Protein interaction through Chromatin Immunoprecipitation (ChIP) sequencing

The interaction between DNA and protein play a very important role in the regulation of gene expression. These interaction can be studied by technique called chromatin immunoprecipitation (ChIP) [8]. A verity of methods such qPCR (quantitative PCR)[9] or Southern blotting[10], DNA microarray are used to identified and quantified the captured DNA fragments population. Many drawbacks including identification of novel fragments by low signal to noise ratio can now resolved by NGS platforms. For example Illumina platform were used for analysis of transcription factor binding sites in the human genome [11].

1.6 MicroRNA (miRNA)

MicroRNA are about 22 nucleotides long, short non-coding RNA's that play a critical role in gene expression during many essential cellular function. MicroRNA are single stranded and the most conserved and prominent among all the non-coding small RNA molecule. The genes encoding miRNAs are much longer than the processed mature MicroRNA was first identified in Caenorhabditis elegans and subsequently found in almost all eukaryotes [12]. Recently, next-generation sequencing technology has been developed and widely applied to genomic studies such as gene expression pattern analysis, genome sequencing and small RNA sequencing. Because of its ultra high-throughput, many new miRNAs with low abundance

could be identified using this technology.

Studying miRNAs is difficult because each miRNA can target 100s of mRNAs directly or indirectly and more than 1 miRNA can converge on one specific mRNA target.

Under normal cellular control, miRNAs are responsible in many biological processes including cellular proliferation and differentiation. miRNAs have been shown to repress the expression of important cancer-related genes and might prove useful in the diagnosis and treatment of cancer. In contrast research, it has been proven that when control mechanisms go wrong, an abnormal amount of miRNA can cause a variety of cancers including pancreatic cancer, thyroid cancer, hepatocellular cancers, breast cancer, lung cancer, colon cancer, and more.

1.7 Biogenesis of microRNA

MicroRNAs are transcribed as RNA polymerase II and their primary transcripts or pri-miRNAs with a cap and poly A tail. These Pri-miRNAs are then cropped by the RNase III enzyme Drosha and its cofactor DGCR8/pasha and processed to short, 70 nucleotide stem loop structures known as precursor-miRNAs (pre-miRNAs). These pre-miRNAs are then processed to mature miRNAs in the cytoplasm by interaction with the endonuclease Dicer, which also initiate the formation of RNA-induced silencing complex (RISC).

When Dicer cleaves the pre-miRNA stem-loop, two complementary short RNA strand are formed. Although both strands of duplexes are necessarily produced in equal amounts by transcription. Based on the thermodynamic stability of each end of this duplex, one of the strands is thought to be a biologically active miRNA or mature miRNA or guide strand, and the other is considered as an inactive strand and a carrier strand called miRNA* (miRNA star) or passenger strand or anti-guide (See Figure 1.3).

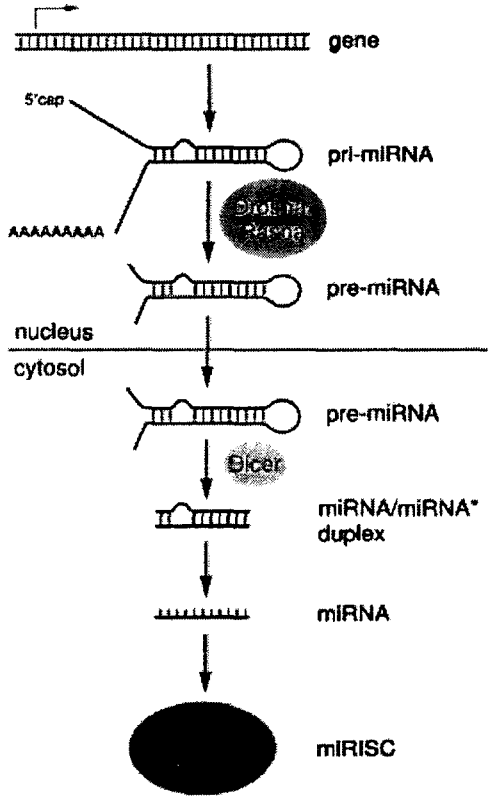


Figure 1.3: microRNA (miRNA) is produce from the precursor-microRNA (pre-miRNA) ,which is formed from a microRNA primary transcript (pri-miRNA).

Generally, the miRNA* strand is typically degraded, whereas the mature miRNA strand is incorporated into the effector complexes, which are known as miRNP, mirgonaute or, more generally, miRISC (miRNA-containing RNA-induced silencing complex).

MicroRNAs are less than one-thousandth the size of an average mRNA molecule, usually coming in at around 22 nucleotides in length. Also, they are non-coding, which means that unlike mRNA, they do not code for a protein. Instead, these RNAs bind to complementary sequences on the 3' untranslated region (UTR) of target messenger RNAs. This can cause mRNA degradation, which has the effect of repressing translation and therefore controlling the expression of a particular gene.

1.8 IsomiRs/Variability in microRNA

The small RNAs that developed from common precursors were grouped into a cluster (See Figure 1.4). Generally the most abundant member is characterized as mature miRNA. The sRNAs that differed from the most abundant mature miRNA by few nucleotides are called IsomiRs. The variability in isomiRs can be observed by variability in either drosa or dicer1 cleavage position within the pre-miRNA hairpin [13].

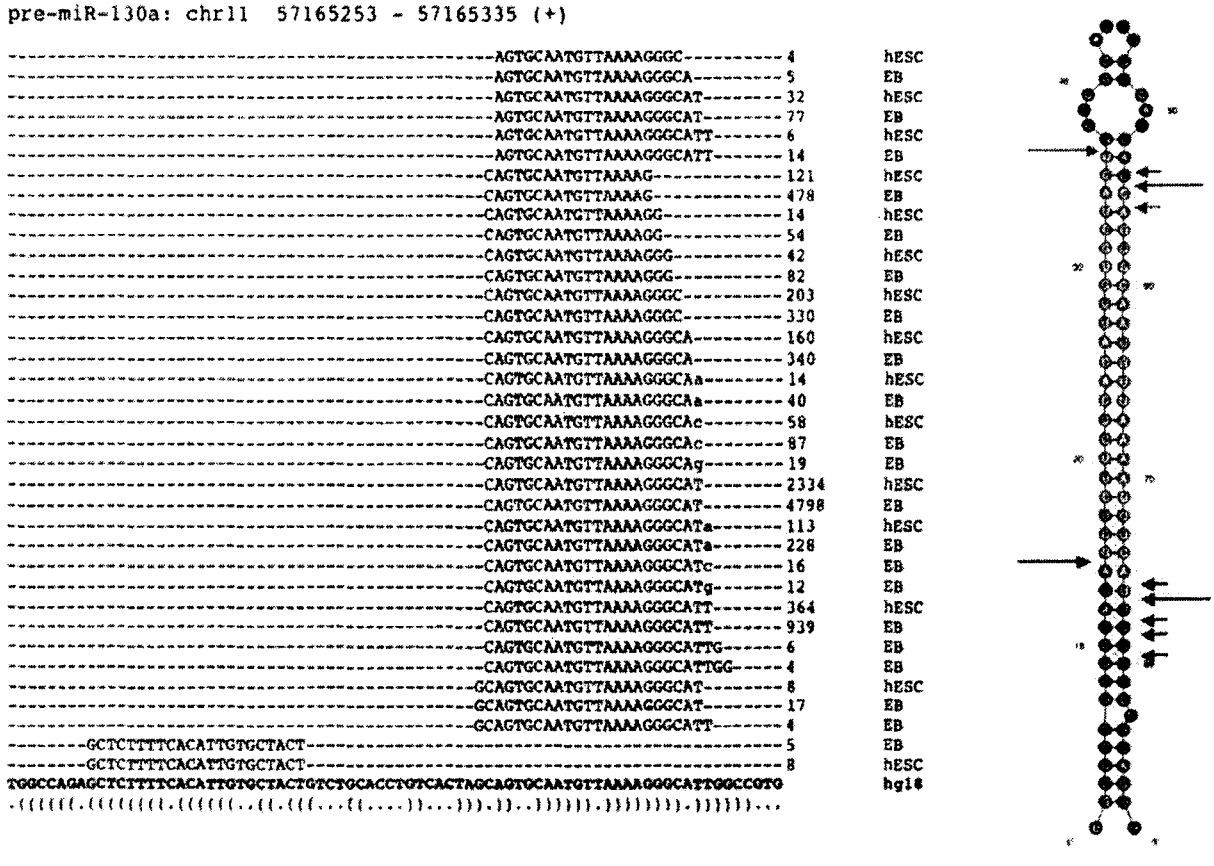


Figure 1.4: Blue sequence is mature miRNA that shows the most abundant sequences and other sequences shows the variability by few nucleotide, called isomiRs. Although, variation at the 3' end is generally much more common than at the 5' end. Sequences representing the miRNA*(star) were observed (highlighted in purple).

1.9 Objectives

1. Many methods have been proposed for normalization of sequence count data. However, different data sets have different normalization requirement. So my objective is to **"test several normalization methods and detect differential expression using R tools"**.

2. I also extend my work for the **"expression analysis of isomiRs from small RNA sequence data"**.

Chapter 2

Methods and Material

2.1 Cell line and Blood samples

The human chronic myeloid leukemia blast crisis cell line, K562 was obtained from National Centre for Cell Sciences, Pune and maintained in RPMI 1640 and DMEM (Gibco) medium, respectively. The medium was supplemented with 10% FBS and penicillin-streptomycin and maintained at 37C with 5% CO_2 in incubator chamber. Buffy coat of healthy blood donors (N4 and N5) were collected from volunteers. Red cell lysis buffer (0.144M NH4Cl, 0.01M NH4HCO3) was added to buffy coat to lyse the remnant RBCs and pure WBC population was obtained by centrifugate at 3000 g [14].

2.2 Data Set used

The sRNA sequencing data containing of peripheral blood leukocytes of two normal individuals (N4, N5), two patients (P1, P2) and tumor cell line K562 were obtained from Illumina high throughput sequencing platform.

Sample	Read
N4	587061
N5	595319
P1	644200
P2	713786
K562	1040348
Annotated sequence(miRNAs registry, release 16)	**Reference Sequence**
Mature miRNA	1223 (1032 mature +191 star)
Precursor miRNA	1048

Table 2.1: Data Sets

2.3 How data from NGS looks like

1. **Data set into fasta format:** The data file into fasta format where the unique header (sequence identity) reserved the information of the sequence length and frequency of that sequence. The sequence ID contained of a running number along with the length and abundance of each individual sequence.

```
>1.22.468414
TGAGGTAGTAGGTTGTGTGGTT
>2.23.316313
TACCACAGGGTAGAACCACGGAC
>3.22.251538
TACCACAGGGTAGAACCACGGA
>4.23.239658
TACCACAGGGTAGAACCACGGAA
```

2. **Data set into Fastq format:** Converted the above fasta format data file into fastq format using a perl code of Maq software.

```
@1.22.468414
TGAGGTAGTAGGTTGTGTGGTT
+
::::::::::::::::::::::
@2.23.316313
TACCACAGGGTAGAACCACGGAC
+
:::::::::::::::::::::::
@3.22.251538
TACCACAGGGTAGAACCACGGA
+
::::::::::::::::::::::
```

3. **Data set into Count format:** Converted the data into digital gene expression format using perl code that was developed in our lab.

miRNA	P1	P2	N4	N5
hsa-let-7a	64475	91829	100473	98863
hsa-let-7b	531019	369116	468414	593030
hsa-let-7c	20225	14133	14163	11823
hsa-let-7d	14530	23794	45176	29141
hsa-let-7d*	166	227	344	267
hsa-let-7e	440	2610	3175	1342

2.4 Alignment of reads using Bowtie software

Bowtie software [15], version 0.12.7, an ultra fast alignment software, has been used to align reads to the precursor miRNAs. Bowtie command for implementing this mapping strategy is:

⇒ **./bowtie hs-pre_mirna data.fastq output_result**

The first argument to bowtie is the basename of the index for the genome to be searched. The second argument is the name of a fastq file containing the reads and the last third one is the output file name. It takes few minute to generate the output depending on the computer memory. Bowtie gives many lines as output, each line is an alignment for a read.

Before implemented the above command, we first created a new index using command:

⇒ **./bowtie-build hsa-1048-pre_mirnas.txt hs-pre_mirna**

Its default outputs one alignment per line. Each line is a combination of some tab separated field from left to right, some of them are as follows:

1. Name of the aligned read
2. Reference strand aligned to, + for forward strand, - for reverse
3. Name of the reference sequence where alignment occurs, or numeric ID if no name was provided
4. 0-based offset into the forward reference strand where leftmost character of the alignment occurs
5. Read sequence (reverse-complimented if orientation is -)

There are some extra columns such as ASCII-encoded these are related to some phred score but those are not required in our experiments.

2.5 Normalization of data

As the total number of reads varies between lanes, read counts must be normalized to allow a systematic comparison across lanes and across samples. Scaling to library size is the most commonly used normalization for RNA-seq data sets. However, most sophisticated normalization procedures have recently been proposed. We evaluate some normalization procedures in here:

1. *Mortazavi et al.* [16] used RPKM methods - Adjust their counts to reads per kilo base per million mapped.
2. *Bolstad et al.* [17,18] used quantile normalization method.
3. *Robinson et al.* [19] used trimmed mean of M-values of normalization method (TMM).

2.6 Methods for Normalization

2.6.1 Trim Mean Value M normalization (TMM) method

TMM method has been suggested to remove RNA composition bias as a number of reads for a gene dependent upon not only on the gene expression's level and length but also depend on the population of RNA from which it originates. TMM equates the overall expression level of the genes between samples by estimation of relative RNA production levels or scaling factor. The TMM [19] method estimates scale factors between samples that can be incorporated into currently used statistical methods for DE analysis. The assumption behind TMM method is similar to the assumption of microarray normalization method that the expression levels of genes are not differentially expressed.

We equate the overall expression levels of genes between samples under the assumption that the majority of them are not DE. A weighted trimmed mean of the log expression ratios (trimmed mean of M values (TMM)) is the simple and robust way to estimate the ratio of RNA production. For sequencing data, we define the gene-wise log-fold-changes as:

$$Mg = log_2 \frac{(Y_{gk}/N_k)}{(Y_{gk'}/N_{k'})} \tag{2.1}$$

and absolute expression level as

$$Ag = \frac{1}{2} log_2((Y_{gk}/N_k) * (Y_{gk'}/N_{k'})) \tag{2.2}$$

Normalization factors across several samples can be calculated by selecting one sample as a reference and calculating the TMM factor for each non-reference sample.

A trimmed mean is the average after removing x% of upper and lower data. The TMM procedure is doubly trimmed, by log-fold-changes M_{gk}^r (sample k relative to sample r for gene g) and by absolute intensity (Ag). By default, we trimmed Mg value by 30% and Ag

value by 5%.

Therefore, the normalization factor for sample k using reference sample r is calculated as

$$log_2(TMM_k^r) = \frac{\sum w_{gk}^r M_{gk}^r}{w_{gk}^r} \tag{2.3}$$

where

$$M_{gk}^r = \frac{log_2(Y_{gk}/N_k)}{log_2(Y_{gk'}/N_{k'})} \; and \; w_{gk}^r = \frac{N_k - Y_{gk}}{N_k Y_{gk}} + \frac{N_r - Y_{gr}}{N_r Y_{gr}}$$

$$Y_{gk}, Y_{gr} > 0$$

The cases whereY_{gk} =0 andY_{gr} =0 are already trimmed for this calculation since log-fold-changes cannot be calculated. G* represent the untrimmed value of Mg and Ag. In the case where reference and sample are same then TMM value should be zero i.e., $TMM_k^r = 0$.

2.6.2 Quantile Normalization

The goal of the quantile method is to make the same distribution for each array in the set of arrays. This method is motivated by the quantile-quantile plot (qq-plot). The qq-plot shows that the distribution of two data vectors would be same if the plot is a straight diagonal, otherwise distribution would be different. This can be extended to n dimensions if all n data vectors have same distribution then plotting the quantiles in n dimensions gives a straight line along the line given by a unit vector $(1/\sqrt{n}, ..., 1/\sqrt{n})$.

Let $q_k = (q_{k_1}, ..., q_{k_n})$ for k = 1, . . . , p be the vector of the k^{th} quantiles for all n arrays $q_k = (q_{k_1}, ..., q_{k_n})$ and $d = (1/\sqrt{n}, ..., 1/\sqrt{n})$ be the unit diagonal. To transform from the quantiles so that they all lie along the diagonal, consider the projection of q onto d.

$$Proj\ \ d_{q_k} = (1/n \sum q_{k_j}, ..., 1/n \sum q_{k_j}) \tag{2.4}$$

The following algorithm for normalizing a set of data vectors for giving them the same distribution.

1. Given n array of length p, form X of dimension $p \times n$ where each array is a column.
2. Sort each column of X individually to give Xsort.
3. Take the mean across rows of Xsort and assign this mean to each element in the row to get X'sort.
4. Get X normalized by rearranging each column of the X'sort to have the same ordering as original X.

2.6.3 Transcripts Parts Per Million (TPM) based Normalization

There is a normalization method called tag per million or transcript per million (TPM). The number of reads of a transcript per sequence was divided by the total clone count of the sample and multiplied by 10^6. The total clone count is the sum of the frequencies of all the unique sequences per transcripts.

TPM according to the tag count in each library as follows:

$$TPM_{ji} = C_{ji} * 10^6 / T_j, \tag{2.5}$$

where TPM_{ji} is the TPM for transcript i in any library j,
C_{ji} is the count of transcript i in library j,
T_j is the total transcript count in library j.

2.7 Differential Expression of genes

Modern molecular biology data present major challenges for the statistical methods that are used to detect differential expression such as requirement of multiple testing procedures. For microarrays, the abundance of particular transcript is measured as a fluorescence intensity which is observed as continuous response, whereas for a digital gene expression data the abundance is observed as a count. Therefore we cannot use the same procedure directly for digital gene expression data that are successful for microarray data. There are many statistical methods to compare RNA-Seq data and identify differentially expressed genes.

2.8 Methods for Differential expression

There are many methods for identifying the differential gene expression between two samples. Some of the methods applicable for inferring differential expression and discussed in here are: Likelihood ratio test [20], T-Test [21].

2.8.1 T-statistics

Suppose that the data consist of measurements y_{gi} under two conditions, where i (=1, 2, . . . , k) represents the i^{th} array, g (=1, 2, . . . , G) denotes the g^{th} gene, and k_1 & k_2 are the number of arrays for each condition, that is, k = k_1 + k_2.

The two sample T-statistic with two independent normal samples without assuming the equal variances between two samples could be written as follows;

$$tg = \frac{diff}{Se_g}, \qquad Se_g = \sqrt{\frac{s^2_{g_1}}{k_1} + \frac{s^2_{g_2}}{k_2}} \tag{2.6}$$

Let the sample means and the sample variances of y_{gi}'s for gene g under two conditions be

denoted as $y_{g_1}, s^2_{g_1}$ and $y_{g_2}, s^2_{g_2}$ respectively. Here, *diff* is the difference between y_{g_1} & y_{g_2} , and s_g & Se_g represent the pooled standard deviation and the standard error of the diff across the replicates for the gene, respectively.

2.8.2 Likelihood based method

Let f(X,) be the density or probability mass function of a random sample of size n with variable $X = x_1, x_2, ..., x_n$ with unknown parameter $\theta_1, \theta_2, ..., \theta_n$ can assume. We want to test the null hypothesis:

$$H_0 : (\theta_1, \theta_2, ..., \theta_n) \epsilon \Theta$$

against the the alternative hypotheses of the type:

$$H_1 : (\theta_1, \theta_2, ..., \theta_n) \epsilon \Theta - \Theta_0$$

The likelihood function of the sample observation is given by

$$L = L(x_1, x_2, ..., x_n|\theta_1, \theta_2, \theta_3...., \theta_n) = \prod_{i=1}^{n} f(x_i; \theta_2, \theta_3...., \theta_n) \tag{2.7}$$

The likelihood ratio test is defined as the quotient of the two maxima (maximum values of the likelihood function for variation of the parameters in θ and θ_0 respectively) and is given by

$$\Lambda = \frac{L(\hat{\Theta_0})}{L(\hat{\Theta})} = \frac{Sup_{\theta \epsilon \Theta_0} L(x, \theta)}{Sup_{\theta \epsilon \Theta} L(x, \theta)}, \tag{2.8}$$

where $L(\hat{\Theta_0})$ and $L(\hat{\Theta})$ are the maxima of the likelihood function with respect to the parameters in the regions Θ_0 and Θ respectively.

2.9 Existing tools for analysis of RNA-seq data

Many tools have been developed to analyze the RNA-seq data. These tools are based on some statistical test or distribution to decide whether for a given gene, an observed difference

for a read count is significant. If the reads were independently sampled from the population, then read counts would follow such distribution, which can be approximated the Poisson distribution. Consequently, Poisson distribution has been used to test for differential expression. In the Poisson distribution, its mean and variance is defined by a single parameter (its variance is equal to mean). It is noted that the assumption of Poisson distribution is too tight because it ignores the extra variation due to actual differences in samples. It predicts smaller variation than what is seen in the data. Therefore resulting statistical test does not control type -I error. To remove this over dispersion problem, another distribution has been proposed to model count data with negative binomial (NB) distribution [22]. The negative binomial distribution have two parameters, which are uniquely determined by mean μ and variance σ^2.

2.9.1 edgeR

edgeR is an existing Bioconductor [23] software package for identifying differential expression of replicated count data. edgeR (empirical analysis of digital gene expression in R) is designed for the analysis of replicated count-based expression data and is an implementation of methodology developed by Robinson and Smyth [24].

This software is equally applicable for RNA-seq, Tag-seq, SAGE, CAGE, Illumina/Solexa, 454 or ABI Solid experiments. In fact this software may be useful in other experiments where counts are observed.

The edgeR [25], model the RNA-Seq data as a negative binomial (NB) distributed. This model is able to separate biological from technical variation by using different distributions depend on situation.

edgeR provides two ways of estimating the dispersion(s), the quantile-adjusted conditional maximum likelihood (qCML) method and the Cox-Reid profile-adjusted likelihood (CR) method. In general, we apply the qCML method to experiments with single factor and

the CR method to experiments with multiple factors. When negative binomial models are fitted and dispersion estimates are obtained, and then proceed with testing procedures for determining differential expression. provides two ways of testing differential expression, the exact test and the generalized linear model (GLM) likelihood ratio test.

edgeR available via Bioconductor and the home page of edgeR:

http://www.bioconductor.org/packages/release/bioc/html/edgeR.html

2.9.2 DESeq

DESeq is also an existing R package which analyzes discrete data from Next generation sequencing such as RNA-Seq and test for differential expression. DESeq [26] uses a negative-binomial distribution, with mean and variance linked by a local regression and was developed for the analysis of digital gene expression.

DESeq is available via Bioconductor and the home page of DESeq:

http://bioconductor.org/packages/release/bioc/html/DESeq.html

2.10 Normalization and Differential expression for Normal and Patient samples

We first excluded miRNAs whose frequencies or expressions were below than ten in all four samples and used DESeq and edgeR for carrying out the analysis for normalization and differential expression.

2.11 Identification of the known miRNA IsomiR clusters or families

As mentioned in the introduction section, isomiRs are variant forms of the miRNAs caused by alternative Dicer cutting [13]. To obtain the isomiRs for every known miRNA, an alignment of the reads to the pre-miRNA hairpins is necessary. This alignment facilitates identification of isomiRs of both the 5p and 3p (mature miRNA sequences derived from the 5' and the 3' region of the pre-miRNA hairpin) mature miRNAs (See Figure 1.4).

2.11.1 Alignment of the reads to the reference pre-miRNAs

The alignment was done by the Bowtie software. The reference pre-miRNAs were downloaded from miRBase release 16 comprising of 1048 pre-miRNAs and 1223 mature miRNA sequences [27]. Reads from each of the five samples were aligned to the reference pre-miRNAs. The output was an alignment of the deep sequencing reads to the Homo sapiens reference pre-miRNAs.

2.11.2 Obtaining clusters or families of the isomiRs

The alignment output was parsed by a perl scripts developed specifically to get a family or a cluster of IsomiRs, that are actually reads that match at the same location but differ by a few nucleotides. Such clusters were obtained for every miRNA from both the regions (5' and 3' region) of the hairpin using the position information given in the output of the alignment. Each cluster of isomiRs was then sorted according to their frequency or expression value, to get the most abundant to the rarest occurring isomiRs. Reads having lengths below seventeen nucleotides were removed.

2.11.3 Identification of the reference miRNA from the IsomiR cluster or family

The next step was to identify the reference mature miRNA from every existing isomiR cluster and to check if it was the most abundant sequence or not. A perl code was developed for this purpose [See appendix A for perl code].

2.12 Method for creating expression profile for IsomiRs

In order to determine the biologically significant differences among the Patient and the Normal samples it was necessary to identify the similar and differentially expressing isomiRs. This was done through the following steps:

1. The Bowtie alignment output, as explained in section 2.4 was used for each of the samples.
2. The frequency of each of the isomiRs of every known miRNA was obtained by parsing the output by a perl code [See appendix A]..
3. A list of each isomiR sequence of every known miRNA along with their respective frequencies was created for all the samples thus obtaining an IsomiR expression profile.

2.13 Normalization and Differential expression for IsomiRs

Using the isomiR expression profile, comparisons of the expression values of every IsomiR were done among the Normals and the Patient samples to identify the differentially expressed IsomiRs using the likelihood based method. These differentially expressed isomiRs were detected from the normalized data based on Quantile normalization methods already explained in section 2.6.2.

Chapter 3

Results and Discussion

3.1 Evaluation of normalization for normal and patient samples

Various normalization techniques have already been discussed in section 2.4 and 2.5. In here, we demonstrate the effect of various normalization on Normal-Patient dataset. Figure 3.1 displays the box-plot of un-normalized data, data after quantile normalization, TMM normalization and TPM normalization.

Implemented quantile normalization procedure for RNA Seq data is inspired from the microarrays. Quantile normalization method is a good procedure to reduce the non-biological errors from samples. In TMM method, estimated normalization factors should ensure that genes with the same expression level in two samples are not detected as DE. TPM method is just to normalize the sequence number by the total sequence count in each library.

edgeR software provides a multidimensional scaling (MDS) plot by which relation between samples, before and after normalization can be evaluated. Figure 3.2 provides MDS plot between the two normal and two patient samples before normalization and after TMM, RLE and Quantile normalization.

Variation within samples were also estimated by using DESeq software. Variation is calculated using simple mean and variance within the samples and shown by SCV plot (See Figure 3.3). So, it is observed that at low count, variation is high and at higher count it decreases.

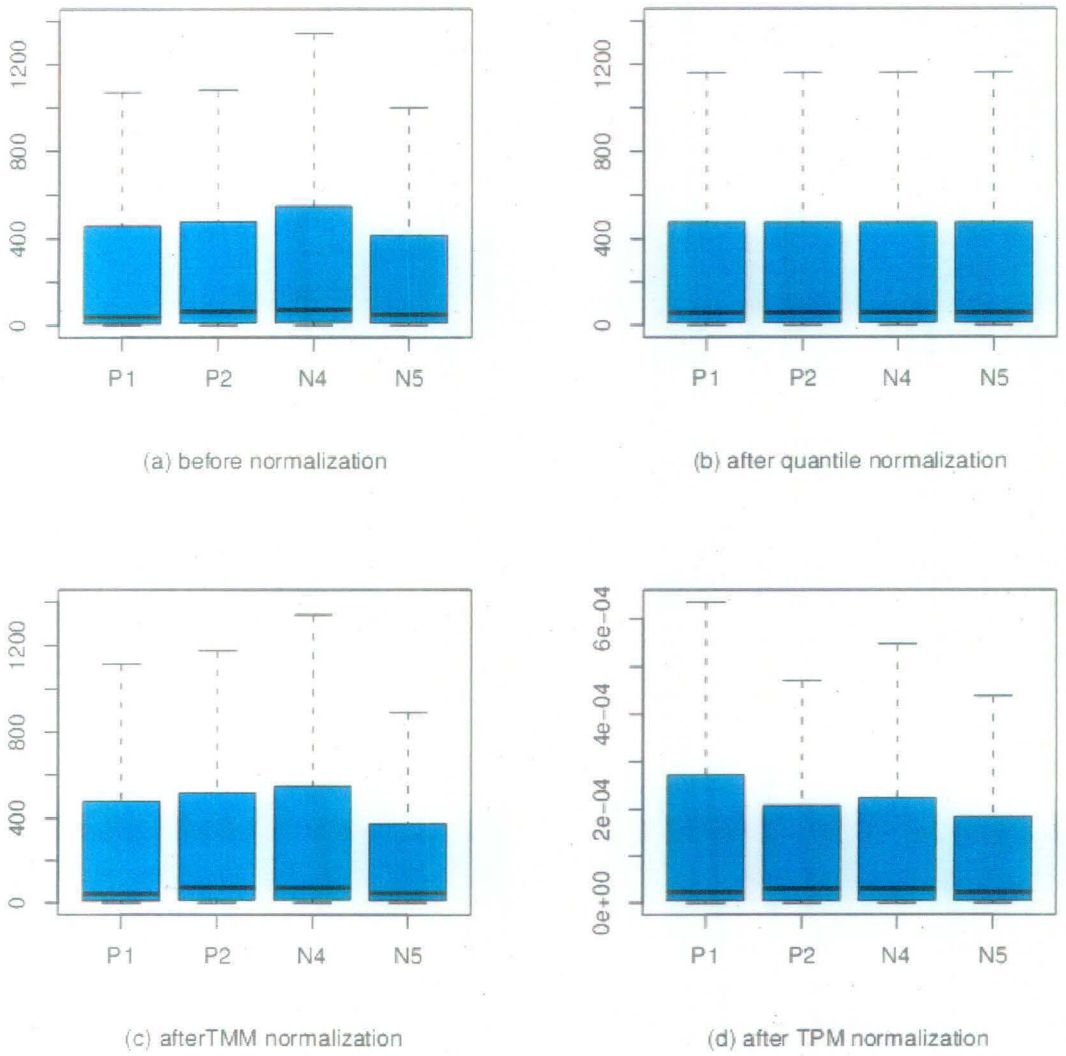


Figure 3.1: Boxplot to show the normalization effect on the normal (N4,N5) and patient (P1,P2) samples before and after using (b) quantile (c) TMM and (c) TPM normalization methods.

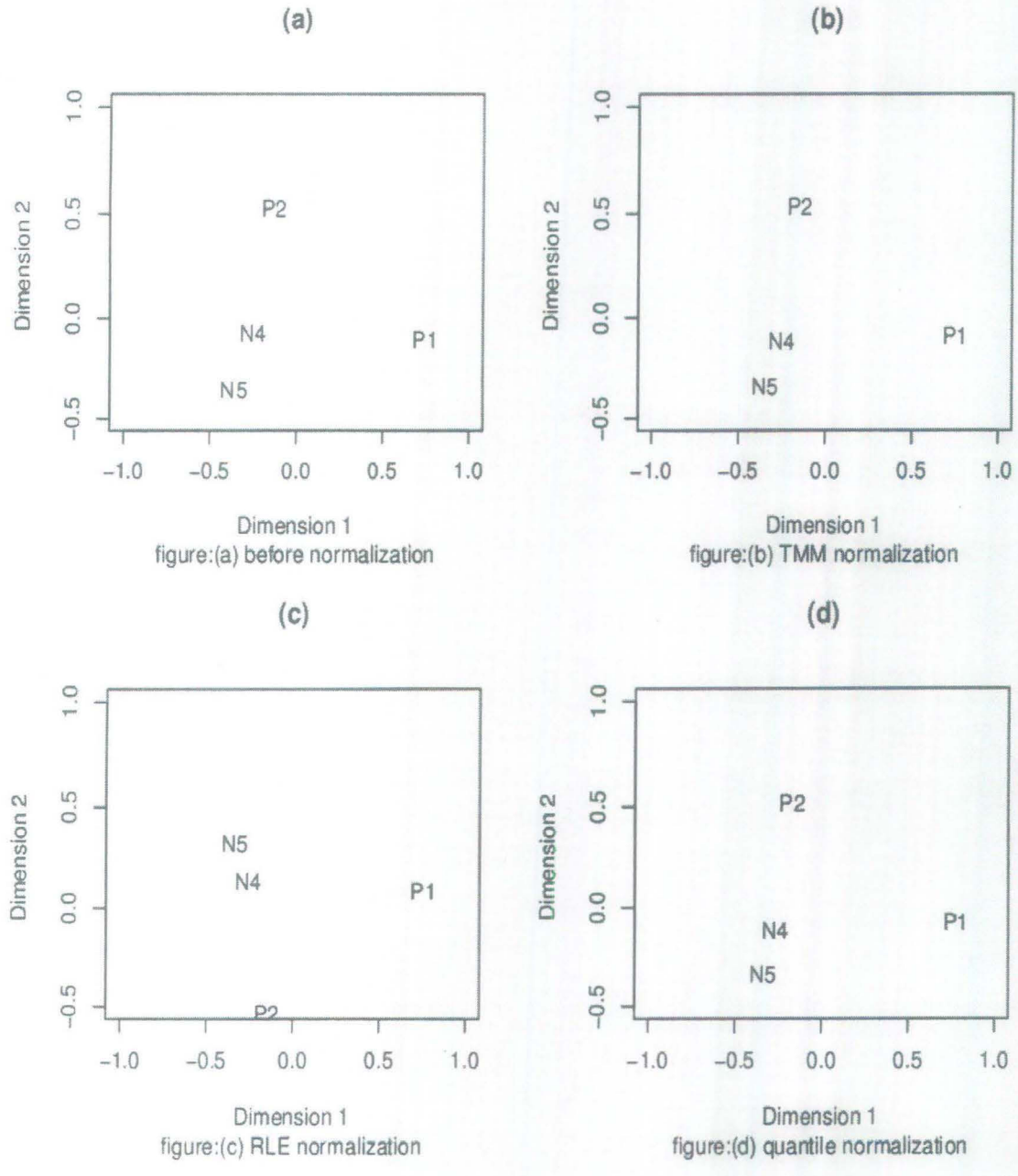


Figure 3.2: MDS plot showing the relations between the samples in two dimensions.

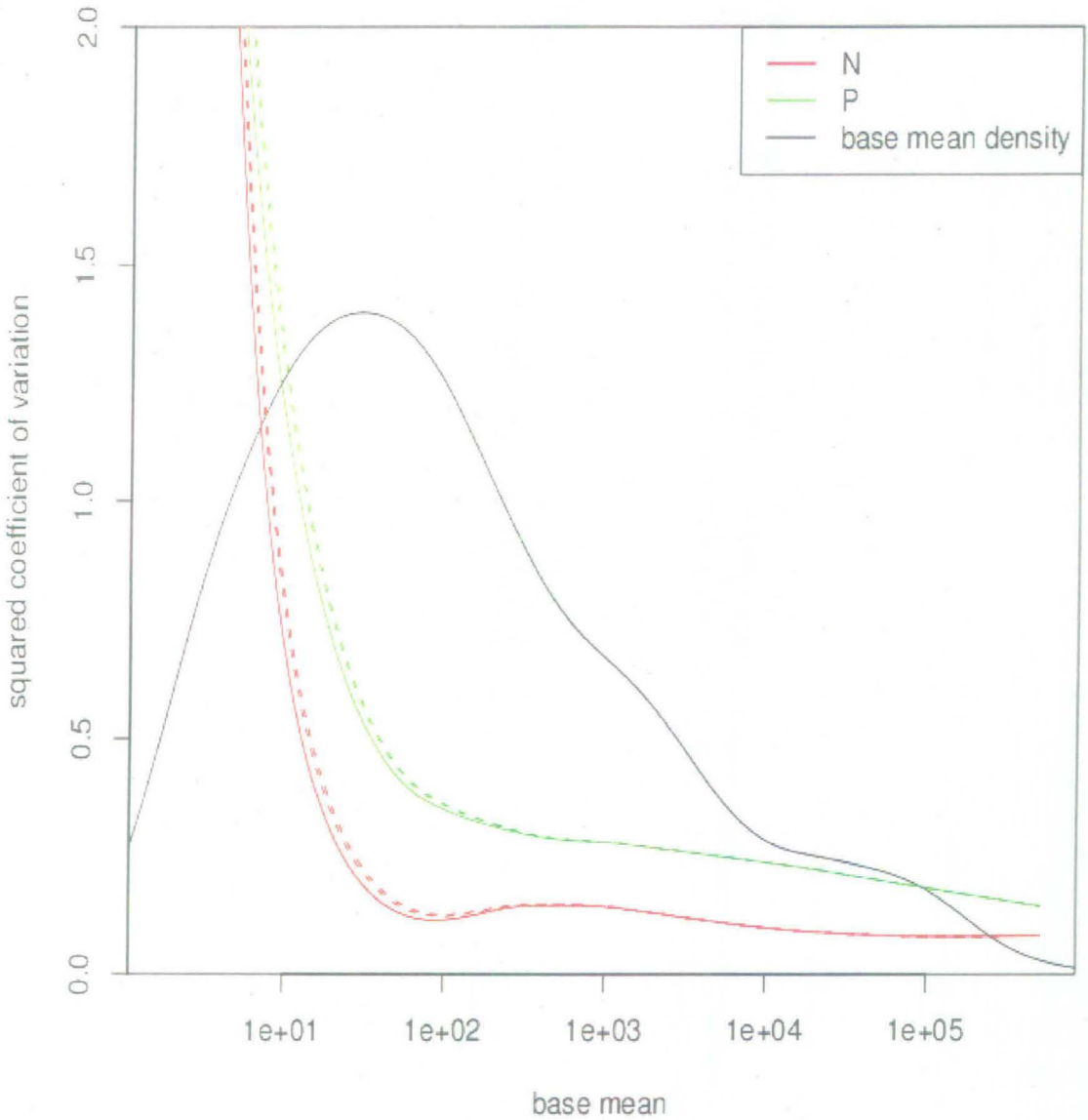


Figure 3.3: Plot to show the estimated variances (as squared coefficients of variation (SCV), i.e., variance over squared mean). In this plot, x axis is the base mean and y axis the squared coefficient of variation (SCV) i.e., the ratio of the variance at base level to the square of the base mean. The solid lines are the SCV for the raw variances.

3.2 Differentially Expressed Genes and IsomiRs

1. **From edgeR software:** Differentially expressed genes were identified by using the edgeR software on the normalized data. edgeR calculates a score of the p-value and fold change for each gene and sorted out the number of up-regulated and down-regulated genes on the basis of p-value and fold change. The number of such differentially expressed genes are shown in Table 3.1 and the top 10 differentially expressed genes are given in Table 3.2-3.4 and the list of all differentially expressed genes (miRNA) are given in the appendix A.

Table 3.1: Number of differentially expressed genes with the percentage of up- regulated and down-regulated genes using different normalization methods

No. of DE genes	up-regulated genes	down-regulated genes	Method used for normalization
39	18 (46% of DEG)	21 (54% of DEG)	Quantile
46	24 (52% of DEG)	22 (48% of DEG)	RLE
39	18 (46% of DEG)	21 (54% of DEG)	TMM

The top 10 differentially expressed genes after all the three normalization are almost similar. These genes are, sorted by the p-value, hsa-miR-1246, hsa-miR-136, hsa-miR-944, hsa-miR-362-3p, hsa-miR-3615, hsa-miR-1271, hsa-miR-29b-1*, hsa-miR-150* , hsa-miR-3613-5p and hsa-miR-3121.

Table 3.2: List of top 10 DEG miRNA sorted by p-value using Quantile normalization.

miRNA	logConc	logFC	PValue	FDR
hsa-miR-1246	-31.12341	37.78528	2.659895e-06	0.0009442627
hsa-miR-944	-33.21702	-33.59807	2.196085e-03	0.1341078528
hsa-miR-362-3p	-33.24951	33.53309	2.405448e-03	0.1341078528
hsa-miR-136	-33.24535	33.54141	2.405448e-03	0.1341078528
hsa-miR-1271	-33.25600	-33.52011	2.520893e-03	0.1341078528
hsa-miR-29b-1*	-33.27898	-33.47414	2.644380e-03	0.1341078528
hsa-miR-3615	-33.27613	33.47984	2.644380e-03	0.1341078528
hsa-miR-150*	-33.41443	-33.20324	4.032347e-03	0.1517635183
hsa-miR-3613-5p	-33.41948	-33.19315	4.032347e-03	0.1517635183
hsa-miR-3121	-33.43282	-33.16646	4.275029e-03	0.1517635183

Table 3.3: List of top 10 DEG miRNA sorted by p-value using RLE normalization.

miRNA	logConc	logFC	PValue	FDR
hsa-miR-1246	-31.18411	37.66388	1.849856e-06	0.0006566988
hsa-miR-136	-33.21907	33.59397	1.757356e-03	0.1140007553
hsa-miR-944	-33.22928	-33.57354	1.842536e-03	0.1140007553
hsa-miR-3615	-33.24609	33.53993	1.933614e-03	0.1140007553
hsa-miR-362-3p	-33.25111	33.52989	1.933614e-03	0.1140007553
hsa-miR-1271	-33.27828	-33.47554	2.135666e-03	0.1140007553
hsa-miR-29b-1*	-33.29968	-33.43274	2.247902e-03	0.1140007553
hsa-miR-150*	-33.44504	-33.14202	3.534816e-03	0.1336223880
hsa-miR-3121	-33.45971	-33.11268	3.764011e-03	0.1336223880
hsa-miR-3613-5p	-33.45111	-33.12989	3.764011e-03	0.1336223880

Table 3.4: List of top 10 DEG miRNA sorted by p-value using TMM normalization.

miRNA	logConc	logFC	PValue	FDR
hsa-miR-1246	-31.16836	37.69538	2.014405e-06	0.0007151139
hsa-miR-136	-33.22423	33.58364	1.944452e-03	0.1263968907
hsa-miR-944	-33.24291	-33.54629	2.039425e-03	0.1263968907
hsa-miR-362-3p	-33.24976	33.53260	2.039425e-03	0.1263968907
hsa-miR-3615	-33.25212	33.52787	2.141050e-03	0.1263968907
hsa-miR-1271	-33.27767	-33.47677	2.249941e-03	0.1263968907
hsa-miR-29b-1*	-33.30132	-33.42946	2.492333e-03	0.1263968907
hsa-miR-150*	-33.43264	-33.16684	3.702500e-03	0.1398614223
hsa-miR-3613-5p	-33.43725	-33.15760	3.702500e-03	0.1398614223
hsa-miR-3121	-33.45257	-33.12696	3.939758e-03	0.1398614223

2. **From DESeq software**: We found, 12 most significantly differentially expressed miRNAs with strongly down-regulated and up regulated miRNAs by using DESeq software.

Table 3.5: Identification of most significant Differentially expressed miRNA by DESeq

id	baseMean	baseMeanA	baseMeanB	foldChange	log2FoldChange	pval	padj	resVarA	resVarB
hsa-miR-1246	206.87	413.74	0	0	-Inf	1.89E-022	6.72E-020	18.37	0
hsa-miR-193a-3p	54.51	98.09	10.92	0.11	-3.17	2.15E-005	0	1.18	0.59
hsa-miR-424	1041.55	1838.94	244.16	0.13	-2.91	2.92E-005	0	12.91	0.08
hsa-miR-495	521.45	89.81	953.09	10.61	3.41	0	0.02	0.15	2.09
hsa-miR-369-3p	90.17	15.72	164.62	10.47	3.39	0	0.03	0.17	0.04
hsa-miR-130a	2952.69	4733.19	1172.19	0.25	-2.01	0	0.03	12.84	0
hsa-miR-720	403.16	655.71	150.61	0.23	-2.12	0	0.03	6.48	1.13
hsa-miR-136	12.33	24.66	0	0	-Inf	0	0.03	0.05	0
hsa-miR-3615	11.88	23.76	0	0	-Inf	0	0.04	0.1	0
hsa-miR-362-3p	11.77	23.53	0	0	-Inf	0	0.04	0.28	0
hsa-miR-196b	224.68	366.53	82.84	0.23	-2.15	0	0.04	10.17	0.07
hsa-miR-1277	86.79	29.65	143.93	4.85	2.28	0	0.09	0.63	16.24

Table 3.6: Identification of most significant down-regulated miRNA by DESeq

id	baseMean	baseMeanA	baseMeanB	foldChange	log2FoldChange	pval	padj	resVarA	resVarB
hsa-miR-1246	206.87	413.74	0	0	-Inf	1.89E-022	6.72E-020	18.37	0
hsa-miR-136	12.33	24.66	0	0	-Inf	0	0.03	0.05	0
hsa-miR-3615	11.88	23.76	0	0	-Inf	0	0.04	0.1	0
hsa-miR-362-3p	11.77	23.53	0	0	-Inf	0	0.04	0.28	0
hsa-miR-193a-3p	54.51	98.09	10.92	0.11	-3.17	2.15E-005	0	1.18	0.59
hsa-miR-424	1041.55	1838.94	244.16	0.13	-2.91	2.92E-005	0	12.91	0.08
hsa-miR-196b	224.68	366.53	82.84	0.23	-2.15	0	0.04	10.17	0.07
hsa-miR-720	403.16	655.71	150.61	0.23	-2.12	0	0.03	6.48	1.13
hsa-miR-130a	2952.69	4733.19	1172.19	0.25	-2.01	0	0.03	12.84	0
hsa-miR-1277	86.79	29.65	143.93	4.85	2.28	0	0.09	0.63	16.24
hsa-miR-369-3p	90.17	15.72	164.62	10.47	3.39	0	0.03	0.17	0.04
hsa-miR-495	521.45	89.81	953.09	10.61	3.41	0	0.02	0.15	2.09

Table 3.7: Identification of most significant up-regulated miRNA by DESeq

id	baseMean	baseMeanA	baseMeanB	foldChange	log2FoldChange	pval	padj	resVarA	resVarB
hsa-miR-495	521.45	89.81	953.09	10.61	3.41	0	0.02	0.15	2.09
hsa-miR-369-3p	90.17	15.72	164.62	10.47	3.39	0	0.03	0.17	0.04
hsa-miR-1277	86.79	29.65	143.93	4.85	2.28	0	0.09	0.63	16.24
hsa-miR-130a	2952.69		4733.19	1172.19 0.25	-2.01	0	0.03	12.84	0
hsa-miR-720	403.16	655.71	150.61	0.23	-2.12	0	0.03	6.48	1.13
hsa-miR-196b	224.68	366.53	82.84	0.23	-2.15	0	0.04	10.17	0.07
hsa-miR-424	1041.55	1838.94	244.16	0.13	-2.91	2.92E-005	0	12.91	0.08
hsa-miR-193a-3p	54.51	98.09	10.92	0.11	-3.17	2.15E-005	0	1.18	0.59
hsa-miR-1246	206.87	413.74	0	0	-Inf	1.89E-022	6.72E-020	18.37	0
hsa-miR-136	12.33	24.66	0	0	-Inf	0	0.03	0.05	0
hsa-miR-3615	11.88	23.76	0	0	-Inf	0	0.04	0.1	0
hsa-miR-362-3p	11.77	23.53	0	0	-Inf	0	0.04	0.28	0

DESeq calculated, its mean expression level (at the base scale) as a joint estimate from both conditions, and estimated separately for each condition, the fold change from the first to the second condition, the logarithm (to basis 2) of the fold change, and the p

value for the statistical significance of this change, for each gene. The padj column contains the p values, adjusted for multiple testing with the Benjamini-Hochberg procedure (See the standard R function p.adjust), which controls false discovery rate (FDR). The last two columns show the ratio of the single gene estimates for the base variance to the fitted value (See Table 3.5-3.7). This may help to notice false hits due to variance outliers. List of top 20 differentially expressed genes are shown in Appendix (A.4).

3. **From Likelihood ratio test:** Using the IsomiR expression profile as explained in

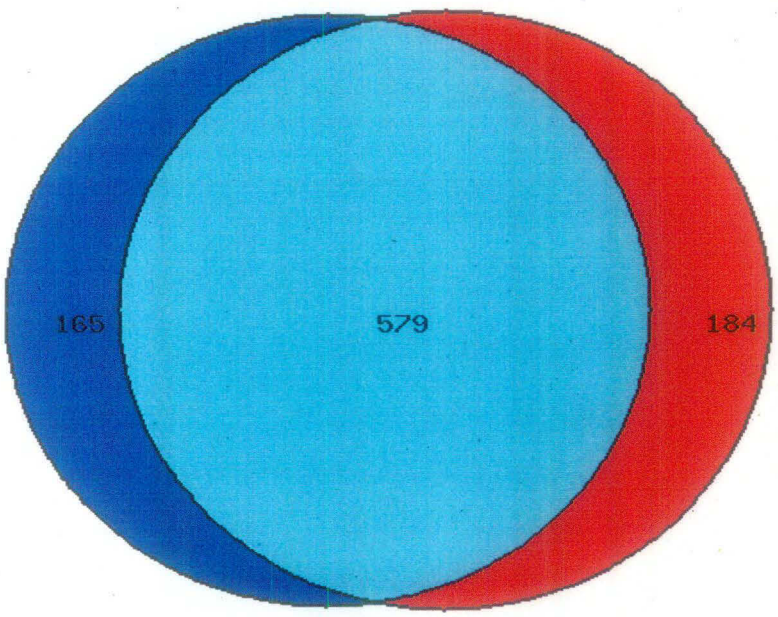


Figure 3.4: Differentially expressed isomiRs among Normals (N4 & N5), shown in blue color and Patients (P1 & P2), shown in red color and common isomiRs are shown in cyan color.

detail in the "Methods" section, total 928 differentially expressed isomiRs were detected between normal and patient samples by applying the likelihood ratio test on the quantile normalized data set. Among the list of differentially expressed isomiRs , we determined some normal specific, some patients specific miRNAs and some isomiRs were common

in both normal and patient samples (See Figure 3.4). These differentially expressed isomiRs sequences can give the better ability to detect differential miRNA expression level. List of top 20 differentially expressed isomiRs are given in Appendix (A.5).

3.3 Identification and Examination of the most abundant isomiR

As discussed in the "Methods"chapter, the most abundant IsomiR was first detected from amongst its IsomiR family members. The most abundant isomiR was then checked to see if it was the reference miRNA as given in the miRBase or not. Those cases where the reference miRNA was the most abundant one were named as "YES"and those cases where some other isomiR was the most abundant, were named as "NO". The Table 3.1 lists the number of "YES"and "NO"cases sample wise.

Hence "YES"has given to those isomiRs which are present in miRBase as reference sequence with highest abundant and "NO"to those sequences which does not correspond to highest abundant but occur in miRBase dataset i.e., this most abundant sequence did not correspond exactly to the current miRBase reference sequence [28]. Numbers of such cases are shown in Table 3.8. List of identified most abundant isomiR are shown in Appendix (A.6-A.10).

The results suggest the following:

a) Tissue and Environment based expression of the isomiRs

The most abundant isomiR sequence varies with different tissues under consideration. Their expression is governed by a number of factors [29].

b) Length preference of the reference miRNAs

Most of the reference miRNAs in miRbase have lengths around 21 to 22nts, and are in miRBase inspite of not being the most abundant isomiR.

Table 3.8: Classification of the most abundant miRNA

Sample	No. of cases	Match with most abundant IsomiRs
N4	240	Yes
N4	238	No
N5	240	Yes
N5	204	No
P1	263	Yes
P1	189	No
P2	242	Yes
P2	225	No
K562	317	Yes
K562	160	No

These could draw questions such as:

1. Is there a length bias in selection of the reference miRNA ?
2. Or are the original submissions to the miRBase incorrect ?

During expression analysis of the known miRNAs, choosing the reference miRNA may not always give the correct representation of the miRNA profile. Selecting the most abundant miRNA is the better alternative and should be used for downstream analysis including differential expression analysis and selecting diagnostic/prognostic markers.

3.4 IsomiRs present in specific condition

Using the IsomiR expression profile as explained in section 2.12, the Normal specific and Patient specific isomiRs were determined. IsomiRs having expression below twenty were

removed. No. of such cases are given in Table 3.9 and list of such isomiRs are given in Appendix (A.11).

Table 3.9: Number of normal and patient specific isomiRs

Number of IsomiRs	Present (In sample)
15	Normal (N4 &N5)
20	Patient (P1 & P2)

3.5 Most abundant star miRNA in Normal and Patient samples

During expression analysis of IsomiRs, some most abundant isomiRs, which are present in miRBase as reference sequence of miRNAs* are detected. These star sequences are having higher frequency to their counterparts. The most abundant star sequences may be varies with different tissues under consideration. No. of such cases are shown in Table 3.10.

Table 3.10: Number of most abundant miRNA* sequences corresponding to their counterparts.

Sample data	Most abundant miRNA* sequences
N4	17
N5	14
P1	12
P2	16
K562	16

Chapter 4

Summary and Conclusion

As explained in earlier section [2.4] that appropriate normalization reduce systematic errors and improve the detection of differentially expressed genes. Thus the conclusion from section [2.4] and [2.5] can be summarized as follow: different types of normalization method have been evaluated and visualized by the plots. Many statistical tests are exist to identify the differentially expressed genes but its depend on the data set that what type of distribution data follow, for example, if data has more variation in its mean and variance then we can not apply Poisson distribution on that data. We can also classify and compare differentially expressed genes by using couple of R tools. The list of miRNAs that were found to be differentially expressed can provide the information on the biology of human leukocytes and also serve as new markers of carcinogenesis.

Conclusions from the section [3.3], [3.4] and [3.5] can be summarized as follow: the list of most abundant isomiR can give the better analysis of the known miRNA. Selecting the most abundant miRNA is the better alternative and should be used for downstream analysis including differential expression analysis and selecting diagnostic/prognostic markers.

Appendix A

List of results

Differentially expressed genes using edgeR

Table A.1: Differentially expressed genes after Quantile normalization

miRNA	logConc	logFC	PValue	FDR
hsa-miR-1246	-31.12341	37.785282	2.659895e-06	0.0009442627
hsa-miR-944	-33.21702	-33.598071	2.196085e-03	0.1341078528
hsa-miR-362-3p	-33.24951	33.533088	2.405448e-03	0.1341078528
hsa-miR-136	-33.24535	33.541415	2.405448e-03	0.1341078528
hsa-miR-1271	-33.25600	-33.520114	2.520893e-03	0.1341078528
hsa-miR-29b-1*	-33.27898	-33.474142	2.644380e-03	0.1341078528
hsa-miR-3615	-33.27613	33.479839	2.644380e-03	0.1341078528
hsa-miR-150*	-33.41443	-33.203240	4.032347e-03	0.1517635183
hsa-miR-3613-5p	-33.41948	-33.193154	4.032347e-03	0.1517635183

Continued on next page

Table A.1 – continued from previous page

miRNA	logConc	logFC	PValue	FDR
hsa-miR-3121	-33.43282	-33.166464	4.275029e-03	0.1517635183
hsa-miR-505	-33.61889	-32.794322	7.249017e-03	0.2311813235
hsa-miR-223*	-33.62799	-32.776125	7.814580e-03	0.2311813235
hsa-miR-504	-33.70146	-32.629193	9.150350e-03	0.2474121502
hsa-miR-329	-33.74705	-32.538006	1.083588e-02	0.2474121502
hsa-miR-3138	-33.79325	32.445604	1.184791e-02	0.2474121502
hsa-miR-4286	-33.77894	32.474225	1.184791e-02	0.2474121502
hsa-miR-940	-33.77894	32.474225	1.184791e-02	0.2474121502
hsa-miR-3183	-33.82247	-32.387172	1.299986e-02	0.2563860909
hsa-miR-582-5p	-33.88489	32.262326	1.583447e-02	0.2958545655
hsa-miR-224	-33.99097	32.050172	2.203832e-02	0.3840321378
hsa-miR-509-3-5p	-34.03460	31.962906	2.488096e-02	0.3840321378
hsa-miR-551a	-34.05578	31.920549	2.488096e-02	0.3840321378
hsa-miR-96	-34.05578	31.920549	2.488096e-02	0.3840321378
hsa-miR-3677	-34.10035	-31.831417	2.827502e-02	0.3961215289
hsa-miR-493	-34.10035	-31.831417	2.827502e-02	0.3961215289
hsa-miR-495	-12.77326	-3.330698	2.913642e-02	0.3961215289
hsa-miR-671-5p	-34.15926	31.713582	3.236770e-02	0.3961215289
hsa-miR-369-3p	-15.27982	-3.291965	3.366532e-02	0.3961215289
hsa-miR-136*	-34.21929	-31.593525	3.735761e-02	0.3961215289
hsa-miR-154*	-34.21929	-31.593525	3.735761e-02	0.3961215289
hsa-miR-122	-34.18401	31.664087	3.735761e-02	0.3961215289
hsa-miR-3676	-34.18401	31.664087	3.735761e-02	0.3961215289
hsa-miR-3127	-34.21716	31.597793	3.735761e-02	0.3961215289

Continued on next page

Table A.1 – continued from previous page

miRNA	logConc	logFC	PValue	FDR
hsa-miR-1274a	-34.29325	-31.445606	4.351758e-02	0.3961215289
hsa-miR-1299	-34.29325	-31.445606	4.351758e-02	0.3961215289
hsa-miR-3190	-34.29325	-31.445606	4.351758e-02	0.3961215289
hsa-miR-378*	-34.28010	31.471903	4.351758e-02	0.3961215289
hsa-miR-204	-34.27176	-31.488595	4.351758e-02	0.3961215289
hsa-miR-3175	-34.24159	31.548926	4.351758e-02	0.3961215289

Table A.2: Differentially expressed genes after RLE normalization

miRNA	logConc	logFC	PValue	FDR
hsa-miR-1246	-31.18411	37.663879	1.849856e-06	0.0006566988
hsa-miR-136	-33.21907	33.593975	1.757356e-03	0.1140007553
hsa-miR-944	-33.22928	-33.573544	1.842536e-03	0.1140007553
hsa-miR-3615	-33.24609	33.539929	1.933614e-03	0.1140007553
hsa-miR-362-3p	-33.25111	33.529893	1.933614e-03	0.1140007553
hsa-miR-1271	-33.27828	-33.475541	2.135666e-03	0.1140007553
hsa-miR-29b-1*	-33.29968	-33.432741	2.247902e-03	0.1140007553
hsa-miR-150*	-33.44504	-33.142024	3.534816e-03	0.1336223880
hsa-miR-3121	-33.45971	-33.112680	3.764011e-03	0.1336223880
hsa-miR-3613-5p	-33.45111	-33.129887	3.764011e-03	0.1336223880
hsa-miR-223*	-33.65010	-32.731918	6.650673e-03	0.1829131748
hsa-miR-505	-33.64410	-32.743917	6.650673e-03	0.1829131748

Continued on next page

Table A.2 – continued from previous page

miRNA	logConc	logFC	PValue	FDR
hsa-miR-4286	-33.66686	32.698379	7.213477e-03	0.1829131748
hsa-miR-940	-33.66686	32.698379	7.213477e-03	0.1829131748
hsa-miR-504	-33.74354	-32.545030	9.363681e-03	0.2027520321
hsa-miR-329	-33.77070	-32.490712	1.028038e-02	0.2027520321
hsa-miR-3138	-33.76986	32.492382	1.028038e-02	0.2027520321
hsa-miR-582-5p	-33.77407	32.483967	1.028038e-02	0.2027520321
hsa-miR-3183	-33.82951	-32.373097	1.132792e-02	0.2116532374
hsa-miR-551a	-33.94734	32.137434	1.743059e-02	0.2946599240
hsa-miR-96	-33.94734	32.137434	1.743059e-02	0.2946599240
hsa-miR-495	-12.85136	-3.404822	2.192116e-02	0.3121464614
hsa-miR-224	-34.04854	31.935022	2.231564e-02	0.3121464614
hsa-miR-671-5p	-34.05249	31.927119	2.231564e-02	0.3121464614
hsa-miR-369-3p	-15.36816	-3.371553	2.406881e-02	0.3121464614
hsa-miR-3127	-34.11141	31.809290	2.549929e-02	0.3121464614
hsa-miR-509-3-5p	-34.09147	31.849178	2.549929e-02	0.3121464614
hsa-miR-3677	-34.10834	-31.815426	2.549929e-02	0.3121464614
hsa-miR-493	-34.10834	-31.815426	2.549929e-02	0.3121464614
hsa-miR-378*	-34.17553	31.681040	2.935765e-02	0.3473988903
hsa-miR-193a-3p	-16.00143	3.142508	3.681784e-02	0.3637004335
hsa-miR-582-3p	-34.24588	31.540356	3.995582e-02	0.3637004335
hsa-miR-136*	-34.25947	-31.513173	3.995582e-02	0.3637004335
hsa-miR-154*	-34.25947	-31.513173	3.995582e-02	0.3637004335
hsa-miR-122	-34.23823	31.555658	3.995582e-02	0.3637004335
hsa-miR-3676	-34.23823	31.555658	3.995582e-02	0.3637004335

Continued on next page

Table A.2 – continued from previous page

miRNA	logConc	logFC	PValue	FDR
hsa-miR-1274a	-34.30210	-31.427918	3.995582e-02	0.3637004335
hsa-miR-1299	-34.30210	-31.427918	3.995582e-02	0.3637004335
hsa-miR-3190	-34.30210	-31.427918	3.995582e-02	0.3637004335
hsa-miR-424	-11.65706	2.912488	4.539811e-02	0.3653841383
hsa-miR-129*	-34.35601	31.320080	4.734555e-02	0.3653841383
hsa-miR-204	-34.31169	-31.408723	4.734555e-02	0.3653841383
hsa-miR-1254	-34.38015	-31.271802	4.734555e-02	0.3653841383
hsa-miR-3175	-34.29469	31.442719	4.734555e-02	0.3653841383
hsa-miR-19a	-34.32376	31.384586	4.734555e-02	0.3653841383
hsa-miR-375	-34.32376	31.384586	4.734555e-02	0.3653841383

Table A.3: Differentially expressed genes after TMM normalization

miRNA	logConc	logFC	PValue	FDR
hsa-miR-1246	-31.16836	37.695382	2.014405e-06	0.0007151139
hsa-miR-136	-33.22423	33.583641	1.944452e-03	0.1263968907
hsa-miR-944	-33.24291	-33.546290	2.039425e-03	0.1263968907
hsa-miR-362-3p	-33.24976	33.532596	2.039425e-03	0.1263968907
hsa-miR-3615	-33.25212	33.527870	2.141050e-03	0.1263968907
hsa-miR-1271	-33.27767	-33.476767	2.249941e-03	0.1263968907
hsa-miR-29b-1*	-33.30132	-33.429463	2.492333e-03	0.1263968907
hsa-miR-150*	-33.43264	-33.166835	3.702500e-03	0.1398614223

Continued on next page

Table A.3 – continued from previous page

miRNA	logConc	logFC	PValue	FDR
hsa-miR-3613-5p	-33.43725	-33.157603	3.702500e-03	0.1398614223
hsa-miR-3121	-33.45257	-33.126960	3.939758e-03	0.1398614223
hsa-miR-505	-33.63936	-32.753395	6.917014e-03	0.2046283282
hsa-miR-223*	-33.64978	-32.732553	6.917014e-03	0.2046283282
hsa-miR-4286	-33.69096	32.650182	8.144721e-03	0.2065268459
hsa-miR-940	-33.69096	32.650182	8.144721e-03	0.2065268459
hsa-miR-504	-33.71479	-32.602530	8.875675e-03	0.2098732243
hsa-miR-329	-33.76819	-32.495734	9.702328e-03	0.2098732243
hsa-miR-3138	-33.77424	32.483624	1.064146e-02	0.2098732243
hsa-miR-582-5p	-33.79788	32.436340	1.064146e-02	0.2098732243
hsa-miR-3183	-33.85092	-32.330275	1.294404e-02	0.2418492604
hsa-miR-551a	-33.97061	32.090892	1.794250e-02	0.3033137223
hsa-miR-96	-33.97061	32.090892	1.794250e-02	0.3033137223
hsa-miR-671-5p	-34.07538	31.881347	2.291300e-02	0.3363184696
hsa-miR-224	-34.03360	31.964918	2.291300e-02	0.3363184696
hsa-miR-495	-12.83821	-3.371262	2.407867e-02	0.3363184696
hsa-miR-3677	-34.12861	-31.774889	2.614677e-02	0.3363184696
hsa-miR-493	-34.12861	-31.774889	2.614677e-02	0.3363184696
hsa-miR-509-3-5	p -34.07669	31.878724	2.614677e-02	0.3363184696
hsa-miR-369-3p	-15.34815	-3.345300	2.652653e-02	0.3363184696
hsa-miR-3127	-34.13406	31.763989	3.006105e-02	0.3638990734
hsa-miR-122	-34.22410	31.583916	3.485231e-02	0.3638990734
hsa-miR-3676	-34.22410	31.583916	3.485231e-02	0.3638990734
hsa-miR-136*	-34.23305	-31.566015	3.485231e-02	0.3638990734

Continued on next page

Table A.3 – continued from previous page

miRNA	logConc	logFC	PValue	FDR
hsa-miR-154*	-34.23305	-31.566015	3.485231e-02	0.3638990734
hsa-miR-378*	-34.19791	31.636286	3.485231e-02	0.3638990734
hsa-miR-3175	-34.28083	31.470441	4.079124e-02	0.3799897593
hsa-miR-204	-34.28557	-31.460970	4.079124e-02	0.3799897593
hsa-miR-582-3p	-34.26793	31.496245	4.079124e-02	0.3799897593
hsa-miR-193a-3p	-15.98274	3.065602	4.284895e-02	0.3799897593
hsa-miR-129*	-34.34246	31.347197	4.825843e-02	0.3799897593
hsa-miR-19a	-34.34543	31.341241	4.825843e-02	0.3799897593
hsa-miR-375	-34.34543	31.341241	4.825843e-02	0.3799897593
hsa-miR-1274a	-34.32134	-31.389435	4.825843e-02	0.3799897593
hsa-miR-1299	-34.32134	-31.389435	4.825843e-02	0.3799897593
hsa-miR-3190	-34.32134	-31.389435	4.825843e-02	0.3799897593

Table A.4: **Top 20 differentially expressed genes using DESeq**

miRNA	bM	bMA	bMB	fC	log_2fC	pval	padj	rVarA	rVarB
hsa-miR-122	2.92	5.84	0	0	-Inf	1	1	1.58	0
hsa-miR-1260b	2.25	4.49	0	0	-Inf	1	1	1.22	0
hsa-miR-1271	11.37	0	22.7	Inf	Inf	1	1	0	0.18
hsa-miR-129*	2.47	4.94	0	0	-Inf	1	1	1.34	0
hsa-miR-150*	8.98	0	17.9	Inf	Inf	1	1	0	0.61
hsa-miR-181c	17.62	18.09	17.2	0.95	-0.08	1	1	0.72	0.31
hsa-miR-186	1836.25	1741.9	1930.6	1.11	0.15	1	1	0.44	1.47
hsa-miR-191*	2.25	4.49	0	0	-Inf	1	1	1.22	0
hsa-miR-19a	2.75	5.5	0	0	-Inf	1	1	1.49	0
hsa-miR-19b-1*	2.02	4.04	0	0	-Inf	1	1	1.09	0
hsa-miR-212	2.25	4.49	0	0	-Inf	1	1	1.22	0
hsa-miR-223*	6.8	0	13.6	Inf	Inf	1	1	0	0.12
hsa-miR-224	3.82	7.64	0	0	-Inf	1	1	2.06	0
hsa-miR-26b*	2.25	4.49	0	0	-Inf	1	1	1.22	0
hsa-miR-29b-1*	11.04	0	22.1	Inf	Inf	1	1	0	0.44
hsa-miR-3121	8.82	0	17.6	Inf	Inf	1	1	0	0.06
hsa-miR-3124	2.02	4.04	0	0	-Inf	1	1	1.09	0
hsa-miR-3127	3.66	7.33	0	0	-Inf	1	1	1.98	0
hsa-miR-3175	2.7	5.39	0	0	-Inf	1	1	1.46	0
hsa-miR-3183	5.38	0	10.8	Inf	Inf	1	1	0	3.28

Table A.5: **Top 20 differentially expressed isomiRs using Likelihood ratio test**

miRNA	isomir	N4	N5	P1	P2
hsa-let-7b	TGANGTAGTAGGTTGTGTGGTAT	28.5	41.5	12	10.75
hsa-mir-155	TTAATGCTAATCGTGATAGGGGTA	75.75	95.5	17	11
hsa-mir-140	TACCACAGGGTAGAACCCCGGACA	48.5	39.75	18.25	11
hsa-mir-320a	AAAAGATGGGTTGAGAGGGCGA	33.75	29.25	11.5	11.25
hsa-mir-23a	ATCACATTGTCAGGGATTTC	97.5	71.25	18.5	11.25
hsa-mir-140	TACCATAGGGTAGAACCACGGAAA	80.5	90	13.75	11.75
hsa-mir-140	TACCACAGGGTAGCACCACGGAA	53.5	51.5	13.75	12.75
hsa-mir-140	TACCACAGGGTACAACCACGGAA	34.5	56.25	14	12.75
hsa-mir-140	TACCACAGGGTAGGACCACGGAA	32.75	53.5	12.5	13.75
hsa-mir-342	GGGGTGCTATCTGTGATTGAGGG	51	45.25	13	14
hsa-mir-3184	TCGTCTCGCTCTCTGCCCCTCA	47.25	54.75	20.25	14
hsa-mir-140	TACCACAGGGTAGAACCCCGGAA	81	107.75	12.5	14.5
hsa-mir-140	TACCACAGGGTAGAACCACGNAT	396	44.75	15.5	14.5
hsa-mir-548w	AAAAGTAACTGCGGTTTTTG	44.75	73.5	13	15
hsa-mir-29c	TAGCACCATTTGAAATCGGTTTT	53	82.25	15.75	15
hsa-mir-30e	CTTTCAGTCGGATGTTTACAGCT	59.25	99.25	28.5	15
hsa-mir-140	TACCACAGGGTAGACCCACGGAA	47.75	65.25	16.75	15.5
hsa-mir-29a	TAGCATTATCTGAAATCGGTTA	113	67.75	21.25	15.75
hsa-mir-140	TACCACAGGGTAGACCCACGGA	59.25	67.5	13.75	16.5
hsa-mir-140	TACCACAGCGTAGAACCACGGA	44.5	82.25	13.75	16.5

Perl Codes

1. Perl code for creating expression profile

```
#!/user/bin/perl
# generate multiple files for N4 sample--------;
for($W=0;$W<940;$W++)
{
  $f= "N4_".$W.".txt";
  open(FILE,$f);
  @D =<FILE>;

  $i=0;
  $j=0;
  $t=0;

 foreach $D(@D)
 {
  chomp $D;
  @C = split(/\s+/,$D);
  @arr1[$i++] = $C[0];
  @arr2[$j++] = $C[1];
  @arr3[$t++] = $C[2];
   }
  $length1=$i;
  print "$i\n";
#generate multiple files for N5 sample ------;
  $f= "N5_".$W.".txt";
```

```
open(FILE1,$f);
@D1 =<FILE1>;
$s=@D1;

$COUNT = 0;
$k=0;
$l=0;
$m=0;
$n=0;

foreach $D1(@D1)
{
  chomp $D1;
  @C1 = split(/\s+/,$D1);
  @arr11[$k++]=$C1[0];
  @arr12[$l++]=$C1[1];
  @arr13[$m++]=$C1[2];
  @arr14[$n++]=$C1[3];
 }
$length2=$k;
print "$k\n";

#generate multiple files for P1 sample ------;
$f= "P1_".$W.".txt";
open(FILE2,$f);
@D2 =<FILE2>;
$sj=@D2;
```

```
$u=0;
$v=0;
$w=0;

foreach $D2(@D2)
{
chomp $D2;
@C2 = split(/\s+/,$D2);
@arr21[$u++]=$C2[0];
@arr22[$v++]=$C2[1];
@arr23[$w++]=$C2[2];
 }
$length3=$u;
print "$u\n";

#generate multiple files for P2 sample ------;
$f= "P2_".$W.".txt";
open(FILE3,$f);
@D3 =<FILE3>;
$sk=@D3;

$p=0;
$q=0;
$r=0;

foreach $D3(@D3)
```

```
 {
 chomp $D3;
 @C3 = split(/\s+/,$D3);
 @arr31[$p++]=$C3[0];
 @arr32[$q++]=$C3[1];
 @arr33[$r++]=$C3[2];
 }
 $length4=$p;
 print "$p\n";

#print "enter the output files name:";

 $f="nn_".$W.".txt";
 open(FH1,">$f");
 $f="pp_".$W.".txt";
 open(FH2,">$f");

#<<<<<<<<<<<<<<<<<< for checking N4 and N5 samples >>>>>>>>>>>>>>>
 for($i=0;$i<$length1;$i++)
 {
    $flag=0;
    $j=0;
     while($j<$length2 && $flag==0)
     {
       if($arr3[$i] eq $arr13[$j])
 {
  print FH1 "$arr3[$i]\t$arr2[$i]\t$arr1[$i]\t$arr11[$j]\n";
```

```
    $flag=1
   }
   $j++;
}
 if($flag==0)
 {
  print FH1 "$arr3[$i]\t$arr2[$i]\t$arr1[$i]\t0\n";
 }
}
#-------------------------------------------------------------------
   for($j=0;$j<$length2;$j++)
    {
    $flag=0;
    $i=0;
      while($i<$length1 && $flag==0)
      {
if($arr3[$i] eq $arr13[$j])
{
 $flag=1
}
 $i++;
      }
      if($flag==0)
      {
      print FH1 "$arr13[$j]\t$arr12[$j]\t0\t$arr11[$j]\n";
      }
    }
```

```
#<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
#<<<<<<<<<<<<<<<<<< for checking P1 and P2 samples >>>>>>>>>>>

  for($i=0;$i<$length3;$i++)
  {
  $flag=0;
  $j=0;
 while($j<$length4 && $flag==0)
 {
  if($arr23[$i] eq $arr33[$j])
  {
    print FH2 "$arr23[$i]\t$arr22[$i]\t$arr21[$i]\t$arr31[$j]\n";
    $flag=1
  }
  $j++;
  }
  if($flag==0)
    {
    print FH2 "$arr23[$i]\t$arr22[$i]\t$arr21[$i]\t0\n";
      }
  }
#-----------------------------------------------------------------------
  for($j=0;$j<$length4;$j++)
  {
    $flag=0;
    $i=0;
    while($i<$length3 && $flag==0)
```

```
    {
      if($arr23[$i] eq $arr33[$j])
      {
      $flag=1
      }
    $i++;
    }
    if($flag==0)
    {
    print FH2 "$arr33[$j]\t$arr32[$j]\t0\t$arr31[$j]\n";
    }
  }
  close FH1;
  close FH2;
#-------------------------------------------------------------
#<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
## for the two output files ..............................

  $f= "nn_".$W.".txt";
  open(FILE4,$f);
  @DD1 =<FILE4>;
  $sii=@DD1;

  $ii=0;
  $jj=0;
  $tt=0;
  $ss=0;
```

```
    foreach $DD1(@DD1)
    {
    chomp $DD1;
    @C4 = split(/\s+/,$DD1);
    @arry1[$ii++]=$C4[0];
    @arry2[$jj++]=$C4[1];
    @arry3[$tt++]=$C4[2];
    @arry4[$ss++]=$C4[3];
     }
      $len1=$ii;
      print "$ii\n";
#>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
   $f= "pp_".$W.".txt";
   open(FILE5,$f);
   @DD2 =<FILE5>;

   $uu=0;
   $vv=0;
   $ww=0;
   $xx=0;
   foreach $DD2(@DD2)
   {
   chomp $DD2;
   @C5 = split(/\s+/,$DD2);
   @arry11[$uu++]=$C5[0];
   @arry12[$vv++]=$C5[1];
   @arry13[$ww++]=$C5[2];
```

```
  @arry14[$xx++]=$C5[3];
  }
  $len2=$uu;
  print "$uu\n";
#>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>.
  $f1="HSA-".$W.".txt";
  open(FH3,">$f1");
#<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<

  for($i=0;$i<$len1;$i++)
  {
    $flag=0;
    $j=0;
    while($j<$len2 && $flag==0)
    {
      if($arry1[$i] eq $arry11[$j])
      {
print FH3 "$arry2[$i]\t$arry1[$i]\t$arry3[$i]\t$arry4[$i]
              \t$arry13[$j]\t$arry14[$j]\n";
$flag=1
      }
      $j++;
    }
      if($flag==0){
print FH3 "$arry2[$i]\t$arry1[$i]\t$arry3[$i]\t$arry4[$i]\t0\t0\n";
        }
  }
```

```
#-------------------------------------------------------------------

for($j=0;$j<$len2;$j++)
{
  $flag=0;
  $i=0;
  while($i<$len1 && $flag==0)
  {
    if($arry1[$i] eq $arry11[$j])
    {
      $flag=1
     }
    $i++;
  }
  if($flag==0)
    {
    print FH3 "$arry12[$j]\t$arry11[$j]\t0\t0\t$arry13[$j]\t$arry14[$j]\n";
    }
}
close FH3;
```

2. **Perl code for identifying the reference miRNA with most abundant isomiR**

```
#.....code for comparing two files.....####
#!/user/bin/perl
#print"enter the sequence file name:";
my $f = $ARGV[0];
open(FILE,$f);
```

```
@D =<FILE>;
$si=@D;

$COUNT = 0;
$i=0;
$j=0;
$t=0;
$u=0;
$v=0;
$w=0;
$x=0;
$y=0;
 foreach $D(@D)
 {
        chomp $D;
@C = split(/\s+/,$D);
@arr1[$i++]=$C[0];
@arr2[$j++]=$C[1];
@arr23[$t++]=$C[2];
@arr24[$u++]=$C[3];
@arr25[$v++]=$C[4];
@arr26[$w++]=$C[5];
@arr27[$x++]=$C[6];
@arr28[$y++]=$C[7];
 }
 $length1=$i;
 print "$i\n";
```

```
  my $f = $ARGV[1];
  open(FILE1,$f);
  @D1 =<FILE1>;
  $s=@D1;
  print "$s\n";

  $COUNT = 0;
  $k=0;$o=0;
  $p=0;$n=0;
  $q=0;$m=0;
  $r=0;$l=0;
  $s=0;

  foreach $D1(@D1)
   {
chomp $D1;
@C1 = split(/\s+/,$D1);
    @arr3[$k++]=$C1[0]; @arr7[$o++]=$C1[4];
    @arr4[$l++]=$C1[1]; @arr8[$p++]=$C1[5];
    @arr5[$m++]=$C1[2]; @arr9[$q++]=$C1[6];
    @arr6[$n++]=$C1[3]; @arr10[$r++]=$C1[7];
    @arr11[$s++]=$C1[8];
   }
  $length2=$k;
 #print "enter the output file name:";
  $outfile1=$ARGV[2];
```

```
open(FH1,">$outfile1");


for($i=0;$i<$length2;$i++)
{
$flag=0;
$j=0;
  while($j<$length1 && $flag==0)
  {
    if($arr26[$j] eq $arr7[$i])
     {
       print FH1 "$arr3[$j]\t$arr4[$i]\t$arr5[$i]\t$arr6[$i]\t$arr7[$i]\t
                  $arr8[$i]\t$arr9[$i]\t$arr10[$i]\tyes\n";
        $flag=1
      }
      $j++;
  }
  if($flag==0)
    {
       print FH1 "$arr3[$j]\t$arr4[$i]\t$arr5[$i]\t$arr6[$i]\t$arr7[$i]\t
                  $arr8[$i]\t$arr9[$i]\t$arr10[$i]\tNo\n";
      }
}
```

Table A.6: **Identification of most abundant IsomiRs for N4 sample**

read-name	Reference	Position	IsomiRs	matching-abundance	highest-abundance	mature	match-with-highest
20.22.100473	hsa-let-7a-3	3	TGAGGTAGTAGGTTGTATAGTT	100473	100473	hsa-let-7a	yes
1.22.468414	hsa-let-7b	5	TGAGGTAGTAGGTTGTGTGGTT	468414	468414	hsa-let-7b	yes
97.22.14163	hsa-let-7c	10	TGAGGTAGTAGGTTGTATGGTT	14163	14163	hsa-let-7c	yes
42.22.45176	hsa-let-7d	7	AGAGGTAGTAGGTTGCATAGTT	45176	64241	hsa-let-7d	No
1679.22.344	hsa-let-7d	61	CTATACGACCTGCTGCCTTTCT	344	344	hsa-let-7d*	yes
328.22.3175	hsa-let-7e	7	TGAGGTAGGAGGTTGTATAGTT	3175	4612	hsa-let-7e	No
51236.22.6	hsa-let-7e	52	CTATACGGCCTCCTAGCTTTCC	6	6	hsa-let-7e*	yes
16.22.133929	hsa-let-7f-2	7	TGAGGTAGTAGATTGTATAGTT	133929	133929	hsa-let-7f	yes
423022.22.1	hsa-let-7f-2	57	CTATACAGTCTACTGTCTTTCC	1	12	hsa-let-7f-2*	No
7.22.162389	hsa-let-7g	4	TGAGGTAGTAGTTTGTACAGTT	162389	162389	hsa-let-7g	yes
434041.21.1	hsa-let-7g	61	CTGTACAGGCCACTGCCTTGC	1	11	hsa-let-7g*	No
84.22.19423	hsa-let-7i	5	TGAGGTAGTAGTTTGTGCTGTT	19423	19423	hsa-let-7i	yes
7371.22.56	hsa-let-7i	61	CTGCGCAAGCTACTGCCTTGCT	56	56	hsa-let-7i*	yes
168.22.7089	hsa-mir-1-2	52	TGGAATGTAAAGAAGTATGTAT	7089	7089	hsa-miR-1	yes
4119.22.110	hsa-mir-100	12	AACCCGTAGATCCGAACTTGTG	110	110	hsa-miR-100	yes
122269.22.3	hsa-mir-101-1	10	CAGTTATCACAGTGCTGATGCT	3	6482	hsa-miR-101*	No
165.21.7210	hsa-mir-101-2	48	TACAGTACTGTGATAACTGAA	7210	10117	hsa-miR-101	No
18.23.109253	hsa-mir-103-2	47	AGCAGCATTGTACAGGGCTATGA	109253	109253	hsa-miR-103	yes
59296.23.5	hsa-mir-103-2	10	AGCTTCTTTACAGTGCTGCCTTG	5	25	hsa-miR-103-2*	No
11206.23.33	hsa-mir-106a	12	AAAAGTGCTTACAGTGCAGGTAG	33	33	hsa-miR-106a	yes
275.22.3928	hsa-mir-106b	51	CCGCACTGTGGGTACTTGCTGC	3928	3928	hsa-miR-106b*	yes
646.21.1226	hsa-mir-106b	11	TAAAGTGCTGACAGTGCAGAT	1226	1226	hsa-miR-106b	yes
100.23.13838	hsa-mir-107	49	AGCAGCATTGTACAGGGCTATCA	13838	13838	hsa-miR-107	yes
2884.23.171	hsa-mir-10a	21	TACCCTGTAGATCCGAATTTGTG	171	935	hsa-miR-10a	No
46186.22.7	hsa-mir-122	14	TGGAGTGTGACAATGGTGTTTG	7	29	hsa-miR-122	No
92985.21.4	hsa-mir-1228	0	GTGGGCGGGGGCAGGTGTGTG	4	86	hsa-miR-1228*	No
426559.23.1	hsa-mir-1229	46	CTCTCACCACTGCCCTCCCACAG	1	5	hsa-miR-1229	No
23291.20.14	hsa-mir-124-3	52	TAAGGCACGCGGTGAATGCC	14	151	hsa-miR-124	No
27566.22.11	hsa-mir-1249	10	ACGCCCTTCCCCCCCTTCTTCA	11	11	hsa-miR-1249	yes
15572.21.22	hsa-mir-1250	23	ACGGTGCTGGATGTGGCCTTT	22	22	hsa-miR-1250	yes
77273.24.4	hsa-mir-1254	17	AGCCTGGAAGCTGGAGCCTGCAGT	4	34	hsa-miR-1254	No
7917.23.51	hsa-mir-1255a	27	AGGATGAGCAAAGAAAGTAGATT	51	51	hsa-miR-1255a	yes
3902.22.118	hsa-mir-1255b-2	5	CGGATGAGCAAAGAAAGTGGTT	118	118	hsa-miR-1255b	yes
27808.22.11	hsa-mir-1256	34	AGGCATTGACTTCTCACTAGCT	11	11	hsa-miR-1256	yes
3615.24.129	hsa-mir-125a	14	TCCCTGAGACCCTTTAACCTGTGA	129	314	hsa-miR-125a-5p	No
7555.22.54	hsa-mir-125a	52	ACAGGTGAGGTTCTTGGGAGCC	54	54	hsa-miR-125a-3p	yes
6016.22.72	hsa-mir-125b-1	14	TCCCTGAGACCCTAACTTGTGA	72	72	hsa-miR-125b	yes
4098.21.111	hsa-mir-126	14	CATTATTACTTTTGGTACGCG	111	118	hsa-miR-126*	No
10408.22.37	hsa-mir-126	51	TCGTACCGTGAGTAATAATGCG	37	37	hsa-miR-126	yes
361308.19.1	hsa-mir-1260b	9	ATCCCACCACTGCCACCAT	1	167	hsa-miR-1260b	No
15670.22.22	hsa-mir-1262	13	ATGGGTGAATTTGTAGAAGGAT	22	22	hsa-miR-1262	yes
3714.18.125	hsa-mir-1268	4	CGGGCGTGGTGGTGGGGG	125	125	hsa-miR-1268	yes
1293.22.482	hsa-mir-127	56	TCGGATCCGTCTGAGCTTGGCT	482	482	hsa-miR-127-3p	yes

read-name	Reference	Position	IsomiRs	matching-abundance	highest-abundance	mature	match-with-highest
20497.22.16	hsa-mir-127	22	CTGAAGCTCAGAGGGCTCTGAT	16	39	hsa-miR-127-5p	No
21660.23.15	hsa-mir-1270-2	12	CTGGAGATATGGAAGAGCTGTGT	15	15	hsa-miR-1270	yes
14646.22.24	hsa-mir-1271	14	CTTGGCACCTAGCAAGCACTCA	24	24	hsa-miR-1271	yes
11422.17.33	hsa-mir-1274b	14	TCCCTGTTCGGGCGCCA	33	153	hsa-miR-1274b	No
7489.17.55	hsa-mir-1275	17	GTGGGGGAGAGGCTGTC	55	55	hsa-miR-1275	yes
1872.22.297	hsa-mir-1277	46	TACGTAGATATATATGTATTTT	297	297	hsa-miR-1277	yes
3266.22.146	hsa-mir-1278	49	TAGTACTGTGCATATCATCTAT	146	146	hsa-miR-1278	yes
335.21.3041	hsa-mir-128-1	49	TCACAGTGAACCGGTCTCTTT	3041	3041	hsa-miR-128	yes
14300.22.25	hsa-mir-1284	29	TCTATACAGACCCTGGCTTTTC	25	25	hsa-miR-1284	yes
6975.22.60	hsa-mir-1285-1	50	TCTGGGCAACAAAGTGAGACCT	60	131	hsa-miR-1285	No
10937.22.35	hsa-mir-1287	15	TGCTGGATCAGTGGTTCGAGTC	35	46	hsa-miR-1287	No
166625.22.2	hsa-mir-129-1	48	AAGCCCTTACCCCAAAAAGTAT	2	7	hsa-miR-129*	No
15529.22.22	hsa-mir-129-2	56	AAGCCCTTACCCCAAAAAGCAT	22	22	hsa-miR-129-3p	yes
31007.21.10	hsa-mir-129-2	14	CTTTTTGCGGTCTGGGCTTGC	10	10	hsa-miR-129-5p	yes
15431.24.23	hsa-mir-1291	13	TGGCCCTGACTGAAGACCAGCAGT	23	713	hsa-miR-1291	No
562319.25.1	hsa-mir-1292	2	TGGGAACGGGTTCCGGCAGACGCTG	1	11	hsa-miR-1292	No
2342.22.222	hsa-mir-1294	47	TGTGAGGTTGGCATTGTTGTCT	222	222	hsa-miR-1294	yes
157620.21.3	hsa-mir-1295	48	TTAGGCCGCAGATCTGGGTGA	3	8	hsa-miR-1295	No
99166.22.4	hsa-mir-1296	15	TTAGGGCCCTGGCTCCATCTCC	4	4	hsa-miR-1296	yes
158208.22.3	hsa-mir-1298	17	TTCATTCGGCTGTCCAGATGTA	3	26	hsa-miR-1298	No
46667.22.7	hsa-mir-1299	61	TTCTGGAATTCTGTGTGAGGGA	7	7	hsa-miR-1299	yes
6449.24.66	hsa-mir-1301	47	TTGCAGCTGCCTGGGAGTGACTTC	66	67	hsa-miR-1301	No
319477.18.1	hsa-mir-1306	54	ACGTTGGCTCTGGTGGTG	1	96	hsa-miR-1306	No
592.22.1426	hsa-mir-1307	79	ACTCGGCGTGGCGTCGGTCGTG	1426	1426	hsa-miR-1307	yes
569.22.1498	hsa-mir-130a	54	CAGTGCAATGTTAAAAGGGCAT	1498	1498	hsa-miR-130a	yes
424.22.2155	hsa-mir-130b	50	CAGTGCAATGATGAAAGGGCAT	2155	2155	hsa-miR-130b	yes
19182.22.17	hsa-mir-132	22	ACCGTGGCTTTCGATTGTTACT	17	17	hsa-miR-132*	yes
67432.22.5	hsa-mir-132	58	TAACAGTCTACAGCCATGGTCG	5	16	hsa-miR-132	No
257010.22.2	hsa-mir-1323	10	TCAAAACTGAGGGGCATTTTCT	2	5	hsa-miR-1323	No
279794.22.2	hsa-mir-133a-2	58	TTTGGTCCCCTTCAACCAGCTG	2	3	hsa-miR-133a	No
2694.22.186	hsa-mir-134	7	TGTGACTGGTTGACCAGAGGGG	186	186	hsa-miR-134	yes
256645.23.2	hsa-mir-135a-1	16	TATGGCTTTTTATTCCTATGTGA	2	36	hsa-miR-135a	No
22811.22.14	hsa-mir-136	48	CATCATCGTCTCAAATGAGTCT	14	14	hsa-miR-136*	yes
175952.23.2	hsa-mir-136	14	ACTCCATTTGTTTTGATGATGGA	2	17	hsa-miR-136	No
11270.23.33	hsa-mir-138-1	22	AGCTGGTGTTGTGAATCAGGCCG	33	33	hsa-miR-138	yes
7855.22.52	hsa-mir-139	6	TCTACAGTGCACGTGTCTCCAG	52	157	hsa-miR-139-5p	No
17900.22.19	hsa-mir-139	43	GGAGACGCGGCCCTGTTGGAGT	19	179	hsa-miR-139-3p	No
92.21.15555	hsa-mir-140	61	TACCACAGGGTAGAACCACGG	15555	316313	hsa-miR-140-3p	No
21512.22.15	hsa-mir-140	22	CAGTGGTTTTACCCTATGGTAG	15	29	hsa-miR-140-5p	No
707.23.1076	hsa-mir-142	51	TGTAGTGTTTCCTACTTTATGGA	1076	1445	hsa-miR-142-3p	No
822.21.857	hsa-mir-142	15	CATAAAGTAGAAAGCACTACT	857	1719	hsa-miR-142-5p	No
318.21.3321	hsa-mir-143	60	TGAGATGAAGCACTGTAGCTC	3321	5330	hsa-miR-143	No
21836.22.15	hsa-mir-143	26	GGTGCAGTGCTGCATCTCTGGT	15	106	hsa-miR-143*	No
1499.22.393	hsa-mir-144	14	GGATATCATCATATACTGTAAG	393	652	hsa-miR-144*	No

read-name	Reference	Position	IsomiRs	matching-abundance	highest-abundance	mature	match-with-highest
34913.20.9	hsa-mir-144	51	TACAGTATAGATGATGTACT	9	64	hsa-miR-144	No
3430.23.138	hsa-mir-145	15	GTCCAGTTTTCCCAGGAATCCCT	138	138	hsa-miR-145	yes
140462.22.3	hsa-mir-145	53	GGATTCCTGGAAATACTGTTCT	3	8	hsa-miR-145*	No
715.22.1062	hsa-mir-146a	20	TGAGAACTGAATTCCATGGGTT	1062	1062	hsa-miR-146a	yes
370.22.2658	hsa-mir-146b	8	TGAGAACTGAATTCCATAGGCT	2658	16229	hsa-miR-146b-5p	No
502315.22.1	hsa-mir-147b	48	GTGTGCGGAAATGCTTCTGCTA	1	1	hsa-miR-147b	yes
19.22.106559	hsa-mir-148a	43	TCAGTGCACTACAGAACTTTGT	106559	106559	hsa-miR-148a	yes
23805.22.13	hsa-mir-148a	5	AAAGTTCTGAGACACTCCGACT	13	13	hsa-miR-148a*	yes
147.22.8142	hsa-mir-148b	62	TCAGTGCATCACAGAACTTTGT	8142	8142	hsa-miR-148b	yes
151354.23.3	hsa-mir-149	14	TCTGGCTCCGTGTCTTCACTCCC	3	3	hsa-miR-149	yes
255.22.4228	hsa-mir-150	15	TCTCCCAACCCTTGTACCAGTG	4228	4228	hsa-miR-150	yes
12568.22.29	hsa-mir-150	50	CTGGTACAGGCCTGGGGGACAG	29	5366	hsa-miR-150*	No
992.21.674	hsa-mir-151	46	CTAGACTGAAGCTCCTTGAGG	674	674	hsa-miR-151-3p	yes
1165.21.546	hsa-mir-151	10	TCGAGGAGCTCACAGTCTAGT	546	607	hsa-miR-151-5p	No
284.21.3775	hsa-mir-152	53	TCAGTGCATGACAGAACTTGG	3775	3775	hsa-miR-152	yes
25393.22.12	hsa-mir-1537	39	AAAACCGTCTAGTTACAGTTGT	12	12	hsa-miR-1537	yes
22409.22.14	hsa-mir-154	50	AATCATACACGGTTGACCTATT	14	14	hsa-miR-154*	yes
728.23.1031	hsa-mir-155	3	TTAATGCTAATCGTGATAGGGGT	1031	1202	hsa-miR-155	No
218329.22.2	hsa-mir-155	42	CTCCTACATATTAGCATTAACA	2	2	hsa-miR-155*	yes
1149.22.556	hsa-mir-15a	13	TAGCAGCACATAATGGTTTGTG	556	6803	hsa-miR-15a	No
81985.22.4	hsa-mir-15a	50	CAGGCCATATTGTGCTGCCTCA	4	4	hsa-miR-15a*	yes
407.22.2266	hsa-mir-15b	19	TAGCAGCACATCATGGTTTACA	2266	4355	hsa-miR-15b	No
62481.22.5	hsa-mir-15b	57	CGAATCATTATTTGCTGCTCTA	5	18	hsa-miR-15b*	No
49.22.37492	hsa-mir-16-2	9	TAGCAGCACGTAAATATTGGCG	37492	37492	hsa-miR-16	yes
203929.22.2	hsa-mir-16-2	52	CCAATATTACTGTGCTGCTTTA	2	13	hsa-miR-16-2*	No
713.22.1062	hsa-mir-17	50	ACTGCAGTGAAGGCACTTGTAG	1062	1062	hsa-miR-17*	yes
1840.23.304	hsa-mir-17	13	CAAAGTGCTTACAGTGCAGGTAG	304	304	hsa-miR-17	yes
2913.22.168	hsa-mir-181a-1	63	ACCATCGACCGTTGATTGTACC	168	168	hsa-miR-181a*	yes
243.23.4392	hsa-mir-181a-2	38	AACATTCAACGCTGTCGGTGAGT	4392	4392	hsa-miR-181a	yes
3816.22.121	hsa-mir-181a-2	76	ACCACTGACCGTTGACTGTACC	121	160	hsa-miR-181a-2*	No
541.23.1587	hsa-mir-181b-1	35	AACATTCATTGCTGTCGGTGGGT	1587	2515	hsa-miR-181b	No
13178.22.27	hsa-mir-181c	26	AACATTCAACCTGTCGGTGAGT	27	1592	hsa-miR-181c	No
102614.22.3	hsa-mir-181c	64	AACCATCGACCGTTGAGTGGAC	3	185	hsa-miR-181c*	No
696.23.1096	hsa-mir-181d	35	AACATTCATTGTTGTCGGTGGGT	1096	3427	hsa-miR-181d	No
6890.24.61	hsa-mir-182	22	TTTGGCAATGGTAGAACTCACACT	61	105	hsa-miR-182	No
28959.22.11	hsa-mir-183	26	TATGGCACTGGTAGAATTCACT	11	11	hsa-miR-183	yes
10.22.154077	hsa-mir-185	14	TGGAGAGAAAGGCAGTTCCTGA	154077	154077	hsa-miR-185	yes
350391.22.1	hsa-mir-185	49	AGGGGCTGGCTTTCCTCTGGTC	1	111	hsa-miR-185*	No
512.22.1675	hsa-mir-186	11	CAAAGAATTCTCCTTTTGGGCT	1675	2640	hsa-miR-186	No
150836.22.3	hsa-mir-187	70	TCGTGTCTTGTGTTGCAGCCGG	3	10	hsa-miR-187	No
85367.21.4	hsa-mir-188	53	CTCCCACATGCAGGGTTTGCA	4	4	hsa-miR-188-3p	yes
203013.21.2	hsa-mir-188	14	CATCCCTTGCATGGTGGAGGG	2	7	hsa-miR-188-5p	No
48219.23.6	hsa-mir-18a	46	ACTGCCCTAAGTGCTCCTTCTGG	6	6	hsa-miR-18a*	yes
67477.23.5	hsa-mir-18a	5	TAAGGTGCATCTAGTGCAGATAG	5	17	hsa-miR-18a	No

read-name	Reference	Position	IsomiRs	matching-abundance	highest-abundance	mature	match-with-highest
5234.21.85	hsa-mir-1908	11	CGGCGGGGACGGCGATTGGTC	85	85	hsa-miR-1908	yes
406045.22.1	hsa-mir-1909	48	CGCAGGGGCCGGGTGCTCACCG	1	1	hsa-miR-1909	yes
53.23.34960	hsa-mir-191	15	CAACGGAATCCCAAAAGCAGCTG	34960	34960	hsa-miR-191	yes
139171.22.3	hsa-mir-191	57	GCTGCGCTTGGATTTCGTCCCC	3	21	hsa-miR-191*	No
62.21.28299	hsa-mir-192	23	CTGACCTATGAATTGACAGCC	28299	28299	hsa-miR-192	yes
220244.22.2	hsa-mir-192	66	CTGCCAATTCCATAGGTCACAG	2	5	hsa-miR-192*	No
536.22.1603	hsa-mir-193a	20	TGGGTCTTTGCGGGCGAGATGA	1603	1603	hsa-miR-193a-5p	yes
13189.22.27	hsa-mir-193a	54	AACTGGCCTACAAAGTCCCAGT	27	27	hsa-miR-193a-3p	yes
8181.22.49	hsa-mir-193b	13	CGGGGTTTTGAGGGCGAGATGA	49	49	hsa-miR-193b*	yes
165313.22.2	hsa-mir-193b	50	AACTGGCCCTCAAAGTCCCGCT	2	2	hsa-miR-193b	yes
8720.22.46	hsa-mir-194-1	14	TGTAACAGCAACTCCATGTGGA	46	46	hsa-miR-194	yes
12650.21.29	hsa-mir-195	14	TAGCAGCACAGAAATATTGGC	29	98	hsa-miR-195	No
5075.22.88	hsa-mir-196a-2	24	TAGGTAGTTTCATGTTGTTGGG	88	137	hsa-miR-196a	No
5346.22.83	hsa-mir-196b	14	TAGGTAGTTTCCTGTTGTTGGG	83	152	hsa-miR-196b	No
2343.22.222	hsa-mir-197	47	TTCACCACCTTCTCCACCCAGC	222	222	hsa-miR-197	yes
10323.23.37	hsa-mir-199a-1	5	CCCAGTGTTCAGACTACCTGTTC	37	37	hsa-miR-199a-5p	yes
79.22.20468	hsa-mir-199a-2	69	ACAGTAGTCTGCACATTGGTTA	20468	20468	hsa-miR-199a-3p	yes
8939.23.44	hsa-mir-199b	25	CCCAGTGTTTAGACTATCTGTTC	44	11548	hsa-miR-199b-5p	No
273867.23.2	hsa-mir-19a	18	TGTGCAAATCTATGCAAAACTGA	2	5	hsa-miR-19a	No
116838.23.3	hsa-mir-19b-1	15	AGTTTTGCAGGTTTGCATCCAGC	3	3	hsa-miR-19b-1*	yes
3101.23.155	hsa-mir-19b-2	61	TGTGCAAATCCATGCAAAACTGA	155	155	hsa-miR-19b	yes
61714.22.5	hsa-mir-200b	20	CATCTTACTGGGCAGCATTGGA	5	50	hsa-miR-200b*	No
93854.22.4	hsa-mir-200b	56	TAATACTGCCTGGTAATGATGA	4	7	hsa-miR-200b	No
2452.23.207	hsa-mir-200c	43	TAATACTGCCGGGTAATGATGGA	207	207	hsa-miR-200c	yes
66814.22.5	hsa-mir-203	64	GTGAAATGTTTAGGACCACTAG	5	5	hsa-miR-203	yes
25346.22.13	hsa-mir-204	32	TTCCCTTTGTCATCCTATGCCT	13	13	hsa-miR-204	yes
29276.22.11	hsa-mir-206	52	TGGAATGTAAGGAAGTGTGTGG	11	11	hsa-miR-206	yes
2451.23.207	hsa-mir-20a	7	TAAAGTGCTTATAGTGCAGGTAG	207	207	hsa-miR-20a	yes
109106.22.3	hsa-mir-20a	43	ACTGCATTATGAGCACTTAAAG	3	3	hsa-miR-20a*	yes
16294.23.21	hsa-mir-20b	5	CAAAGTGCTCATAGTGCAGGTAG	21	21	hsa-miR-20b	yes
177240.22.2	hsa-mir-20b	43	ACTGTAGTATGGGCACTTCCAG	2	7	hsa-miR-20b*	No
60.22.30955	hsa-mir-21	7	TAGCTTATCAGACTGATGTTGA	30955	49979	hsa-miR-21	No
12171.21.30	hsa-mir-21	45	CAACACCAGTCGATGGGCTGT	30	98	hsa-miR-21*	No
6941.22.60	hsa-mir-210	65	CTGTGCGTGTGACAGCGGCTGA	60	60	hsa-miR-210	yes
7660.22.54	hsa-mir-2110	7	TTGGGGAAACGGCCGCTGAGTG	54	483	hsa-miR-2110	No
4598.22.97	hsa-mir-2115	57	CATCAGAATTCATGGAGGCTAG	97	97	hsa-miR-2115*	yes
30208.22.10	hsa-mir-2115	20	AGCTTCCATGACTCCTGATGGA	10	10	hsa-miR-2115	yes
510274.21.1	hsa-mir-212	70	TAACAGTCTCCAGTCACGGCC	1	17	hsa-miR-212	No
74218.22.4	hsa-mir-214	70	ACAGCAGGCACAGACAGGCAGT	4	4	hsa-miR-214	yes
33350.21.9	hsa-mir-215	26	ATGACCTATGAATTGACAGAC	9	78	hsa-miR-215	No
515810.23.1	hsa-mir-217	34	TACTGCATCAGGAACTGATTGGA	1	7	hsa-miR-217	No
27697.22.11	hsa-mir-219-1	61	AGAGTTGAGTCTGGACGTCCCG	11	17	hsa-miR-219-1-3p	No
5852.22.74	hsa-mir-219-2	61	AGAATTGTGGCTGGACATCTGT	71	74	hsa-miR-219-2-3p	yes
326584.21.1	hsa-mir-219-2	18	TGATTGTCCAAACGCAATTCT	1	10	hsa-miR-219-5p	No

read-name	Reference	Position	IsomiRs	matching-abundance	highest-abundance	mature	match-with-highest
165.22.1863	hsa-mir-22	52	AAGCTGCCAGTTGAAGAACTGT	1863	1863	hsa-miR-22	yes
1049.22.634	hsa-mir-22	14	AGTTCTTCAGTGGCAAGCTTTA	634	634	hsa-miR-22*	yes
64.23.27953	hsa-mir-221	64	AGCTACATTGTCTGCTGGGTTTC	27953	27953	hsa-miR-221	yes
818.22.863	hsa-mir-221	24	ACCTGGCATACAATGTAGATTT	863	863	hsa-miR-221*	yes
683.21.1126	hsa-mir-222	68	AGCTACATCTGGCTACTGGGT	1126	2685	hsa-miR-222	No
423951.22.1	hsa-mir-222	30	CTCAGTAGCCAGTGTAGATCCT	1	7	hsa-miR-222*	No
39.22.51436	hsa-mir-223	67	TGTCAGTTTGTCAAATACCCCA	51436	55380	hsa-miR-223	No
22956.22.14	hsa-mir-223	25	CGTGTATTTGACAAGCTGAGTT	14	439	hsa-miR-223*	No
373604.21.1	hsa-mir-224	7	CAAGTCACTAGTGGTTCCGTT	1	349	hsa-miR-224	No
112417.24.3	hsa-mir-2277	17	AGCGCGGGCTGAGCGCTGCCAGTC	3	16	hsa-miR-2277-5p	No
5059.21.88	hsa-mir-2355	10	ATCCCCAGATACAATGGACAA	88	518	hsa-miR-2355-5p	No
5116.22.87	hsa-mir-2355	53	ATTGTCCTTGCTGTTTGGAGAT	87	87	hsa-miR-2355-3p	yes
118.21.11160	hsa-mir-23a	44	ATCACATTGCCAGGGATTTCC	11160	21009	hsa-miR-23a	No
2965.22.165	hsa-mir-23a	8	GGGGTTCCTGGGGATGGGATTT	165	165	hsa-miR-23a*	yes
2108.21.252	hsa-mir-23b	57	ATCACATTGCCAGGGATTACC	252	538	hsa-miR-23b	No
9024.22.44	hsa-mir-23b	19	TGGGTTCCTGGCATGCTGATTT	44	44	hsa-miR-23b*	yes
83.22.19599	hsa-mir-24-2	49	TGGCTCAGTTCAGCAGGAACAG	19599	19599	hsa-miR-24	yes
29.22.70760	hsa-mir-25	51	CATTGCACTTGTCTCGGTCTGA	70760	70760	hsa-miR-25	yes
652.21.1210	hsa-mir-25	13	AGGCGGAGACTTGGGCAATTG	1210	2669	hsa-miR-25*	No
26.22.80956	hsa-mir-26a-1	9	TTCAAGTAATCCAGGATAGGCT	80956	80956	hsa-miR-26a	yes
37709.22.8	hsa-mir-26a-1	48	CCTATTCTTGGTTACTTGCACG	8	8	hsa-miR-26a-1*	yes
399688.22.1	hsa-mir-26a-2	51	CCTATTCTTGATTACTTGTTTC	1	7217	hsa-miR-26a-2*	No
308.21.3409	hsa-mir-26b	11	TTCAAGTAATTCAGGATAGGT	3409	28934	hsa-miR-26b	No
401906.22.1	hsa-mir-26b	16	CCTGTTCTCCATTACTTGGCTC	1	7	hsa-miR-26b*	No
1842.21.304	hsa-mir-27a	50	TTCACAGTGGCTAAGTTCCGC	304	3286	hsa-miR-27a	No
7024.22.59	hsa-mir-27a	9	AGGGCTTAGCTGCTTGTGAGCA	59	59	hsa-miR-27a*	yes
1215.21.519	hsa-mir-27b	60	TTCACAGTGGCTAAGTTCTGC	519	519	hsa-miR-27b	yes
32923.22.9	hsa-mir-27b	18	AGAGCTTAGCTGATTGGTGAAC	9	45	hsa-miR-27b*	No
332.22.3073	hsa-mir-28	53	CACTAGATTGTGAGCTCCTGGA	3073	4019	hsa-miR-28-3p	No
3035.22.159	hsa-mir-28	13	AAGGAGCTCACAGTCTATTGAG	159	159	hsa-miR-28-5p	yes
188267.21.2	hsa-mir-296	13	AGGGCCCCCCCTCAATCCTGT	2	14	hsa-miR-296-5p	No
148996.22.3	hsa-mir-299	38	TATGTGGGATGGTAAACCGCTT	3	3	hsa-miR-299-3p	yes
272672.22.2	hsa-mir-299	6	TGGTTTACCGTCCCACATACAT	2	3	hsa-miR-299-5p	No
27.22.80708	hsa-mir-29a	41	TAGCACCATCTGAAATCGGTTA	80708	80708	hsa-miR-29a	yes
876.23.781	hsa-mir-29b-1	50	TAGCACCATTTGAAATCAGTGTT	781	781	hsa-miR-29b	yes
16481.21.21	hsa-mir-29b-1	9	GCTGGTTTCATATGGTGGTTTAGA	21	21	hsa-miR-29b-1*	yes
222024.22.2	hsa-mir-29b-2	10	CTGGTTTCACATGGTGGCTTAG	2	16	hsa-miR-29b-2*	No
177.22.6691	hsa-mir-29c	53	TAGCACCATTTGAAATCGGTTA	6691	6691	hsa-miR-29c	yes
68889.22.5	hsa-mir-29c	15	TGACCGATTTCTCCTGGTGTTC	5	37	hsa-miR-29c*	No
383404.23.1	hsa-mir-301a	50	CAGTGCAATAGTATTGTCAAAGC	1	4	hsa-miR-301a	No
523211.23.1	hsa-mir-3065	9	TCAACAAAATCACTGATGCTGGA	1	18	hsa-miR-3065-5p	No
229794.22.2	hsa-mir-3074	49	GATATCAGCTCAGTAGGCACCG	2	72818	hsa-miR-3074	No
9797.22.40	hsa-mir-30a	5	TGTAAACATCCTCGACTGGAAG	40	424	hsa-miR-30a	No
16424.22.21	hsa-mir-30a	46	CTTTCAGTCGGATGTTTGCAGC	21	21	hsa-miR-30a*	yes

read-name	Reference	Position	IsomiRs	matching-abundance	highest-abundance	mature	match-with-highest
1309.22.475	hsa-mir-30b	16	TGTAAACATCCTACACTCAGCT	475	475	hsa-miR-30b	yes
86102.22.4	hsa-mir-30b	54	CTGGGAGGTGGATGTTTACTTC	4	51	hsa-miR-30b*	No
1513.23.388	hsa-mir-30c-1	16	TGTAAACATCCTACACTCTCAGC	388	1967	hsa-miR-30c	No
23004.22.14	hsa-mir-30c-1	55	CTGGGAGAGGGTTGTTTACTCC	14	166	hsa-miR-30c-1*	No
587.22.1449	hsa-mir-30d	5	TGTAAACATCCCCGACTGGAAG	1449	14156	hsa-miR-30d	No
131916.22.3	hsa-mir-30d	45	CTTTCAGTCAGATGTTTGCTGC	3	8	hsa-miR-30d*	No
669.22.1153	hsa-mir-30e	58	CTTTCAGTCGGATGTTTACAGC	1153	3556	hsa-miR-30e*	No
2280.22.229	hsa-mir-30e	16	TGTAAACATCCTTGACTGGAAG	229	4393	hsa-miR-30e	No
1800.21.312	hsa-mir-31	7	AGGCAAGATGCTGGCATAGCT	312	377	hsa-miR-31	No
151115.22.3	hsa-mir-31	43	TGCTATGCCAACATATTGCCAT	3	3	hsa-miR-31*	yes
5007.21.89	hsa-mir-3120	50	CACAGCAAGTGTAGACAGGCA	89	89	hsa-miR-3120	yes
11746.22.24	hsa-mir-3121	16	TAAATAGAGTAGGCAAAGGACA	24	24	hsa-miR-3121	yes
251907.22.2	hsa-mir-3122	9	GTTGGGACAAGAGGACGGTCTT	2	8	hsa-miR-3122	No
10308.21.8	hsa-mir-3124	6	TTCGCGGGCGAAGGCAAAGTC	8	8	hsa-miR-3124	yes
360937.23.1	hsa-mir-3127	10	ATCAGGGCTTGTGGAATGGGAAG	1	9	hsa-miR-3127	No
537828.23.1	hsa-mir-3128	4	TCTGGCAAGTAAAAAACTCTCAT	1	3	hsa-miR-3128	No
4909.21.91	hsa-mir-3130-1	43	GCTGCACCGGAGACTGGGTAA	91	91	hsa-miR-3130-3p	yes
46452.24.7	hsa-mir-3138	47	TGTGGACAGTGAGGTAGAGGGAGT	7	57	hsa-miR-3138	No
113069.22.3	hsa-mir-3140	59	AGCTTTTGGGAATTCAGGTAGT	3	37	hsa-miR-3140	No
87923.19.4	hsa-mir-3141	9	GAGGGCGGGTGGAGGAGGA	4	13	hsa-miR-3141	No
203231.22.2	hsa-mir-3146	49	CATGCTAGGATAGAAAGAATGG	2	4	hsa-miR-3146	No
100232.23.4	hsa-mir-3149	50	TTTGTATGGATATGTGTGTGTAT	4	28	hsa-miR-3149	No
9358.21.42	hsa-mir-3150b	52	TGAGGAGATCGTCGAGGTTGG	42	42	hsa-miR-3150b	yes
490775.21.1	hsa-mir-3151	9	GGTGGGGCAATGGGATCAGGT	1	1	hsa-miR-3151	yes
19110.22.17	hsa-mir-3154	53	CAGAAGGGGAGTTGGGAGCAGA	17	22	hsa-miR-3154	No
123191.21.3	hsa-mir-3155	51	CCAGGCTCTGCAGTGGGAACT	3	3	hsa-miR-3155	yes
94606.22.4	hsa-mir-3159	9	TAGGATTACAAGTGTCGGCCAC	4	34	hsa-miR-3159	No
328742.22.1	hsa-mir-3160-1	53	AGAGCTGAGACTAGAAAGCCCA	1	1	hsa-miR-3160	yes
85792.23.4	hsa-mir-3161	9	CTGATAAGAACAGAGGCCCAGAT	4	13	hsa-miR-3161	No
7418.22.56	hsa-mir-3164	9	TGTGACTTTAAGGGAAATGGCG	56	56	hsa-miR-3164	yes
352385.22.1	hsa-mir-3165	9	AGGTGGATGCAATGTGACCTCA	1	1	hsa-miR-3165	yes
416173.22.1	hsa-mir-3175	9	CGGGGAGAGAACGCAGTGACGT	1	1	hsa-miR-3175	yes
58477.22.5	hsa-mir-3179-1	51	AGAAGGGGTGAAATTTAAACGT	5	5	hsa-miR-3179	yes
137206.22.3	hsa-mir-3183	9	GCCTCTCTCGGAGTCGCTCGGA	3	14	hsa-miR-3183	No
149370.21.3	hsa-mir-3186	52	TCACGCGGAGAGATGGCTTTG	3	18	hsa-miR-3186-3p	No
40150.23.8	hsa-mir-3190	46	TGTGGAAGGTAGACGGCCAGAGA	8	39	hsa-miR-3190	No
155793.23.3	hsa-mir-3191	45	TGGGGACGTAGCTGGCCAGACAG	3	3	hsa-miR-3191	yes
35246.23.9	hsa-mir-3192	9	TCTGGGAGGTTGTAGCAGTGGAA	9	9	hsa-miR-3192	yes
38990.22.8	hsa-mir-3198	48	GTGGAGTCCTGGGGAATGGAGA	8	26	hsa-miR-3198	No
45382.22.7	hsa-mir-32	5	TATTGCACATTACTAAGTTGCA	7	30	hsa-miR-32	No
198668.22.2	hsa-mir-32	46	CAATTTAGTGTGTGTGATATTT	2	17	hsa-miR-32*	No
22412.22.14	hsa-mir-3200	12	AATCTGAGAAGGCGCACAAGGT	14	59	hsa-miR-3200-5p	No
376092.22.1	hsa-mir-3200	53	CACCTTGCGCTACTCAGGTCTG	1	1	hsa-miR-3200-3p	yes
268371.22.2	hsa-mir-3202-1	8	TGGAAGGGAGAAGAGCTTTAAT	2	8	hsa-miR-3202	No
13.22.148957	hsa-mir-320a	47	AAAAGCTGGGTTGAGAGGGCGA	148957	148957	hsa-miR-320a	yes

read-name	Reference	Position	IsomiRs	matching-abundance	highest-abundance	mature	match-with-highest
457.22.1917	hsa-mir-320b-2	71	AAAAGCTGGGTTGAGAGGGCAA	1917	1922	hsa-miR-320b	No
7541.20.54	hsa-mir-320c-2	30	AAAAGCTGGGTTGAGAGGGT	54	54	hsa-miR-320c	yes
16748.19.20	hsa-mir-320d-2	29	AAAAGCTGGGTTGAGAGGA	20	20	hsa-miR-320d	yes
26023.21.12	hsa-mir-323	50	CACATTACACGGTCGACCTCT	12	25	hsa-miR-323-3p	No
2502.23.202	hsa-mir-324	15	CGCATCCCCTAGGGCATTGGTGT	202	202	hsa-miR-324-5p	yes
176587.20.2	hsa-mir-324	52	ACTGCCCCAGGTGCTGCTGG	2	30	hsa-miR-324-3p	No
83417.20.4	hsa-mir-326	59	CCTCTGGGCCCTTCCTCCAG	4	152	hsa-miR-326	No
10345.22.37	hsa-mir-328	17	CTGGCCCTCTCTGCCCTTCCGT	37	37	hsa-miR-328	yes
23820.22.13	hsa-mir-329-1	49	AACACACCTGGTTAACCTCTTT	13	13	hsa-miR-329	yes
537.23.1602	hsa-mir-330	56	GCAAAGCACACGGCCTGCAGAGA	1602	1602	hsa-miR-330-3p	yes
1197.21.530	hsa-mir-331	60	GCCCCTGGGCCTATCCTAGAA	530	530	hsa-miR-331-3p	yes
4423.23.102	hsa-mir-335	15	TCAAGAGCAATAACGAAAAATGT	102	108	hsa-miR-335	No
32345.22.10	hsa-mir-335	51	TTTTTCATTATTGCTCCTGACC	10	10	hsa-miR-335*	yes
129579.22.3	hsa-mir-337	60	CTCCTATATGATGCCTTTCTTC	3	5	hsa-miR-337-3p	No
35148.22.9	hsa-mir-338	41	TCCAGCATCAGTGATTTTGTTG	9	242	hsa-miR-338-3p	No
2478.23.205	hsa-mir-339	49	TGAGCGCCTCGACGACAGAGCCG	205	226	hsa-miR-339-3p	No
9755.23.40	hsa-mir-339	14	TCCCTGTCCTCCAGGAGCTCACG	40	60	hsa-miR-339-5p	No
1201.21.528	hsa-mir-33a	5	GTGCATTGTAGTTGCATTGCA	528	528	hsa-miR-33a	yes
26628.20.12	hsa-mir-33b	15	GTGCATTGCTGTTGCATTGC	12	12	hsa-miR-33b	yes
324.22.3213	hsa-mir-340	15	TTATAAAGCAATGAGACTGATT	3213	3213	hsa-miR-340	yes
268.23.4037	hsa-mir-342	60	TCTCACACAGAAATCGCACCCGT	4037	4037	hsa-miR-342-3p	yes
2294.21.227	hsa-mir-342	18	AGGGGTGCTATCTGTGATTGA	227	624	hsa-miR-342-5p	No
15830.22.22	hsa-mir-345	17	GCTGACTCCTAGTCCAGGGCTC	22	107	hsa-miR-345	No
69852.22.5	hsa-mir-34a	21	TGGCAGTGTCTTAGCTGGTTGT	5	9	hsa-miR-34a	No
6041.23.71	hsa-mir-34c	12	AGGCAGTGTAGTTAGCTGATTGC	71	71	hsa-miR-34c-5p	yes
4384.23.103	hsa-mir-3605	21	TGAGGATGGATAGCAAGGAAGCC	103	123	hsa-miR-3605-5p	No
371393.21.1	hsa-mir-3609	50	CAAAGTGATGAGTAATACTGGCTG	1	1	hsa-miR-3609	yes
2045.22.263	hsa-mir-361	5	TTATCAGAATCTCCAGGGGTAC	263	564	hsa-miR-361-5p	No
2237.23.234	hsa-mir-361	11	TCCCCCAGGTGTGATTCTGATTT	234	234	hsa-miR-361-3p	yes
12381.22.30	hsa-mir-3613	15	TGTTGTACTTTTTTTTTTGTTC	30	30	hsa-miR-3613-5p	yes
2544.23.199	hsa-mir-3614	52	TAGCCTTCAGATCTTGGTGTTTT	199	1011	hsa-miR-3614-3p	No
50409.23.6	hsa-mir-3614	14	CCACTTGGATCTGAAGGCTGCCC	6	116	hsa-miR-3614-5p	No
45659.21.7	hsa-mir-3615	50	TCTCTCGGCTCCTCGCGGCTC	7	55	hsa-miR-3615	No
29605.22.10	hsa-mir-3617	9	AAAGACATAGTTGCAAGATGGG	10	10	hsa-miR-3617	yes
16147.24.21	hsa-mir-362	4	AATCCTTGGAACCTAGGTGTGAGT	21	21	hsa-miR-362-5p	yes
72189.22.4	hsa-mir-362	41	AACACACCTATTCAAGGATTCA	4	4	hsa-miR-362-3p	yes
2536.22.199	hsa-mir-363	49	AATTGCACGGTATCCATCTGTA	199	713	hsa-miR-363	No
177867.23.2	hsa-mir-3647	61	AGAAAATTTTTGTGTGTCTGATC	2	23	hsa-miR-3647-3p	No
3081.22.156	hsa-mir-365-2	67	TAATGCCCCTAAAAATCCTTAT	156	156	hsa-miR-365	yes
33164.22.9	hsa-mir-365-2	28	AGGGACTTTCAGGGGCAGCTGT	9	18	hsa-miR-365*	No
214206.18.2	hsa-mir-3652	0	CGGCTGGAGGTGTGAGGA	2	2	hsa-miR-3652	yes
161902.22.2	hsa-mir-3667	8	AAAGACCCATTGAGGAGAAGGT	2	7	hsa-miR-3667-5p	No
316216.22.1	hsa-mir-3667	44	ACCTTCCTCTCCATGGGTCTTT	1	2	hsa-miR-3667-3p	No
62319.20.5	hsa-mir-3676	59	CCGTGTTTCCCCCACGCTTT	5	189	hsa-miR-3676	No

read-name	Reference	Position	IsomiRs	matching-abundance	highest-abundance	mature	match-with-highest
38027.22.8	hsa-mir-3677	38	CTCGTGGGCTCTGGCCACGGCC	8	17	hsa-miR-3677	No
13892.23.26	hsa-mir-3679	5	TGAGGATATGGCAGGGAAGGGGA	26	33	hsa-miR-3679-5p	No
2560.21.197	hsa-mir-369	43	AATAATACATGGTTGATCTTT	197	197	hsa-miR-369-3p	yes
952.23.704	hsa-mir-3690	9	ACCTGGACCCAGCGTAGACAAAG	704	704	hsa-miR-3690	yes
13780.22.26	hsa-mir-370	47	GCCTGCTGGGGTGGAACCTGGT	26	185	hsa-miR-370	No
1342.22.456	hsa-mir-374a	41	CTTATCAGATTGTATTGTAATT	456	456	hsa-miR-374a*	yes
3198.22.149	hsa-mir-374a	11	TTATAATACAACCTGATAAGTG	149	241	hsa-miR-374a	No
709.22.1070	hsa-mir-374b	10	ATATAATACAACCTGCTAAGTG	1070	1070	hsa-miR-374b	yes
8334.22.48	hsa-mir-374b	40	CTTAGCAGGTTGTATTATCATT	48	48	hsa-miR-374b*	yes
40111.22.8	hsa-mir-375	39	TTTGTTCGTTCGGCTCGCGTGA	8	8	hsa-miR-375	yes
193146.21.2	hsa-mir-376a-1	43	ATCATAGAGGAAAATCCACGT	2	7	hsa-miR-376a	No
60359.22.5	hsa-mir-376b	61	ATCATAGAGGAAAATCCATGTT	5	8	hsa-miR-376b	No
10729.21.35	hsa-mir-376c	42	AACATAGAGGAAATTCCACGT	35	35	hsa-miR-376c	yes
42245.22.7	hsa-mir-377	44	ATCACACAAAGGCAACTTTTGT	7	7	hsa-miR-377	yes
58726.22.5	hsa-mir-377	6	AGAGGTTGCCCTTGGTGAATTC	5	5	hsa-miR-377*	yes
187.21.6241	hsa-mir-378	42	ACTGGACTTGGAGTCAGAAGG	6241	11217	hsa-miR-378	No
425150.22.1	hsa-mir-378	4	CTCCTGACTCCAGGTCCTGTGT	1	21	hsa-miR-378*	No
6791.21.62	hsa-mir-379	5	TGGTAGACTATGGAACGTAGG	62	62	hsa-miR-379	yes
521027.22.1	hsa-mir-381	48	TATACAAGGGCAAGCTCTCTGT	1	21	hsa-miR-381	No
1030.22.646	hsa-mir-382	10	GAAGTTGTTCGTGGTGGATTCG	646	646	hsa-miR-382	yes
35659.22.9	hsa-mir-3909	70	TGTCCTCTAGGGCCTGCAGTCT	9	11	hsa-miR-3909	No
40595.20.7	hsa-mir-3910-2	49	AAAGGCATAAAACCAAGACA	7	23	hsa-miR-3910	No
157156.22.3	hsa-mir-3911	11	TGTGTGGATCCTGGAGGAGGCA	3	10	hsa-miR-3911	No
55924.22.6	hsa-mir-3913-1	22	TTTGGGACTGATCTTGATGTCT	6	6	hsa-miR-3913	yes
3940.26.116	hsa-mir-3916	19	AAGAGGAAGAAATGGCTGGTTCTCAG	116	116	hsa-miR-3916	yes
171528.21.2	hsa-mir-3918	18	ACAGGGCCGCAGATGGAGACT	2	2	hsa-miR-3918	yes
58187.22.5	hsa-mir-3920	50	ACTGATTATCTTAACTCTCTGA	5	5	hsa-miR-3920	yes
24712.22.13	hsa-mir-3928	36	GGAGGAACCTTGGAGCTTCGGC	13	14	hsa-miR-3928	No
24999.22.13	hsa-mir-3934	24	TCAGGTGTGGAAACTGAGGCAG	13	28	hsa-miR-3934	No
7026.23.59	hsa-mir-409	14	AGGTTACCCGAGCAACTTTGCAT	59	59	hsa-miR-409-5p	yes
7475.22.55	hsa-mir-409	46	GAATGTTGCTCGGTGAACCCCT	55	86	hsa-miR-409-3p	No
2826.21.175	hsa-mir-410	49	AATATAACACAGATGGCCTGT	175	175	hsa-miR-410	yes
39339.21.8	hsa-mir-411	15	TAGTAGACCGTATAGCGTACG	8	8	hsa-miR-411	yes
7226.23.57	hsa-mir-421	47	ATCAACAGACATTAATTGGGCGC	57	184	hsa-miR-421	No
12.23.152535	hsa-mir-423	16	TGAGGGGCAGAGAGCGAGACTTT	152535	152535	hsa-miR-423-5p	yes
447.23.1966	hsa-mir-423	52	AGCTCGGTCTGAGGCCCCTCAGT	1966	1966	hsa-miR-423-3p	yes
1473.22.402	hsa-mir-424	10	CAGCAGCAATTCATGTTTTGAA	402	2151	hsa-miR-424	No
2475.21.205	hsa-mir-424	47	CAAAACGTGAGGCGCTGCTAT	205	205	hsa-miR-424*	yes
802.23.890	hsa-mir-425	13	AATGACACGATCACTCCCGTTGA	890	890	hsa-miR-425	yes
7030.22.59	hsa-mir-425	54	ATCGGGAATGTCGTGTCCGCCC	59	519	hsa-miR-425*	No
557949.17.1	hsa-mir-4306	64	TGGAGAGAAAGGCAGTA	1	27	hsa-miR-4306	No
10463.21.37	hsa-mir-431	19	TGTCTTGCAGGCCGTCATGCA	37	37	hsa-miR-431	yes
382810.22.1	hsa-mir-431	62	CAGGTCGTCTTGCAGGGCTTCT	1	6	hsa-miR-431*	No
2007.23.271	hsa-mir-432	13	TCTTGGAGTAGGTCATTGGGTGG	271	544	hsa-miR-432	No
1816.22.310	hsa-mir-433	63	ATCATGATGGGCTCCTCGGTGT	310	310	hsa-miR-433	yes

read-name	Reference	Position	IsomiRs	matching-abundance	highest-abundance	mature	match-with-highest
29538.22.11	hsa-mir-450a-2	22	TTTTGCGATGTGTTCCTAATAT	11	11	hsa-miR-450a	yes
46801.22.7	hsa-mir-450b	10	TTTTGCAATATGTTCCTGAATA	7	16	hsa-miR-450b-5p	No
230.22.4723	hsa-mir-451	16	AAACCGTTACCATTACTGAGTT	4723	8415	hsa-miR-451	No
23844.22.13	hsa-mir-452	13	AACTGTTTGCAGAGGAAACTGA	13	61	hsa-miR-452	No
29917.22.10	hsa-mir-454	23	ACCCTATCAATATTGTCTCTGC	10	10	hsa-miR-454*	yes
256170.23.2	hsa-mir-454	63	TAGTGCAATATTGCTTATAGGGT	2	19	hsa-miR-454	No
5192.22.86	hsa-mir-484	7	TCAGGCTCAGTCCCCTCCCGAT	86	86	hsa-miR-484	yes
14033.22.25	hsa-mir-485	8	AGAGGCTGGCCGTGATGAATTC	25	28	hsa-miR-485-5p	No
14720.22.24	hsa-mir-485	45	GTCATACACGGCTCTCCTCTCT	24	24	hsa-miR-485-3p	yes
337.22.3029	hsa-mir-486	3	TCCTGTACTGAGCTGCCCCGAG	3029	3029	hsa-miR-486-5p	yes
5511.21.80	hsa-mir-486	45	CGGGGCAGCTCAGTACAGGAT	80	1164	hsa-miR-486-3p	No
47553.22.6	hsa-mir-487a	48	AATCATACAGGGACATCCAGTT	6	15	hsa-miR-487a	No
4937.22.90	hsa-mir-487b	50	AATCGTACAGGGTCATCCACTT	90	90	hsa-miR-487b	yes
79144.22.4	hsa-mir-491	15	AGTGGGGAACCCTTCCATGAGG	4	15	hsa-miR-491-5p	No
222638.22.2	hsa-mir-491	49	CTTATGCAAGATTCCCTTCTAC	2	2	hsa-miR-491-3p	yes
10011.22.39	hsa-mir-493	15	TTGTACATGGTAGGCTTTCATT	39	39	hsa-miR-493*	yes
263216.22.2	hsa-mir-493	56	TGAAGGTCTACTGTGTGCCAGG	2	4	hsa-miR-493	No
29105.22.11	hsa-mir-494	47	TGAAACATACACGGGAAACCTC	11	245	hsa-miR-494	No
782.22.920	hsa-mir-495	19	AAACAAACATGGTGCACTTCTT	920	1166	hsa-miR-495	No
97193.22.4	hsa-mir-496	55	TGAGTATTACATGGCCAATCTC	4	8	hsa-miR-496	No
28083.21.11	hsa-mir-497	23	CAGCAGCACACTGTGGTTTGT	11	12	hsa-miR-497	No
2821.21.176	hsa-mir-499	32	TTAAGACTTGCAGTGATGTTT	176	176	hsa-miR-499-5p	yes
37191.22.8	hsa-mir-500a	51	ATGCACCTGGGCAAGGATTCTG	8	183	hsa-miR-500a*	No
5486.22.80	hsa-mir-501	50	AATGCACCCGGGCAAGGATTCT	80	80	hsa-miR-501-3p	yes
302678.22.1	hsa-mir-501	13	AATCCTTTGTCCCTGGGTGAGA	1	9	hsa-miR-501-5p	No
4150.22.109	hsa-mir-502	51	AATGCACCTGGGCAAGGATTCA	109	109	hsa-miR-502-3p	yes
13851.23.26	hsa-mir-503	5	TAGCAGCGGGAACAGTTCTGCAG	26	290	hsa-miR-503	No
12451.22.29	hsa-mir-504	12	AGACCCTGGTCTGCACTCTATC	29	82	hsa-miR-504	No
4609.22.97	hsa-mir-505	14	GGGAGCCAGGAAGTATTGATGT	97	259	hsa-miR-505*	No
19496.22.17	hsa-mir-505	50	CGTCAACACTTGCTGGTTTCCT	17	21	hsa-miR-505	No
266795.22.2	hsa-mir-509-2	54	TGATTGGTACGTCTGTGGGTAG	2	2	hsa-miR-509-3p	yes
67690.22.5	hsa-mir-509-3	9	TACTGCAGACGTGGCAATCATG	5	5	hsa-miR-509-3-5p	yes
578964.22.1	hsa-mir-513c	13	TTCTCAAGGAGGTGTCGTTTAT	1	1	hsa-miR-513c	yes
682.22.1127	hsa-mir-532	19	CATGCCTTGAGTGTAGGACCGT	1127	1127	hsa-miR-532-5p	yes
1867.22.297	hsa-mir-532	56	CCTCCCACACCCAAGGCTTGCA	297	297	hsa-miR-532-3p	yes
101631.25.3	hsa-mir-541	9	AAAGGATTCTGCTGTCGGTCCCACT	3	6	hsa-miR-541*	No
9800.22.40	hsa-mir-542	52	TGTGACAGATTGATAACTGAAA	40	40	hsa-miR-542-3p	yes
3970.22.115	hsa-mir-543	46	AAACATTCGCGGTGCACTTCTT	115	115	hsa-miR-543	yes
119681.22.3	hsa-mir-548a-3	60	CAAAACTGGCAATTACTTTTGC	3	13	hsa-miR-548a-3p	No
161038.22.2	hsa-mir-548a-3	24	AAAAGTAATTGCGAGTTTTACC	2	2	hsa-miR-548a-5p	yes
35954.22.8	hsa-mir-548c	24	AAAAGTAATTGCGGTTTTTGCC	8	114	hsa-miR-548c-5p	No
161054.22.2	hsa-mir-548d-1	24	AAAAGTAATTGTGGTTTTTGCC	2	8	hsa-miR-548d-5p	No
197267.22.2	hsa-mir-548d-1	60	CAAAAACCACAGTTTCTTTTGC	2	5	hsa-miR-548d-3p	No

read-name	Reference	Position	IsomiRs	matching-abundance	highest-abundance	mature	match-with-highest
923.22.736	hsa-mir-548c	52	AAAAACTGAGACTACTTTTGCA	736	736	hsa-miR-548c	yes
616.22.1346	hsa-mir-548j	28	AAAAGTAATTGCGGTCTTTGGT	1346	1346	hsa-miR-548j	yes
1489.22.394	hsa-mir-548k	31	AAAAGTACTTGCGGATTTTGCT	394	394	hsa-miR-548k	yes
14935.22.23	hsa-mir-548l	14	AAAAGTATTTGCGGGTTTTGTC	23	54	hsa-miR-548l	No
24238.22.13	hsa-mir-548n	9	CAAAAGTAATTGTGGATTTTGT	13	13	hsa-miR-548n	yes
9109.22.43	hsa-mir-548o	86	CCAAAACTGCAGTTACTTTTGC	43	43	hsa-miR-548o	yes
12901.21.28	hsa-mir-548t	9	CAAAAGTGATCGTGGTTTTTG	28	28	hsa-miR-548t	yes
21462.23.15	hsa-mir-54u	48	CAAAGACTGCAATTACTTTTGCG	15	15	hsa-miR-548u	yes
25309.22.13	hsa-mir-550a-1	60	TGTCTTACTCCCTCAGGCACAT	13	13	hsa-miR-550a*	yes
52327.21.6	hsa-mir-551a	60	GCGACCCACTCTTGGTTTCCA	6	6	hsa-miR-551a	yes
3748.22.124	hsa-mir-574	60	CACGCTCATGCACACACCCACA	124	124	hsa-miR-574-3p	yes
69446.23.5	hsa-mir-574	24	TGAGTGTGTGTGTGTGAGTGTGT	5	16	hsa-miR-574-5p	No
4157.22.109	hsa-mir-576	15	ATTCTAATTTCTCCACGTCTTT	109	109	hsa-miR-576-5p	yes
5617.22.78	hsa-mir-576	54	AAGATGTGGAAAAATTGGAATC	78	78	hsa-miR-576-3p	yes
254947.21.2	hsa-mir-577	15	TAGATAAAATATTGGTACCTG	2	40	hsa-miR-577	No
55538.23.6	hsa-mir-582	15	TTACAGTTGTTCAACCAGTTACT	6	13	hsa-miR-582-5p	No
147080.22.3	hsa-mir-582	52	TAACTGGTTGAACAACTGAACC	3	22	hsa-miR-582-3p	No
13135.22.28	hsa-mir-584	15	TTATGGTTTGCCTGGGACTGAG	28	4884	hsa-miR-584	No
14816.22.24	hsa-mir-589	23	TGAGAACCACGTCTGCTCTGAG	24	144	hsa-miR-589	No
17265.24.20	hsa-mir-589	60	TCAGAACAAATGCCGGTTCCCAGA	20	100	hsa-miR-589*	No
8339.22.48	hsa-mir-590	15	GAGCTTATTCATAAAAGTGCAG	48	48	hsa-miR-590-5p	yes
67502.21.5	hsa-mir-590	55	TAATTTTATGTATAAGCTAGT	5	7	hsa-miR-590-3p	No
678.22.1129	hsa-mir-598	60	TACGTCATCGTTGTCATCGTCA	1129	1129	hsa-miR-598	yes
35985.23.8	hsa-mir-618	15	AAACTCTACTTGTCCTTCTGAGT	8	10	hsa-miR-618	No
118677.22.3	hsa-mir-624	15	TAGTACCAGTACCTTGTGTTCA	3	3	hsa-miR-624*	yes
6131.22.70	hsa-mir-625	51	GACTATAGAACTTTCCCCCTCA	70	70	hsa-miR-625*	yes
24134.21.13	hsa-mir-625	14	AGGGGGAAAGTTCTATAGTCC	13	47	hsa-miR-625	No
9166.22.41	hsa-mir-628	22	ATGCTGACATATTTACTAGAGG	41	41	hsa-miR-628-5p	yes
535348.21.1	hsa-mir-628	60	TCTAGTAAGAGTGGCAGTCGA	1	3	hsa-miR-628-3p	No
1920.21.287	hsa-mir-629	21	TGGGTTTACGTTGGGAGAACT	287	287	hsa-miR-629	yes
53495.22.6	hsa-mir-629	60	GTTCTCCCAACGTAAGCCCAGC	6	18	hsa-miR-629*	No
114927.25.3	hsa-mir-638	15	AGGGATCGCGGGCGGGTGGCGGCCT	3	3	hsa-miR-638	yes
161897.24.2	hsa-mir-641	15	AAAGACATAGGATAGAGTCACCTC	2	12	hsa-miR-641	No
495510.22.1	hsa-mir-642a	15	GTCCCTCTCCAAATGTGTCTTG	1	12	hsa-miR-642a	No
326720.22.1	hsa-mir-642b	46	AGACACATTTGGAGAGGGACCC	1	4	hsa-miR-642b	No
32313.22.10	hsa-mir-651	15	TTTAGGATAAGCTTGACTTTTG	10	11	hsa-miR-651	No
9055.21.43	hsa-mir-652	60	AATGGCGCCACTAGGGTTGTG	43	312	hsa-miR-652	No
54041.22.6	hsa-mir-654	50	TATGTCTGCTGACCATCACCTT	6	6	hsa-miR-654-3p	yes
156275.22.3	hsa-mir-654	15	TGGTGGGCCGCAGAACATGTGC	3	5	hsa-miR-654-5p	No
37081.22.8	hsa-mir-655	60	ATAATACATGGTTAACCTCTTT	8	20	hsa-miR-655	No
302326.21.1	hsa-mir-656	42	AATATTATACAGTCAACCTCT	1	1	hsa-miR-656	yes
4537.22.99	hsa-mir-660	15	TACCCATTGCATATCGGAGTTG	99	138	hsa-miR-660	No
4766.23.94	hsa-mir-664	18	TATTCATTTATCCCCAGCCTACA	94	94	hsa-miR-664	yes
7560.24.54	hsa-mir-664	10	ACTGGCTAGGGAAAATGATTGGAT	54	1441	hsa-miR-664*	No

read-name	Reference	Position	IsomiRs	matching-abundance	highest-abundance	mature	match-with-highest
59325.23.5	hsa-mir-671	28	AGGAAGCCCTGGAGGGGCTGGAG	5	18	hsa-miR-671-5p	No
95495.21.4	hsa-mir-671	67	TCCGGTTCTCAGGGCTCCACC	4	6	hsa-miR-671-3p	No
566463.23.1	hsa-mir-675	9	TGGTGCGGAGAGGGCCCACAGTG	1	1	hsa-miR-675	yes
86244.21.4	hsa-mir-676	42	CTGTCCTAAGGTTGTTGAGTT	4	4	hsa-miR-676	yes
3589.22.130	hsa-mir-7-1	65	CAACAAATCACAGTCTGCCATA	130	130	hsa-miR-7-1*	yes
1802.23.312	hsa-mir-7-2	31	TGGAAGACTAGTGATTTTGTTGT	312	392	hsa-miR-7	No
47431.23.6	hsa-mir-708	10	AAGGAGCTTACAATCTAGCTGGG	6	6	hsa-miR-708	yes
9355.17.42	hsa-mir-720	26	TCTCGCTGGGGCCTCCA	42	115	hsa-miR-720	No
159.22.7545	hsa-mir-744	10	TGCGGGGCTAGGGCTAACAGCA	7545	7545	hsa-miR-744	yes
13335.20.27	hsa-mir-760	48	CGGCTCTGGGTCTGTGGGGA	27	145	hsa-miR-760	No
97856.21.4	hsa-mir-765	68	TGGAGGAGAAGGAAGGTGATG	4	12	hsa-miR-765	No
5906.22.73	hsa-mir-766	64	ACTCCAGCCCCACAGCCTCAGC	73	73	hsa-miR-766	yes
8539.22.47	hsa-mir-769	29	TGAGACCTCTGGGTTCTGAGCT	47	47	hsa-miR-769-5p	yes
9712.23.40	hsa-mir-769	68	CTGGGATCTCCGGGGTCTTGGTT	40	40	hsa-miR-769-3p	yes
6135.21.70	hsa-mir-873	10	GCAGGAACTTGTGAGTCTCCT	70	70	hsa-miR-873	yes
30938.22.10	hsa-mir-874	46	CTGCCCTGGCCCGAGGGACCGA	10	11	hsa-miR-874	No
39977.22.8	hsa-mir-876	10	TGGATTTCTTTGTGAATCACCA	8	8	hsa-miR-876-5p	yes
4655.20.96	hsa-mir-877	0	GTAGAGGAGATGGCGCAGGG	96	1096	hsa-miR-877	No
4835.21.93	hsa-mir-889	48	TTAATATCGGACAACCATTGT	93	93	hsa-miR-889	yes
60140.22.5	hsa-mir-9-1	54	ATAAAGCTAGATAACCGAAAGT	5	5	hsa-miR-9*	yes
3145.23.152	hsa-mir-9-2	15	TCTTTGGTTATCTAGCTGTATGA	152	152	hsa-miR-9	yes
571.23.1488	hsa-mir-92a-1	10	AGGTTGGGATCGGTTGCAATGCT	1488	2387	hsa-miR-92a-1*	No
71.22.24347	hsa-mir-92a-2	47	TATTGCACTTGTCCCGGCCTGT	24347	24347	hsa-miR-92a	yes
4813.22.93	hsa-mir-92b	60	TATTGCACTCGTCCCGGCCTCC	93	142	hsa-miR-92b	No
36947.22.8	hsa-mir-92b	19	AGGGACGGGACGCGGTGCAGTG	8	29	hsa-miR-92b*	No
276.23.3921	hsa-mir-93	10	CAAAGTGCTGTTCGTGCAGGTAG	3921	3921	hsa-miR-93	yes
14487.22.24	hsa-mir-93	49	ACTGCTGAGCTAGCACTTCCCG	24	74	hsa-miR-93*	No
82632.23.4	hsa-mir-935	55	CCAGTTACCGCTTCCGCTACCGC	4	24	hsa-miR-935	No
271533.24.2	hsa-mir-939	14	TGGGGAGCTGAGGCTCTGGGGGTG	2	3	hsa-miR-939	No
36190.21.8	hsa-mir-940	59	AAGGCAGGGCCCCCGCTCCCC	8	8	hsa-miR-940	yes
1669.23.346	hsa-mir-941-3	70	CACCCGGCTGTGTGCACATGTGC	346	346	hsa-miR-941	yes
262666.22.2	hsa-mir-942	12	TCTTCTCTGTTTTGGCCATGTG	2	17	hsa-miR-942	No
29634.22.10	hsa-mir-944	53	AAATTATTGTACATCGGATGAG	10	10	hsa-miR-944	yes
13138.22.28	hsa-mir-95	48	TTCAACGGGTATTTATTGAGCA	28	28	hsa-miR-95	yes
580.22.1456	hsa-mir-98	21	TGAGGTAGTAAGTTGTATTGTT	1456	1456	hsa-miR-98	yes
4584.22.97	hsa-mir-99a	12	AACCCGTAGATCCGATCTTGTG	97	163	hsa-miR-99a	No
723.22.1042	hsa-mir-99b	6	CACCCGTAGAACCGACCTTGCG	1042	1042	hsa-miR-99b	yes
81171.22.4	hsa-mir-99b	44	CAAGCTCGTGTCTGTGGGTCCG	4	72	hsa-miR-99b*	No

Table A.7: **Identification of most abundant IsomiRs for N5 sample**

read-name	Reference	Position	IsomiRs	matching-abundance	highest-abundance	mature	match-with-highest
18.22.98863	hsa-let-7a-2	4	TGAGGTAGTAGGTTGTATAGTT	98863	98863	hsa-let-7a	yes
199993.21.2	hsa-let-7a-3	51	CTATACAATCTACTGTCTTTC	2	14	hsa-let-7a*	No
1.22.593030	hsa-let-7b	5	TGAGGTAGTAGGTTGTGTGGTT	593030	593030	hsa-let-7b	yes
107.22.11823	hsa-let-7c	10	TGAGGTAGTAGGTTGTATGGTT	11823	11823	hsa-let-7c	yes
54.22.29141	hsa-let-7d	7	AGAGGTAGTAGGTTGCATAGTT	29141	34781	hsa-let-7d	No
1785.22.267	hsa-let-7d	61	CTATACGACCTGCTGCCTTTCT	267	267	hsa-let-7d*	yes
544.22.1342	hsa-let-7e	7	TGAGGTAGGAGGTTGTATAGTT	1342	1342	hsa-let-7e	yes
11.22.136781	hsa-let-7f-2	7	TGAGGTAGTAGATTGTATAGTT	136781	136781	hsa-let-7f	yes
6.22.192102	hsa-let-7g	4	TGAGGTAGTAGTTTGTACAGTT	192102	192102	hsa-let-7g	yes
85.22.15473	hsa-let-7i	5	TGAGGTAGTAGTTTGTGCTGTT	15473	15473	hsa-let-7i	yes
35405.22.8	hsa-let-7i	61	CTGCGCAAGCTACTGCCTTGCT	8	8	hsa-let-7i*	yes
76.22.17583	hsa-mir-1-1	45	TGGAATGTAAAGAAGTATGTAT	17583	17583	hsa-miR-1	yes
8863.22.41	hsa-mir-100	12	AACCCGTAGATCCGAACTTGTG	41	41	hsa-miR-100	yes
282.21.3347	hsa-mir-101-2	48	TACAGTACTGTGATAACTGAA	3347	3968	hsa-miR-101	No
21.23.78618	hsa-mir-103-1	47	AGCAGCATTGTACAGGGCTATGA	78618	78618	hsa-miR-103	yes
25705.23.11	hsa-mir-106a	12	AAAAGTGCTTACAGTGCAGGTAG	11	11	hsa-miR-106a	yes
277.22.3501	hsa-mir-106b	51	CCGCACTGTGGGTACTTGCTGC	3501	3501	hsa-miR-106b*	yes
763.21.839	hsa-mir-106b	11	TAAAGTGCTGACAGTGCAGAT	839	839	hsa-miR-106b	yes
150.23.7546	hsa-mir-107	49	AGCAGCATTGTACAGGGCTATCA	7546	72985	hsa-miR-107	No
2349.23.195	hsa-mir-10a	21	TACCCTGTAGATCCGAATTTGTG	195	792	hsa-miR-10a	No
242382.21.2	hsa-mir-1197	56	TAGGACACATGGTCTACTTCT	2	2	hsa-miR-1197	yes
566268.22.1	hsa-mir-122	14	TGGAGTGTGACAATGGTGTTTG	1	11	hsa-miR-122	No
411712.22.1	hsa-mir-124-3	14	CGTGTTCACAGCGGACCTTGAT	1	1	hsa-miR-124*	yes
516359.20.1	hsa-mir-124-3	52	TAAGGCACGCGGTGAATGCC	1	49	hsa-miR-124	No
65638.19.4	hsa-mir-1246	10	AATGGATTTTTGGAGCAGG	4	201	hsa-miR-1246	No
21263.21.14	hsa-mir-1250	23	ACGGTGCTGGATGTGGCCTTT	14	14	hsa-miR-1250	yes
30965.24.9	hsa-mir-1254	17	AGCCTGGAAGCTGGAGCCTGCAGT	9	14	hsa-miR-1254	No
3984.23.103	hsa-mir-1255a	27	AGGATGAGCAAAGAAAGTAGATT	103	111	hsa-miR-1255a	No
4129.22.91	hsa-mir-1255b-2	5	CGGATGAGCAAAGAAAGTGGTT	91	91	hsa-miR-1255b	yes
28336.22.10	hsa-mir-1256	34	AGGCATTGACTTCTCACTAGCT	10	10	hsa-miR-1256	yes
3902.24.106	hsa-mir-125a	11	TCCCTGAGACCCTTTAACCTGTGA	106	330	hsa-miR-125a-5p	No
19071.22.16	hsa-mir-125a	52	ACAGGTGAGGTTCTTGGGAGCC	16	16	hsa-miR-125a-3p	yes
5715.22.69	hsa-mir-125b-2	16	TCCCTGAGACCCTAACTTGTGA	69	69	hsa-miR-125b	yes
6496.21.59	hsa-mir-126	14	CATTATTACTTTTGGTACGCG	59	80	hsa-miR-126*	No
25312.22.12	hsa-mir-126	51	TCGTACCGTGAGTAATAATGCG	12	23	hsa-miR-126	No
11787.22.29	hsa-mir-1262	13	ATGGGTGAATTTGTAGAAGGAT	29	29	hsa-miR-1262	yes
1041.18.545	hsa-mir-1268	4	CGGGCGTGGTGGTGGGGG	545	545	hsa-miR-1268	yes
1036.22.549	hsa-mir-127	56	TCGGATCCGTCTGAGCTTGGCT	549	549	hsa-miR-127-3p	yes
419677.22.1	hsa-mir-127	22	CTGAAGCTCAGAGGGCTCTGAT	1	8	hsa-miR-127-5p	No
10246.23.35	hsa-mir-1270-1	12	CTGGAGATATGGAAGAGCTGTGT	35	35	hsa-miR-1270	yes

Table A.8: **Identification of most abundant IsomiRs for P1 sample**

read-name	Reference	Position	IsomiRs	matching-abundance	highest-abundance	mature	match-with-highest
71151.21.1	hsa-let-7a-1	56	CTATACAATCTACTGTCTTTC	4	4	hsa-let-7a*	yes
28.22.64475	hsa-let-7a-3	3	TGAGGTAGTAGGTTGTATAGTT	64475	64475	hsa-let-7a	yes
1.22.531019	hsa-let-7b	5	TGAGGTAGTAGGTTGTGTGGTT	531019	531019	hsa-let-7b	yes
74.22.20225	hsa-let-7c	10	TGAGGTAGTAGGTTGTATGGTT	20225	20225	hsa-let-7c	yes
91.22.14530	hsa-let-7d	7	AGAGGTAGTAGGTTGCATAGTT	14530	15107	hsa-let-7d	No
2757.22.166	hsa-let-7d	61	CTATACGACCTGCTGCCTTTCT	166	166	hsa-let-7d*	yes
1263.22.440	hsa-let-7e	7	TGAGGTAGGAGGTTGTATAGTT	440	440	hsa-let-7e	yes
22.22.78258	hsa-let-7f-1	6	TGAGGTAGTAGATTGTATAGTT	78258	78258	hsa-let-7f	yes
23.22.76875	hsa-let-7g	4	TGAGGTAGTAGTTTGTACAGTT	76875	76875	hsa-let-7g	yes
86.22.15389	hsa-let-7i	5	TGAGGTAGTAGTTTGTGCTGTT	15389	15389	hsa-let-7i	yes
10474.22.34	hsa-let-7i	61	CTGCGCAAGCTACTGCCTTGCT	34	34	hsa-let-7i*	yes
141.22.8454	hsa-mir-1-1	45	TGGAATGTAAAGAAGTATGTAT	8454	8454	hsa-miR-1	yes
9878.22.36	hsa-mir-100	12	AACCCGTAGATCCGAACTTGTG	36	36	hsa-miR-100	yes
252.21.4120	hsa-mir-101-1	46	TACAGTACTGTGATAACTGAA	4120	4120	hsa-miR-101	yes
30.23.58150	hsa-mir-103-2	47	AGCAGCATTGTACAGGGCTATGA	58150	58150	hsa-miR-103	yes
13248.23.25	hsa-mir-106a	12	AAAAGTGCTTACAGTGCAGGTAG	25	25	hsa-miR-106a	yes
226.22.4791	hsa-mir-106b	51	CCGCACTGTGGGTACTTGCTGC	4791	4791	hsa-miR-106b*	yes
419.21.1977	hsa-mir-106b	11	TAAAGTGCTGACAGTGCAGAT	1977	1977	hsa-miR-106b	yes
223.23.4855	hsa-mir-107	49	AGCAGCATTGTACAGGGCTATCA	4855	48092	hsa-miR-107	No
6823.23.56	hsa-mir-10a	21	TACCCTGTAGATCCGAATTTGTG	56	256	hsa-miR-10a	No
76696.23.4	hsa-mir-10b	26	TACCCTGTAGAACCGAATTTGTG	4	12	hsa-miR-10b	No
640469.22.1	hsa-mir-1180	40	TTTCCGGCTCGCGTGGGTGTGT	1	13	hsa-miR-1180	No
239168.22.2	hsa-mir-122	14	TGGAGTGTGACAATGGTGTTTG	2	5	hsa-miR-122	No
41195.19.7	hsa-mir-1224	0	GTGAGGACTCGGGAGGTGG	7	17	hsa-miR-1224-5p	No
531826.26.1	hsa-mir-1226	0	GTGAGGGCATGCAGGCCTGGATGGGG	1	1	hsa-miR-1226*	yes
48270.20.6	hsa-mir-124-1	52	TAAGGCACGCGGTGAATGCC	6	92	hsa-miR-124	No
136091.26.2	hsa-mir-1244-3	51	AAGTAGTTGGTTTGTATGAGATGGTT	2	2	hsa-miR-1244	yes
8678.19.42	hsa-mir-1246	10	AATGGATTTTTGGAGCAGG	42	1895	hsa-miR-1246	No
296234.27.1	hsa-mir-1248	3	ACCTTCTTGTATAAGCACTGTGCTAAA	1	6	hsa-miR-1248	No
143136.22.2	hsa-mir-1249	40	ACGCCCTTCCCCCCCTTCTTCA	2	2	hsa-miR-1249	yes
18262.21.17	hsa-mir-1250	23	ACGGTGCTGGATGTGGCCTTT	17	17	hsa-miR-1250	yes
147194.22.2	hsa-mir-1252	4	AGAAGGAAATTGAATTCATTTA	2	2	hsa-miR-1252	yes
3192.23.136	hsa-mir-1255a	27	AGGATGAGCAAAGAAAGTAGATT	136	136	hsa-miR-1255a	yes
3339.22.129	hsa-mir-1255b-2	5	CGGATGAGCAAAGAAAGTGGTT	129	129	hsa-miR-1255b	yes
38679.22.7	hsa-mir-1256	34	AGGCATTGACTTCTCACTAGCT	7	7	hsa-miR-1256	yes
17939.24.18	hsa-mir-125a	14	TCCCTGAGACCCTTTAACCTGTGA	18	47	hsa-miR-125a-5p	No
51255.22.5	hsa-mir-125a	52	ACAGGTGAGGTTCTTGGGAGCC	5	5	hsa-miR-125a-3p	yes
7661.22.49	hsa-mir-125b-1	14	TCCCTGAGACCCTAACTTGTGA	49	49	hsa-miR-125b	yes
14361.21.23	hsa-mir-126	14	CATTATTACTTTTGGTACGCG	23	40	hsa-miR-126*	No
29960.22.10	hsa-mir-126	51	TCGTACCGTGAGTAATAATGCG	10	13	hsa-miR-126	No

Table A.9: **Identification of most abundant IsomiRs for P2 sample**

read-name	Reference	Position	IsomiRs	matching-abundance	highest-abundance	mature	match-with-highest
525788.21.1	hsa-let-7a-1	56	CTATACAATCTACTGTCTTTC	1	2	hsa-let-7a*	No
19.22.91829	hsa-let-7a-2	4	TGAGGTAGTAGGTTGTATAGTT	91829	91829	hsa-let-7a	yes
1.22.369116	hsa-let-7b	5	TGAGGTAGTAGGTTGTGTGGTT	369116	369116	hsa-let-7b	yes
109.22.14133	hsa-let-7c	10	TGAGGTAGTAGGTTGTATGGTT	14133	14133	hsa-let-7c	yes
68.22.23794	hsa-let-7d	7	AGAGGTAGTAGGTTGCATAGTT	23794	27035	hsa-let-7d	No
2483.22.227	hsa-let-7d	61	CTATACGACCTGCTGCCTTTCT	227	227	hsa-let-7d*	yes
383.22.2610	hsa-let-7e	7	TGAGGTAGGAGGTTGTATAGTT	2610	4321	hsa-let-7e	No
6.22.159409	hsa-let-7f-2	7	TGAGGTAGTAGATTGTATAGTT	159409	159409	hsa-let-7f	yes
8.22.137770	hsa-let-7g	4	TGAGGTAGTAGTTTGTACAGTT	137770	137770	hsa-let-7g	yes
274935.21.2	hsa-let-7g	61	CTGTACAGGCCACTGCCTTGC	2	2	hsa-let-7g*	yes
79.22.19837	hsa-let-7i	5	TGAGGTAGTAGTTTGTGCTGTT	19837	19837	hsa-let-7i	yes
18167.22.24	hsa-let-7i	61	CTGCGCAAGCTACTGCCTTGCT	24	26	hsa-let-7i*	No
165.22.7906	hsa-mir-1-1	45	TGGAATGTAAAGAAGTATGTAT	7906	7906	hsa-miR-1	yes
12207.22.37	hsa-mir-100	12	AACCCGTAGATCCGAACTTGTG	37	37	hsa-miR-100	yes
238.21.4659	hsa-mir-101-1	16	TACAGTACTGTGATAACTGAA	4659	5064	hsa-miR-101	No
478839.22.1	hsa-mir-101-1	10	CAGTTATCACAGTGCTGATGCT	1	2	hsa-miR-101*	No
18.23.94367	hsa-mir-103-2	47	AGCAGCATTGTACAGGGCTATGA	94367	94367	hsa-miR-103	yes
74722.23.5	hsa-mir-103-2	10	AGCTTCTTTACAGTGCTGCCTTG	5	17	hsa-miR-103-2*	No
11613.23.39	hsa-mir-106a	12	AAAAGTGCTTACAGTGCAGGTAG	39	39	hsa-miR-106a	yes
273.22.4015	hsa-mir-106b	51	CCGCACTGTGGGTACTTGCTGC	4015	4015	hsa-miR-106b*	yes
665.21.1334	hsa-mir-106b	11	TAAAGTGCTGACAGTGCAGAT	1334	1334	hsa-miR-106b	yes
118.23.12605	hsa-mir-107	19	AGCAGCATTGTACAGGGCTATCA	12605	89803	hsa-miR-107	No
5046.23.99	hsa-mir-10a	21	TACCCTGTAGATCCGAATTTGTG	99	396	hsa-miR-10a	No
201238.21.2	hsa-mir-1179	14	AAGCATTCTTTCATTGGTTGG	2	7	hsa-miR-1179	No
31899.22.13	hsa-mir-122	11	TGGAGTGTGACAATGGTGTTTG	13	13	hsa-miR-122	yes
56821.19.7	hsa-mir-1224	0	GTGAGGACTCGGGAGGTGG	7	17	hsa-miR-1224-5p	No
18274.20.24	hsa-mir-124-2	61	TAAGGCACGCGGTGAATGCC	24	24	hsa-miR-124	yes
918.19.864	hsa-mir-1246	10	AATGGATTTTTGGAGCAGG	864	895	hsa-miR-1246	No
93954.22.4	hsa-mir-1249	40	ACGCCCTTCCCCCCCTTCTTCA	4	4	hsa-miR-1249	yes
25421.21.16	hsa-mir-1250	23	ACGGTGCTGGATGTGGCCTTT	16	16	hsa-miR-1250	yes
61533.24.6	hsa-mir-1254	17	AGCCTGGAAGCTGGAGCCTGCAGT	6	94	hsa-miR-1254	No
8095.23.59	hsa-mir-1255a	27	AGGATGAGCAAAGAAAGTAGATT	59	59	hsa-miR-1255a	yes
3919.22.132	hsa-mir-1255b-1	2	CGGATGAGCAAAGAAAGTGGTT	132	132	hsa-miR-1255b	yes
19384.22.22	hsa-mir-1256	34	AGGCATTGACTTCTCACTAGCT	22	22	hsa-miR-1256	yes
11047.24.42	hsa-mir-125a	14	TCCCTGAGACCCTTTAACCTGTGA	42	157	hsa-miR-125a-5p	No
41471.22.9	hsa-mir-125a	52	ACAGGTGAGGTTCTTGGGAGCC	9	9	hsa-miR-125a-3p	yes
5847.22.85	hsa-mir-125b-2	16	TCCCTGAGACCCTAACTTGTGA	85	85	hsa-miR-125b	yes
2935.21.184	hsa-mir-126	14	CATTATTACTTTTGGTACGCG	184	184	hsa-miR-126*	yes
50124.22.8	hsa-mir-126	51	TCGTACCGTGAGTAATAATGCG	8	58	hsa-miR-126	No
238165.18.2	hsa-mir-1260	13	ATCCCACCTCTGCCACCA	2	2956	hsa-miR-1260	No

Table A.10: **Identification of most abundant IsomiRs for K562 sample**

read-name	Reference	Position	IsomiRs	matching-abundance	highest-abundance	mature	match-with-highest
2.22.61968	hsa-let-7a-1	5	TGAGGTAGTAGGTTGTATAGTT	61968	61968	hsa-let-7a	yes
849.22.619	hsa-let-7b	5	TGAGGTAGTAGGTTGTGTGGTT	619	619	hsa-let-7b	yes
318.22.1347	hsa-let-7c	10	TGAGGTAGTAGGTTGTATGGTT	1347	1347	hsa-let-7c	yes
1243.22.440	hsa-let-7d	7	AGAGGTAGTAGGTTGCATAGTT	440	440	hsa-let-7d	yes
31.22.13238	hsa-let-7e	7	TGAGGTAGGAGGTTGTATAGTT	13238	13238	hsa-let-7e	yes
3584.22.154	hsa-let-7e	52	CTATACGGCCTCCTAGCTTTCC	1	1	hsa-let-7e*	yes
357273.22.1	hsa-let-7f-1	6	TGAGGTAGTAGATTGTATAGTT	42404	42404	hsa-let-7f	yes
5.22.42404	hsa-let-7g	4	TGAGGTAGTAGTTTGTACAGTT	3897	3897	hsa-let-7g	yes
112.22.3897	hsa-let-7i	5	TGAGGTAGTAGTTTGTGCTGTT	344	344	hsa-let-7i	yes
1616.22.344	hsa-mir-1-1	45	TGGAATGTAAAGAAGTATGTAT	245	245	hsa-miR-1	yes
2264.22.245	hsa-mir-100	12	AACCCGTAGATCCGAACTTGTG	9	9	hsa-miR-100	yes
32670.22.9	hsa-mir-101-2	18	TACAGTACTGTGATAACTGAA	2118	2118	hsa-miR-101	yes
195.21.2118	hsa-mir-103-2	47	AGCAGCATTGTACAGGGCTATGA	22135	22135	hsa-miR-103	yes
54486.19.5	hsa-mir-103-2	10	AGCTTCTTTACAGTGCTGCCTTG	3	3	hsa-miR-103-2*	yes
3584.22.154	hsa-mir-105-1	12	TCAAATGCTCAGACTCCTGTGGT	2	11	hsa-miR-105	No
3584.22.154	hsa-mir-105-1	50	ACGGATGTTTGAGCATGTGCTA	1	46	hsa-miR-105*	No
3584.22.154	hsa-mir-106a	12	AAAAGTGCTTACAGTGCAGGTAG	3	28	hsa-miR-106a	No
3584.22.154	hsa-mir-106b	11	TAAAGTGCTGACAGTGCAGAT	334	969	hsa-miR-106b	No
65497.21.4	hsa-mir-106b	51	CCGCACTGTGGGTACTTGCTGC	202	202	hsa-miR-106b*	yes
19.23.22135	hsa-mir-107	49	AGCAGCATTGTACAGGGCTATCA	2745	2745	hsa-miR-107	yes
3584.22.154	hsa-mir-10a	21	TACCCTGTAGATCCGAATTTGTG	932	2125	hsa-miR-10a	No
3584.22.154	hsa-mir-10a	62	CAAATTCGTATCTAGGGGAATA	3	24	hsa-miR-10a*	No
78893.23.3	hsa-mir-1180	40	TTTCCGGCTCGCGTGGGTGTGT	4	4	hsa-miR-1180	yes
140128.23.2	hsa-mir-1185-1	14	AGAGGATACCCTTTGTATGTT	1	1	hsa-miR-1185	yes
1025270.22.1	hsa-mir-1197	56	TAGGACACATGGTCTACTTCT	2	2	hsa-miR-1197	yes
86115.23.3	hsa-mir-122	14	TGGAGTGTGACAATGGTGTTTG	8	8	hsa-miR-122	yes
3584.22.154	hsa-mir-124-3	52	TAAGGCACGCGGTGAATGCC	5	36	hsa-miR-124	No
20896.23.17	hsa-mir-124-3	14	CGTGTTCACAGCGGACCTTGAT	1	1	hsa-miR-124*	yes
3584.22.154	hsa-mir-1246	10	AATGGATTTTTGGAGCAGG	235	648	hsa-miR-1246	No
3584.22.154	hsa-mir-1248	3	ACCTTCTTGTATAAGCACTGTGCTAAA	2	90	hsa-miR-1248	No
1667.21.334	hsa-mir-1252	4	AGAAGGAAATTGAATTCATTTA	1	1	hsa-miR-1252	yes
2704.22.202	hsa-mir-1254	17	AGCCTGGAAGCTGGAGCCTGCAGT	38	38	hsa-miR-1254	yes
154.23.2745	hsa-mir-1255a	27	AGGATGAGCAAAGAAAGTAGATT	183	183	hsa-miR-1255a	yes
516.23.932	hsa-mir-1255b-2	5	CGGATGAGCAAAGAAAGTGGTT	95	95	hsa-miR-1255b	yes
80217.22.3	hsa-mir-1256	34	AGGCATTGACTTCTCACTAGCT	6	6	hsa-miR-1256	yes
3584.22.154	hsa-mir-125a	14	TCCCTGAGACCCTTTAACCTGTGA	157	306	hsa-miR-125a-5p	No
3584.22.154	hsa-mir-125a	52	ACAGGTGAGGTTCTTGGGAGCC	4	6	hsa-miR-125a-3p	No
65164.22.4	hsa-mir-125b-1	14	TCCCTGAGACCCTAACTTGTGA	51	51	hsa-miR-125b	yes
538182.21.1	hsa-mir-126	14	CATTATTACTTTTGGTACGCG	252	252	hsa-miR-126*	yes
3584.22.154	hsa-mir-126	51	TCGTACCGTGAGTAATAATGCG	116	156	hsa-miR-126	No

Table A.11: **IsomiRs present in specific condition**

miRNA	Sequence	N4	N5	P1	P2
hsa-miR-1250	ACGGTGCTGGATGTGGCCTTT	22	14	0	0
hsa-miR-1271	CTTGGCACCTAGCAAGCACTCA	24	23	0	0
hsa-miR-1277	TACGTAGATATATATGTATTTT	297	42	0	0
hsa-miR-1294	TGTGAGGTTGGCATTGTTGTCT	222	290	0	0
hsa-miR-138	AGCTGGTGTTGTGAATCAGGCCG	33	19	0	0
hsa-miR-1908	CGGCGGGGACGGCGATTGGTC	85	112	0	0
hsa-miR-193b*	CGGGGTTTTGAGGGCGAGATGA	49	41	0	0
hsa-miR-3121	TAAATAGAGTAGGCAAAGGACA	24	14	0	0
hsa-miR-379	TGGTAGACTATGGAACGTAGG	62	30	0	0
hsa-miR-3909	TGTCCTCTAGGGCCTGCAGTCT	9	20	0	0
hsa-miR-431	TGTCTTGCAGGCCGTCATGCA	37	30	0	0
hsa-miR-542-3p	TGTGACAGATTGATAACTGAAA	40	27	0	0
hsa-miR-618	AAACTCTACTTGTCCTTCTGAGT	8	42	0	0
hsa-miR-889	TTAATATCGGACAACCATTGT	93	61	0	0
hsa-miR-95	TTCAACGGGTATTTATTGAGCA	28	68	0	0
hsa-miR-1	TGGAATGTAAAGAAGTATGTAT	0	0	8454	7906
hsa-miR-1255a	AGGATGAGCAAAGAAAGTAGATT	0	0	136	59
hsa-miR-1274b	TCCCTGTTCGGGCGCCA	0	0	2684	393
hsa-miR-1301	TTGCAGCTGCCTGGGAGTGACTTC	0	0	24	89
hsa-miR-136	ACTCCATTTGTTTTGATGATGGA	0	0	22	25
hsa-miR-155	TTAATGCTAATCGTGATAGGGGT	0	0	545	594
hsa-miR-221	AGCTACATTGTCTGCTGGGTTTC	0	0	8526	21527
hsa-miR-30c	TGTAAACATCCTACACTCTCAGC	0	0	447	342
hsa-miR-3154	CAGAAGGGGAGTTGGGAGCAGA	0	0	74	11
hsa-miR-335	TCAAGAGCAATAACGAAAAATGT	0	0	80	134
hsa-miR-365	TAATGCCCCTAAAAATCCTTAT	0	0	64	40
hsa-miR-3928	GGAGGAACCTTGGAGCTTCGGC	0	0	20	19
hsa-miR-425	AATGACACGATCACTCCCGTTGA	0	0	952	532
hsa-miR-451	AAACCGTTACCATTACTGAGTT	0	0	6863	5984
hsa-miR-484	TCAGGCTCAGTCCCCTCCCGAT	0	0	55	109
hsa-miR-502-3p	AATGCACCTGGGCAAGGATTCA	0	0	74	81
hsa-miR-550a*	TGTCTTACTCCCTCAGGCACAT	0	0	29	19
hsa-miR-660	TACCCATTGCATATCGGAGTTG	0	0	151	110
hsa-miR-766	ACTCCAGCCCCACAGCCTCAGC	0	0	51	111
hsa-miR-940	AAGGCAGGGCCCCCGCTCCCC	0	0	22	7

Note

*Complete list of results has been compiled and can be requested from the author.

Bibliography

[1] John S. Mattick , and Igor V. Makunin: Non-coding RNA. Hum. Mol. Genet. 15: R17-R29. 2006

[2] Tolia NH, Joshua-Tor L: Slicer and the Argonautes. Nature Chemical Biology (2007) Jan;3(1):36-43

[3] Neil Hall: Advanced sequencing technologies and their wider impact in microbiology.The Journal of Experimental Biology(2007) 209, 1518-1525.

[4] Richard Williams, Sergio G Peisajovich, Oliver J Miller, Shlomo Magdassi, Dan S Tawfik, Andrew D Griffiths: Amplification of complex gene libraries by emulsion PCR. Nature Methods (2006) 3 (7): 545550.

[5] Elaine R. Mardis: Next-Generation DNA Sequencing Methods. Genomics and Human Genetics(2008). Vol. 9: 387-402.

[6] Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y: RNA-seq: an assessment of technical reproducibility and comparison with gene expression arrays. Genome Res. 2008 Sep;18(9):1509-17.

[7] Sean R. Eddy: Noncoding RNA genes and the modern RNA world. Nature Reviews Genetics 2, 919-929 (December 2001)

[8] Gordon Robertson, Martin Hirst, Matthew Bainbridge, Misha Bilenky, Yongjun Zhao, Thomas Zeng, Ghia Euskirchen, Bridget Bernier, Richard Varhol, Allen Delaney, Nina

Thiessen, Obi L Griffith, Ann He, Marco Marra, Michael Snyder & Steven Jones: Genome-wide profiles of STAT1 DNA association using chromatin immuno- precipitation and massively parallel sequencing . Nature Methods - 4, 651 - 657 (2007).

[9] VanGuilder HD, Vrana KE, Freeman WM (2008). Twenty-five years of quantitative PCR for gene expression analysis. Biotechniques 44 (5): 619626.

[10] Southern, Edwin Mellor (5 November 1975): Detection of specific sequences among DNA fragments separated by gel electrophoresis. Journal of Molecular Biology 98 (3): 503517.

[11] Johnson DS, Mortazavi A, Myers RM, Wold B. 2007. Genome-wide mapping of in vivo protein-DNA interactions. Science 316:1497502.

[12] Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC: Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans.
Nature 1998, 391:806-811.

[13] Ryan D. Morin, Michael D. OConnor, Malachi Griffith, Florian Kuchenbauer, Allen Delaney, Anna-Liisa Prabhu, Yongjun Zhao, Helen McDonald, Thomas Zeng, Martin Hirst, Connie J. Eaves and Marco A. Marra: Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. Genome Res. 2008 April; 18(4): 610621.

[14] Vaz C, Ahmad HM, Sharma P, Gupta R, Kumar L, Kulshreshtha R, Bhattacharya A: Analysis of microRNA transcriptome by deep sequencing of small RNA libraries of peripheral blood. BMC Genomics 2010, 11:288.

[15] Ben Langmead, Cole Trapnell, Mihai Pop and Steven L Salzberg: Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biology 2009, 10:R25.

[16] Ali Mortazavi, Brian A Williams, Kenneth McCue, Lorian Schaeffer & Barbara Wold: Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nature Methods - 5, 621 - 628 (2008).

[17] Ben Bolstad: Probe Level Quantile Normalization of High Density Oligonucleotide Array Data. Unpublished Manuscript.

[18] B. M. Bolstad, R. A. Irizarry, M. Astrand and T. P. Speed: A Comparison of Normalization Methods for High Density Oligonucleotide Array Data Based on Variance and Bias. Bioinformatics 19(2):185-193.

[19] Mark D Robinson, Alicia Oshlack: A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biology 2010, 11:R25.

[20] George W. Wright and Richard M. Simon: A random variance model for detection of differential gene expression in small microarray experiments. Bioinformatics(2003) 19 (18): 2448-2455.

[21] Xiangqin Cui and Gary A Churchill: Statistical tests for differential expression in cDNA microarray experiments.Genome Biology 2003, 4:210.

[22] Beyer, W. H. CRC Standard Mathematical Tables, 28th ed. Boca Raton, FL: CRC Press, p. 533, 1987.

[23] Robert C Gentleman, Vincent J Carey, Douglas M Bates, Ben Bolstad, Marcel Dettling, Sandrine Dudoit, Byron Ellis, Laurent Gautier, Yongchao Ge, Jeff Gentry, Kurt Hornik, Torsten Hothorn, Wolfgang Huber, Stefano Iacus, Rafael Irizarry, Friedrich Leisch, Cheng Li, Martin Maechler, Anthony J Rossini, Gunther Sawitzki, Colin Smith, Gordon Smyth, Luke Tierney, Jean YH Yang and Jianhua Zhang: Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 2004; 5(10): R80.

[24] Robinson MD, Smyth GK: Small-sample estimation of negative binomial dispersion, with applications to SAGE data. (2008) 9 (2): 321-332.

[25] Mark D. Robinson, Davis J. McCarthy and Gordon K. Smyth: edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics(2010) 26 (1): 139-140.

[26] Simon Anders and Wolfgang Huber: Differential expression analysis for sequence count data. Genome Biology 2010, 11:R106.

[27] Kozomara A, Griffiths-Jones S: miRBase: integrating microRNA annotation and deep-sequencing data. NAR (2011) 39 (suppl 1): D152-D157.

[28] Griffiths-Jones S, Saini HK, van Dongen S, Enright AJ: miRBase: tools for microRNA genomics. NAR 2008 36(Database Issue):D154-D158.

[29] Li Guo, Zuhong Lu: Global expression analysis of miRNA gene cluster and family based on isomiRs from deep sequencing data.Computational Biology and Chemistry (2010).