Detection and characterization of small non-coding RNAs in mouse gametes and embryos prior to zygotic genome activation

Jesús García-López¹, Eduardo Larriba and Jesús del Mazo^{*}

Department of Cellular and Molecular Biology, Centro de Investigaciones Biológicas (CSIC), Ramiro de Maeztu 9, Madrid 28040, Spain.

¹ Present address: Oncology Department. St. Jude Children's Research Hospital, Memphis, TN 38105, USA.

^{*} Author to whom correspondence should be addressed; Tel:+34-918373112#4324

Fax +34 915360432. Email: jdelmazo@cib.csic.es

Abstract

Small non-coding RNAs (ncRNAs) are regulatory elements of gene expression in all cell types and tissues. An ever-increasing number of studies have implicated ncRNAs in differentiation and developmental processes. In mammals, as consequence of fertilization, the content of ncRNAs in the zygote is mostly the result of the maternal material included on oocytes and the potential sperm-borne paternal contributions. The genetic identity program of any individual is the reprogramming of each parental contribution to the zygotic genome activation. In mouse, this activation occurs at 2-cell stage. In this program of early development [the small ncRNAs can play important roles](#). Here, we describe protocols for collection of oocytes, spermatozoa and zygotes in mouse, followed by RNA purification to analyze the different types of small ncRNA by next generation sequencing approaches (NGS). Bioinformatics protocols also describe the methodology able to characterize microRNAs (miRNAs) as the most well-known and widespread regulatory small ncRNA. The comparative analysis allows identifying the changes and background previous to zygotic genome activation.

Key words: Small noncoding RNAs

Oocyte

Spermatozoa

Zygote

miRNAs

NGS

1. Introduction

The correct embryo development requires deep but selective changes in the pattern of gene expression from the mostly based maternal contribution in the oocyte to the transcriptome of the totipotent zygote. After fertilization, this transition involves the replacement of specific genetic information coded abundantly in the maternal RNAs from oocyte, and potentially as sperm-borne, with zygotic genome expression. Consequently, an assessment of the transcriptome landscape as result of zygotic activation, occurring at 2-cell stage in mouse [1], requires comparative analysis of previous transcript background. That is, the whole mRNAs and ncRNAs contribution of both gametes to the zygote and the zygote itself previously to the cleavage and zygotic activation.

Small non-coding RNAs (ncRNAs) are regulatory components of gene expression and genome structure in all cell types and tissues. The role of ncRNAs in the maternal to zygotic transition has been remarked as key regulatory elements in the control of maternal transcript elimination and new gene expression regulation [2]. On the other hand, Yuan et al. [3] recently reported the relative importance of some sperm-borne ncRNAs during embryo development, but not in the fertilization neither in the zygotic activation.

Both specific mRNA and ncRNA present in the zygote have to be functionally conserved in the early preimplantation embryo development. Among the different types of ncRNAs: microRNAs (miRNAs), piwi-interacting RNAs (piRNAs), endogenous siRNAs (endo-siRNAs), the miRNAs are the best characterized ncRNAs in the biogenesis and function [4]. Maternal miRNAs are essential in the zygotic development and the earliest stages of mouse embryonic development [5]. However, some other regulated mRNA and miRNAs are selectively eliminated before or as consequence of zygotic activation. In fact, some miRNAs are actively participating in the selective “erase” of maternal mRNAs [6]. Maternal, or putative paternal mRNAs are clearance by MZT-specific miRNAs (maternal-to-zygotic transition). A clear example is miR-430 in zebrafish (miR-294 orthologous in mouse)

involved in mRNA degradation by deadenylation of maternal transcripts [7].

The NGS approach allows the analysis and characterization of different types of small ncRNAs. Although other types of small ncRNA, which are present in both gametes and zygotes [8,9], could act in the elimination of mRNA [10] facilitating the reprogramming of zygote activation, here we focus on the study of miRNAs through bioinformatics approaches .

In this chapter, we describe, the cytological, molecular and bioinformatics protocols to analyze qualitative and quantitatively the expression profiles of small ncRNAs, focusing on miRNAs, in spermatozoa, oocytes and zygotes. We provide the methodology for isolating both male and female gametes and zygotes, RNA isolation from these cells, and finally, the bioinformatics analysis of miRNAs from data obtained by NGS technology.

Different stages of bioinformatics pipeline present in these protocol, required the use of the terminal (Bash, Bourne again shell), these stages are detailed in format copy and paste to facilitate the use of terminal. In addition, we include the use of different free software programs that have detailed documentation and tutorials. Analysis of miRNA encompassing different steps, including: evaluation of sRNA-Seq reads, trimming and cleaning sequencing reads, and a specific framework for identification and quantification a miRNA and isomiRs. Finally, miRNA regulatory elements and targets genes were predicted using different web-tools (Figure 1).

Fig. 1

An overall pipeline designed to process small ncRNA data after NGS. Workflow is colored based in the different stages of bioinformatics protocol. In *green*, quality control of sequencing libraries and trimming. In *red*, identification of miRNA and isomiR detection and characterization. In *blue*,normalization of miRNAs and isomiRs and differential expression analysis. In *pink*, miRNA target identification. Software used in each phase is

indicated in *brackets*

2. Materials.

2.1 Isolating gametes and zygotes and RNA purification

1. Mice: 4-5-week-old C57BL6 females and adult DBA/6J males bred under specific pathogen-free (SPF), temperature ($22\pm 1^{\circ}\text{C}$), humidity-controlled (50-55%), 12h light/dark cycles with *ad libitum* access to food and water.
2. Pregnant mare serum (PMSG).
3. Human chorionic gonadotropin (HCG).
4. Hyaluronidase solution (0.3 mg/mL) (Sigma H3884).
5. Mineral oil embryo-tested light (Sigma, M8410).
6. M2 medium (Sigma M7167).
7. M16 medium (Sigma M7292).
8. Acidic Tyrode's solution (Sigma, T1788)
9. RNase-free PBS.
10. TRIzol[®] Reagent (Invitrogen).
11. Percoll.
12. RNase free H₂O.
13. Fine forceps.
14. Fine scissors
15. Ethanol (75%)
16. Isopropanol
17. Chloroform
18. Glycogen blue (Thermo Fisher Scientific, AM9516)
19. Stereomicroscope with transmitted light
20. Incubator, humidified at 37 °C, 5% CO₂ in air.
21. Mouth-controlled pipette (consisting in a mouth piece, a tygon tube and a pulled Pasteur pipette).
22. Sterile Petri dishes 35X10 mm. (Nunc).
23. Neubauer hemocytometer

24. Micropipettes and tips.
25. Eppendorf tubes.
26. Glass pipettes.
27. Tube shaker.
28. NanoDrop Spectrophotometer (ND-1000).
29. Microcentrifuge (Eppendorf 5424 or equivalent).
30. Centrifuge (Eppendorf 5403 or equivalent).
31. Heat block.
32. Bioanalyzer 2100 (Agilent Technologies)

2.2. Bioinformatics analysis from NGS data

2.2.1 Hardware

We recommend a High Performance Computing (HPC) especially when planning to analyze a large number of NGS sequencing files. Alternatively, this protocol is designed for a personal computer with the following minimum configuration: 4 core processor with 8 threads and 16 GB of RAM memory.

2.2.2 Software

We suggest organizing both software and databases in different folders (e.g. “Software/” and “Databases/”).

This protocol was designed to be as user-friendly as possible in a personal computer. It also includes the strategies and commands to enable the analysis of different libraries at once using bash scripts.

All software/packages in this protocol contain a complete manual available on-line.

1. Operating System: Unix-based operating system such as Ubuntu or Mac OS (*see Note 1*).
2. Databases files: miRBase: <http://www.mirbase.org/> (Version 21).

3. Isomirage: <http://cru.genomics.iit.it/Isomirage/> [11]
4. R and Bioconductor (DESeq package): <https://cran.r-project.org/bin/linux/ubuntu/> and <https://www.bioconductor.org/install/> (see **Note 2**).
5. Bowtie: <https://sourceforge.net/projects/bowtie-bio/files/bowtie/1.1.2/>
(for Ubuntu users, type in the terminal: `sudo apt-get install bowtie`)
6. SAM tools: <https://sourceforge.net/projects/samtools/files/>
(for Ubuntu users, type in the terminal: `sudo apt-get install samtools`)
7. FastQC and Trim Galore: <http://www.bioinformatics.babraham.ac.uk/projects/>
(for Ubuntu users, type in the terminal: `sudo apt-get install fastqc`)
8. FASTX-toolkit: http://hannonlab.cshl.edu/fastx_toolkit/download.html
(for Ubuntu users, type in the terminal: `sudo apt-get install fastx-toolkit`).
9. miRWalk2.0: <http://zmf.umm.uni-heidelberg.de/apps/zmf/mirwalk2/>

3. Methods

3.1. Setting up the oocytes and zygotes collection

3.1.1. Female hormone induction

1. To obtain large amount of oocytes and zygotes, we induce hormonal superovulation of females. For most mouse strains, 5IU of Pregnant mare's serum gonadotropin (PMSG) injected intraperitoneally (i.p.) induce the maturation of oocytes mimicking follicle-stimulating hormone (FSH).

2. Then, 48h later, human chorionic gonadotropin (hCG) is administrated i.p. to induce the rupture of the matured follicles mimicking the effect of luteinizing hormone (LH).

Generally, 5IU are sufficient for the majority of the strains. The age and weight of the female usually affects the number of oocytes/zygotes recovered. The most appropriate age for inducing superovulation is different for each strain, but commonly varies between 4 to 5 weeks. Also consider that underweight females yield a low number of oocytes/zygotes after superovulation.

3. Finally, for zygote harvesting purposes it is necessary set the mating after hCG i.p. injection. The following day after mating the female will be checked for a copulation plug and the oocytes/zygotes will be harvested between 8-to-10 hours after midnight of hCG administration day.

3.1.2. Making glass pipettes

1. Melt a Pasteur pipette by rotating it in a burner until the glass become soft.
2. Take away the glass from the flame and pull both ends to produce a tube with a thinner diameter. Do not pull the glass while is still in the flame.
3. Cut the extreme of the glass pipette with a diamond pencil.

3.1.3. Preparing microdrops culture plates

1. Dispense 100 μ L drops of M2 medium on the bottom of 35mm plastic dish.
2. Cover the drops with mineral oil in order to avoid desiccation.
3. Place the microdrop culture dish in the incubator at 37°C and 5% CO₂.

3.1.4. Collecting oocytes and zygotes

1. Open the abdominal cavity of female mice and pull the uterus, oviduct and ovary away from the body.
2. First, cut between the oviduct and ovary, and then cut again between uterus and the oviduct close to the oviduct.
3. Transfer the oviduct to a 35-mm petri dish containing M2 medium.
4. Oocytes and zygotes are located in the ampulla, which resembles an enlarged bag on the upper part of the oviduct. Using one pair of forceps to tear the ampulla and the oocytes/zygotes will come out themselves.
5. At that point, oocytes/zygotes are surrounded by cumulus cells. To remove them, add

hyaluronidase solution in the petri dish containing the M2 medium and wait 1-3 minutes.

Cumulus cells will start to disaggregate from oocytes/zygotes facilitating their individual collection.

6. Transfer the oocytes/zygotes to fresh M2 medium and clean up them pipetting up and down several times to remove any trace of hyaluronidase.

7. Finally, transfer the desired number of oocytes/zygotes into a new petri dish containing 1X PBS solution.

8. Use a pipette to transfer the oocytes/zygotes to a new petri dish containing Tyrode's solution to remove the zona pellucida. The oocytes/zygotes have to be monitored constantly under a stereomicroscope during the treatment to prevent prolonged exposure to acidic pH. Once the zona pellucida is removed, transfer them to a new petri dish with different drops containing PBS 1X in order to clean them from Tyrode's solution treatment. Then the oocytes and zygotes are ready to be storage.

9. Collect the oocytes/zygotes on a PCR Eppendorf tube in a final volume of 4 μ L using a P2 micropipette.

3.2. Isolating epididymal spermatozoa

1. After euthanasia, adult mouse males were abdominal dissected and epididymis are collected with fine forceps and scissors and placed in M2 medium in a petri dish (the number of animals depends on the amount of RNA, about 20 mice).

2. Fat, veins and cauda epididymis are discarded. Caputs containing mature spermatozoa are collected.

3. Each 4-5 caputs are immersed in drops of 200 μ l of M2 media covered with mineral oil.

4. By 30G needle fine cuts in the epididymis facilitate the release of motile spermatozoa to the media.

5. Drops of M2 with caputs are incubated at 37°C during 10-15 min to allow swim-out and

dispersion of spermatozoa into the media.

6. The media with suspension of spermatozoa from groups of 3-4 drops are collected in Eppendorf tubes at 37°C until use.

7. Spermatozoa are isolated by Percoll gradient. In 15 ml conical tube, in a horizontal position as much as possible, place carefully successive layers of 3ml of 90% Percoll in PBS followed by 45% and finally 1ml of spermatozoa suspension.

8. Tubes are centrifuged at 800xg during 20 min.

9. Spermatozoa are as pellet that is washed two times with PBS and centrifuged each wash.

10. To eliminate potential somatic cell contamination by osmotic shock the pellet can be resuspended in RNase free water centrifuge 3min at 6000xg discarding supernatant.

11. Resuspend in PBS. Take 1 µL of suspension, diluted 400X and count concentration by Neubauer chamber.

12. Resuspend in 100 µL of Trizol and frozen at -80°C until use for RNA isolation.

3.3. RNA isolation from low amount of cells

1. To isolate RNA for samples with small amount of cells (oocytes and zygotes) (1×10^4 - 1×10^5 cells) add 250 µL of TRIzol, as well as for 5×10^7 spermatozoa. For samples $> 1 \times 10^5$ add $< 1 \times 10^6$ we recommend to add 500 µL of TRIzol.

2. Homogenize the sample pipetting up down several times and incubate the sample 5 min at room temperature.

3. Add 66 µL of chloroform and shake the tube during 30 sec. and incubate at room temperature 2-5 min.

4. Centrifuge the sample at $12,000 \times g$ for 15 minutes at 4°C.

5. Place the aqueous phase into a new Eppendorf tube.

6. Add 5µg of glycogen blue to make visible the RNA pellet in the next steps.

7. Add 166 µL of isopropanol and incubate 2h at -20°C.

8. Centrifuge at 12,000xg for 10 min. at 4°C.
9. Remove the supernatant leaving just the bluish pellet containing precipitated RNA.
10. Wash the RNA pellet, with 250 μ L of 75% ethanol (or 80% if a enrichment of small RNAs is desired). Vortex the sample to mix.
11. Centrifuge the sample at $7500 \times g$ for 5 minutes at 4°C, and discard the supernatant.
12. Vacuum or air dry the RNA pellet for 5–10 minutes. Do not dry the pellet by vacuum centrifuge.
13. Resuspend the RNA pellet in RNase-free water and incubate in a water bath or heat block set at 55–60°C for 15 min.
14. In nanoDrop Spectrophotometer, use absorbance of RNA at 260 and 280 nm. to determine total RNA concentration.
15. Finally, determine the quality of RNA by 2100 Bioanalyzer using supplier procedures (for spermatozoa *see* **Note 3** and Figure 2).

Fig. 2

Electropherograms of total RNAs from male germ cell and tissue assessed by Bioanalyzer. PGCs correspond to Primordial Germ Cells from male embryos of 13.5 postcoitum days; SPZ isolated spermatozoa following the protocols detailed in this chapter; Adult testis homogenate. Note the high proportion of small RNAs and practically absence of 18S and 28S RNAs in SPZ sample.

3.4. Next-Generation Sequencing (NGS) and Quality Control and Trimming of Files from NGS

Sequencing RNA using a small RNASeq protocol in a Massive Parallel Sequencing platform (NGS)(*see Note 4*).

This protocol was designed to work with FASTQ format files (*.fastq), which are commonly generated by Illumina sequencers. This format can also be easily obtained from different Massive Parallel Sequencing platforms (NGS). We recommend to download or copy all the *.fastq files to the same folder (e.g., into a folder called Reads/inside Documents/).

3.4.1. Quality Control

For quality control we used FastQC software (*see* Subheading 2.2.2 , **item 7**). A web-browser will open showing the quality report by opening the report .html file. The graph: “Per base sequence quality” shows a box plot of quality scores in relation to the global sequences. FastQC website details some examples of different quality reports.

1. Run FastQC inside your Reads/ folder:

```
cd Documents/Reads/
```

```
fastqc -t 7 input_file.fastq
```

Where -t is the number of threads that will be used to run FastQC (seven in last command) and input_file is the .fastq file generated by the NGS platform.

2. For running FastQC on multiple files located in the same directory follow:

```
fastqc -t 7 input_file_1.fastq input_file_2.fastq input_file_N.fastq
```

Where -t is the number of threads to run FastQC and input_file_1, 2, N are the different .fastq files to analyze.

3. For running FastQC interactively, type in the terminal:

```
fastqc -t (number of threads)
```

A window of FastQC will open go to “File”, then “Open” and select your *. fastq file and run the analysis.

3.4.2 Adapter trimming

1. To trim adapters and low quality sequences from raw data we recommend to use Trim Galore. Type the following commands in the terminal in your Documents folder:

```
cd Reads/
```

```
/Software/trim_galore -q 28 --small_rna --length 18 input_file.fastq
```

Trim Galore generates a new file with the extension `.trimmed.fastq` maintaining the name of the input file; (in this example, `input_file.trimmed.fastq`). The `-q` flag indicates the minimum Phred score to trim the 3' of the reads (by default is 20). The information provided by FastQC can be used to decide the minimum Phred score to trim the samples. For detection of adapters in the sample use the flag `--small_rna`. By default this flag uses an Illumina Small RNA Adapter (“TGGAATTCTCGG”). If a different adapter is used, include `-a` flag and the sequence adapter (in DNA format) in substitution of `--small_rna` flag. Flag `--length` indicates the minimum length of sequence to keep, in this case the minimum length sequence was 18 nt. By default, Trim Galore generates a `.txt` file with the statistics of the trimming processes.

3.5. Identification of miRNA by mapping against precursor miRBase.

We recommend the use of precursors (`steamloop_mmu_21.fasta`) to filter miRNA reads for the identification of canonical miRNA and Isomirs described in the next steps of this protocol (*see Note 5*).

1. Generate a bowtie index from the precursor miRNA sequences. Open a new terminal and type:

```
cd Documents/Databases/
```

```
fasta_nucleotide_changer -i steamloop_mmu_21.RNA.fasta -o steamloop_mmu_21.fasta -d
```

```
bowtie-build steamloop_mmu_21.fasta steamMir21
```

MiRBase sequences are deposited in RNA nucleotide code, so the first step is the transformation of RNA to DNA using FASTX-toolkit (first command). Bowtie-build builds a Bowtie index from a set of DNA sequences, generating 6 files with extension: `.ebwt`.

Check them typing in the terminal “ls”. To index any file in fasta format only change the first argument for the fasta file and the second argument for name of database (see Note 5).

2. The next step is to map all trimmed sncRNA sequences to miRBase precursors to filter sequences that was not complementary to reported miRNA. For this purpose we use the bowtie aligner. Type in your terminal:

```
cd ../  
mkdir MapPrecursors/  
mkdir SAMPrecursors/  
cd Reads/  
bowtie ../Databases/steamMir21 input_file.trimmed.fastq -p 4 -v 3 -a --strata --best --tryhard --chunkmbs 256 -  
S --al ../MapPrecursors/input_file.steam21.fastq > ../SAMPrecursors/input_file.steam21.sam
```

To run bowtie on multiple fasta sequences in the same directory:

```
for i in `ls *.trimmed.fastq` ; do bowtie ../Databases/steamMir21 -f input_file.trimmed.fasta -p 4 -v 3 -a --strata  
--best --tryhard --chunkmbs 256 -S --al ../MapPrecursors/$i.steam21.fastq > ../SAMPrecursors/$i.steam21.sam  
; done
```

Here, -p indicates the number of threads to run bowtie; -v is the bowtie alignment mode followed by the number of mismatches, in this case mismatches3 mismatches are allowed (to perform a stringent analysis, put 0 to 1 mismatches, change -v 3 by -v 0); last flags are allow bowtie to obtain the best alignments for each read (see bowtie manual, 2.2.2.5). Finally, to control the output of bowtie use the flag --al (directory where the mapped sequences are redirected in this case to MapPrecursors) and -S (to return the mapping file in .sam format instead of .map format). Symbol > is to redirect the output to a specific folder (in this case SAMPrecursors/). SAM format is easy to transform to BAM format.

3.6. miRNA and isomIR identification using Isomirage.

1. Download and install the Isomirage. Go to Isomirage [11] web page, download section (second tab) and download “Custom libraries mouse” (is a zipped file) in the Database folder and “Executable” (is a jar file) in the Software folder. (See

).

2. Unzip the Custom libraries; make a bowtie database with bowtie-build and run the alignment. To setup the database of Isomirage type in the terminal, in the documents folder:

```
cd Database/
```

```
unzip mmu21t.zip
```

3. Decompression of zip generate two files (you can type “ls” in the terminal to list them), one is a genome file that is needed for running Isomirage and the second is a fasta file with sequences of different miRNA forms. Next steep is to generate a bowtie index database and map the fastq reads to such database for Isomirage analysis. In the Database folder type in the terminal:

```
Bowtie-build mmu21t.fasta mmu21t
```

```
cd ..
```

```
mkdir IsomirageAnalysis/
```

```
cd MapPrecursors/
```

```
bowtie ../Databases/steamMir21 input_file.trimmed.steam21.fastq -p 4 -v 0 --best --tryhard --chunkmbs 256 >
../IsomirageAnalysis/input_file.steam21.map
```

4. For running multiple fasta sequences in the same directory:

```
for i in `ls *. trimmed.steam21.fastq` ; do bowtie ../Databases/steamMir21 $i -p 4 -v 0 --best --tryhard --
chunkmbs 256 > ../IsomirageAnalysis/$i.steam21.map ; done
```

5. Description of the bowtie flags is detailed in the 3.6.2 section. After the mapping process, a map file for each fastq file will be generated. When bowtie finishes, a report in the terminal indicating the total sequences, mapped sequences, not mapped sequences and number of

alignments will be reported.

6. Isomirage runs in Java, and has a user-friendly interface. To run Isomirage, type in the terminal in the Documents/ folder:

```
cd Software/  
java -jar Isomirage.jar
```

To start the analysis, go to Files in the menu, and select a .map file (in the Documents/IsomirageAnalysis folder) using the option *Open bowtie map*. It is possible to include different files of mapping with the option “*Add more map files*” from the Files menu. Finally, select a genome file for the analysis in File menu and “*Choose genome*”. In this case is a mmu21t.genome file in the Documents/Database/. Start the analysis using “*Run*” option in the analysis section. During analysis, Isomirage indicates the different phases of the analysis in the progress window.

7. Isomirage generates different output files (in this case located at Documents/IsomirageAnalysis folder) in .txt format (it is possible to open the files using any spreadsheet software). Two types of reports are generated based in the first column: one called _data.txt (name of the miRNA, trimmed variants and sequence) and the other called _summary.txt (containing names of the miRNA and trimmed variants) and 3 types of normalization methods, raw counts, Robust Multi-array Average (RMA) counts and Reads Per Kilobase Million (RPKM) normalized counts. Isomirage always produces six output files independently of the number of samples analyzed. If different samples are processed at the same time, Isomirage output consistent in the six files in which, each column represent a value for each file. Header of column is renamed with the name of the map file.

8. Generate a list of raw counts only containing the canonical miRNA sequences using: input_file.trimmed.steam21_linked.txt_raw_summary.txt using the next bash scripts , by typing in the terminal:

```
grep "miRNA" input_file.trimmed.steam21_linked.txt_raw_summary.txt | awk -F'_' 'BEGIN{print  
"miRNA""\t""Counts"} {print $1""$3}' | sed 's/t//g' >
```

```
input_file.trimmed.steam21_raw_summary.CANONICAL.txt
```

9. Obtain the list of miRNA and isomiR variants by:

```
awk -F '\t' '$2 >= 1 {print $1\"$2}' input_file.trimmed.steam21_linked.txt_raw_summary.txt | awk -F ' ' '{print $1\"$2\"$3\"$4}' > input_file.trimmed.steam21_raw_summary.FILTERCOUNTS.csv
```

This script transforms the output of Isomirage to tabular format: miRNA names, and modifications of sequences. To change the threshold of counts to consider any sequence, modify the value of ≥ 1 . For example for 5 counts ≥ 5 .

10. To classify the different isomiRs , apply:

```
awk -F '\t' '$2 >= 1 {print $1}' input_file.trimmed.steam21_linked.txt_raw_summary.txt | awk -F ' ' '{print $2\"$3}' | sort | uniq -c -d > input_file.trimmed.steam21_raw_summary.NUMBEROFMODIFICATIONS.csv
```

3.7. Differential expression analysis using DESeq package.

For differential expression analysis we recommend to process all samples at the same time.

When processing different samples at once using Isomirage, inputs_raw_summary.txt file is ready to run DESeq on it. We performed the analysis with two files, input_file1 and input_file2. We use for differential expression analysis the file:

```
input_files.trimmed.steam21.map_linked.txt_raw_summary.txt
```

1. For differential expression analysis of two libraries analyzed with Isomirage, in the folder containing the input_file.trimmed.steam21.map_linked.txt_raw_summary.txt, type in the terminal:

```
R
```

Alternatively, run R in Rstudio (see Note 2).

2. In the R terminal or in the Rstudio script screen:

```
library( "DESeq" )
```

Call DESeq and the associated packages.

```
countsTable <- read.delim(“input_file.trimmed.steam21.map_raw_summary.txt”)
```

```
rs <- rowSums(countsTable)
```

```
use <- (rs > 1)
```

```
countTableFilt <- countsTable[ use, ]
```

These codes are referred to input counts table in R (in this case the input table is:

input_file.trimmed.steam21.map_raw_summary.txt), and perform the filter threshold as one count.

3. Raw counts from the different sequencing libraries are normalized adjusting for different library sizes among the treatments (see Note 7):

```
conds <- factor( c( "input_file1","input_file1" ) )  
cds <- newCountDataSet( countTableFilt, conds )  
cds <- estimateSizeFactors( cds )  
write.table(counts(cds,normalized=TRUE), file =  
"input_file.trimmed.steam21.map_NORCOUNTS_summary.txt", sep = "\t")
```

Normalized counts are generate in txt file called:

input_file.trimmed.steam21.map_NORCOUNTS_summary.txt, that can be open using any spreadsheet software.

4. For differential expression analysis type:

```
cds <- estimateDispersions( cds, method="blind", fitType="local", sharingMode="fit-only" )  
res <- nbinomTest( cds, "input_file1", "input_file2")  
write.table(res, file = "input_file1vsinput_file2.txt", sep = "\t")
```

In these cases do not use the p-values or q-values to filter the genes because no replicates are present. The differential expression is contained in:

input_file1vsinput_file2.txt.

3.8. Uploading the NGS data to GEO-NCBI.

We recommend the submission of sncRNA-seq data to the Gene Expression Omnibus (GEO) genomics data repository. Submission of sncRNA transcriptomic data in GEO facilitates a GEO accession number(s), which is necessary for the most publication processes. For submission of samples it is necessary to collect all the information concerning the experiment and individual for complete Metadata worksheet, a normalization data (such

as generate in the section 3.8) needed for Matrix worksheet and reads generate by sequencing platform in fastq format.

A nice and user-friendly manual of submission is available in the NCBI portal:

<http://www.ncbi.nlm.nih.gov/geo/info/seq.html>

4. Notes

1. Ubuntu operating system can be downloaded free from:

<http://www.ubuntu.com/download/desktop>. We recommend download the LTS version.

2. R installation includes a step by step on the web. Alternatively R studio can be used that contains graphical environment for R making designed to be as user-friendly environment.

<https://www.rstudio.com/>.

3. Consider that total sperm profile obtained by Bioanalyzer present in spermatozoa does not show any profile as is detected in any cell type. It is known that in spermatozoa the 18S and 28S RNAs almost disappear in the sample but the amount of small RNAs is proportionally higher (Figure 2).

4. For RNA-Seq sequencing different Massive Parallel Sequencing platforms are available (for example Illumina sequencing by synthesis and Ion Torrent semiconductor sequencing platforms). NGS platforms have specific protocol for small RNA-Seq libraries construction and sequencing. Due to the peculiar characteristics of small RNA-seq libraries [12] we recommend to perform the sequencing to get around 10 millions of reads per library in single end sequencing assay (sequencing involves sequencing RNA from only one end) and read

length of 50 cycles. Output format for bioinformatics protocol is FASTQ.

5. Go to miRBase web page, browse section and selected *Mus musculus*. A web-page with all pre- and mature miRNA forms is loaded. Go to the final of the web page, and select in dropped menus: *Steam-loop sequences*, *Unaligned fasta format*, *Select all button* and finally *Fetch sequence* button. A txt web page will be load. Simple download the web page using the save button of your browser. We recommend rename the file as steamloop_mmu_21.RNA.fasta, fasta extension is important for different phases of these protocol. Alternatively, it is possible to download the mature microRNA sequence by selecting “Mature sequence” in the drop menu.

6. Isomirage needs the Java Virtual Machine working in the system. In Ubuntu to check the Java Virtual Machine, type in the terminal:

```
java -version
```

If Java was installed, three lines will be shown similar to:

```
java version "1.8.0_77"
```

```
Java(TM) SE Runtime Environment (build 1.8.0_77-b03)
```

```
Java HotSpot(TM) 64-Bit Server VM (build 25.77-b03, mixed mode)
```

To install Java manually (<http://askubuntu.com/questions/521145/how-to-install-oracle-java-on-ubuntu-14-04>)

7. We recommend using the **cds** values obtained from initial trimmed libraries. For this purpose, it is possible to obtain the read values of all sequencing libraries. When obtained the values of cds from the total library, type `SizeFactors( cds )` to obtain the values. These values can be used by typing in R:

```
cds$SizeFactors <- cds values .
```

Acknowledgments

We thank D. Fernández-Pérez for critical review of the manuscript. The research in JdM's laboratory was supported by a grant from MINECO (BFU2013-42164-R), Spain.

References

1. Schultz RM (1993) Regulation of zygotic gene activation in the mouse. *Bioessays* 15 (8):531-538.
2. Pauli A, Rinn JL, Schier AF (2011) Non-coding RNAs as regulators of embryogenesis. *Nature Rev* 12 (2):136-149.
3. Yuan S, Schuster A, Tang C, Yu T, Ortogero N, Bao J, Zheng H, Yan W (2016) Sperm-borne miRNAs and endo-siRNAs are important for fertilization and preimplantation embryonic development. *Development* 143 (4):635-647.
4. García-López J, Brieño-Enríquez MA, del Mazo J (2013) MicroRNA biogenesis and variability. *BioMol Concepts* 4 (4):367-380
5. Tang F, Kaneda M, O'Carroll D, Hajkova P, Barton SC, Sun YA, Lee C, Tarakhovsky A, Lao K, Surani MA (2007) Maternal microRNAs are essential for mouse zygotic development. *Genes Dev* 21 (6):644-648
6. Giraldez AJ (2010) microRNAs, the cell's Nepenthe: clearing the past during the maternal-to-zygotic transition and cellular reprogramming. *Curr Opin Genet Dev* 20 (4):369-375.
7. Giraldez AJ, Mishima Y, Rihel J, Grocock RJ, Van Dongen S, Inoue K, Enright AJ, Schier AF (2006) Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. *Science* 312 (5770):75-79
8. Garcia-Lopez J, Hourcade Jde D, Alonso L, Cardenas DB, del Mazo J (2014) Global characterization and target identification of piRNAs and endo-siRNAs in mouse gametes and zygotes. *Biochim Biophys Acta* 1839 (6):463-475.

9. Garcia-Lopez J, Alonso L, Cardenas DB, Artaza-Alvarez H, Hourcade JdeD, Martinez S, Brieno-Enriquez MA, del Mazo J (2015) Diversity and functional convergence of small noncoding RNAs in male germ cell differentiation and fertilization. *RNA* 21 (5):946-962.
10. Gou LT, Dai P, Yang JH, Xue Y, Hu YP, Zhou Y, Kang JY, Wang X, Li H, Hua MM, Zhao S, Hu SD, Wu LG, Shi HJ, Li Y, Fu XD, Qu LH, Wang ED, Liu MF (2014) Pachytene piRNAs instruct massive mRNA elimination during late spermiogenesis. *Cell Res* 24 (6):680-700.
11. Muller H, Marzi MJ, Nicassio F (2014) IsomiRage: From Functional Classification to Differential Expression of miRNA Isoforms. *Front Bioeng Biotechnol* 2:38.
12. Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, Szczesniak MW, Gaffney DJ, Elo LL, Zhang X, Mortazavi A (2016) A survey of best practices for RNA-seq data analysis. *Genome Biol.* 17(1):13.