Human tRF & tiRNA Sequencing Service Report

Aksomics Inc.

January 17, 2020

Project:	**Human tRF & tiRNA sequencing**
Customer:	**毛启星**
Company/Institute:	**南京医科大学**
Project Code:	**H1910024**
Organism:	**Human**

Contents

1 General Information and Summary

1.1 Workflow of The Project

1.1.1 RNA Sample Submission & QC

High quality samples are most important for small RNA sequencing projects. The recommended purified total RNA amount is 1∼2μg. Alternatively, we offer to isolate RNA from your sample sources, including cultured cells, tissues, serum or plasma.
Before the sequencing experiment, we check the integrity and quantity of each RNA sample using agarose gel electrophoresis and Nanodrop^{TM} instrument. The RNA QC results are included in the service report.

1.1.2 Pretreatment of tRF & tiRNA

tRNA derived fragments (tRF & tiRNA) are heavily decorated by RNA modifications that interfere with small RNA-seq library construction. We do the following treatments before library preparation for total RNA samples: 3' -aminoacyl (charged) deacylation to 3' -OH for 3' adaptor ligation, 3' -cP (2' ,3' -cyclic phosphate) removal to 3' -OH for 3' adaptor ligation, 5' -OH (hydroxyl group) phosphorylation to 5' -P for 5' -adaptor ligation, m1A and m3C demethylation for efficient reverse transcription.

1.1.3 Library preparation

Sequencing libraries are size-selected for the RNA biotypes to be sequenced using an automated gel cutter. The libraries are qualified and absolutely quantified using Agilent BioAnalyzer 2100.

1.1.4 tRF & tiRNA sequencing

For standard small RNA sequencing on Illumina NextSeq instrument, the sequencing type is 50bp single-read.

1.1.5 The mature and precursor tRNA sequences

tRNA sequences of cytoplasmic were downloaded from GtRNAdb [1,2]. tRNA sequences of mitochondrial were predicted with tRNAscan-SE [3,4] software. To generate the mature tRNA libraries, we removed the predicted intronic sequences (if present) and added an additional 3' -terminal "CCA" to each tRNA. To generate the precursor tRNA libraries, we included 40 nucleotides of flanking genomic sequence on either side of the original tRNA sequence [5].

1.1.6 tRF & tiRNA data analysis

Comprehensive data and result of statistical analyses are provided in the Arraystar tRF & tiRNA-seq data analysis package (see also data analysis workflow):

- Raw sequencing data QC
- Mapping Summary: Alignment information of sequence mapping to the reference genome
- Identification of tRF & tiRNA
- Differential expression analysis of tRF & tiRNA and miRNA
- The scatter plots of differentially expressed tRFs & tiRNAs and miRNA
- Supervised analysis: For each group comparison, we generate a hierarchical clustering heatmap and a volcano plot to display significantly differentially expressed tRFs & tiRNAs and miRNA.

1.1.7 Project report and technical support

The final report contains:

- An easy-to-read data report containing general information and summary of the project, sample and sequencing data QC, and an overview of analysis results with publication-quality graphic illustrations;
- Data analysis files readily adoptable for publication or performing your own analysis;
- A methods section describing the experimental procedures and protocols used;
- Raw sequencing data in FASTQ formats file;
- GEO submission guide for the small RNA sequencing data.

After you receive the project report, we will continue to offer technical support by answering questions you may have or helping with the information of follow-up studies, such as qPCR validation of the results.

1.2 Sample Description

1.2.1 Sample Groups

Organism: Human

Sample Type: RNA

Sample Count: 10

Table 1. Sample groups

Sample	Group	QC	Lib	Status
2018L0071TT	T	Pass	OK	OK
2018L0030TT	T	Pass	OK	OK
2018L0121TT	T	Pass	OK	OK
2018L0049TT	T	Pass	OK	OK
2018L0100TT	T	Pass	OK	OK
2018L0071TN	N	Pass	OK	OK
2018L0030TN	N	Pass	OK	OK
2018L0121TN	N	Pass	OK	OK
2018L0049TN	N	Pass	OK	OK
2018L0100TN	N	Pass	OK	OK

Sample: Samples name
Group: A comma-separated list of group names
QC: RNA sample quality control
Lib: Sequencing library quality assessment
Status: Final status in the experiment

1.2.2 Comparison Setup

Table 2. Comparison Setup

Test Group	Control Group	Fold Change	P Value	Q Value
T	N	1.5	0.05	1

Test Group:	set up test group for comparison
Control Group:	set up control group for comparison
Fold Change:	set up cutoff of fold change for comparison
P Value:	set up cutoff of p-value for comparison
Q Value:	set up cutoff of q-value for comparison

1.3 Experiment Workflow

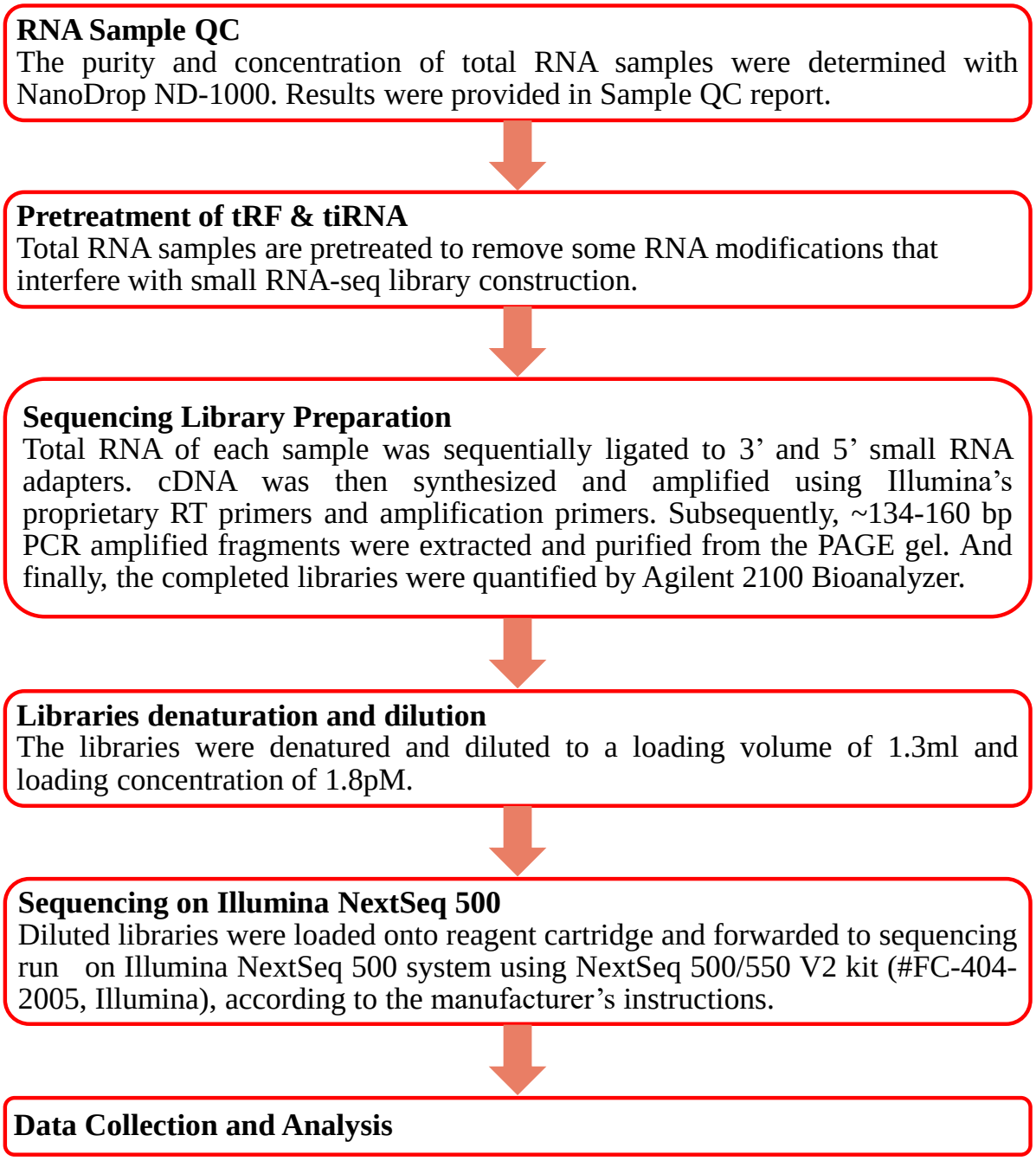

Figure 1. tRF & tiRNA-seq experiment workflow

Total RNA samples are qualified by agarose gel electrophoresis and quantified using Nanodrop. We use the commercial kit for tRF & tiRNA-seq library preparation, which includes 3' -adapter and 5' -adapter ligation adaptor ligation, cDNA synthesis and library PCR amplification. The prepared tRF & tiRNA-seq libraries are finally quantified using Agilent BioAnalyzer 2100, then sequenced using Illumina NextSeq 500.

1.4 Data Analysis Workflow

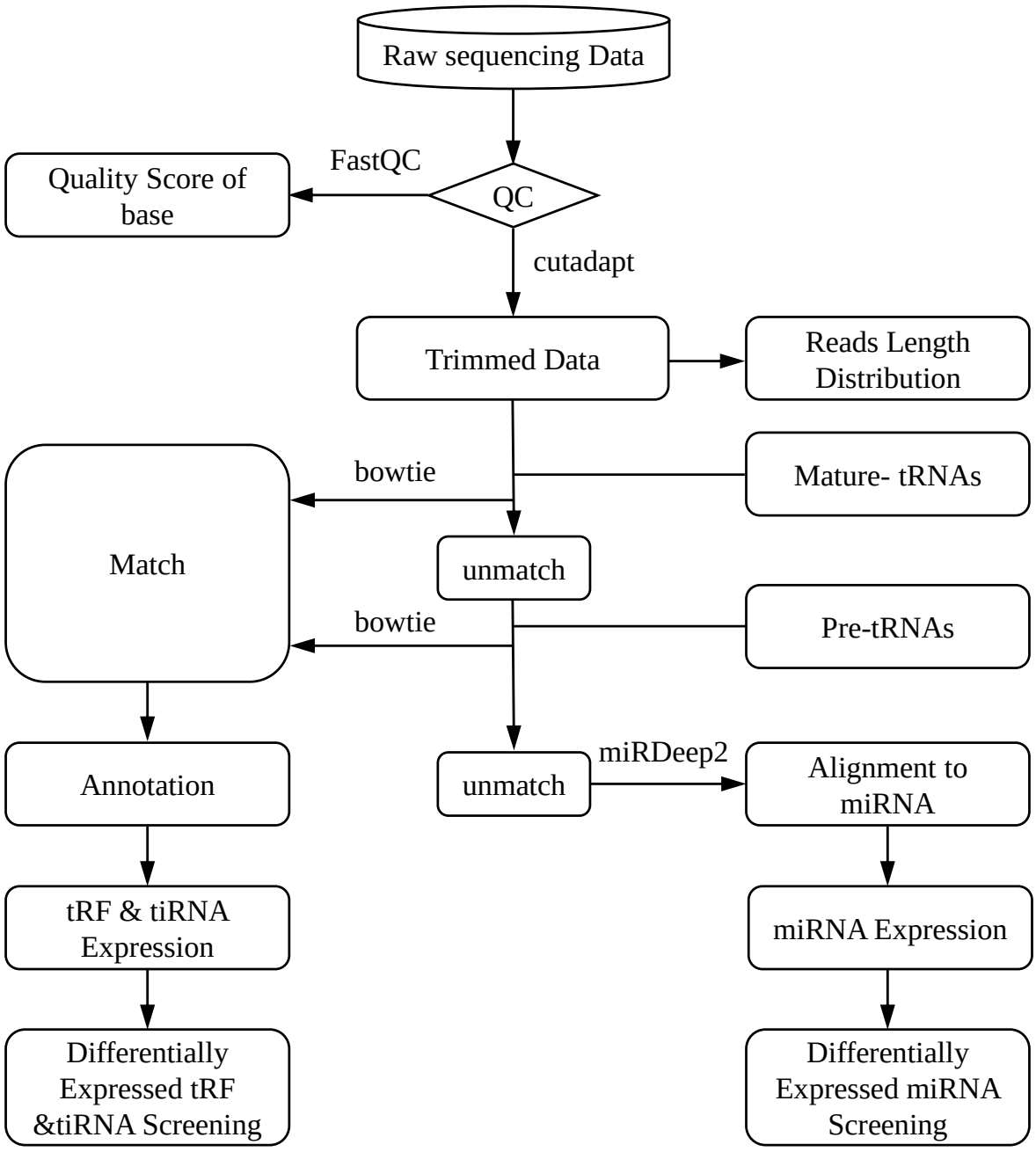

Figure 2. tRF & tiRNA-seq data analysis workflow

Raw sequencing data generated form Illumina NextSeq 500 that pass the Illumina chastity filter are used to the following analysis. Trimmed reads (trimmed 5' , 3' -adaptor bases) are aligned allowing for 1 mismatch only to the mature tRNA sequences, then reads that do not map are aligned allowing for 1 mismatch only to precursor tRNA sequences with bowtie software [6]. The remaining reads are aligned allowing for 1 mismatch only to miRNA reference sequences with miRDeep2 [7]. Based on alignment statistical analysis (mapping ratio, read length, fragment sequence bias), we determine whether the results can be used for subsequent data analysis. If so, the expression profiling and differentially expressed tRFs & tiRNAs and miRNAs are calculated. Principal Component Analysis (PCA), Correlation Analysis, Pie plots, Venn plots, Hierarchical clustering, Scatter plots and Volcano plots are performed for the expressed tRF & tiRNA in R or perl environment for statistical computing and graphics.

1.5 Summary

Dear 毛启星,

We have completed your small sequencing project. Total RNA from each sample was quantified using a NanoDrop ND-1000 instrument. Total RNA samples were first pretreated as following to remove some RNA modifications that interfere with small RNA-seq library construction: 3' -aminoacyl (charged) deacylation to 3' -OH for 3' -adaptor ligation, 3' -cP (2' ,3' -cyclic phosphate) removal to 3' -OH for 3' -adaptor ligation, 5' -OH (hydroxyl group) phosphorylation to 5' -P for 5' -adaptor ligation, m1A and m3C demethylation for efficient reverse transcription. Then pretreated total RNA was used to prepare the sequencing library in the following steps: 1) 3' -adapter ligation; 2) 5' -adapter ligation; 3) cDNA synthesis; 4) PCR amplification; 5) size selection of 134∼160bp PCR amplified fragments (corresponding to 14∼40nt small RNA size range). The libraries were denatured as single-stranded DNA molecules, captured on Illumina flow cells, amplified in situ as sequencing clusters and sequenced for 50 cycles on Illumina NextSeq 500 system per the manufacturer' s instructions.

Image analysis and base calling are performed using Solexa pipeline v1.8 (Off-Line Base Caller software, v1.8). Sequencing quality are examined by FastQC [8] and trimmed reads (pass Illumina quality filter, trimmed 5' ,3' -adaptor bases by cutadapt [9]) are aligned allowing for 1 mismatch only to the mature tRNA sequences, then reads that do not map are aligned allowing for 1 mismatch only to precursor tRNA sequences with bowtie software [6]. The remaining reads are aligned allowing for 1 mismatch only to miRNA reference sequences with miRDeep2 [7]. The abundance of tRF & tiRNA and miRNA are evaluated using their sequencing counts and is normalized as counts per million of total aligned reads (CPM). The tRFs & tiRNAs and miRNAs differentially expressed are screened based on the count value with R package edgeR [10]. Principal Component Analysis (PCA), Correlation Analysis, Pie plots, Venn plots, Hierarchical clustering, Scatter plots and Volcano plots are performed in R or perl environment for statistical computing and graphics of the expressed tRF & tiRNA.

Respectfully,
Aksomics Services.
Aksomics Inc.

2 tRF & tiRNA-seq Data Analysis

2.1 Introduction of tRF & tiRNA

Transfer RNA (tRNA) is an adaptor molecule that decodes mRNA and translates protein. Recent studies have demonstrated that tRNAs also serve as a major source of small non-coding RNAs having distinct and varied functions. tRNA derived fragments (tRF & tiRNA) can be broadly classified into two main groups: tRNA related fragments (tRF) generated from mature or precursor tRNA and tRNA halves (tiRNA) generated by specific cleavage in the anticodon loops of mature tRNA, with characteristic sizes, nucleotide compositions, functions and biogenesis [11–13]. tRF & tiRNA have been suggested to play roles in cell proliferation, priming of viral reverse transcriptases, regulation of gene expression, RNA processing, modulation of the DNA damage response, tumor suppression, and neurodegeneration [14]. tRF & tiRNA are classified into various types depending on where they map on the precursor or mature tRNA transcript (Figure. 3) [14]. On the basis of their mapped positions, these tRF & tiRNA are of five types: tRF-5, tRF-3, tRF-1, tRF-2 and tiRNA [14–16]:

- The tRF-5 are generated from 5' ends of the mature tRNA. They are divided into subclasses of tRF-5: tRF-5a (14∼16nt), tRF-5b (22∼24nt) and tRF-5c (28∼32nt) by specific lengths.
- The tRF-3 are generated from 3' ends of the mature tRNA. The tRF-3 are mainly either 17∼18nt or 19∼22nt, based on which they are subclassified as either tRF-3a or tRF-3b.
- The tRF-1 (14∼33nt) are generated from 3'ends of the precursor tRNA. Analysis of tRNA fragments that originate from precursor tRNA trailer sequences indicate that the 5' end of tRF-1 matches with the cut site of RNase Z and the 3' end matches with an RNA polymerase III (RNA pol III) transcription termination signal.
- The tRF-2 contains only the anticodon stem and loop tRNA.
- The tRNA halves (tiRNA) are generated by specific cleavage in the anticodon loops of mature tRNA and therefore are 31∼40 bases long. There are two subclasses of tiRNA based on whether they include the sequence 5' or 3' of the anticodon cleavage site: tiRNA-5 start from 5' end of mature tRNA and end in the anticodon loop, whereas tiRNA-3 start from the anti-codon loop and end at 3' end of mature tRNA.

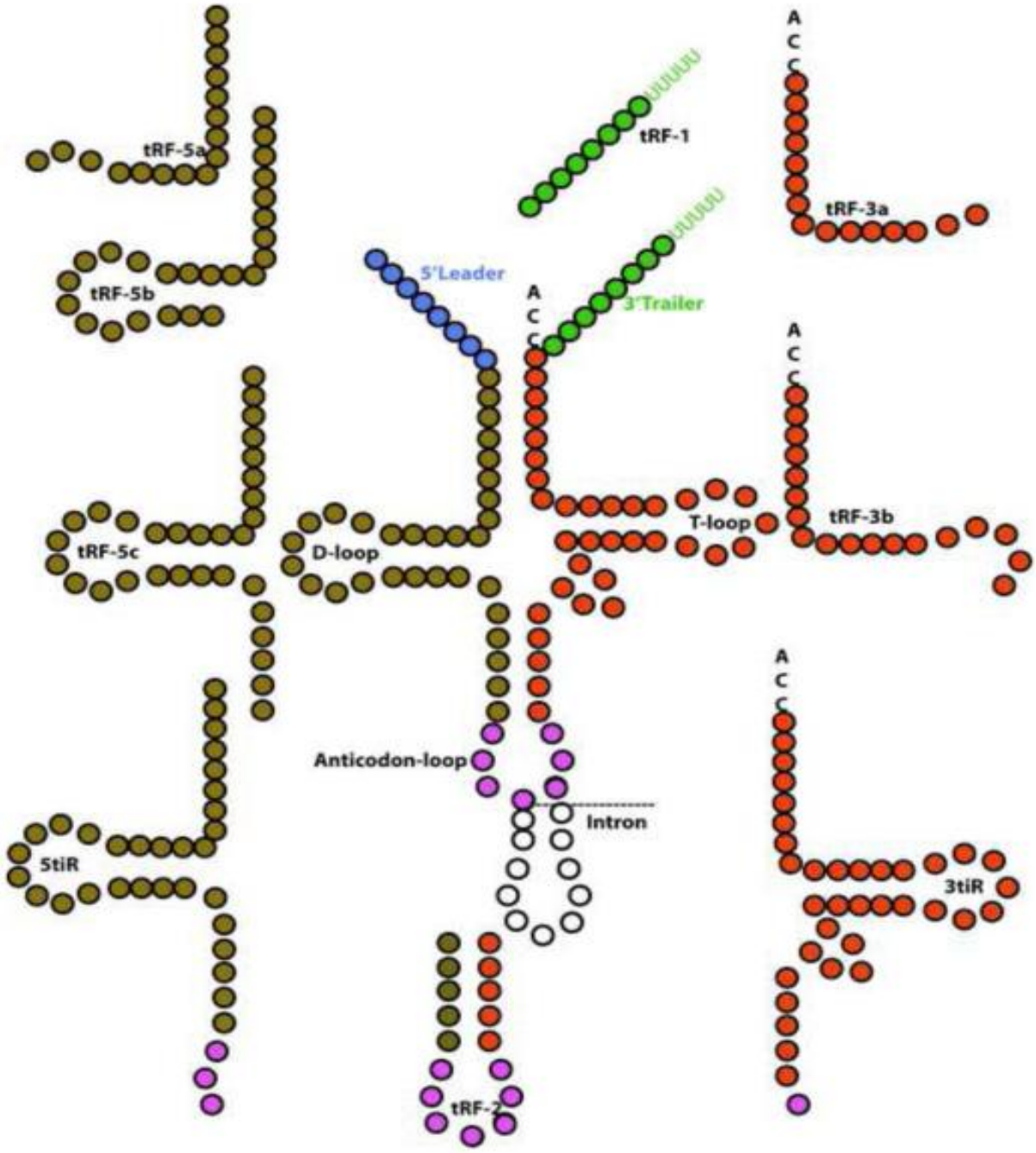

Figure 3. Classification of tRF & tiRNA. Fragments from tRNAs color coded to indicate area of origin. White circles indicate an intron present in some tRNAs (e.g. chr16.trna4-ProAGG) that is normally spliced out by TSEN and CLP1. tRF-2 is a class proposed in the review [14].

2.2 Sequencing Quality Control

Raw data files in FASTQ format were generated from the Illumina sequencer. To examine the sequencing quality, the quality score plot of each sample was plotted. Quality score Q is logarithmically related to the base calling error probability (P):

$$Q = -10log_{10}(P) \quad (1)$$

For example, Q30 means the incorrect base calling probability to be 0.001 or 99.9% base calling accuracy.

Note: All the sequencing quality control plots can be found in Sequence_QC folder.

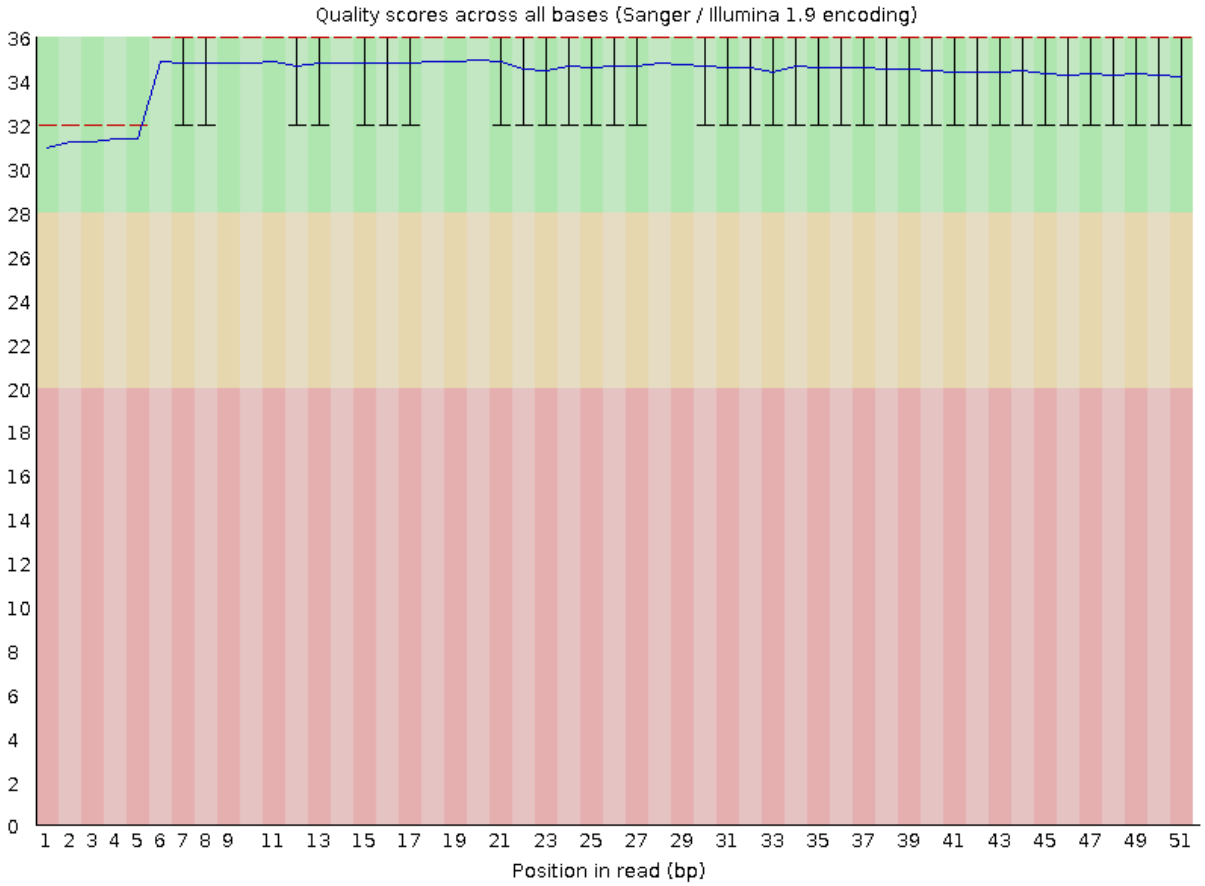

Figure 4. tRF & tiRNA-seq quality score plot. The position in the read is plotted on the X-axis and the Q value is plotted on the Y-axis. The red line is the median Q score, and the blue line is the mean Q score. The boxplot represents the inter-quartile range, while the whiskers represent the 10% and 90% points. A Q score above 30 (>99.9% correct) is considered high quality data

Table 3. Quality score.

Sample	TotalRead	TotalBase	BaseQ30	BaseQ30 (%)
2018L0071TT	9382436	478504236	451656578	94.39
2018L0030TT	5308015	270708765	255074443	94.22
2018L0121TT	8012874	408656574	386701657	94.63
2018L0049TT	6672665	340305915	320705698	94.24
2018L0100TT	6863698	350048598	331090369	94.58
2018L0071TN	5655520	288431520	259523526	89.98
2018L0030TN	7295367	372063717	335747651	90.24
2018L0121TN	6126473	312450123	294215961	94.16
2018L0049TN	5109603	260589753	236703765	90.83
2018L0100TN	9916547	505743897	476723100	94.26

Sample: Sample name
TotalRead: Raw sequencing reads after quality filtering
TotalBase: Number of bases after quality filtering
BaseQ30: Number of bases of Q score more than 30 after quality filtering
BaseQ30 (%): The proportion of bases (Q$\geqslant$30) number after quality filtering

2.3 Mapping

2.3.1 Reads Length Distribution

After Illumina quality control, the sequencing reads were 5', 3'-adaptor trimmed, and discarded reads (length < 14nt or length > 40nt) with cutadapt [9], and were recorded in FASTA format. A bar figure is used to show the sequence read length distribution.

Note: The figures of reads length distribution can be found in Plots/ReadsLength folder.
The table of reads length distribution is provided in Plots/ReadsLength/Table_ReadsLength.xlsx.

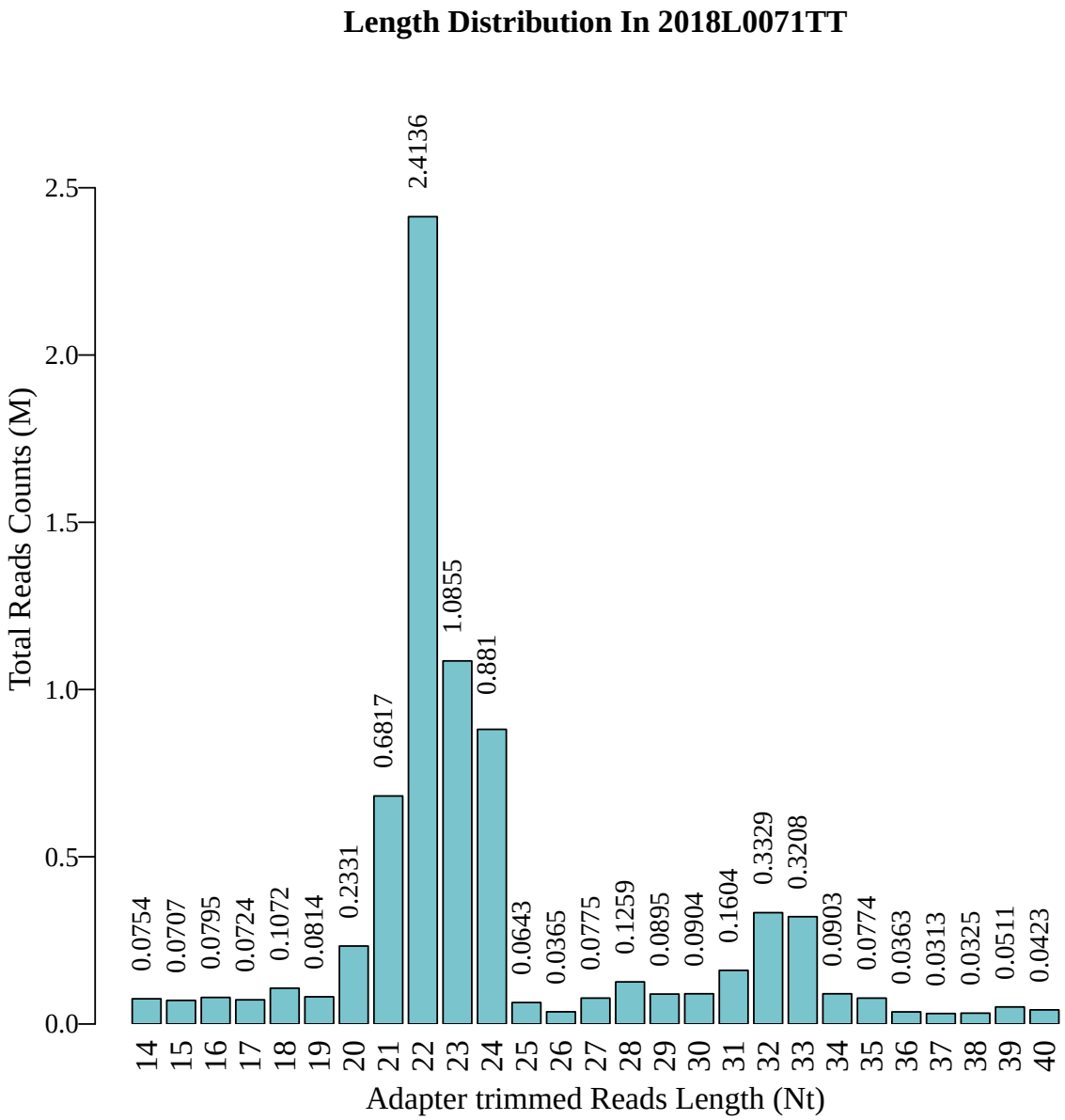

Figure 5. Reads length distribution. Bar Chart showing the total read counts against the lengths of the trimmed reads.

2.3.2 Mapping Summary

Trimmed reads in FASTA format are aligned allowing for 1 mismatch only to the mature tRNA sequences, then reads that do not map are aligned allowing for 1 mismatch only to precursor tRNA sequences with bowtie software [6]. The remaining reads are aligned allowing for 1 mismatch only to miRNA reference sequences with miRDeep2 [7]. The reads statistical information is show in the table below. The aligned percentage depends on multiple factors, including sample quality, library quality and sequencing quality.

Table 4. Mapping summary

Sample	Trimmed	Mat-tRNA	Mat-tRNA(%)	Pre-tRNA	Pre-tRNA(%)	miRNA	miRNA(%)
2018L0071TT	7440881	707812	9.51	17623	0.24	4882568	65.62
2018L0030TT	3814671	633974	16.62	27861	0.73	2378910	62.36
2018L0121TT	5987099	381896	6.38	14170	0.24	4736981	79.12
2018L0049TT	5280286	733958	13.90	14984	0.28	2580910	48.88
2018L0100TT	5716398	2279199	39.87	22285	0.39	2333120	40.81
2018L0071TN	4711940	127817	2.71	16348	0.35	3643438	77.32
2018L0030TN	6133386	1405727	22.92	34910	0.57	3531293	57.57
2018L0121TN	5294990	761729	14.39	44078	0.83	3327222	62.84
2018L0049TN	4197033	184751	4.40	46715	1.11	3304085	78.72
2018L0100TN	8674930	1276372	14.71	55689	0.64	5837224	67.29

Sample:	sample name
Trimmed:	Reads number after 5’ ,3’ -adaptor trimmed and discarded reads (length < 14nt or > 40nt)
Mat-tRNA:	Reads number aligned to mature tRNA
Mat-tRNA(%):	The proportion of reads number aligning to mature tRNA
Pre-tRNA:	Reads number aligned to precursor tRNA
Pre-tRNA(%):	The proportion of reads number aligning to precursor tRNA
miRNA:	Reads number aligned to miRNA
miRNA(%):	The proportion of reads number aligning to miRNA

2.4 Expression profiling of tRF & tiRNA

The abundance of tRF & tiRNA is evaluated using their sequencing counts and is normalized as counts per million of total aligned reads (CPM). For each tRF & tiRNA, estimated the expression level using the mapped reads number and given an ID. Then the tRF & tiRNA are filtered if CPM less than 20 in all samples. The abundance of tRF & tiRNA can be calculated with the formula:

$$Count = \sum_{i=1}^{n} \frac{c_i}{m_i} \tag{2}$$

i: The i-th read aligned to the tRF & tiRNA region.
n: The number of the reads aligned to the tRF & tiRNA region
c_i: The count of the i-th read
m_i: The number of tRF & tiRNA generated from the i-th read (m_i possibly occur great than one, only when allowing for more than 1 mismatch).

The CPM value of tRF & tiRNA can be calculated with the formula:

$$CPM = \frac{10^6 Count}{N} \tag{3}$$

N: the total number of reads mapped onto all of the mature or precursor tRNA.

tRF & tiRNA ID will be divided into four parts by the horizontal line, eg: “tRF-Gly-TCC-001”. The following is a brief description of these fields:

(1) tRF or tiRNA, The type of tRF & tiRNA, eg: tRF;
(2) amino acid, Type of amino acid, eg: Gly;
(3) anticodon, Type of anticodon, eg: TCC;
(4) Serial, The serial number of tRF & tiRNA derived from the tRNA of the same anticodon, eg: 001;

Note: The table of tRF & tiRNA expression profiles can be found in Express folder.
The following table only shows a part of tRF & tiRNA expression profiles Express/Expression_tRF.xlsx.

Table 5. Part of the results of the tRF & tiRNA expression profile data file content.

tRF_ID	tRF_Seq	Type	Length
tRF-Pro-TGG-001	AAAGACTTTTTCTCTGACCA	tRF-3b	20
tRF-Lys-TTT-012	AACACCTCTTTACAGTGACCA	tRF-3b	21
tRF-Val-CAC-015	AACCGGGCAGAAGCACCA	tRF-3a	18
tRF-Val-AAC-001	AACCGGGCGGAAACACCA	tRF-3a	18
tRF-Arg-CCT-002	AAGAAAGGCCGAATTT	tRF-1	16

tRF_ID: The ID of tRF & tiRNA.
tRF_Seq: The sequence of tRF & tiRNA.
Type: The type of tRF & tiRNA.
Length: The length of the tRF & tiRNA sequence.

2.5 The Analysis of Expression Level

2.5.1 The Correlation Coefficient of Samples

The correlation coefficient of samples is an important evaluation criterion of the reliability and reasonability of the sample selection, the correlation coefficient more close to one and the two compared samples are more similar. According to the expression level of each sample, calculated the correlation coefficient between any two of all the samples. The following correlation figure is plot with the CPM values of tRF & tiRNA for all samples.

Note: The correlation plots for all samples can be found in Plots/pca_and_correlation folder. The heatmap for the relations is as below: the table of Pearson correlation of tRF & tiRNA level expression between all sample Plots/pca_and_correlation/correlation.txt

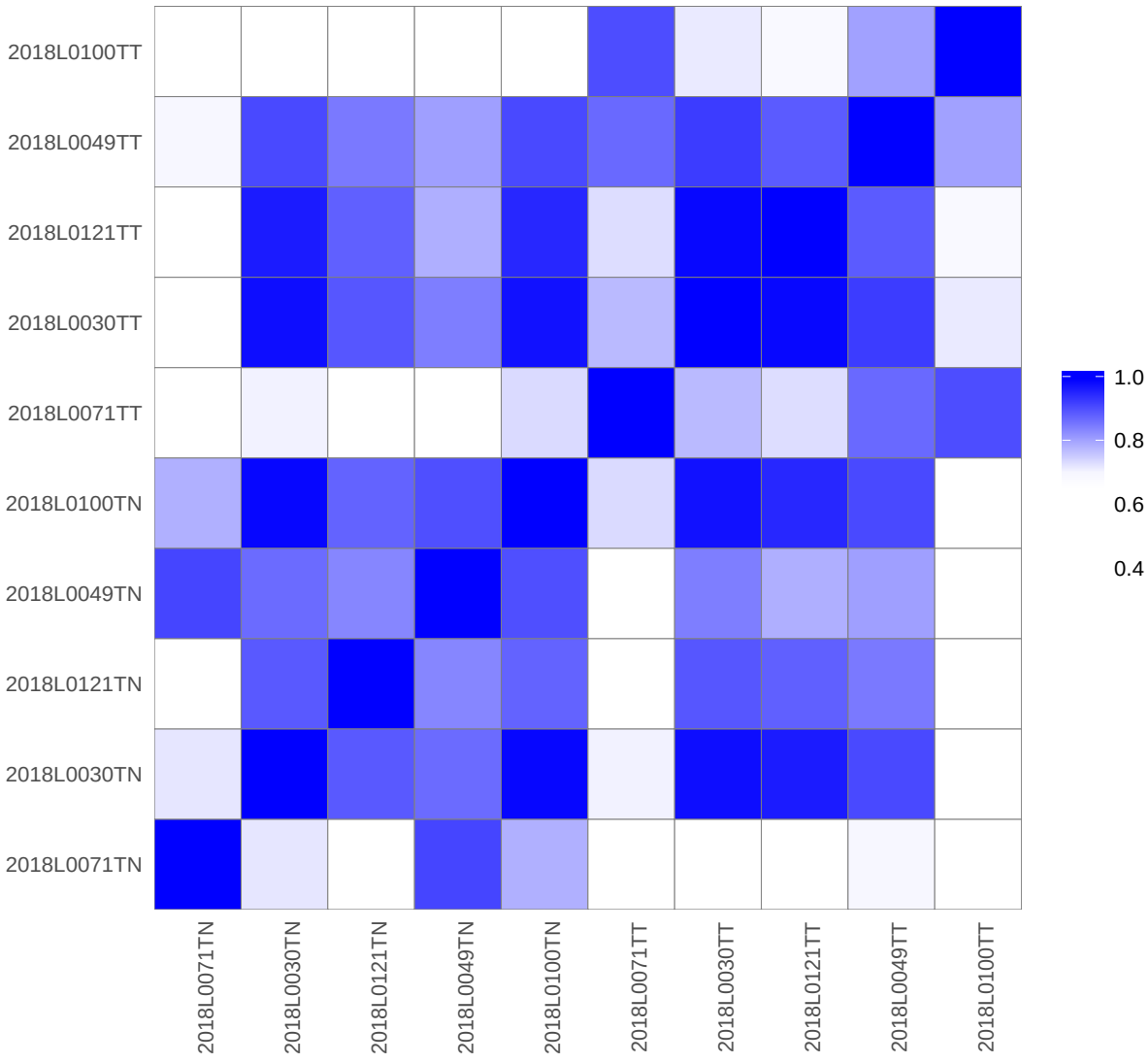

Figure 6. Heatmap of correlation coefficient from all samples. The color in the panel represents the correlation coefficient of the two samples. Blue represents the two samples with a high correlation coefficient, and the white is represents the low similarity of the two samples. This figure was plot with R gplots package.

2.5.2 Principal Component Analysis

Principal Component Analysis (PCA) is a statistic method used in unsupervised analysis to reduce the dimension of large data sets, and it is a useful tool to explore sample classes based on the expression. The PCA was performed with tRF & tiRNA that have the ANOVA *pvalue* $\leqslant 0.05$ on CPM value (Not available for samples with no replicates). The result from PCA shows a distinguishable tRF & tiRNA expression profiling among the samples. The below figure is an overview of samples clustering and plot with R scatterplot3d package.

Note: The figure of PCA can be found in Plots/pca_and_correlation folder.

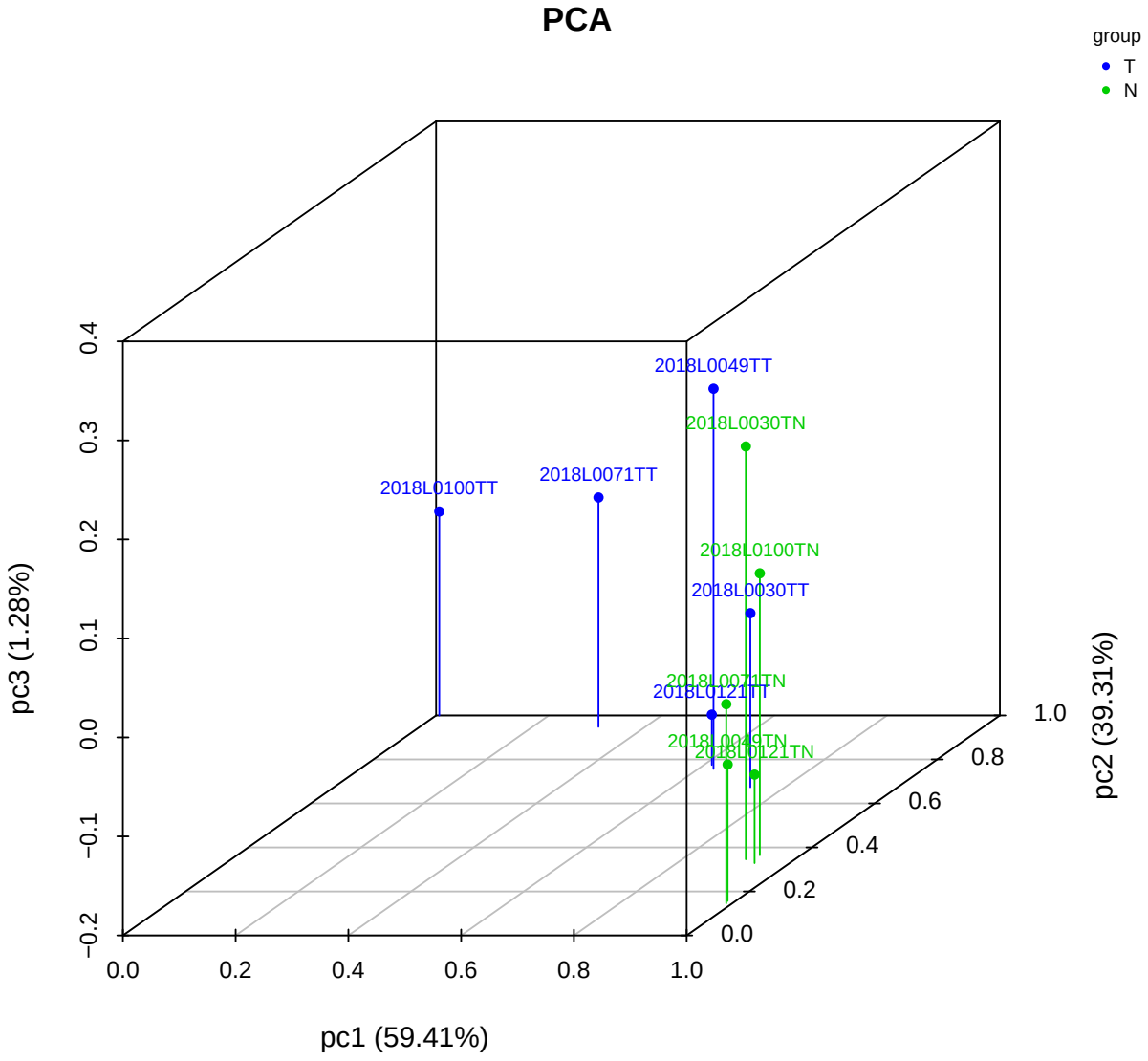

Figure 7. Primary component analyze. the X, Y and Z axis represents the three main factors which affected the expression level of the sample. The colored point represents the corresponding sample, and the location of it shows the main character of the sample. Space distance represents the similarity of data size.

2.5.3 Venn Diagram of Commonly Expressed and Specifically Expressed tRF & tiRNA

Venn diagram shows the commonly expressed and specifically expressed tRFs & tiRNAs. The commonly expressed tRFs & tiRNAs represent the CPM values which were more than 20 in both two groups, and the specifically expressed tRFs & tiRNAs represent the CPM values which were more than 20 in one group while less than 20 in the other group. The venn diagram is plot with R VennDiagram package.

Note: The venn diagram of the commonly expressed and specifically expressed tRFs & tiRNAs can be found in Plots/Venn folder.

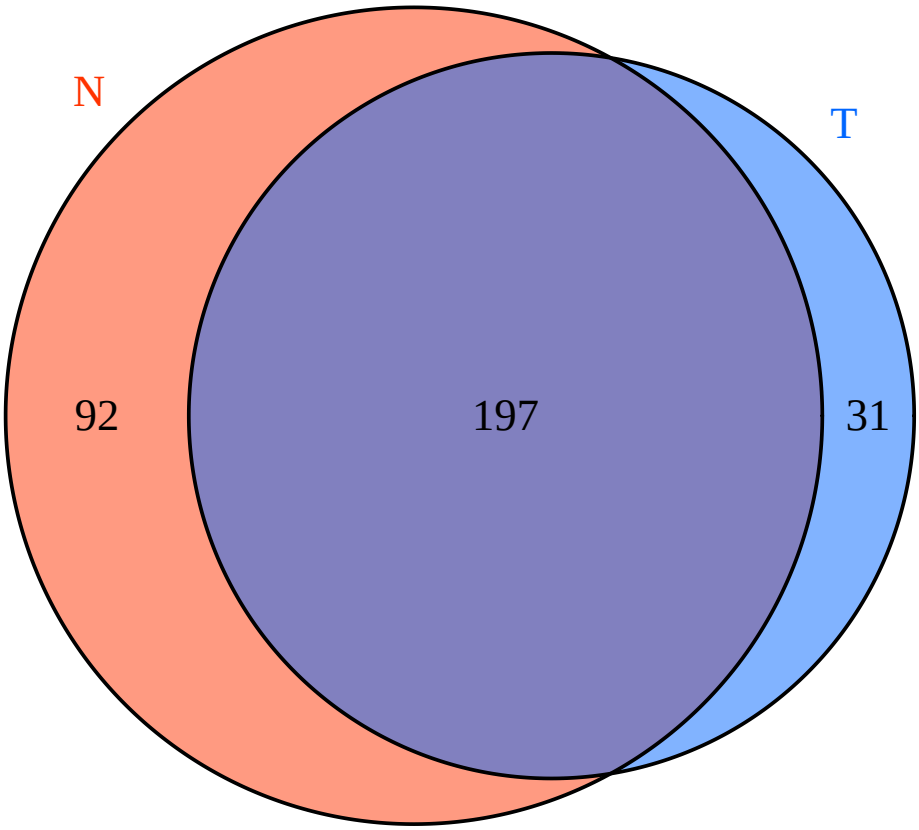

Figure 8. Venn diagram based on number of commonly expressed and specifically expressed tRF & tiRNA. This diagram shows the number of tRF & tiRNA which expressed in both of two groups and also indicated the number of specific expression tRF & tiRNA.

2.5.4 Venn Diagram of Known and Detected tRF & tiRNA

Venn diagram shows the known tRFs from tRFdb [17,18] and the detected tRF & tiRNA in this project. The commonly tRF & tiRNA represent these there are in both two groups, and the specifically tRF & tiRNA represent these is occur in one group while is not occur in the other group. The venn diagram is plot with R VennDiagram package.

Note: The follow venn diagram is provided in Plots/Venn/Venn_tRFs_Known_vs_Detected.pdf.

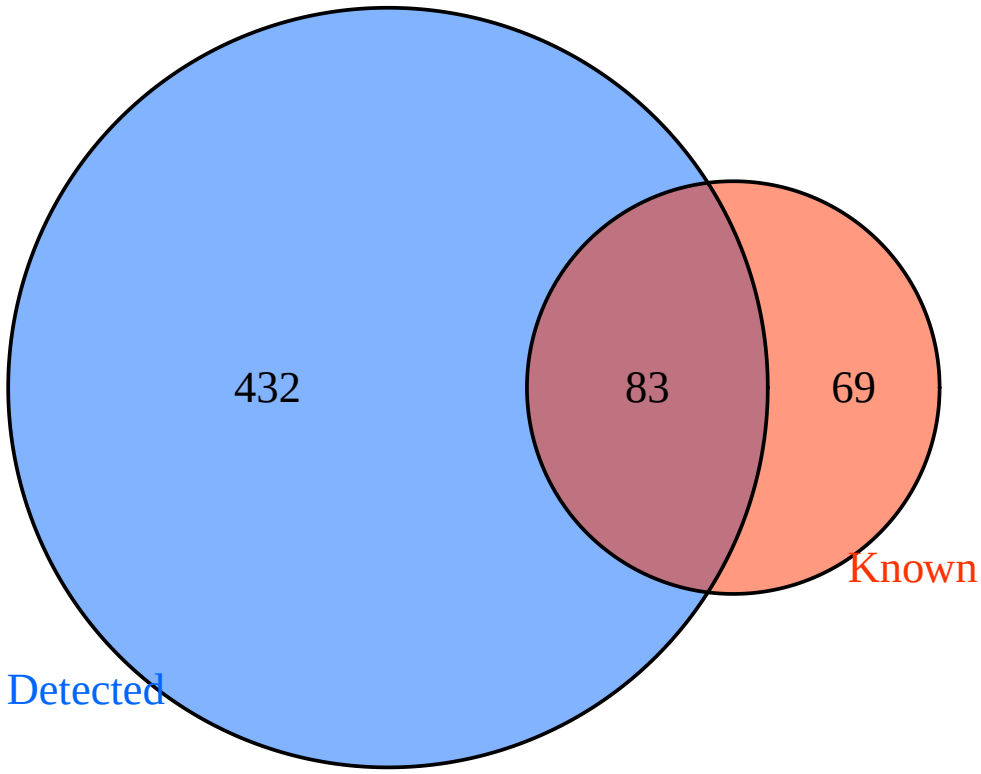

Figure 9. Venn diagram based on number of tRF & tiRNA known and detected. This diagram shows the number of tRF & tiRNA detected in this project and collected in the tRFdb.

2.6 Pie Chart of Each Subtype tRF & tiRNA

A pie chart is a circular statistical graphic which is divided into slices to illustrate numerical proportion. In a pie chart, the arc length of each slice (and consequently its central angle and area), is proportional to the quantity it represents. tRF & tiRNA are separated subtypes by their sites and length. The pie chart show the distribution of the number for each subtype tRF & tiRNA which the CPM of the sample or the average CPM of the group is not less than 20. For each sample and group at least two duplicate samples, the Pie Chart is plot with R pie package.

Note: The pie chart of subtype tRF & tiRNA can be found in Plots/Pie.

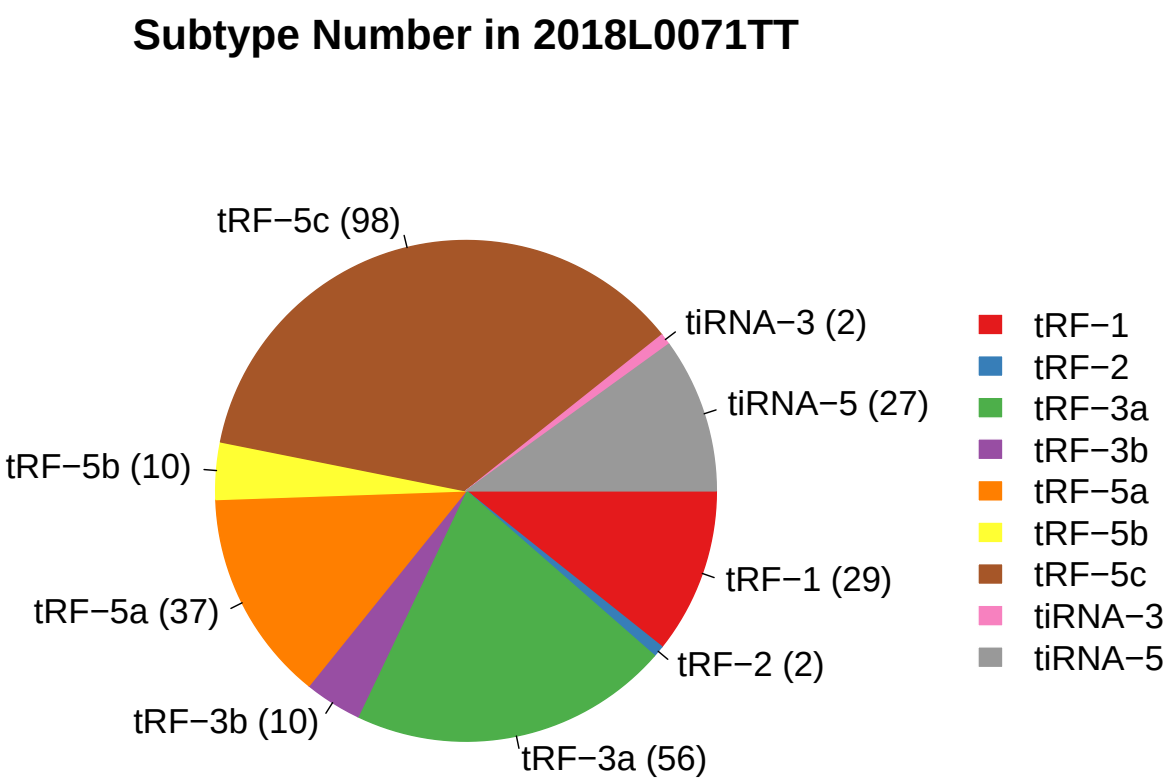

Figure 10. Pie char of the distribution of subtype tRF & tiRNA. The values in bracket are represented the number of subtype tRF & tiRNA. The color represents the subtype tRF & tiRNA.

2.7 Stacked Bar Chart

The stacked bar chart stacks bars that represent different groups on top of each other. The height of the resulting bar shows the combined result of the groups.

2.7.1 The Number of Subtype tRF & tiRNA against tRNA Isodecoders

tRNA isodecoders share the same anticodon but have differences in their body sequence. The number of subtype tRF & tiRNA, which the CPM of the sample or the average CPM of the group is not less than 20, can be counted against tRNA isodecoders. The stacked bar chart stacks bars that represent different tRNA isodecoders on top of each other. The height of the resulting bar shows the combined result of tRNA isodecoders. For each sample and group at least two duplicate samples, the Stacked Bar Chart is plot with R barplot package.

Note: The stacked bar chart of the number of subtype tRF & tiRNA against tRNA isodecoders can be found in Plots/SubtypeIsodecoder folder. The tables corresponding the following stacked bar chrat is provided in Plots/SubtypeIsodecoder/SubtypeIsodecoder.xlsx.

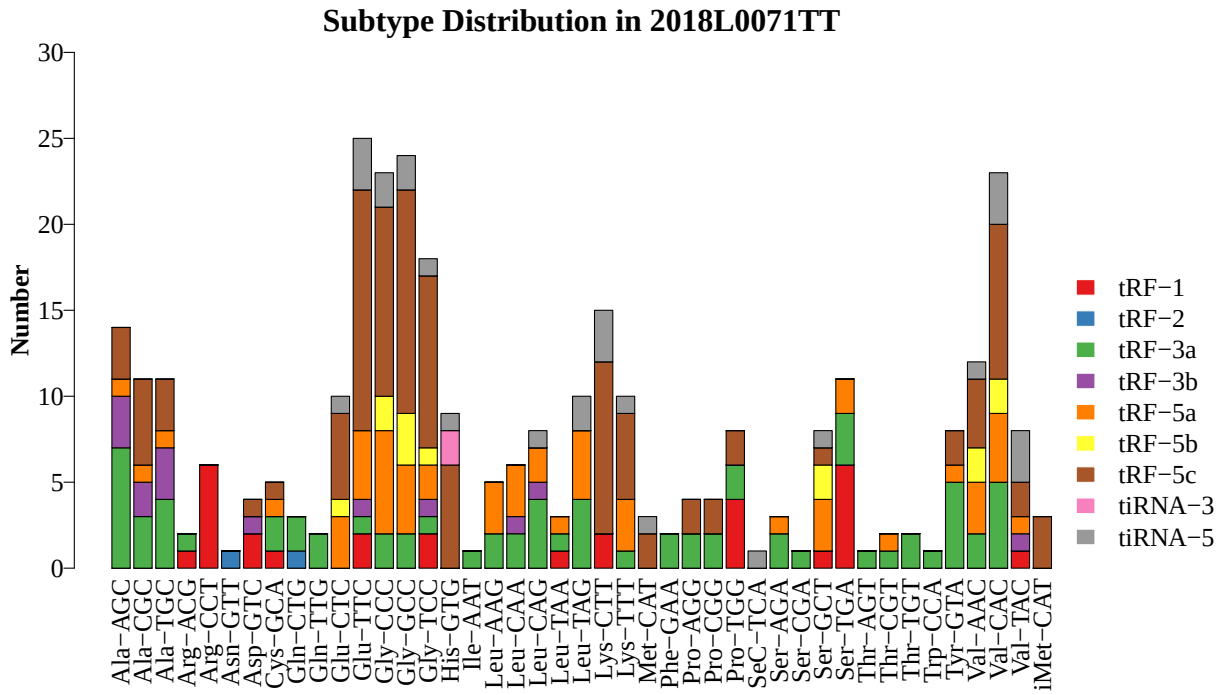

Figure 11. The number of subtype tRF & tiRNA against tRNA isodecoders. The X axes represents tRNA isodecoders and the Y axes show the number of all subtype tRF & tiRNA against tRNA isodecoders. The color represents the subtype tRF & tiRNA.

2.7.2 The Frequency of Subtype against Length of the tRF & tiRNA

The frequency of subtype tRF & tiRNA, which the CPM of the sample or the average CPM of the group is not less than 20, can be calculated against the length of the sequence. The stacked bar chart stacks bars that represent different length of tRF & tiRNA on top of each other. The height of the resulting bar shows the combined result of length of the tRF & tiRNA. For each sample and group at least two duplicate samples, the Stacked Bar Chart is plot with R barplot package.

Note: The stacked bar chart of the frequency of subtype against length of the tRF & tiRNA can be found in Plots/SubtypeLength folder. The tables corresponding the following stacked bar chrat is provided in Plots/SubtypeLength/SubtypeLength.xlsx.

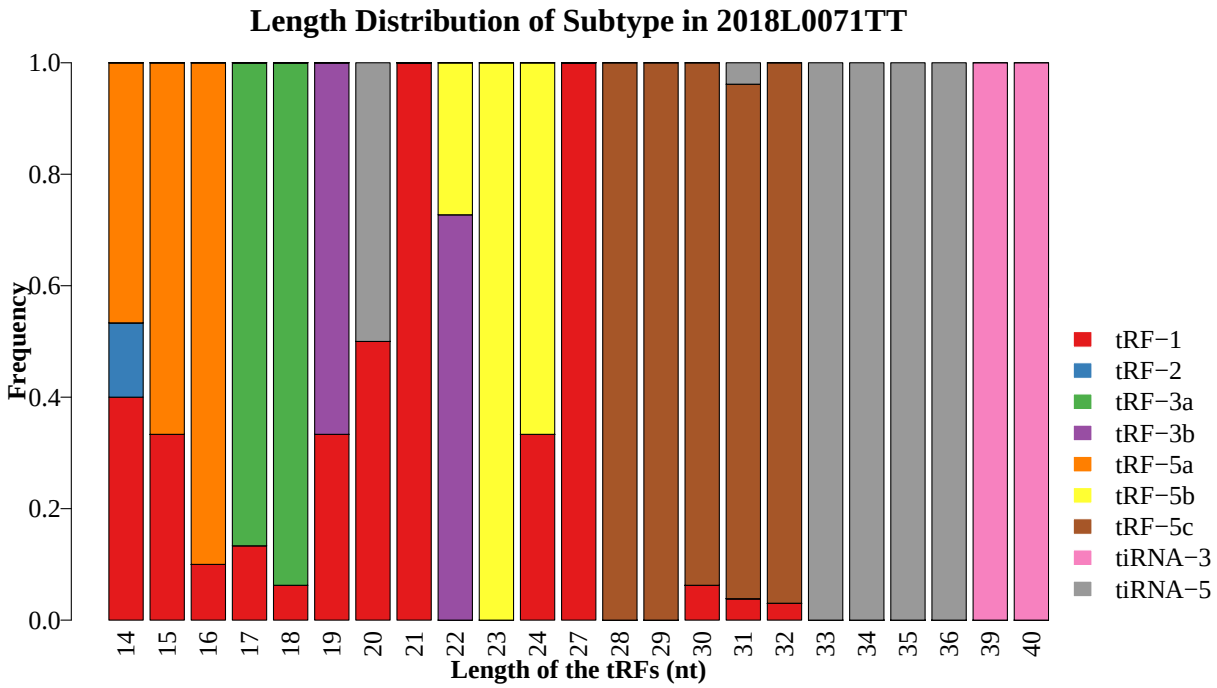

Figure 12. The Frequency of Subtype against Length of the tRF & tiRNA. The X axes represents length of tRF & tiRNA and the Y axes show the frequency of the subtype against length of tRF & tiRNA. The color represents the subtype tRF & tiRNA.

2.8 Differential Expression Analysis of tRF & tiRNA

Differentially expressed tRFs & tiRNAs analyses was performed with R package edgeR [10]. Fold change (cutoff 1.5), P value (cutoff 0.05 performed only when have replicate) were used for screening differentially expressed tRFs & tiRNAs.

Note: The table of the differentially expressed tRFs & tiRNAs can be found in Diff_Express folder. The following table only shows a part of significantly differentially expressed tRFs & tiRNAs Diff_Express/Differentially_Expressed_tRF.xlsx.

Table 6. Part of data file content of the significantly differentially expressed tRFs & tiRNAs.

tRF_ID	log_2FC	FC	p_value	q_value	Test_CPM	Control_CPM
tRF-Ser-TGA-003	6.14	70.47	9.70e-05	4.71e-03	4.78	-1.35
tRF-Val-CAC-005	5.64	49.89	1.95e-06	5.03e-04	6.51	0.87
tRF-Ser-TGA-024	5.52	45.76	8.10e-03	5.79e-02	2.93	-2.59
tRF-Arg-CCT-002	4.74	26.81	1.19e-03	1.95e-02	3.40	-1.34
tRF-Val-CAC-003	4.65	25.14	1.51e-05	1.94e-03	7.58	2.93

tRF_ID:	The ID of tRF & tiRNA
log_2FC:	The difference in mean between two groups (T_CPM - N_CPM)
FC:	Fold Change $2^{log_2 FC}$
p_value:	The p-value of the exact test by negative binomial distribution
q_value:	The FDR adjusted p-value. The q-value will be set as 1 if any group in the comparison has no repilicate
Test_CPM:	The average of log scaled CPM of tRF &tiRNA in T group. The values calculated by Negative Binomial Generalized Linear Model.
Control_CPM:	The average of log scaled CPM of tRF & tiRNA in N group. The values calculated by Negative Binomial Generalized Linear Model.

2.9 The Scatter Plots of Differentially Expressed tRF & tiRNA

The scatter plot is a visualization method used for assessing the tRF & tiRNA expression variation (or reproducibility) between the two compared (groups of) samples. The following scatter plot was generated form log_2 scaled CPM values of tRF & tiRNA.

Note: The scatter plots of tRF & tiRNA can be found in Plots/Scatter_and_Volcano folder.

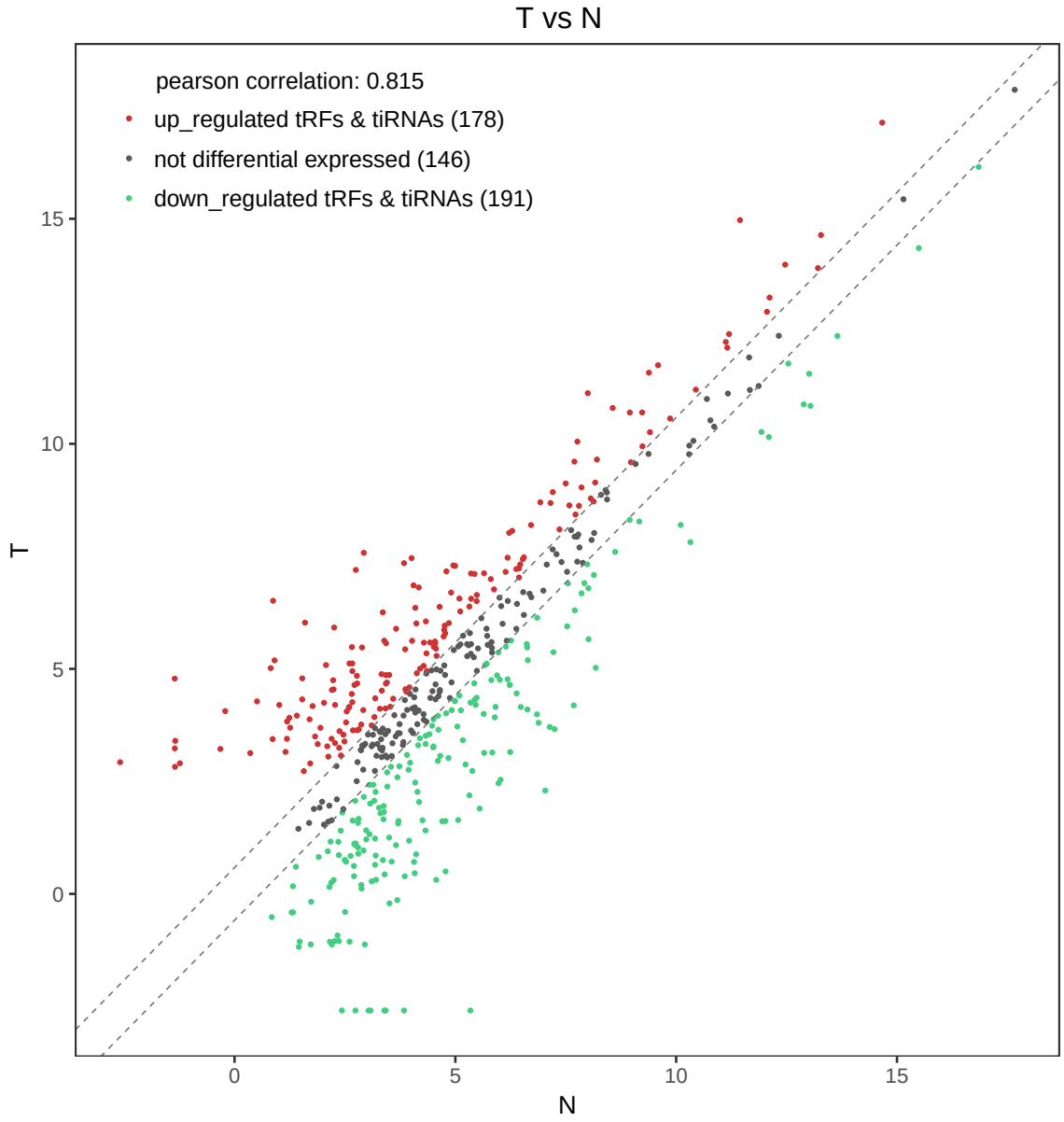

Figure 13. The scatter plot between two groups for tRF & tiRNA. The CPM values of all tRFs & tiRNAs are plot. The values of X and Y axes in the scatter plot are the averaged CPM values of each group (log_2 scaled). tRFs & tiRNAs above the top line (red dots, up-regulation) or below the bottom line (green dots, down-regulation) indicate more than 1.5 fold change between the two compared groups. Gray dots indicate non-differentially expressed tRFs & tiRNAs.

2.10 The Volcano Plots of Differentially Expressed tRF & tiRNA

(Note: this part will be provided only when the two group have replicate)

The volcano plot provides a visualization method to perform a quick visual identification of the tRF & tiRNA displaying large magnitude changes which are also statistically significant. The plot is constructed by plotting -$log_{10}(q\_value)$ on the y-axis, and tRF & tiRNA expression $log_2(Fold\ Change)$ between the two experimental groups on the x-axis. The volcano plot is plot based on tRF & tiRNA with CPM values.

Note: The volcano plots of tRF & tiRNA can be found in Plots/Scatter_and_Volcano folder.

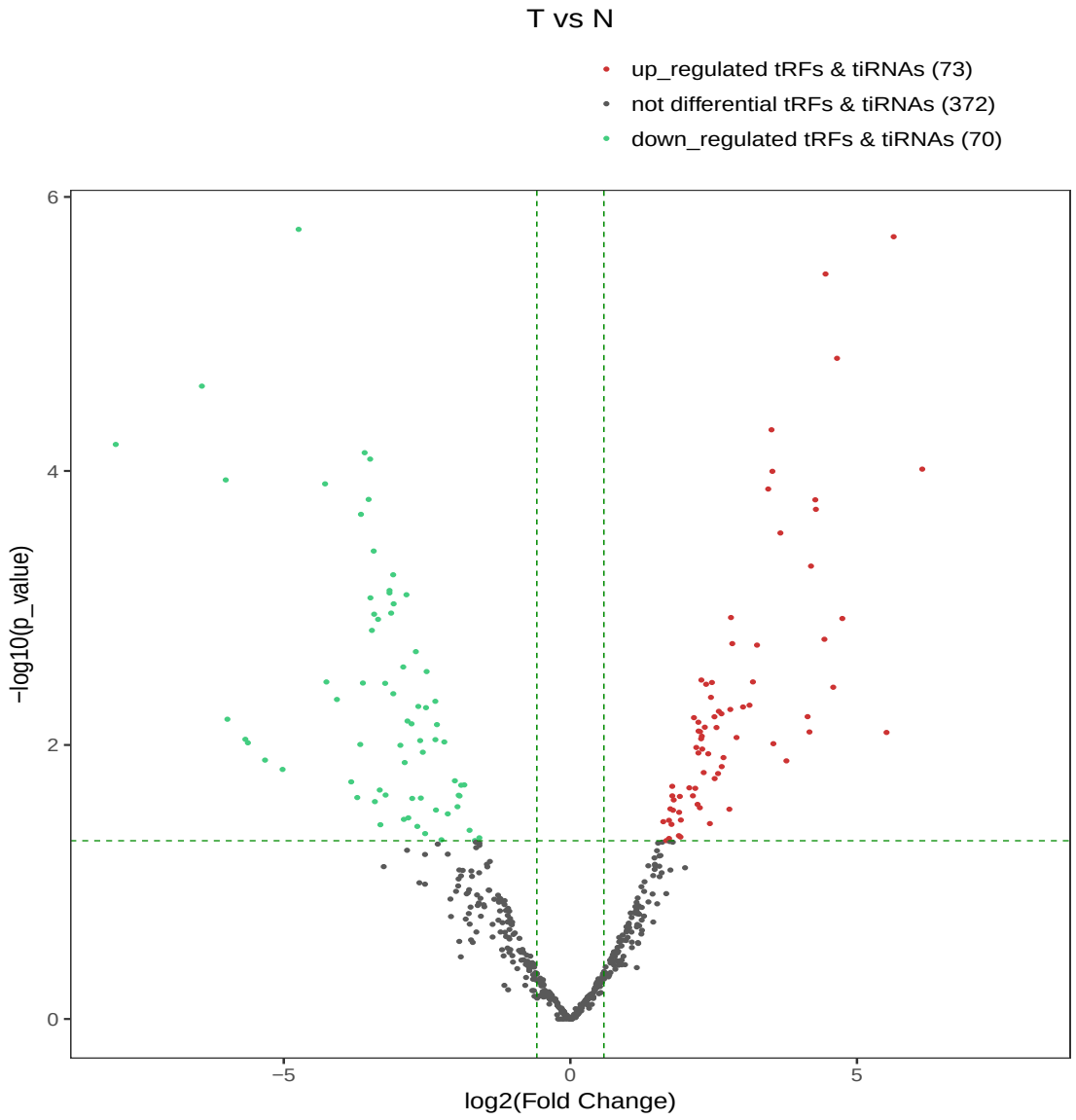

Figure 14. The volcano plot of tRF & tiRNA. The values of X and Y axes in the volcano plot are log_2 transformed fold change and -log_{10} transformed p-values between the two groups, respectively. Red/Green circles indicate statistically significant differentially expressed tRFs & tiRNAs with fold change no less than 1.5 and p-value $\leqslant$ 0.05 (Red: up-regulated; Green: down-regulated). Gray circles indicate non-differentially expressed tRFs & tiRNAs, with FC and/or q-value are not meeting the cutoff thresholds.

2.11 Hierarchical Clustering of Differentially Expressed tRF & tiRNA

Hierarchical clustering is one of the most widely used clustering methods for analyzing tRF & tiRNA expression data. Cluster analysis arranges samples into groups based on their tRF & tiRNA expression level (CPM values), which allows us to hypothesize the relationships among samples. The dendrogram shows the relationships among tRF & tiRNA expression patterns of samples.

Hierarchical clustering is performed using the differentially expressed tRFs & tiRNAs (if number is more than two). Each row represents one tRF & tiRNA and all selected are categorized into no more than 10 clusters based on K-means clustering, each column represents one sample. The result from hierarchical clustering shows a distinguishable tRF & tiRNA expression profiling among samples. Hierarchical clustering is performed based on differentially expressed tRF & tiRNA with R heatmap2 package.

Note: All the heatmap can be found in Plots/Heatmap folder. The group information of the tRF & tiRNA in each comparison and can also be found in Plots/Heatmap folder and format *_kmeans_grouping.txt.

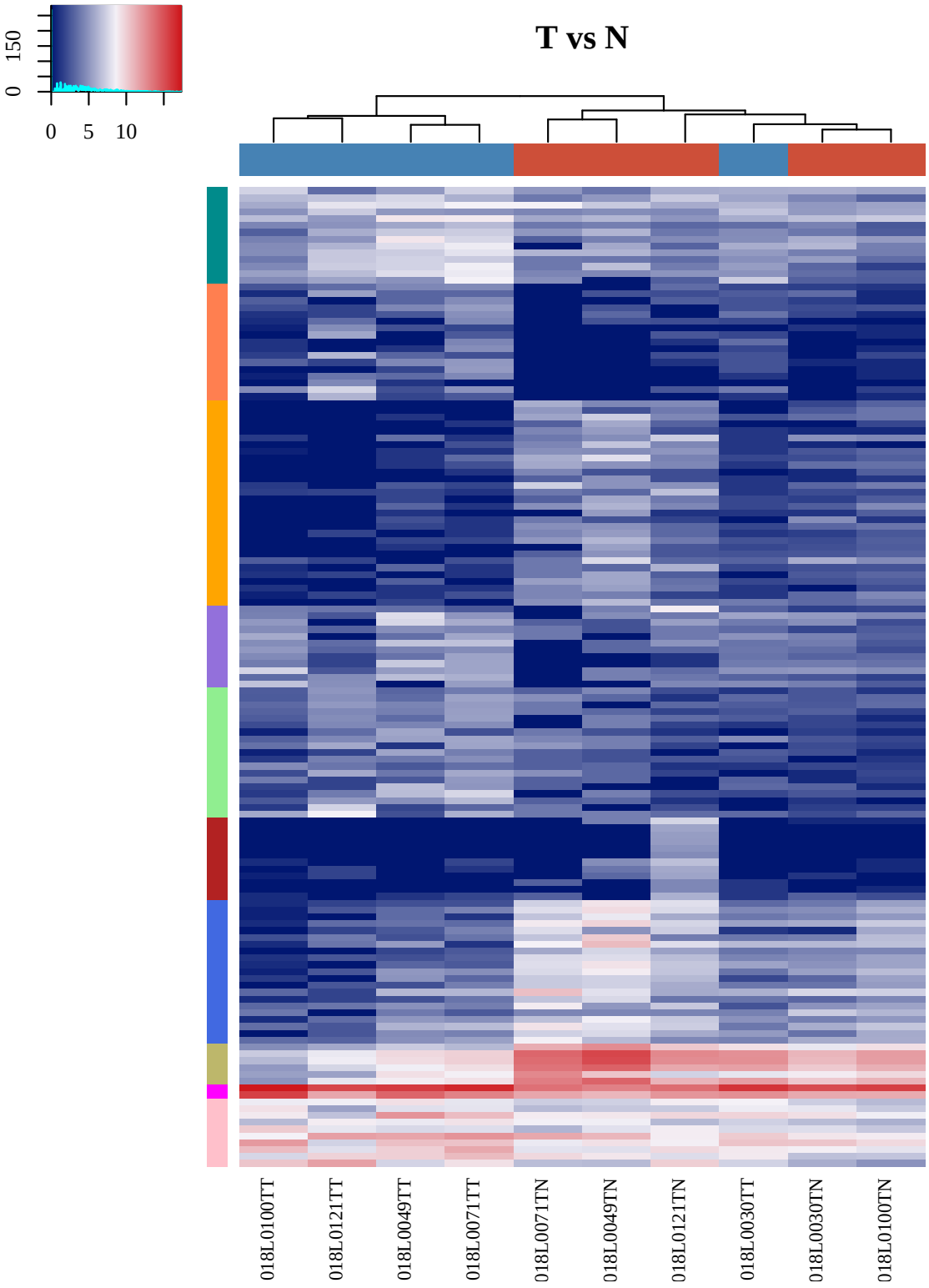

Figure 15. The unsupervised hierarchical clustering heatmap for tRF & tiRNA. The color in the panel represents the relative expression level (log_2-transformed). The color scale is show below: blue represents an expression level below the mean, and red represents an expression level above the mean. The colored bar top at the top panel showed the samples group, and the colored bar at the right side of the panel indicated the divisions which were performed using K-means.

2.12 miRNA Analysis

The reads that do not map to the mature or precursor tRNA sequences are aligned allowing for 1 mismatch only to miRNA reference sequences with miRDeep2 [7]. The expression profiling of miRNA can be calculated based on counts of reads mapped. The differentially expressed and miRNAs are screened based on the count value with R package edgeR [10]. Hierarchical clustering, Scatter plots and Volcano plots are performed in R or perl environment for statistical computing and graphics of the expressed miRNA.

2.12.1 Expression profiling of miRNA

The abundance of miRNA is evaluated using their sequencing counts and is normalized as counts per million of total aligned reads (CPM).

Note: The table of miRNA expression profiles can be found in miRNA folder.
The following table only shows a part of miRNA expression profiles miRNA/Expression_miRNA.xlsx.

Table 7. Part of the results of the miRNA expression profile data file content.

Mature_ID	Mature_Acc	Mature_Seq	Seed	High_Confidence
hsa-miR-548ac	MIMAT0018938	CAAAAACCGGCAAUUACUUUUG	AAAAACC	NO
hsa-miR-548bb-3p	MIMAT0035704	CAAAAACCAUAGUUACUUUUGC	AAAAACC	NO
hsa-miR-548d-3p	MIMAT0003323	CAAAAACCACAGUUUCUUUUGC	AAAAACC	Yes
hsa-miR-548h-3p	MIMAT0022723	CAAAAACCGCAAUUACUUUUGCA	AAAAACC	Yes
hsa-miR-548z	MIMAT0018446	CAAAAACCGCAAUUACUUUUGCA	AAAAACC	NO

Mature_ID:	The ID of mature miRNA annotated in miRBase22.
Mature_Acc:	The accession number of mature miRNA from miRBase.
Mature_Seq:	The sequence of mature miRNA.
Seed	Positions 2-8 of the mature miRNA.
High_Confidence	The high confidence microRNAs are robustly supported as authentic miRNA genes.

2.12.2 Differential expression analysis of miRNA

Differentially expressed miRNA analyses was performed with R package edgeR [10]. Fold change (cutoff 1.5), P value (cutoff 0.05 performed only when have replicate) were used for screening differentially expressed miRNA. The following table only shows a part of significantly differentially expressed miRNA (miRNA/Differentially_Expressed_miRNA.xlsx).

Note: The table of the differentially expressed miRNA can be found in miRNA folder.

Table 8. Part of data file content of the significantly differentially expressed miRNAs

Mature_ID	log_2FC	FC	p_value	q_value	Test_CPM	Control_CPM
hsa-miR-1269a	9.60	775.40	2.06e-04	9.72e-03	7.49	-2.10
hsa-miR-196a-3p	7.63	197.64	1.24e-02	1.44e-01	4.29	-3.34
hsa-miR-1269b	6.54	93.22	2.56e-05	2.47e-03	1.65	-4.90
hsa-miR-2682-5p	6.41	85.22	6.49e-03	9.58e-02	1.52	-4.90
hsa-miR-4661-5p	5.70	51.95	6.87e-05	4.41e-03	0.80	-4.90

Mature_ID:	The ID of mature miRNA annotated in miRBase22.
log_2FC:	The difference in mean between two groups (T_CPM - N_CPM)
FC:	Fold Change $2^{log_2 FC}$
p_value:	The p-value of the exact test by negative binomial distribution
q_value:	The FDR adjusted p-value. The q-value will be set as 1 if any group in the comparison has no repilicate
Test_CPM:	The average of log scaled CPM of miRNA in T group. The values calculated by Negative Binomial Generalized Linear Model.
Control_CPM:	The average of log scaled CPM of miRNA in N group. The values calculated by Negative Binomial Generalized Linear Model.

2.12.3 The Scatter Plots of Differentially Expressed miRNA

The scatter plot is a visualization method used for assessing the miRNA expression variation (or reproducibility) between the two compared (groups of) samples, and is generated form log_2 scaled CPM values of miRNA.

Note: The scatter plots of miRNA can be found in miRNA/Plots folder and format scatter_*.png / pdf.

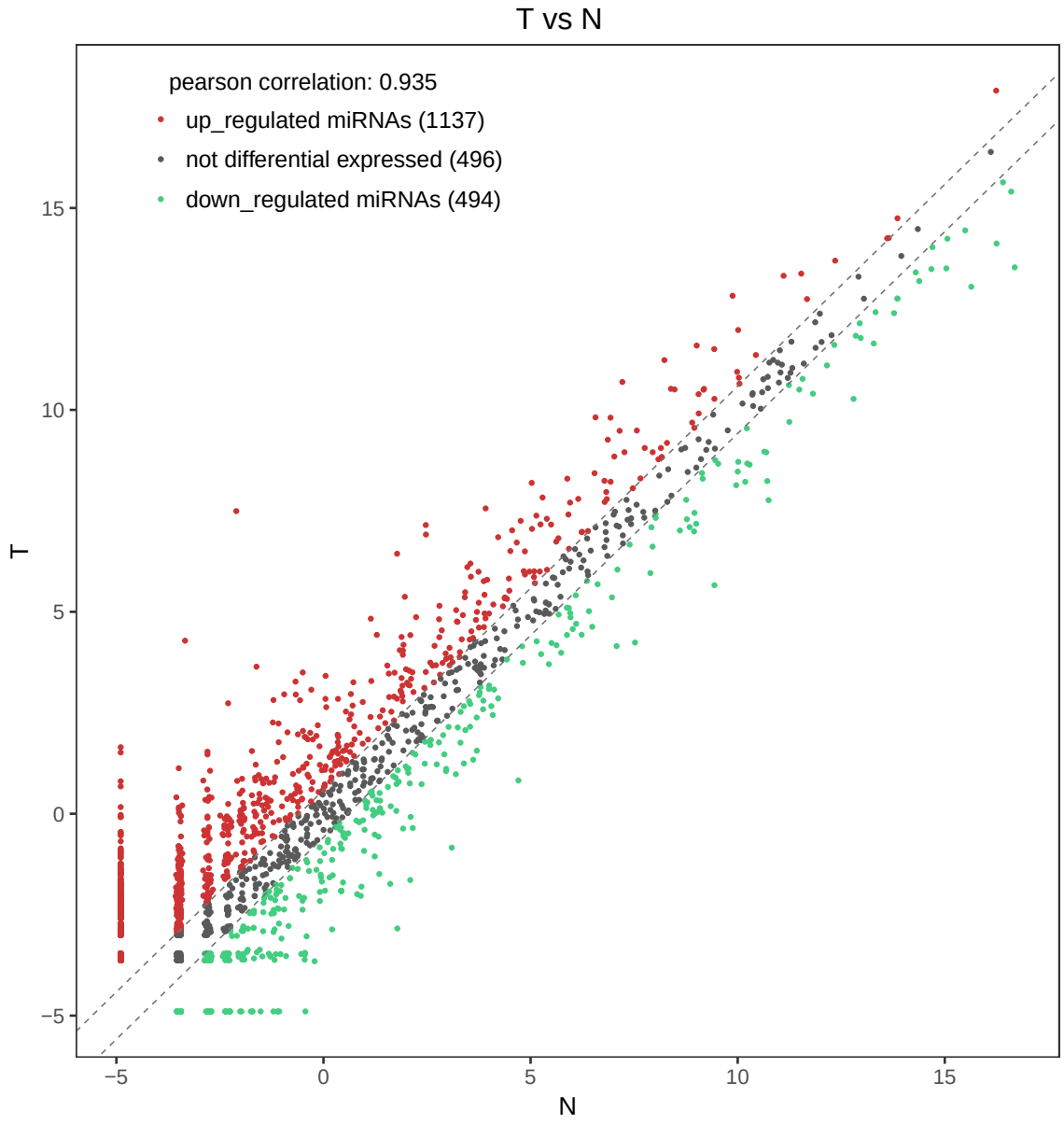

Figure 16. The scatter plot between two groups for miRNA. The CPM values of all miRNAs are plot. The values of X and Y axes in the scatter plot are the averaged CPM values of each group (log_2 scaled). miRNAs above the top line (red dots, up-regulation) or below the bottom line (green dots, down-regulation) indicate more than 1.5 fold change between the two compared groups. Gray dots indicate non-differentially expressed miRNAs.

2.12.4 Volcano Plots of Differentially Expressed miRNA

(Note: this part will be provided only when the two group have replicate)

The volcano plot is constructed by plotting -$log_{10}(q - value)$ on the y-axis, and miRNA expression log_2 fold change between the two experimental groups on the x-axis. The volcano plot is plot based on miRNA with CPM values.

Note: The volcano plots of miRNA can be found in miRNA/Plots folder and format volcano_*.png / pdf.

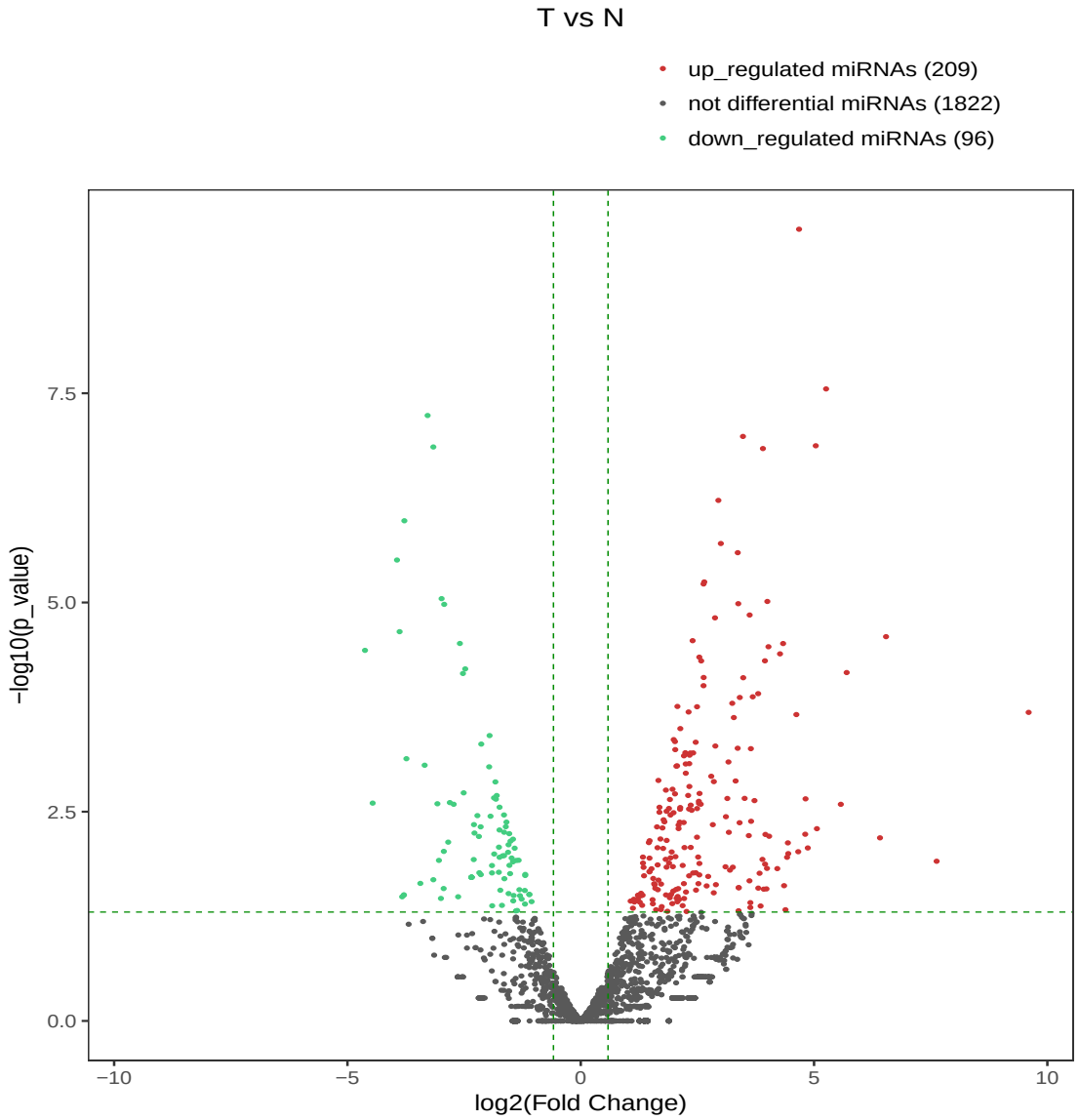

Figure 17. The volcano plot of miRNA. The values of X and Y axes in the volcano Plot are log_2 transformed fold change and -log_{10} transformed p-values between the two groups, respectively. Red/Green circles indicate statistically significant differentially expressed miRNAs with fold change no less than 1.5 and p-value $\leqslant 0.05$ (Red: up-regulated; Green: down-regulated). Gray circles indicate non-differentially expressed miRNAs, with FC and/or q-value are not meeting the cutoff thresholds.

2.12.5 Hierarchical Clustering of Differentially Expressed miRNA

Hierarchical clustering is performed using the differentially expressed miRNAs (if number is more than two). Each row represents one miRNA and all selected are categorized into no more than 10 clusters based on K-means clustering, each column represents one sample. The result from hierarchical clustering shows a distinguishable miRNA expression profiling among samples.

Note: All the heatmap can be found in miRNA/Plots folder and format heatmap_*.png / pdf. The group information of the miRNA in each comparison can also be found in miRNA/Plots folder and format *_kmeans_grouping.txt.

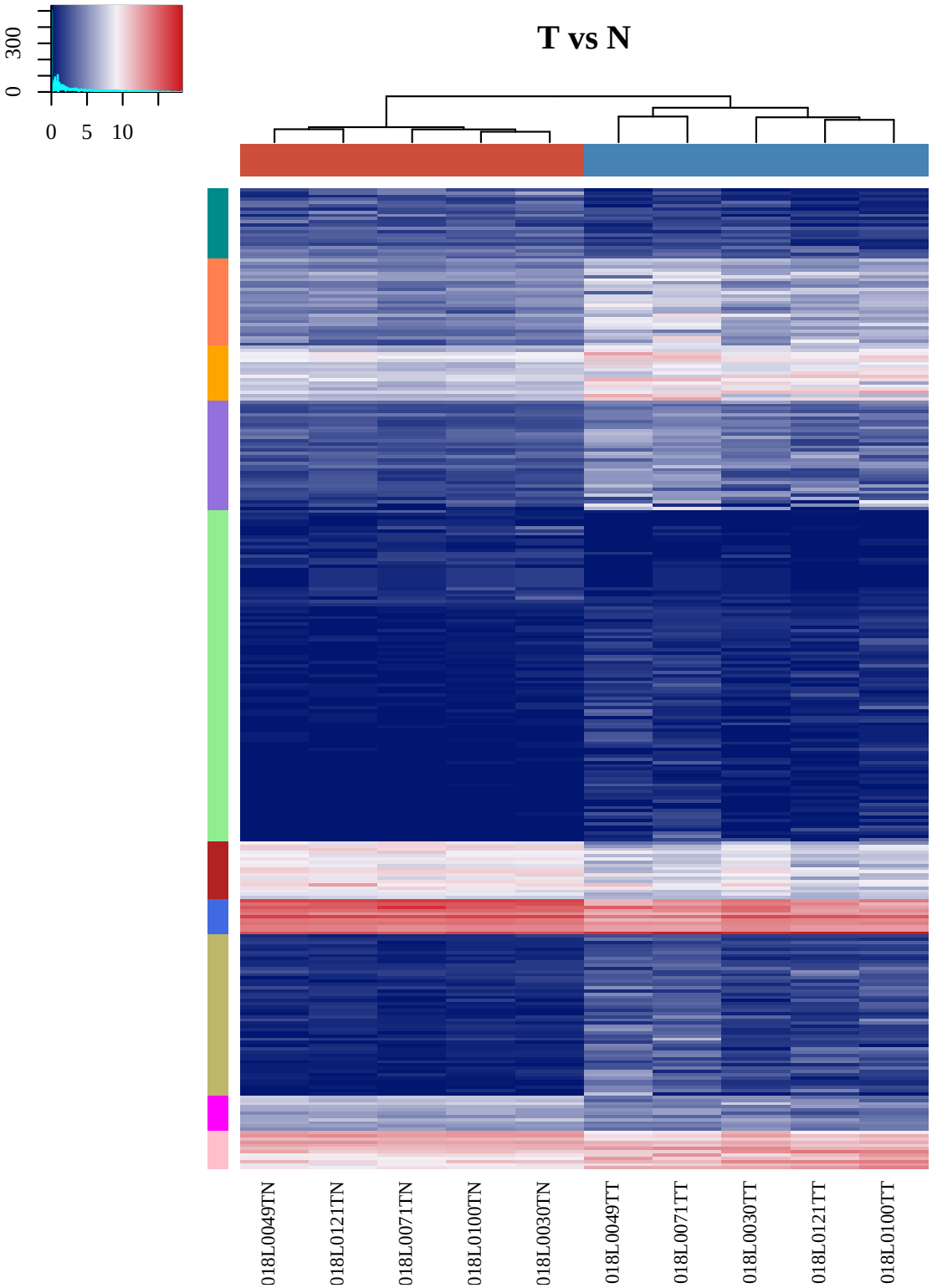

Figure 18. The unsupervised hierarchical clustering heatmap for miRNA. The color in the panel represents the relative expression level (log_2-transformed). The color scale is show below: blue represents an expression level below the mean, and red represents an expression level above the mean. The colored bar top at the top panel showed the samples group, and the colored bar at the right side of the panel indicated the divisions which were performed using K-means.

3 Methods

Library preparation

Reagents: NEBNext® Multiplex Small RNA Library Prep Set for Illumina®
Equipment: NanoDrop ND-1000
Agilent 2100 Bioanalyzer

Procedures:
Agarose electrophoresis was used to check the integrality of total RNA samples, and then the samples were quantified on the NanoDrop ND-1000 instrument. Total RNA samples were first pretreated as following to remove some RNA modifications that interfere with small RNA-seq library construction: 3' -aminoacyl (charged) deacylation to 3' -OH for 3' adaptor ligation, 3' -cP (2' ,3' -cyclic phosphate) removal to 3' -OH for 3' adaptor ligation, 5' -OH (hydroxyl group) phosphorylation to 5' -P for 5' -adaptor ligation, m1A and m3C demethylation for efficient reverse transcription. Pretreated total RNA of each sample was taken for tRF & tiRNA-seq library preparation. Library preparation procedures included: 1) 3' -adapter ligation; 2) 5' -adapter ligation; 3) cDNA synthesis; 4) PCR amplification; 5) size selection of ∼134-160bp PCR amplified fragments (corresponding to ∼14-40nt small RNAs). The completed libraries were quantified by Agilent 2100 Bioanalyzer. Libraries were mixed in equal amounts according to the quantification results, and used for sequencing on the instrument.

Sequencing

Reagent: NextSeq 500/550 V2 kit (#FC-404-2005, Illumina)
Equipment: Illumina NextSeq 500

Procedures:
The DNA fragments in well mixed libraries were denatured with 0.1M NaOH to generate single-stranded DNA molecules, and loaded onto the reagent cartridge at 1.8pM concentration. The sequencing run was performed on NextSeq system using NextSeq 500/550 V2 kit (#FC-404-2005, Illumina) according to the manufacturer' s instructions. Sequencing was carried out by running 50 cycles.

Data Analysis

Image analysis and base calling are performed using Solexa pipeline v1.8 (Off-Line Base Caller software, v1.8). Sequencing quality are examined by FastQC [8] and trimmed reads (pass Illumina quality filter, trimmed 5' , 3' -adaptor bases by cutadapt [9]) are aligned allowing for 1 mismatch only to the mature tRNA sequences, then reads that do not map are aligned allowing for 1 mismatch only to precursor tRNA sequences with bowtie software [6]. The remaining reads are aligned allowing for 1 mismatch only to miRNA reference sequences with miRDeep2 [7]. The expression profiling of tRF & tiRNA and miRNA can be calculated based on counts of reads mapped. The differentially expressed tRFs & tiRNAs and miRNAs are screened based on the count value with R package edgeR [10]. Principal Component Analysis (PCA), Correlation Analysis, Pie plots, Venn plots, Hierarchical clustering, Scatter plots and Volcano plots are performed in R or perl environment for statistical computing and graphics of the expressed tRF & tiRNA.

4 Appendix

4.1 Quality Control

4.1.1 Sample Quality Control

RNA quality control
Agarose gel electrophoresis was employed to check the integrality of total RNA samples.
The NanoDrop ND-1000 instrument was used for accurate measurement concentration (abs 260) and protein contamination (ratio abs260/abs280) of total RNA samples.

Quality assessment of sequencing library
The Agilent 2100 Bioanalyzer was used for accurate assessment of the quality and concentration of sequencing library.

Note: All the detailed results are provided in Sample_QC_Report.pdf.

4.1.2 Sequencing Quality Control

The raw data files of FASTQ format are generated from the Illumina sequencer. To examine the sequencing quality, the quality score plot of each sample was plot. Quality score Q is logarithmically related to the base-calling error probability (P):

$$Q = -10log_{10}(P) \tag{4}$$

For example, Q30 means the incorrect base calling probability to be 0.001 or 99.9% base calling accuracy.

Note: All the sequencing quality control plots can be found in Sequence_QC folder.

4.2 Raw Sequence Data

Raw sequence data were generated as clean reads from Illumina NextSeq by real-time base calling and quality filtering. The clean reads were recorded in FASTQ format, containing the read information, sequences and quality encoding.

The clean reads are located in:
Format: FASTQ
Directory: Sequence/Raw_Data
Files: *.fastq.gz (compressed)

4.3 Trimmed Sequence Data

Subsequently, the 5’ , 3’ -adapter sequence are trimmed from the clean reads by cutadapt and the reads with lengths shorter than 14nt or longer than 40 were discarded. The trimmed reads are collapsed into FASTA format. The identifier line, which begin with ‘>’ , gives a name that are divided into three parts by underline. The first part of the character indicate sample name, the second part of the integer indicate rank the i-th out of all sequences, the third part is made up of character and integer. For example, the identifier line ‘>samplei_1_x123’ , the character ‘samplei’ indicate sample name, the integer ‘1’ indicate rank the first out of all sequences, the character ‘x’ as an identifier and the subsequent integer ‘123’ indicate the sequence have occur 123 time.

The trimmed reads are located in:
Format: FASTA
Directory: Sequence/Trimmed_Data
Files: *.fa.gz (compressed)

4.4 Submitting tRF & tiRNA Sequence Data to GEO

tRF & tiRNA sequencing profiling data may be required by the journal publication policy to be uploaded into Gene Expression Omnibus (GEO) before publishing the manuscript. Please follow the instructions in 'small_RNA-seq_GEO_submission.pdf' for detail.

4.5 Aligned Results

The trimmed reads (14nt ≤ length ≤ 40nt) are aligned allowing for 1 mismatch only to the mature tRNA sequences, then reads that do not map are aligned allowing for 1 mismatch only to precursor tRNA sequences with bowtie software [6]. The remaining reads are aligned allowing for 1 mismatch only to miRNA reference sequences with miRDeep2 [7].

4.5.1 Bowtie Output

The alignment result can be save to a file in TXT format for each sample. Bowtie outputs one alignment per line. Each line is a collection of 8 fields separated by tabs; from left to right, the fields are:

(1) Name of read that aligned.

(2) Reference strand aligned to, + for forward strand, - for reverse.

(3) Name of 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-complemented if orientation is -).

(6) ASCII-encoded read qualities (reversed if orientation is -). The encoded quality values are on the Phred scale and the encoding is ASCII-offset by 33 (ASCII char !).

(7) If -M was specified and the prescribed ceiling was exceeded for this read, this column contains the value of the ceiling, indicating that at least that many valid alignments were found in addition to the one reported.

(8) Comma-separated list of mismatch descriptors. If there are no mismatches in the alignment, this field is empty. A single descriptor has the format offset: reference-base > read-base. The offset is expressed as a 0-based offset from the high-quality (5') end of the read.

For more details, see Bowtie Output Format

Directory: Alignment
Files: Alignment_*.gz (compressed)

4.5.2 Showing Sequence Aligned to tRNAs

The sequences are aligned to the relative position in the corresponding mature or precursor tRNA and output to the alignment file in TXT format. The following table is just the example table explaining content of alignment file and show simulated data.

Table 9. The example table explaining content of the alignment file.

tRNA-Ala-AGC-1-1	chr13:21242569-21242641	Ala-AGC	At: 33-35						
GGGGGTGTAGCTCAGTGGTAGAGCGCGTGCTTAGCATGCACGAGGCCCTGGGTTCGATCCCCAGCACCTCCACCA									
))))))).))))).....((((.)))))).....(((((.....)))))).....((((((((((((....									
GGGGGTGTAGCTCAGT.....				tRF-5a	ExactMatch	27	46	35	29
GGGGGTGTAGCTCAGTGGTAGAGCGCGTGCT.....				tRF-5c	ExactMatch	29	16	12	60
.....TCCCCAGCACCTCCACCA				tRF-3a	ExactMatch	24	9	7	27
.....TCGATCCCCAGCACCTCCACCA				tRF-3b	ExactMatch	23	20	49	351
⋮				⋮	⋮	⋮	⋮	⋮	⋮

The sequences is aligned to the relative position in the corresponding mature or precursor tRNA. The first line is divided into four parts by the tab, it showed the tRNA ID, locus, amino acid-anticodon, and position of anticodon, respectively. The second and third line show the tRNA sequence and secondary structure, respectively. The each line from the fourth to blank are divided into four parts by tab, it showed the sequences, the type of tRF & tiRNA, information of align, and time of the same sequences, respectively. If the type of tRF & tiRNA is character “Other” , it indicated that the sequence aligned to tRNA sequencing but does not conform to the definition of subtype tRF & tiRNA (**see 2.1 section**).

Directory: Alignment
File: Alignment_to_tRNAs.gz (compressed)

4.5.3 The arf Format

The arf format is a proprietary file format generated and processed by miRDeep2 [7]. It contains information of reads mapped to a reference genome. Each line in such a file contains 13 columns.

(1) read identifier

(2) length of read sequence

(3) start position in read sequence that is mapped

(4) end position in read sequence that is mapped

(5) read sequence

(6) identifier of the genome-part to which a read is mapped to. This is either a scaffold id or a chromosome name

(7) length of the genome sequence a read is mapped to

(8) start position in the genome where a read is mapped to

(9) end position in the genome where a read is mapped to

(10) genome sequence to which a read is mapped

(11) genome strand information. Plus means the read is aligned to the sense-strand of the genome. Minus means it is aligned to the antisense-strand of the genome

(12) Number of mismatches in the read mapping

(13) Edit string that indicates matches by lowercase 'm' and mismatches by uppercase 'M'

Directory: miRNA
Files: miRNA_mapped.arf.gz (compressed)

4.6 Software Version

Softwares used in this project:

- Perl: v5.16.3
- Python: 2.7.5
- R: 3.5.1
- FastQC: v0.11.7
- Cutadapt: 1.17
- Bowtie: 1.2.2
- miRDeep2: 2.0.0.8

4.7 Summary Table of Files for Data Delivery

The delivered data files may be too large to view in Microsoft Excel, the recommended softwares for viewing the data are show in the following table. If the file is compressed into gz format, you can be uncompressed by softwares such as tar, 7-zip or gzip.

Note: We keep all your sequence and related files for only one year by our company policies.

Table 10. Summary of report folder

Data Folder	Files	Description	Software
Sequence/Raw_Data	*.fastq.gz	Compressed raw data	WordPad
Sequence/Trimmed_Data	*.fa.gz	Compressed rrimmed data	WordPad
Sequence_QC	*.png	Reads sequence score plot	Photo Viewer
Alignment	*.gz	Alignment results	WordPad
Express	Expression.xlsx	Expression profiles	Excel
Diff_Express	*.xlsx	Differentially expressed data	Excel
Plots	ReadsLength	Plots	Photo Viewer
	pca_and_correlation		
	Venn		
	Pie		
	SubtypeIsodecoder		
	SubtypeLength		
	Scatter_and_Volcano		
	Heatmap		
miRNA	miRNA_mapped.arf.gz	Result of alignment to miRNA	Excel
	*.xlsx	Expressed (Differentially) data	Excel
	Plots	Plots	Photo Viewer

5 Databases and References

5.1 Databases

GEO GEO is a public functional genomics data repository supporting MIAME-compliant data submissions, and is freely accessible at http://www.ncbi.nlm.nih.gov/geo/.

GtRNAdb The genomic tRNA database contains tRNA gene predictions made by tRNAscan-SE on complete or nearly complete genomes. The database can be searched by sequence or gene features, and is available at http://gtrnadb.ucsc.edu/

tRFdb tRFdb is a relational database of transfer RNA related fragments and can be available at http://genome.bioch.virginia.edu/trfdb/. With over 100 small RNA libraries analyzed, the database currently contains the sequences and read counts of the three classes of tRFs (tRF-5, tRF-3 and tRF-1) for eight species. The database can be searched by tRF ID or tRF sequence, and the results can be limited by organism. In each organism, tRFs are named in the order they are identified, with the frst tRF-5 named 5001, the frst tRF-3 named 3001 and the frst tRF-1 named 1001. In the case of tRF-5s and tRF-3s, there are multiple distinct subclasses. When there are two or more tRF-5s that differ only in length: an 'a', 'b' or 'c' is appended for tRF-5s of ∼15, ∼22 or ∼31 bases. All tRF-5as, -bs and -cs share a common seed sequence. Similarly, when there are two distinct tRF-3s mapping to the same tRNA, the tRF-3s of length ∼18 have an 'a' appended, while tRF-3s of length ∼22 have a 'b' appended. The latter is of particular importance since tRF-3-as and tRF-3-bs have different 5' ends and therefore different seed sequences.

MINIbase MINTbase is a web-based framework that serves the dual-purpose of being a content repository that comprises nuclear and mitochondrial tRFs found in multiple human tissues and a tool for the interactive exploration of these newly discovered molecules, and is freely accessible at http://cm.jefferson.edu/MINTbase/.

miRBase The miRBase database is a searchable database of published miRNA sequences and annotation, and can be available at http://www.mirbase.org/index.shtml.

5.2 References

[1] Chan, P.P. and T.M. Lowe, GtRNAdb: a database of transfer RNA genes detected in genomic sequence. Nucleic Acids Res, 2009. 37(Database issue): p. D93-7.

[2] Chan, P.P. and T.M. Lowe, GtRNAdb 2.0: an expanded database of transfer RNA genes identified in complete and draft genomes. Nucleic Acids Res, 2016. 44(D1): p. D184-9.

[3] Lowe, T.M. and P.P. Chan, tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. Nucleic Acids Res, 2016. 44(W1): p. W54-7.

[4] Lowe, T.M. and S.R. Eddy, tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res, 1997. 25(5): p. 955-64.

[5] Selitsky, S.R. and P. Sethupathy, tDRmapper: challenges and solutions to mapping, naming, and quantifying tRNA-derived RNAs from human small RNA-sequencing data. BMC Bioinformatics, 2015. 16: p. 354.

[6] Langmead, B., et al., Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol, 2009. 10(3): p. R25.

[7] Friedlander, M.R., et al., miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. Nucleic Acids Res, 2012. 40(1): p. 37-52.

[8] Andrews, S., FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/. 2010.

[9] Martin, M., Cutadapt removes adapter sequences from high-throughput sequencing reads. 2011, 2011. 17(1).

[10] Robinson, M.D., D.J. McCarthy, and G.K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics, 2010. 26(1): p. 139-40.

[11] Anderson, P. and P. Ivanov, tRNA fragments in human health and disease. FEBS Lett, 2014. 588(23): p. 4297-304.

[12] Pliatsika, V., et al., MINTbase: a framework for the interactive exploration of mitochondrial and nuclear tRNA fragments. Bioinformatics, 2016. 32(16): p. 2481-9.

[13] Zheng, L.L., et al., tRF2Cancer: A web server to detect tRNA-derived small RNA fragments (tRFs) and their expression in multiple cancers. Nucleic Acids Res, 2016. 44(W1): p. W185-93.

[14] Kumar, P., C. Kuscu, and A. Dutta, Biogenesis and Function of Transfer RNA-Related Fragments (tRFs). Trends Biochem Sci, 2016. 41(8): p. 679-689.

[15] Saikia, M., et al., Genome-wide identification and quantitative analysis of cleaved tRNA fragments induced by cellular stress. J Biol Chem, 2012. 287(51): p. 42708-25.

[16] Li, S. and G.F. Hu, Emerging role of angiogenin in stress response and cell survival under adverse conditions. J Cell Physiol, 2012. 227(7): p. 2822-6.

[17] Kumar, P., et al., Meta-analysis of tRNA derived RNA fragments reveals that they are evolutionarily conserved and associate with AGO proteins to recognize specific RNA targets. BMC Biol, 2014. 12: p. 78.

[18] Kumar, P., et al., tRFdb: a database for transfer RNA fragments. Nucleic Acids Res, 2015. 43(Database issue): p. D141-5.