

Supplementary Figure 1



CPSS

Computational Platform for analysis of
Small RNA deep Sequencing data
BioStarCa group

HomeResults & DownloadHelpToolsContact Us

Search

Input Your Job ID Here

Example: 9801014522

Links

DARIO

SeqBuster

miRBase

miRanda

RNAhybrid

TargetScan

miRecords

dbSNP

Faculties

Qinghua Shi

Yang Gao

About Us

Welcome to CPSS

User can prepare their input file from [HERE](#).

Users can start analysis from [HERE \(Single / Paired \)](#).

"Single" means the dataset of one sample, and "Paired" means the datasets of two samples.

Introduction

Key words: Small RNA deep-sequencing, miRNA, piRNA, CPSS

Non-coding small RNAs are RNA molecules that are not translated into proteins, include ribosomal RNAs (rRNAs), transfer RNAs (tRNAs), microRNAs (miRNA), piwi-interacting RNAs (piRNAs) and other RNA species (1). These RNAs are also known as the regulators that carry out a variety of biological functions, have been discovered only 20 years ago (2).

MicroRNAs, first discovered in C. Elegans (3), are short RNA molecules, approximately 22 nucleotides in length. Mature miRNAs are post-transcriptional regulators that bind to complementary sequences on target mRNAs, usually resulting in translational repression, degradation, or both of the targeting mRNAs. The importance of miRNAs is evidenced by their evolutionary conservation across species and by many cellular pathways in which they are involved, including growth and proliferation, apoptosis, and developmental timing (4).

Piwi-interacting RNAs (piRNAs) are typically ~30 bases long, as a distinct population of piwi-associated small RNAs first discovered in Drosophila. The function of piRNAs is currently unknown, but the homology of Piwi proteins to Argonaute proteins and similarities of piRNAs to microRNAs and short-interfering RNAs suggest that piRNAs play a role in regulatory processes during meiosis, particularly those in spermatogenesis (5).

Deep sequencing technologies have found broad applicability in functional genomics research (6). The main advantage of deep sequencing technologies is eliminating the need for in vivo cloning by clonal

Welcome to our cpss

CPSS HELP

For all users, CPSS can analyze the small RNA deep sequencing data in different architecture: single and paired.

By using CPSS, small RNA-NGS data can be analyzed systematically in new platform after a single submission of data by integration of annotation and functional analysis of novel and/or differentially expressed miRNAs. CPSS generates an analysis report including: 1) annotation analysis, which provides a comprehensive analysis for small RNA transcripts, such as the length distribution and genome mapping of the sequencing reads, small RNA annotation, prediction of novel miRNAs, identification of differentially expressed miRNAs, piRNAs and other small RNAs between samples; and detection of miRNA SEBs and subfamilies; 2) functional analysis, which provides the functional analysis of miRNAs, e.g. predicting miRNA target genes by multi-tools, enriching Gene Ontology terms (GO), performing signaling pathways, and protein-protein interaction (PPI) analysis for the predicted genes.

If users provide an email address, our server will send a reminder email to users when the job is done. Users could get the results after receiving the reminder email.

Please enter your email.

2. When uploaded to CPSS, the data should be in a specified file format (*.fa or *.gz). And a SeqClean program was provided in CPSS package (http://seq.clean.edu.cn/doc/seqclean_pkg_instructions.html), which could perform the transformation of raw data from Illumina Genome Analyzer, 454 FLX instrument. The FASTA format (*.fa) files can be further compressed into *.gz format using a software that can be found in <http://www.ftp.org/download.html> for free. Several small RNA-NGS data from different species are provided in CPSS, and users could download these data to test our web server. In addition, the maximum allowable size of data input is not limited.

Sample Upload.

sample.fa.gz (*.gz)

3. Users should choose the reference genome through a scroll bar for sequence reads alignment. Currently, CPSS is used for the analysis of small RNA deep sequencing data from ten organisms, and will be extended to the analysis of data from more species in near future.

Welcome to CPSS

Released Feb2012 version 1.1

CPSS (Computational Platform for analysis of Small RNA deep Sequencing data) is a web server which provides a comprehensively and systematically integrated pipeline for analyzing small RNA deep sequencing data.

- Reads length distribution
- Genome mapping of the sequencing reads
- Small RNA quantification
- Novel microRNA prediction
- MicroRNA target prediction
- GO analysis for the predicted target genes
- Pathway analysis for genes in the enriched GO terms
- Protein-Protein interaction analysis for genes in the enriched GO terms

Please enter your email.

Please enter your email.

Sample upload.

Sample for test from human.fa.gz(*.fa) More test data.

Species selection.

Human hg19

Parameters for mapping tools.

How to avoid inputting large data

By using clean-reads.pl, the size of raw data file is reduced to an acceptable size for the web server because low quality reads and 3/5' adaptor sequences can be filtered and removed.

1. low quality reads:
For Solexa/Illumina 1.0 format, the quality value can be calculated by $Q = (\text{ASCII character code}) - 64$. If $Q < 9$, this reads was defined as low quality reads.
For Sanger format, the quality value can be calculated by $Q = (\text{ASCII character code}) - 33$. If $Q < 15$, this reads was defined as low quality reads.
For Illumina 1.3+ format, the quality value can be calculated by $Q = (\text{ASCII character code}) - 64$. If $Q < 10$, this reads was defined as low quality reads.

2. adaptor sequences:
Solexa use standard 3- and 5-adaptor sequences in their small RNA library so it is not necessary analysis the adaptor sequences. If the adaptor sequence is at 5' of the read, then the read is removed. If the adaptor sequence is at 3' of the read, the adaptor sequence is trimmed from the read sequence and discard reads <16nt after the removal of 3' adaptor.

The data in FASTQ format could be purified into FASTA format. An example sequence in FASTA format as follow:

```
>tagid1_785344 ("tagid1" = unique ID, "785344" = reads count)
TGAGGTAGTAGTATTGATAGTT
```

Usage:
Options:
-I Short reads file in fastq file_type
-T File type, default=1
1 = Solexa/Illumina 1.0 format: encode a Solexa/Illumina quality score from -5 to 62 using ASCII 59 to 126
2 = Sanger format: encode a Phred quality score from 0 to 93 using ASCII 33 to 126
3 = Illumina 1.3+ format: encode a Phred quality score from 0 to 62 using ASCII 64

Welcome to CPSS

Released Feb2012 version 1.1

CPSS (Computational Platform for analysis of Small RNA deep Sequencing data) is a web server which provides a comprehensively and systematically integrated pipeline for analyzing small RNA deep sequencing data.

Please enter your ID.

Example results:

Single sample:

Human fetal ovary: 9801014522.

Mouse fetal ovary: 4101809422.

Paired samples:

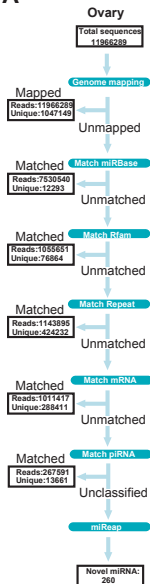
Wild type vs. Spo11 KO mouse ovary: 922041050.

Wild type vs. Dmc1 KO mouse ovary: 560323860.

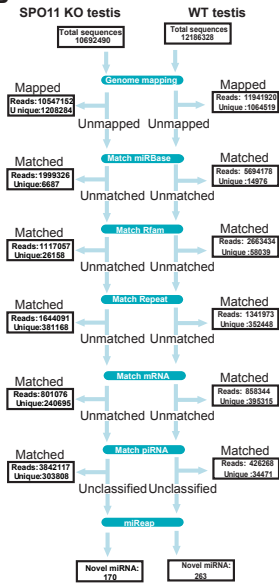
Annotation analysis			
1. General information	1.1. Novel miRNA expression		
	1.1.1. miRNA expression		
2. Details of expression	1.1.1.1. miRNA expression		
	1.1.1.1.1. miRNA expression		
	1.1.1.1.1.1. miRNA expression		
	1.1.1.1.1.1.1. miRNA expression		
3. miRNA modification/editing	1.1.1.1.1.1.1.1. miRNA expression		
	1.1.1.1.1.1.1.1.1. miRNA expression		
Function analysis			
4. miRNA target prediction	1.1.1.1.1.1.1.1.1.1. miRNA expression		
	1.1.1.1.1.1.1.1.1.1.1. miRNA expression		

Supplementary Figure 3

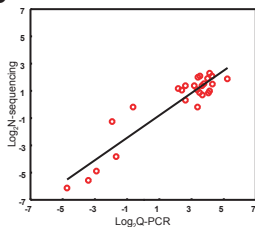
A



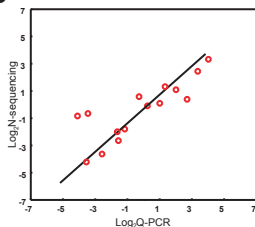
B



C



D



Supplementary Figure 4

