Methods

Sample collection and preparation

RNA isolation, library preparation and sequencing

- **RNA isolation**
- **RNA quantification and qualification (Novogene Experimental Department)**
  - ✧ RNA degradation and contamination was monitored on 1% agarose gels.
  - ✧ RNA purity was checked using the NanoPhotometer® spectrophotometer (IMPLEN, CA, USA).
  - ✧ RNA concentration was measured using Qubit® RNA Assay Kit in Qubit® 2.0 Flurometer (Life Technologies, CA, USA).
  - ✧ RNA integrity was assessed using the RNA Nano 6000 Assay Kit of the Bioanalyzer 2100 system (Agilent Technologies, CA, USA).

- **Library preparation for lncRNA sequencing (Novogene Experimental Department)**

  A total amount of 5 μg RNA per sample was used as input material for the RNA sample preparations. Firstly, ribosomal RNA was removed by Epicentre Ribo-zero™ rRNA Removal Kit (Epicentre, USA), and rRNA free residue was cleaned up by ethanol precipitation. Subsequently, the linear RNA was digested with 3 U of RNase R (Epicentre, USA) per μg of RNA. The sequencing libraries were generated by NEBNext® Ultra™ Directional RNA Library Prep Kit for Illumina® (NEB, USA) following manufacturer's recommendations. Briefly, fragmentation was carried out using divalent cations under elevated temperature in NEBNext First Strand Synthesis Reaction Buffer(5X). First strand cDNA was synthesized using random hexamer primer and M-MuLV Reverse Transcriptase (RNaseH-). Second strand cDNA synthesis was subsequently performed using DNA Polymerase I and RNase H. In the reaction buffer, dNTPs with dTTP were

replaced by dUTP. Remaining overhangs were converted into blunt ends via exonuclease/polymerase activities. After adenylation of 3' ends of DNA fragments, NEBNext Adaptor with hairpin loop structure were ligated to prepare for hybridization. In order to select cDNA fragments of preferentially 150~200 bp in length, the library fragments were purified with AMPure XP system (Beckman Coulter, Beverly, USA). Then 3 μl USER Enzyme (NEB, USA) was used with size-selected, adaptor-ligated cDNA at 37°C for 15 min followed by 5 min at 95℃ before PCR. Then PCR was performed with Phusion High-Fidelity DNA polymerase, Universal PCR primers and Index (X) Primer. At last, products were purified (AMPure XP system) and library quality was assessed on the Agilent Bioanalyzer 2100 system

- **Clustering and sequencing (Novogene Experimental Department)**

The clustering of the index-coded samples was performed on a cBot Cluster Generation System using TruSeq PE Cluster Kit v3-cBot-HS (Illumia) according to the manufacturer's instructions. After cluster generation, the libraries were sequenced on an Illumina Hiseq 4000 platform and 150 bp paired-end reads were generated.

Data analysis (Novogene Gene Regulation Department)

- **Quality control**

Raw data(raw reads) of fastq format were firstly processed through in-house perlscripts. In this step, clean data(clean reads) were obtained by removing reads containing adapter, reads on containing ploy-N and low quality reads from raw data. At the same time, Q20, Q30 and GC content of the clean data were calculated. All the downstream analyses were based on the clean data with high quality.

- **Mapping to the reference genome**

Reference genome and gene model annotation files were downloaded from genome website directly. Index of the reference genome was built using bowtie2 v2.2.8 and paired-end clean reads were aligned to the reference genome using Bowtie (Langmead, B.et al)

- **circRNA identification**

The circRNA were detected and identified using find_circ (Sebastian Memczak et al., 2013) and CIRI2 (Gao et al., 2017). Circos software was used to construct the circos figure.

- **Quantification of gene expression level**

The raw counts were first normalized using TPM (Zhou et al., 2010)

Normalized expression level = (readCount*1,000,000)/libsize (libsize is the sum of circRNA readcount)

- **Differential expression analysis**

For the samples with biological replicates:

Differential expression analysis of two conditions/groups was performed using the DESeq R package (1.10.1). DESeq provide statistical routines for determining differential expression in digital gene expression data using a model based on the negative binomial distribution. The resulting P values were adjusted using the Benjamini and Hochberg's approach for controlling the false discovery rate. Genes with an adjusted P-value found by DESeq were assigned as differentially expressed.

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 DEGseq (2010) R package. Pvalue was adjusted using q value (Storey et al, 2003). qvalue1 was set as the threshold for significantly differential expression.

- **GO and KEGG enrichment analysis**

Gene Ontology (GO) enrichment analysis for host genes of differentially expressed circRNAs were implemented by the GOseq R package, in which gene length bias was corrected (Young et al., 2010). GO terms with corrected Pvalue less than 0.05 were considered significantly enriched by differential expressed genes. KEGG is a database resource for understanding high-level functions and utilities of the biological system (Kanehisa et al., 2008), 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 KOBAS software to test the statistical enrichment of differential expression genes or circRNA host genes in KEGG pathways (Mao, X.et al.1995).

MicroRNA targer site analysis

MicroRNA targer site in exons of circRNA loci were identified using miRanda (animal species) or psRobot (plant species).

CircRNA-miRNA-gene network analysis

Cytoscape software was used to construct the circRNA-miRNA-gene networks.