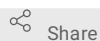Sep 11, 2021

🌐 Identification of differentially expressed long noncoding RNAs and pathways in liver tissues from rats with hepatic fibrosis

Xiong Xiao[1], Yan Wang[1], Xiaozhong Wang[2]

[1]Department of Traditional Chinese Medicine, The Fifth People's Hospital Affiliated to Chengdu University of Traditional Chinese Medicine;

[2]Department of Liver Disease, Traditional Chinese Medicine Hospital Affiliated to Xinjiang Medical University

| 1 | Works for me | | ∞ Share |

dx.doi.org/10.17504/protocols.io.bwptpdnn

TCM of CDWY

xxzaihuaxi

DISCLAIMER

ABSTRACT

To identify long non-coding RNAs (lncRNAs) and their potential roles in hepatic fibrosis in rat liver issues induced by CCl4, lncRNAs and genes were analyzed in fibrotic rat liver tissues by quantitative reverse transcription polymerase chain reaction (qRT-PCR).

DOI

dx.doi.org/10.17504/protocols.io.bwptpdnn

PROTOCOL CITATION

FUNDERS ACKNOWLEDGEMENT

KEYWORDS

function study, hepatic fibrosis, long noncoding RNAs, rat liver issues, qRT-PCR, quantitative reverse transcription polymerase chain reaction

GUIDELINES

Kim, D., G. Pertea, C. Trapnell, H. Pimentel, R. Kelley and S. L. Salzberg (2013). "TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions." Genome Biology **14**(4): R36.
Robinson, M. D., D. J. McCarthy and G. K. Smyth (2010). "edgeR: a Bioconductor package for differential expression analysis of digital gene expression data." Bioinformatics **26**(1): 139-140.
Trapnell, C., B. A. Williams, G. Pertea, A. Mortazavi, G. Kwan, M. J. van Baren, S. L. Salzberg, B. J. Wold and L. Pachter (2010). "Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation." Nat Biotechnol **28**(5): 511-515.
Xie, C., X. Mao, J. Huang, Y. Ding, J. Wu, S. Dong, L. Kong, G. Gao, C. Y. Li and L. Wei (2011). "KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases." Nucleic Acids Research **39**(Web Server issue): 316-322.
Cabili MN, Trapnell C, Goff L, Koziol M, Tazon-Vega B, Regev A, Rinn JL. 2011. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes & development 25: 1915-1927.
Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nature protocols 7: 562-578.
Kong L, Zhang Y, Ye ZQ, Liu XQ, Zhao SQ, Wei L, Gao G. 2007. CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. Nucleic acids research 35: W345-349.
Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26: 139-140.

MATERIALS TEXT

hods: extract (Trizol, Invitrogen), reverse transcription (thermo # k1622, thermo), quantitative PCR (SYBR select master mix, Qigen), primer express 2.0 software design, Shanghai Shenggong Biotechnology Co., Ltd; Low speed and large capacity multi tube centrifuge (Shanghai Anting Scientific Instrument Factory, tdl-5-a), low temperature centrifuge (Eppendorf, 5810R), bio rad mycyler thermal cyeler PCR instrument, PCR ABI stepone plus fluorescence quantitative PCR instrument was used.

DISCLAIMER:

1   RNA-Seq Raw Data Clean and Alignment

2　　Raw reads containing more than 2-N bases were first discarded.

3　　Then adaptors and low-quality bases were trimmed from raw sequencing reads using FASTX-Toolkit (Version 0.0.13). The short reads less than 16nt were also dropped.

4　　After that, clean reads were aligned to the GRch38 genome by tophat2 (Kim, Pertea et al. 2013) allowing 4 mismatches. Uniquely mapped reads were used for gene reads number counting and FPKM calculation (fragments per kilobase of transcript per million fragments mapped) (Trapnell, Williams et al. 2010).

5　　Differentially Expressed Genes (DEG) analysis

6　　The R Bioconductor package edgeR (Robinson, McCarthy et al. 2010) was utilized to screen out the differentially expressed genes (DEGs). A false discovery rate 2 or < 0.5 were set as the cut-off criteria for identifying DEGs.

7　　Functional enrichment analysis

8　　To sort out functional categories of DEGs, Gene Ontology (GO) terms and KEGG pathways were identified using KOBAS 2.0 server (Xie, Mao et al. 2011). Hypergeometric test and Benjamini-Hochberg FDR controlling procedure were used to define the enrichment of each term.

9　　LncRNA Prediction

10　　LncRNA prediction pipeline was followed the method of one previous study (Cabili et al. 2011). Detail prediction pipeline and the filtering thresholds were described as follows:

11　　(1) First, based on the alignment result of RNA-Seq, transcripts were assembled by Cufflinks V2.2 (Trapnell et al. 2012) using default parameters. After the initial assembly, transcripts with FPKM no less than 0.3 were reserved for the following filtering.

12　　(2) Cuffcompare that was embedded in Cufflinks was used to compare the transcripts with known genes of reference genome, and novel transcripts including intergenic, intronic and antisense region were reserved as the candidate lncRNAs. Transcripts adjacent to known coding genes within 1000 bp were regarded as UTRs and also discarded.

13　　(3) To filter the coding potential transcripts, coding potential score (CPS) was evaluated by coding potential calculator (CPC) software (Kong et al. 2007). CPC is a support vector machine-based classifier to assess the protein-coding potential of transcripts based on six biologically meaningful sequence features. Transcripts with CPS below zero were regarded as non-coding RNAs.

14　　(4) Transcripts satisfying the above conditions, with multiple exons no smaller than 200 bases and single exon no

smaller than 1000 bases were reserved as lncRNAs.

15   (5) Finally, we combined known and predicted lncRNAs from all samples together to obtain the final lncRNA set, then we re-calculated the expression level of each lncRNA genes. Antisense reads of lncRNAs were discarded.

16   Differentially Expressed lncRNAs

17   After getting the Expression level of all lncRNAs in all samples, differentially expressed lncRNAs were analyzed by using edgeR (Robinson et al. 2010), one of R packages. For each lncRNA, the p-value was obtained based on the model of negative binomial distribution. The fold changes were also estimated within this package. 0.05 q-value and 2-fold change were set as the threshold to define Differentially Expressed lncRNAs.

18   Cis acting

19   Based on the expression of each mRNA and DElncRNA, correlation coefficient and P-value are obtained for each mRNA-DELncRNA pair.

20   Then we filtered the result by a given threshold, with absolute correlation coefficient no less than 0.6 and P-value less than 0.05. Besides the positive correlation pairs, negative pairs with correlation coefficient less than 0 were also included. The filtered gene pairs format the expression network. For each differentially expressed lncRNA, we obtain expressed genes from its upstream and downstream region within 10000 bases, and these genes overlap with co-expressed genes to obtain lncRNA targets.