Pakistan J. Zool., vol. 56(3), pp 1323-1335, 2024 **DOI**: https://dx.doi.org/10.17582/journal.pjz/20230512070544

Supplementary Material

Modulation of Tau Expression in the Colonic Muscle Layer of a Rat Model with Opioid-Induced Constipation

Jinzhao Li[1], Yawen Zhang[2], Binghan Jia[1], Yuqiong Zhao[1], Huijuan Luo[1], Xiaojie Ren[1], Yuan Li[3], Xiaoyan Bai[1], Jing Ye[3] and Junping Li[1]*

[1]Department of College of Basic Medical Sciences, Ningxia Medical University, No. 1160 of Shengli Street, Ningxia Hui Autonomous Region, Yinchuan 750004, China
[2]Department of General Surgery, People's Hospital of Ningxia Hui Autonomous Region, Yinchuan 750004, China
[3]Department of College of Clinical Medicine, Ningxia Medical University, Yinchuan 750004, China

Jinzhao Li and Yawen Zhang contributed equally to this study.

Supplementary Material 1. Material methods of the gene sequencing.

Experiment protocol

1ug total RNA was prepared for cDNA libraries using cDNA-PCR Sequencing Kit (SQK-PCS109) protocol provided by Oxford Nanopore Technologies (ONT). Briefly, the template switching activity of reverse transcriptases enrich for full-length cDNAs and add defined PCR adapters directly to both ends of the first-strand cDNA. And following cDNA PCR for 14 circles with LongAmp Tag (NEB). The PCR products were then subjected to ONT adaptor ligation using T4 DNA ligase (NEB). Agencourt XP beads was used for DNA purification according to ONT protocol. The final cDNA libraries were added to FLO-MIN109 flowcells and run on PromethION platform at Biomarker Technology Company (Beijing, China).

Data analysis

Oxford nanopore technologies long read processing

Raw reads were first filtered with minimum average read quality score=7 and minimum read length=500bp. Ribosomal RNA were discarded after mapping to rRNA database. Next, full-length, non-chemiric (FLNC) transcripts were determined by searching for primer at both ends of reads. Clusters of FLNC transcripts were obtained after mapping to reference genome with mimimap2, and consensus isoforms were obtained after polishing within each cluster by pinfish.

Remove redundant

Consensus sequences were mapped to reference genome using minimap2. Mapped reads were further collapsed by cDNA_Cupcake package with min-coverage=85% and min-identity=90%. 5' difference was not considered when collapsing redundant transcripts.

Find fusion transcript

The criteria for fusion candidates are that a single transcript must:

(1) must map to 2 or more loci

(2) minimum coverage for each loci is 5% and minimum coverage in bp is >= 1 bp

(3) total coverage is >= 95%

(4) distance between the loci is at least 10kb

Structure analysis

Transcripts were validated against known reference transcript annotations with gffcompare. AS events including IR, ES, AD, AA and MEE were identified by the AStalavista tool. SSR of the transcriptome were identified using MISA. APA analysis was conducted with TAPIS. CDS were predicted by Trans Decoder.

* Corresponding author: lijunping_jp@163.com
0030-9923/2024/0003-1323 $ 9.00/0

Transcription factors prediction

Plant transcription factors were identified with iTAK, while Animal transcription factors were identified from animal TFDB.

lncRNA analysis

Four computational approaches include CPC/CNCI/CPAT/Pfam/ were combined to sort non-protein coding RNA candidates from putative protein-coding RNAs in the transcripts. Putative protein-coding RNAs were filtered out using a minimum length and exon number threshold . Transcripts with lengths more than 200 nt and have more than two exons were selected as lncRNA candidates and further screened using CPC/CNCI/CPAT/Pfam that have the power to distinguish the protein-coding genes from the non-coding genes.

Gene functional annotation

Gene function was annotated based on the following databases:

NR (NCBI non-redundant protein sequences); Pfam (Protein family); KOG/COG/eggNOG (Clusters of Orthologous Groups of proteins); Swiss-Prot (A manually annotated and reviewed protein sequence database);

KEGG (Kyoto Encyclopedia of Genes and Genomes); GO (Gene Ontology).

Quantification of gene/transcript expression levels and Differential expression analysis

Full length reads were mapped to the reference transcriptome sequence. Reads with match quality above 5 were further used to quantify. Expression levels were estimated by reads per gene/transcript per 10,000 reads mapped.

For the samples with biological replicates

Differential expression analysis of two conditions/ groups was performed using the edgeR R package (3.8.6). edgeR provides 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 a P value < 0.05 and foldchange ≥ 1.5 found by edgeR 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 EBSeq R package (1.6.0). The resulting FDR (false discovery rate) were adjusted using the PPDE (posterior probability of being DE). The FDR < 0.05 and foldchange ≥ 1.5 was set as the threshold for significantly differential expression.

Functional enrichment analysis

GO enrichment analysis

Gene Ontology (GO) enrichment analysis of the differentially expressed genes (DEGs) was implemented by the GOseq R packages based Wallenius non-central hyper-geometric distribution (Young et al, 2010), which can adjust for gene length bias in DEGs.

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 KOBAS (Mao et al., 2005) software to test the statistical enrichment of differential expression genes in KEGG pathways.

PPI (Protein protein interaction)

The sequences of the DEGs were blast (blastx) to the genome of a related species (the protein protein interaction of which exists in the STRING database: http://string-db.org/) to get the predicted PPI of these DEGs. Then the PPI of these DEGs were visualized in Cytoscape (Shannon et al., 2003).