

Development of a network visualization and analysis system for malignant tumors based on transcriptome data

GENG ZEZHENG¹, FANG XIAONAN^{1*}, BI CHUANMEI¹, DONG LING¹

¹SHANDONG MANAGEMENT UNIVERSITY

Abstract: Shiny technology has developed rapidly in recent years, as an R package for developing interactive app, through which we can package the written R code into a web app, which can not only save user time, but also accelerate the development of the speed of user-end communication, analyze the transcriptome data of related malignant tumors, and construct a ceRNA network diagram of desired malignant tumors. The code utilizing shiny technology package can facilitate users to map the ceRNA network associated with malignant tumors only through screen operation, significantly improving the efficiency and accuracy of clinical decision support in primary hospitals.

1 Introduction

The recent discovery of ceRNA regulatory networks as novel mechanisms by which RNAs interact with each other in vivo and can regulate the action of coding genes has also expanded the original in vivo understanding of the large number of noncoding RNAs^[1]. A large number of experiments have confirmed that the ceRNA network is widespread in different tumors, and that ceRNAs may influence the prognosis both in terms of early diagnosis and prognosis assessment, as well as targets for cancer therapy.

The web application for visualization built with shiny framework is able to carry out network scaffolding of ceRNA on transcriptome data of human malignant tumors, and can perform differential analysis of RNAs involved in ceRNA network construction, survival analysis and other downstream analysis, can fully mine the relationship between genomics data and obtain the targets and key communication that are of great significance for malignant tumor treatment, and finally find the genes closely related to the corresponding malignant tumors, which greatly improve the efficiency and accuracy of clinical decision-making of medical personnel.

2 System function and process

The functions designed in this system, including the data preprocessing, data normalization, differential gene expression analysis, ceRNA network analysis, univariate survival analysis as well as enrichment analysis, are a complete functional pipeline for ceRNA network construction and prognosis analysis^[2]. Users upload the downloaded transcriptome data of related malignant tumors to the system, so that the above analysis operations can be performed, and the corresponding figures can be drawn for visual display. Figure.1 shows the system flow chart of the system.

In this system, the application of Data by GDC (Genomic Data Commons) of TCGA project web site (<https://portal.gdc.cancer.gov/repository>) provides, Users can download the malignant tumor data they need from the website. Because the speed of the network will determine the smooth degree of downloading data, the system does not provide data downloading function. The downloaded data include RNA sequencing (RNA-seq) data, miRNA data and clinical data^[3]. Once these three types of data have been uploaded to the system, users can proceed with data preprocessing.

*Corresponding author: franknan@126.com

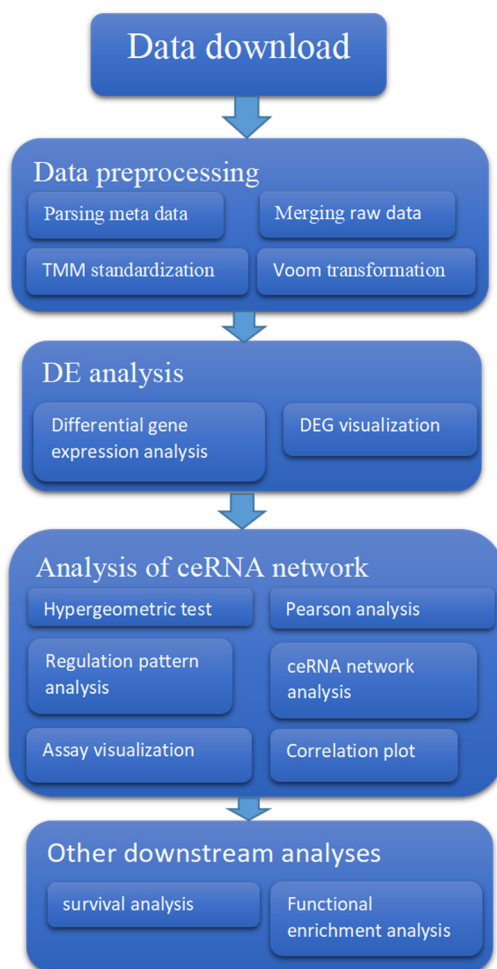


Fig 1. Functional flowchart

3 Introduction to system features and page presentation

3.1 Data preprocessing

In the data preprocessing page, which is divided into two parts function, on the left is to upload data and select the name of the malignant tumor analyzed and on the right is the button for data preprocessing, before performing data analysis, users need to upload data and select cancer names, in the upload data function, users can upload the downloaded corresponding RNA-seq data, miRNA data and clinical data to upload the system, below select the selected cancer name, once the preparation is complete, can preprocess.

Firstly, the parsing of RNA-seq and miRNA metadata was performed, and the patient data with uncertain survival status were filtered out, because the survival status with uncertain data could not provide valid information for the construction of the overall model and could also lead to false results, the filtering out of these data could ensure the authenticity and scientific quality of the model^[4].

Secondly, the merging of RNA-seq, miRNA as well as clinical data is performed, and by clicking on the RNA-

seq data merging and miRNA data merging button users can realize to merge the original technical data of RNA-seq, miRNA into a single expression matrix, and can filter out common clinical information, remove redundant information, and simplify operations.

Finally, the expression matrix was TMM normalized and voom transformed to remove singular type data and transform the data into a format that can be used for differential gene analysis^[5].

3.2 Differential gene expression analysis

In the differential analysis page, the functions of differential analysis button, Heatmap download, data display of differential genes and differential image display are included, and when clicking on the differential expression analysis button, the genes in cancer tissues and normal tissues will be studied in the selected malignant tumor genes of the users, and the differential genes will be filtered according to the criteria of $FDR < 0.01, \log_2|FC| > 2.0$ ^[6]. And the screened differential genes are displayed to the user in the form of a graph, and a volcano plot(Figure.2), bar graph and heat map of the differential genes are drawn. Allows the user a more intuitive view of the differential gene results.

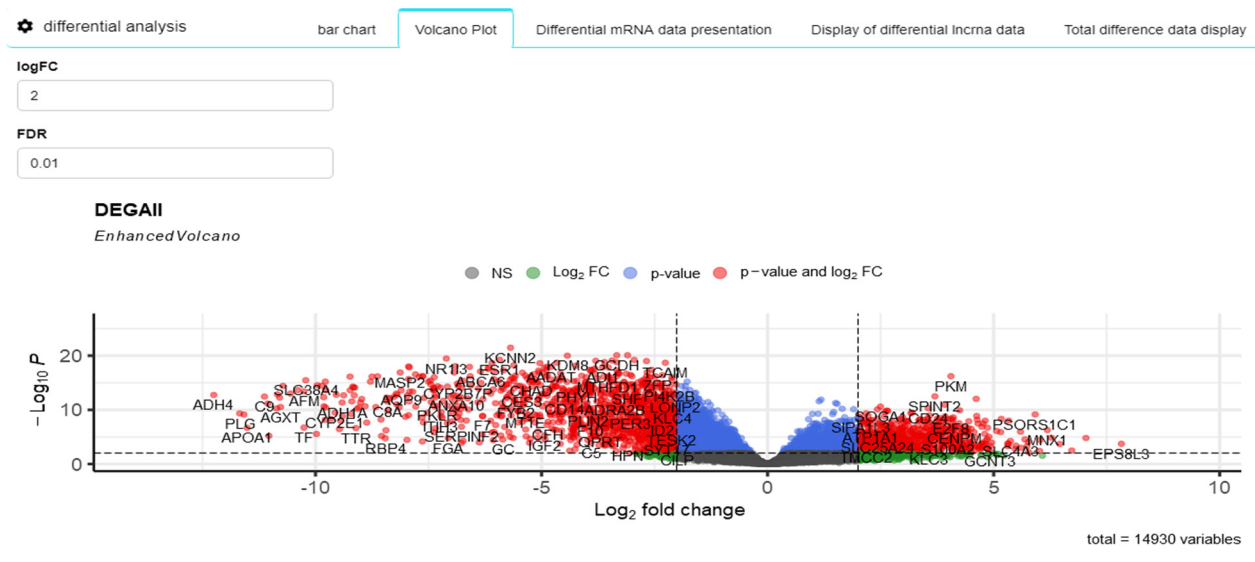


Fig 2. volcano plot

3.3 ceRNA network construction

In the ceRNA network interface, on the left is the operation page, on the right is the display page, the operation page can carry out the ceRNA network analysis button, the value input box of hyperpvalue and corpvalue, edge and node text file download, on the right shows the display of the edge and node files after the

completion of the network construction and the visualization display of a small amount of the already constructed ceRNA network^[7]. Users can control the condition of censoring differential genes by adjusting the coefficients of hyperpvalue and corpvalue, download the edge and node files in the corresponding conditions, and use Cytoscape to draw the constructed ceRNA network.

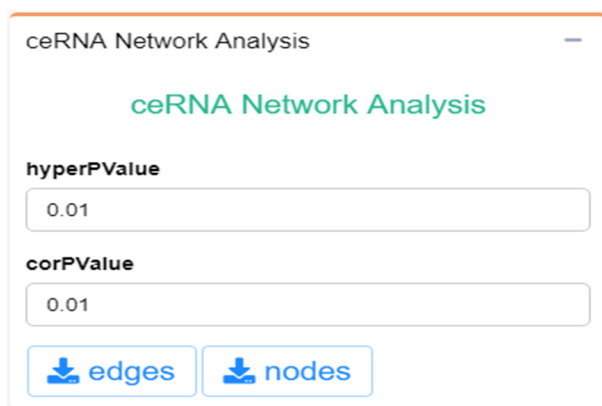


Fig 3. operation page

3.4 Other downstream analyses

In the downstream analysis, mainly univariate survival analysis and functional enrichment analysis were performed, and the two analysis functions can help identify the genes in the ceRNAs network that play an important role in prognosis or are involved in important pathways.

3.4.1 Univariate survival analysis

In the survival analysis interface, users can draw a survival plot(Figure.4) for this gene by selecting different genes, system, apply KM algorithm and median the threshold value to distinguish high expression population from low expression population. And the hazard ratio at the 95% confidence interval can be reported for this gene and the significance of each gene for overall survival can be tested^[8].

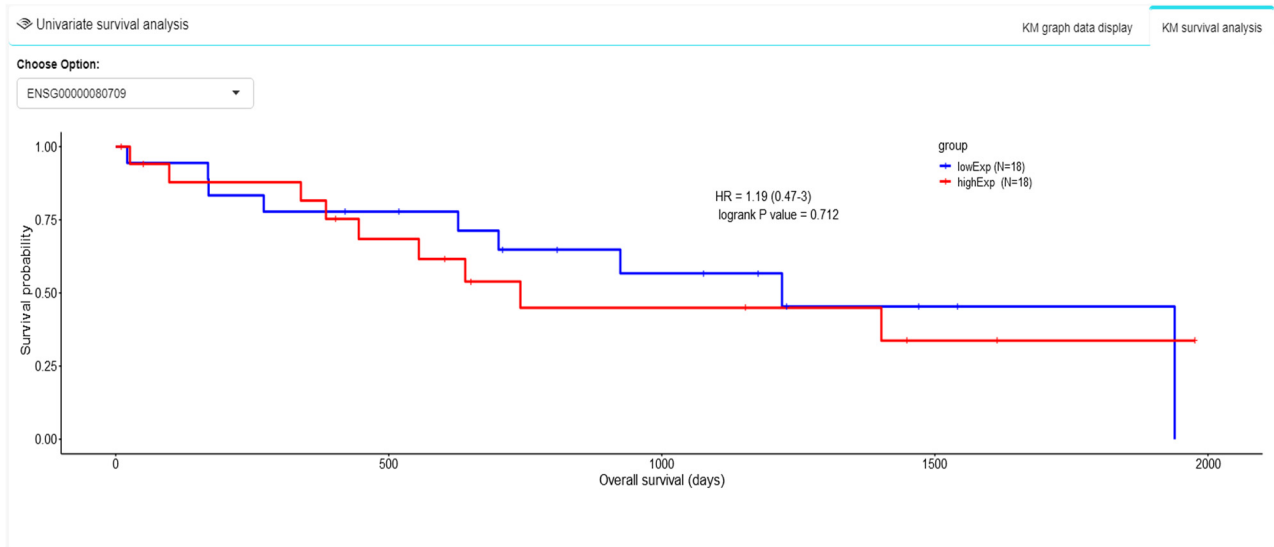


Fig 4. survival plot

3.4.2 Gene enrichment analysis

After the differentially expressed genes obtained from the differential expression analysis were subjected to enrichment analysis, an overview sentence comparing the occurrence of the overall response event could be summed up, such as: something related to a certain signaling pathway and the occurrence of a certain cancer^[9].

Gene enrichment analysis is the identification of the molecular features and the pathways involved in the resulting differentially expressed genes, which is also routinely performed for high-throughput omics data analysis, and `gdcenrichanalysis` function allows gene

ontology (go), Kyoto Encyclopedia of genes and genomes (KEGG), and gene enrichment analysis on a

series of genes simultaneously^[10]. Go enrichment analysis is to utilize the gene annotation information in the go database for analysis, and different genes also have different corresponding *p*-values values, the smaller the *p*-value, indicating the more enriched the gene. Pathway analysis of differential genes was performed by Kyoto Encyclopedia of genes and genomes (KEGG) functional enrichment analysis to understand the significantly altered metabolic pathways under experimental conditions^[11].

The bubble plot(Figure 5) and path charts(Figure 6) were plotted for users in the functional enrichment analysis.

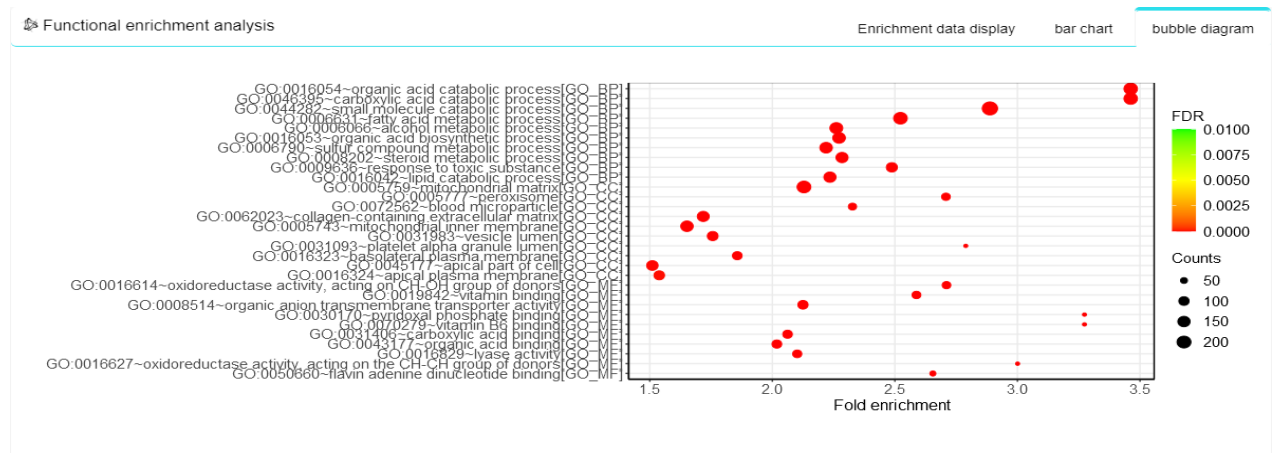


Fig 5. bubble plot

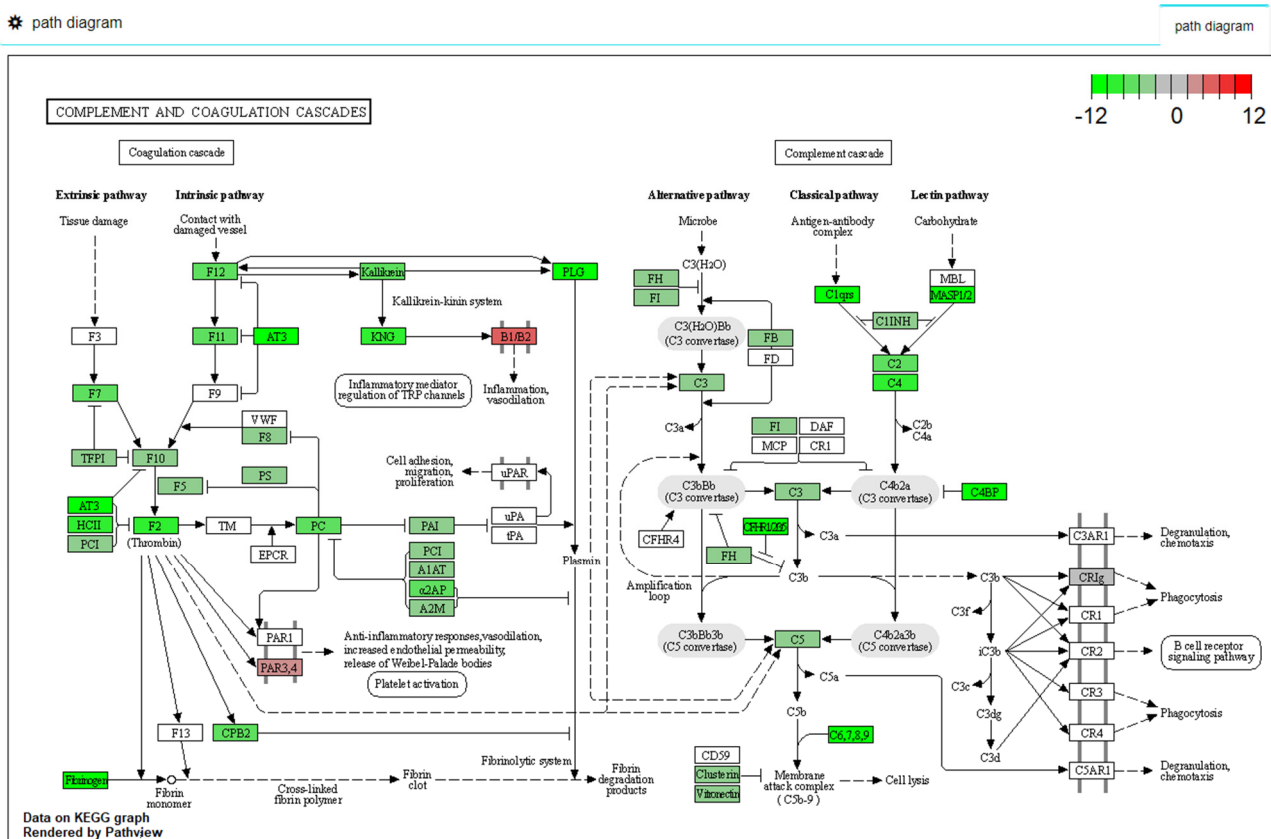


Fig 6. path charts

4 Conclusions

Although the application of information technology in medical is less than other industries, in today's rapidly developing society, with the further development of future computer technology, data mining technology and omics data analysis technology, the use of visual analysis system can greatly assist physicians in clinical decision support, meanwhile, the use of visual analysis system for analysis, after visiting the patient and the phenomenon of seeing precious can also be alleviated to some extent.

In this paper, we developed a system for the ceRNA network construction of malignant tumors based on transcriptome data, through data mining techniques, using RStudio software for trial and model construction, shiny framework package code and Cytoscape for network construction, after survival analysis and enrichment analysis of the differentially expressed genes, it can be found to be significant for the diagnosis and treatment of the studied malignant tumors. The targets and signaling pathways that can make the zero code basic medical staff to complete the network construction of related cancers, complete the purpose of this study.

Acknowledgement

This study was supported by 2019 National College Student Innovation and Entrepreneurship Training Program Project (Approval No. 201914438005), the doctoral initiation fund of Shandong Management University (Approval No. SDMUD201901) and scientific

research pilot program of Shandong Management University (Approval No. QH2020Z05).

References

1. Evolution and Functions of Long Noncoding RNAs[J]. Chris P. Ponting, Peter L. Oliver, Wolf Reik. Cell. 2009 (4).
2. MicroRNA-mRNA interactions in colorectal cancer and their role in tumor progression[J]. Raheleh Amirkhah, Ulf Schmitz, Michael Linnebacher, Olaf Wolkenhauer, Ali Farazmand. Genes Chromosomes Cancer. 2015 (3).
3. Li, Ruidong, Qu, et al. GDCRNATools: an R/Bioconductor package for integrative analysis of lncRNA, miRNA and mRNA data in GDC[J]. Bioinformatics, 2018.
4. Wu Xiongwei, Liu Jie, Zhu Chengzhan, et al. Identification of Potential Biomarkers of Prognosis Related Long Non-Coding RNA (lncRNA) in Pediatric Rhabdoid Tumor of the Kidney Based on ceRNA Networks. 2020, 26
5. H. Atakan Ekiz, Christopher J. Conley, W. Zac Stephens, Ryan M. O'Connell. CIPR: a web-based R/shiny app and R package to annotate cell clusters in single cell RNA sequencing experiments[J]. BMC Bioinformatics, 2020, 21(5).
6. Yiming Yu, Wen Yao, Yuping Wang, Fangfang Huang. shinyChromosome: An R/Shiny Application for Interactive Creation of Non-circular Plots of Whole

- Genomes[J]. *Genomics, Proteomics & Bioinformatics*, 2019,17(5).
7. Salmena L, Poliseno L, Tay Y, et al. A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language?[J]. *Cell*, 2011, 146(3): 353-358.
 8. Tay Y, Rinn J, Pandolfi P P. The multilayered complexity of ceRNA crosstalk and competition[J]. *Nature*, 2014, 505(7483): 344-352.
 9. Li J H, Liu S, Zhou H, et al. starBase v2. 0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data[J]. *Nucleic acids research*, 2014, 42(D1): D92-D97.
 10. Zhou M, Wang X, Shi H, et al. Characterization of long non-coding RNA-associated ceRNA network to reveal potential prognostic lncRNA biomarkers in human ovarian cancer[J]. *Oncotarget*, 2016, 7(11): 12598.
 11. Zhang Y, Xu Y, Feng L, et al. Comprehensive characterization of lncRNA-mRNA related ceRNA network across 12 major cancers[J]. *Oncotarget*, 2016, 7(39): 64148.