

Small RNA-seq Analysis

Methods

Library construction and sequencing

Total RNA was separated by 15% agarose gels to extract the small RNA (18-30 nt). After precipitated by ethanol and centrifugal enrichment of small RNA sample, the library was prepared according to the method and process of Small RNA Sample Preparation Kit (Illumina, RS-200-0048). The main contents are as follows: 1) Connecting the 3' adaptor to the separated small RNA. 2) Connecting the 5' adaptor to the separated small RNA. 3) RT-PCR. 4) Recycling strips of 145-160bp (22-30nt RNA).

RNA concentration of library was measured using Qubit® RNA Assay Kit in Qubit® 2.0 (Thermo Scientific, Waltham, MA) to preliminary quantify and then dilute to 1ng/ l. Insert size was assessed using the Agilent Bioanalyzer 2100 system (Agilent Technologies, CA, USA), and after the insert size consistent with expectations, qualified insert size was accurate quantitative using Taqman fluorescence probe of AB Step One Plus Real-Time PCR system (Library valid concentration>2nM). The qualified libraries were sequenced by an Illumina Hiseq 2500 platform and generate 50 bp single-end reads.

Sequencing data analysis

Cutadapt¹ (v1.16) was used for adapter trimming. Trimmed reads were mapped to genome by bowtie² (v0.11.3) (bowtie -n 1 -e 80 -l 30 -m 5 -best -strata). Novel miRNAs are predicted by miRDeep2³ (v2.0.0.8) upon the detection of potential mature, star and loop sequences from the read pool. Targetfinder⁴ was used to predict small RNA binding sites on target genes from a sequence database.

Differential expression of miRNAs between experimental conditions was calculated using edgeR⁵ (v3.12.1). Go enrichment analysis of the miRNA target genes was performed using the Goseq⁶ (v1.22.0) based Wallenius non-central hyper-geometric distribution, which can adjust for gene length bias. KEGG pathway enrichment analysis was done using KOBAS⁷ (v2.0). Other data statistic and visualizing were performed by in-house R scripts.

Reference

- 1.Martin, M. *et al.* Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal* **17**, pp-10 (2011).
- 2.Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L.*et al.* Ultrafast and memory-efficient alignment of short dNA sequences to the human genome. *Genome Biology* **10**, R25 (2009).
- 3.Friedländer, M. R., Mackowiak, S. D., Li, N., Chen, W. & Rajewsky, N.*et al.* miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic acids research* **40**, 37-52 (2011).

4. Allen, E., Xie, Z., Gustafson, A. M. & Carrington, J. C. *et al.* microRNA-directed phasing during trans-acting siRNA biogenesis in plants. *Cell* **121**, 207–221 (2005).
5. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. *et al.* edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
6. Young, M. D., Wakefield, M. J., Smyth, G. K. & Oshlack, A. *et al.* Gene ontology analysis for rNA-seq: accounting for selection bias. *Genome biology* **11**, R14 (2010).
7. Xie, C. *et al.* KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic acids research* **39**, W316–W322 (2011).