1.Sample collection and preparation

1.1 RNA quantification and qualification

RNA degradation and contamination, especially DNA contamination, was monitored on 1.5% agarose gels. RNA concentration and purity was measured using the NanoDrop 2000 Spectrophotometer (Thermo Fisher Scientific, Wilmington, DE). RNA integrity was assessed using the RNA Nano 6000 Assay Kit of the Agilent Bioanalyzer 2100 System (Agilent Technologies, CA, USA).

1.2 Library preparation for sRNA sequencing

A total amount of 2.5 ng RNA per sample was used as input material for the RNA sample preparations. Sequencing libraries were generated using NEBNext^{R} Ultra^{TM} small RNA Sample Library Prep Kit for Illumina^{R}(NEB, USA) following manufacturer's recommendations and index codes were added to attribute sequences to each sample. Briefly, first of all, ligated the 3′ SR Adaptor. Mixed 3′ SR Adaptor for Illumina, RNA and Nuclease-Free Water, mixture system after incubation for 2 minutes at 70 degrees in a preheated thermal cycler. Tube was transfer to ice. Then, add 3′ Ligation Reaction Buffer (2X) and 3′ Ligation Enzyme Mix ligate the 3′ SR Adaptor. Incubated for 1 hour at 25°C in a thermal cycler. To prevent adaptor-dimer formation, the SR RT Primer hybridizes to the excess of 3′ SR Adaptor (that remains free after the 3′ ligation reaction) and transforms the single stranded DNA adaptor into a double-stranded DNA molecule. dsDNAs are not substrates for ligation mediated. The second, ligated the 5′ SR Adaptor. Then, reverse transcription synthetic first chain. Last, PCR amplification and Size Selection. PAGE gel was used to electrophoresis fragment screening purposes, rubber cutting recycling as the pieces get small RNA libraries. At last, PCR products were purified (AMPure XP system) and library quality was assessed on the Agilent Bioanalyzer 2100 system.

1.3 Clustering and sequencing

The clustering of the index-coded samples was performed on a cBot Cluster Generation System using TruSeq PE Cluster Kitv3-cBot-HS(Illumia) according to the manufacturer's instructions. After cluster generation, the library preparations were sequenced on an Illumina platform and reads were generated.

2.Data analysis

2.1 Quality control

Raw data (raw reads) of fastq format were firstly processed through in-house perl scripts. In this step, clean data(clean reads) were obtained by removing reads containing adapter, reads containing ploy-N and low quality reads from raw data. And reads were trimmed and cleaned by removing the sequences smaller than 15 nt or longer than 35 nt. At the same time, Q20, Q30, GC-content of the clean data were calculated. All the downstream analyses were based on clean data with high quality.

2.2 miRNA analysis

2.2.1 Comparative analysis

Use Bowtie tools soft, The Clean Reads respectively with Silva database, GtRNAdb database, Rfam database and Repbase database sequence alignment, filter ribosomal RNA (rRNA), transfer RNA (tRNA), small nuclear RNA (snRNA), small nucleolar RNA (snoRNA) and other ncRNA and repeats. The remaining reads were used to detect known miRNA and novel miRNA predicted by comparing with known miRNAs from miRBase. Randfold tools soft was used for novel miRNA secondary structure prediction.

2.2.2 Target gene functional annotation

Gene function was annotated based on the following databases:
NR(non-redundant protein sequence database);
Swiss-Prot(A manually annotated, non-redundant protein sequence database);GO(Gene Ontology database);
COG(The database of Clusters of Orthologous Groups of proteins);
KOG(The database of Clusters of Protein homology);
Pfam(The database of Homologous protein family);
KEGG(The database of Kyoto Encyclopedia of Genes and Genomes).

2.2.3 Quantification of miRNA expression levels

miRNA expression levels were estimated for each sample:
1) sRNA were mapped back onto the precursor sequence.
2) Readcount for each miRNA was obtained from the mapping results

2.2.4 Differential expression analysis

1)For the samples with biological replicates:
Differential expression analysis of two conditions/groups was performed using the DESeq R package. DESeq provide statistical routines for determining differential expressionin digital gene expression data using a model based on the negative binomial distribution. The resulting Pvalues were adjusted using the

Benjamini and Hochberg's approach for controlling the false discovery rate. Genes with an adjusted P-value <0.01 and absolute value of log2(Fold change) >1 found by DESeq were assigned as differentially expressed.
2)For the samples without biological replicates:

Prior to differential gene expression analysis, for each sequenced library, the read counts were adjusted by edgeR program package through one scaling normalized factor. Differential expression analysis of two samples was performed using the EBseq (2010) R package. The resulting FDR (false discovery rate) were adjusted using the PPDE (posterior probability of being DE) . The FDR < 0.05 & |log2 (FoldChange) | ≥1 was set as the threshold for significantly differential expression.

2.2.5 Target gene functional enrichment analysis

1) GO enrichment analysis

Gene Ontology (GO) enrichment analysis of the target gene of differenti ally expressed miRNAs was implemented by the clusterProfiler R packages. Enrichment analysis uses hypergeometric testing to find GO entries that are significantly enriched compared to the entire genome background.
2) KEGG pathway enrichment analysis

KEGG (Kanehisa et al., 2008) is a database resource for understanding high-level functions and utilities of the biological system, such as the cell, the organism and the ecosystem, from molecular-level information, especially large-scale molecular datasets generated by genome sequencing and other high-throughput experimental technologies (http://www.genome.jp/kegg/). We used clusterProfiler R packages to find KEGG pathway that are significantly enriched compared to the entire genome background.