

Analysis of Competing Endogenous RNA Network and Prediction of Prognosis in Acute Lymphoblastic Leukemia Patients of Phase II and III

Jianzhi Deng¹, Xiaohui Cheng¹ and Yuehan Zhou^{2,*}

¹Guilin University of Technology, Guilin 541004, China

²Guilin Medical University, Guilin 541004, China

*Corresponding author

Keywords: TARGET, Acute lymphoblastic leukemia, ceRNA, COX model, ROC, Medical big data, Progressive biomarker.

Abstract. In this paper, we focused on the competing endogenous RNA (ceRNA) of lncRNA-mRNA-miRNA regulatory network and the target mRNAs as biomarker predicting progression of acute lymphoblastic leukemia (ALL) based on the Therapeutically Applicable Research to Generate Effective Treatment (TARGET) database. By screening the RNA-seq, 112 differentially expressed lncRNAs (DElncRNAs), 73 differentially expressed miRNAs (DEmiRNAs) and 204 differentially expressed mRNAs (DEmRNAs) were found by edgeR package (adjusted P-value < 0.01 and |LogFoldChange| > 4). Among them, the regulatory of 12 DElncRNAs, 5 DEmRNAs and 15 DEmiRNAs was found for the construction of the ceRNA network. 3 DEmRNAs in the ceRNA network, PRKAA2, FOXF2 and TFAP2C, constructed a COX model for the progressive prediction of ALL and was verified by the receiver operating characteristic (ROC) curve, area under the ROC curve (AUC) and kaplan-meier survival curve in a 5-year analysis.

Introduction

Acute lymphoblastic leukemia (ALL) is a fatal hematological system disease with speedy development and high lethal rate [1, 2]. And with the progression of ALL, some genes would be dysregulated. In this work, we compared the RNA expression between stage II and III, and tried to find out the dysregulate RNAs of ALL and the related ceRNA network based on the downloaded ALL phase II and phase III RNA-seq data from The Therapeutically Applicable Research to Generate Effective Treatment (TARGET) database[3]. DEmRNAs, DEmiRNAs and DElncRNAs between stage III and in stage II were screened. The regulatory of lncRNA-mRNA-miRNA was found by using the online database. Based on the differentially expressed RNAs, the ceRNA network was constructed [4]. COX predicted model of the target mRNAs which were shown in the ceRNA network were analysed with their related clinical data by kaplan-meier method and proved by receiver operating characteristic (ROC) analysis[5, 6].

Materials and Methods

Data Download and Matrix Merging

The data of ALL patients were downloaded from the TARGET website (<http://ocg.cancer.gov/programs/target>) including gene expression and miRNA-seq of phase II, phase III with clinical files. The gene expression and miRNA-seq of each patient was saved as a txt file. And all the clinical data was saved in one excel files. The information of the downloaded files was shown in Table 1.

Firstly, the gene expression data of each file was merged into one single gene expression data matrix file. And the lncRNA expression matrix and mRNA expression matrix were extracted from the gene expression data matrix. Secondly, the miRNA expression matrix was merged by the all the downloaded miRNA-seq files. Thirdly, 192 clinical data of the patients sample was selected according the downloaded gene expression files.

Table 1. Information of the downloaded files

Gene data	Clinical number	gene number	Sample number
Gene Expression	192	lncRNA 12067	Stage II 216
		mRNA 19082	Stage III 96
miRNA-seq	192	miRNA 1871	Stage II 231
			Stage III 39

DEGs Screening

DElncRNAs, DEmRNAs and DEmiRNAs between Phase II ALL patients and Phase III ALL patients were selected by edgeR R package with the cutoff of adjusted P-value < 0.01 and |LogFoldChange| > 4 [7, 8]. The DElncRNAs, DEmRNAs and DEmiRNAs were the RNA that expressed higher or lower in the phase III samples rather than the phase II samples. And the DElncRNAs, DEmRNAs and DEmiRNAs were used for the analysis of ceRNA network.

ceRNA Network Analysis

By searching the mircode database (www.mircode.org), the regulatory between lncRNA and miRNA was found. And starBase database (starbase.sysu.edu.cn) was used to search the 3'-UTR and 5'-UTR of DEmiRNAs. Target mRNAs of the DEmiRNAs were screened that both in miRDB (www.mirdb.org), miRTarBase (mirtarbase.mbc.nctu.edu.tw) and TargetScan (www.targetscan.org). After combining the DEGs and their dys-regulated state, the regulatory of lncRNA-mRNA-miRNA was found. And the ceRNA network was also built by cytoscape software [9].

ROC and Survival Analysis

According to the target mRNA in the ceRNA network, the related survival analysis and ROC analysis were done [10]. The ROC curves and AUC were analyzed by using the multivariable COX regression model firstly. In the multivariable COX regression model, only one of the co-expression DEmRNAs would kept when others were regarded as redundancy. The survival KM-plot curves based on the target mRNA was drawn by survival R package under kaplan-meier method[11, 12].

Results

DEGs Screening

After screening the merged gene expression or miRNA-seq matrix files, 112 DElncRNAs, 73 DEmiRNAs and 204 DEmRNAs were found by edgeR package with the cutoff of adjusted P-value < 0.01 and |LogFoldChange| > 4.

ceRNA Network

12 DElncRNAs were found interacted with 15 DEmiRNAs in 56 regulation pairs when analyzed the 112 DElncRNAs and 73 DEmiRNAs screened above by searching the mircode database. And 5 of the 204 DEmRNAs which simultaneously found in miRDB, miRTarBase and TargetScan, were the target mRNAs of the 15 DEmiRNAs. CeRNA network of the 12 DElncRNAs, 5 DEmRNAs and 15 DEmiRNAs regulatory was shown in Figure X. The differentially expressed genes in ceRNA network were listed in Table 2. PRKAA2, BDNF, FBXL7, FOXF2 and TFAP2C in Figure.1 were the target of 5 DEmiRNAs. And the relationships were listed in Table 3.

Table 2. Differentially expressed genes in ceRNA network

Type	Gene name
lncRNA	KIAA0087, C20orf166-AS1, C11orf44, C1orf132, ZNF385D-AS1, NAALADL2-AS2, ENOX1-AS1, BCYRN1, SNHG6, SNHG9, CACNA1C-IT3, DIO3OS
miRNA	hsa-mir-1271, hsa-mir-182, hsa-mir-183, hsa-mir-214, hsa-mir-217, hsa-mir-4465, hsa-mir-383, hsa-mir-372, hsa-mir-22, hsa-mir-192, hsa-mir-215, hsa-mir-206, hsa-mir-216a, hsa-mir-137, hsa-mir-216b
mRNA	PRKAA2, BDNF, FBXL7, FOXF2, TFAP2C

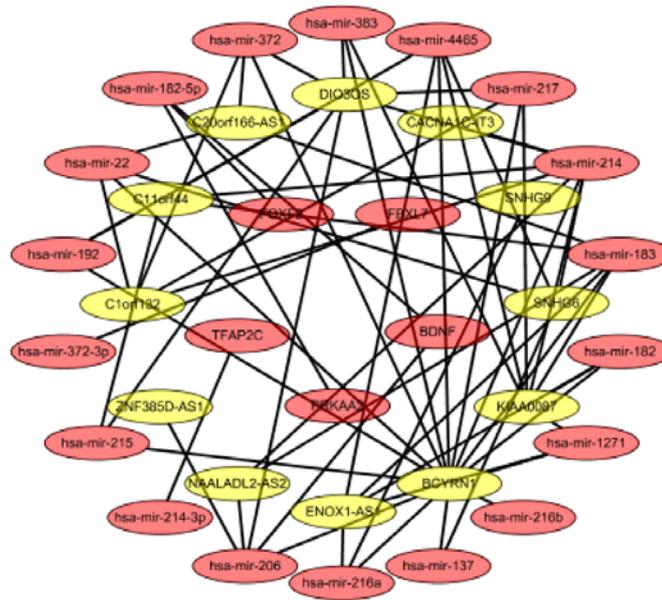


Figure 1. ceRNA network

Table 3. miRNAs and target mRNAs

miRNA	Target mRNA
hsa-mir-182-5p	PRKAA2, BDNF, FOXF2
hsa-mir-372-3p	FBXL7
hsa-mir-214-3p	TFAP2C
hsa-mir-206	BDNF

In the COX model, PRKAA2, FOXF2 and TFAP2C were analysed after excluding the redundantly co-expressed mRNAs, BDNF and FBXL7. And the value of the COX analysis was listed in Table 4.

Table 4. The COX model value of the ROC analysis

gene	coef	z	Pr(> z)
TFAP2C	0.150079	3.643013	0.000269
PRKAA2	-0.13583	-2.20529	0.027434
FOXF2	0.104932	1.744582	0.081058

ROC and Survival Analysis

The ROC analysis was drawn based on the target mRNA in tableX. And the ROC curve and AUC value(AUC=0.627) was shown in Figure 2. The survival data was analysed by kaplan-meier method. And the survival kmplot curve was drawn in Figure 3(P=2.014e-02). Compared with the phase II, it was 20% higher risk in 5 years survival time.

Discussion

ALL is a highly fatal disease of the blood system. In this paper, we aimed to find out the mRNA biomarker predicting progression of ALL and the ceRNA network of DElncRNAs, DEmRNAs and DEMiRNAs. Our research was based on the RNA-seq files downloaded from TARGET database. 112 DElncRNAs, 204 DEmRNAs and 73 DEMiRNAs between the phase II and phase III were screened from the RNA expression matrix of over 300 ALL patients' samples by the edgeR R package (adjusted P-value < 0.01 and |LogFoldChange| > 4).

In order to construct the ceRNA network, 12 DElncRNAs were interacted with 15 DEMiRNAs in 56 regulation pairs by searching the mircode database. Meanwhile, 5 DEmRNAs were found as the

target mRNAs of the screened DEmiRNAs. And the regulatory in the ceRNA network was made by the 12 DElncRNAs, 5 DEemRNAs and 15 DEmiRNAs.

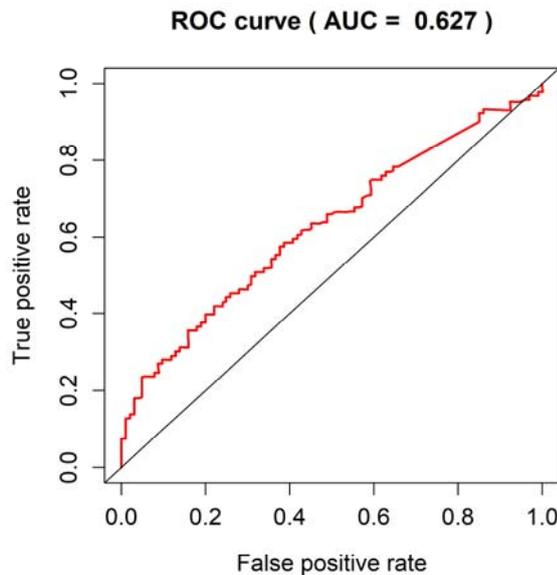


Figure 2. ROC curve

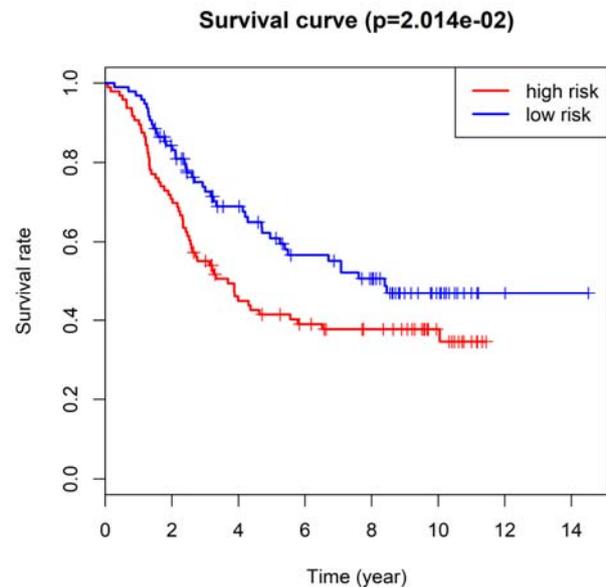


Figure 3. Survival curve

For further research, the 5 target mRNAs (PRKAA2, BDNF, FBXL7, FOXF2 and TFAP2C) in the ceRNA network were combined with the clinical data and analysed by the multivariable COX regression model. In a 5-year survival analysis, the ROC curve with an AUC of 0.627 and a survival curve with p value of 2.014e-02 proved that the COX model of PRKAA2, FOXF2 and TFAP2C was effective. These target mRNAs related to the progress of ALL could be the progressive biomarker of ALL.

Acknowledgments

This work was supported by National Natural Science Foundation of China (No. 81660031, 81360090), Guangxi Natural Science Foundation of China (2014GXNSFBA118151), and Guangxi key Laboratory Foundation of Embedded Technology and Intelligent System.

References

- [1] Kaymak Cihan, M., Karabulut, H.G., Yurur Kutlay, N., Ilgin Ruhi, H., Tukun, A., and Olcay, L.: 'Association Between N363S and BclI Polymorphisms of the Glucocorticoid Receptor Gene (NR3C1) and Glucocorticoid Side Effects During Childhood Acute Lymphoblastic Leukemia Treatment', *Turk J Haematol*, 2017, 34, (2), pp. 151-158
- [2] Li, S., Bian, H., Cao, Y., Juan, C., Cao, Q., Zhou, G., and Fang, Y.: 'Identification of novel lncRNAs involved in the pathogenesis of childhood acute lymphoblastic leukemia', *Oncol Lett*, 2019, 17, (2), pp. 2081-2090
- [3] Newton, J.G., Horan, J.T., Newman, S., Rossi, M.R., Ketterling, R.P., and Park, S.I.: 'CD36-positive B-lymphoblasts Predict Poor Outcome in Children With B-lymphoblastic Leukemia', *Pediatr Dev Pathol*, 2017, 20, (3), pp. 224-231
- [4] Salmena, L., Poliseno, L., Tay, Y., Kats, L., and Pandolfi, P.P.: 'A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language?', *Cell*, 2011, 146, (3), pp. 353-358
- [5] Olaso, V., Cordoba, J., Molina, J.M., Lopez, B., Siles, M.S., Monte, E., Calvo, F., Bau, I., Beltran, B., Martinez, B., and Gobernado, M.: '[ROC curve analysis of factors predictive of response to

- treatment with interferon plus ribavirin in patients with chronic hepatitis C relapse after previous interferon treatment]’, *Rev Esp Quimioter*, 2002, 15, (4), pp. 366-373
- [6] Lukasiewicz, E., Gorfine, M., Neumann, A.U., and Freedman, L.S.: ‘Combining longitudinal discriminant analysis and partial area under the ROC curve to predict non-response to treatment for hepatitis C virus’, *Stat Methods Med Res*, 2011, 20, (3), pp. 275-289
- [7] Jianzhi Deng, X.C., Yuehan Zhou: ‘Gene clustering, enrichment and survival analysis of differentially expressed genes in Low Grade Glioma between different genders by big data analysis’. *Proc. ICDMML2019, Hong Kong, 2019.4.28 2019 pp. Pages*
- [8] Law, C.W., Alhamdoosh, M., Su, S., Dong, X., Tian, L., Smyth, G.K., and Ritchie, M.E.: ‘RNA-seq analysis is easy as 1-2-3 with limma, Glimma and edgeR’, *F1000Res*, 2016, 5
- [9] Jianzhi Deng, M.Y., Xiaohui Cheng, Yuehan Zhou: ‘PRLR is a target of oncogenic miR-204 in Low Grade Glioma and involved in a ceRNA network by Computational Biology’. *Proc. AMMSO2019, guilin, 2019.4.22 2019 pp. Pages*
- [10] Liao, P., Wu, H., and Yu, T.: ‘ROC Curve Analysis in the Presence of Imperfect Reference Standards’, *Stat Biosci*, 2017, 9, (1), pp. 91-104
- [11] Goel, M.K., Khanna, P., and Kishore, J.: ‘Understanding survival analysis: Kaplan-Meier estimate’, *Int J Ayurveda Res*, 2010, 1, (4), pp. 274-278
- [12] Chuang, S.K., Tian, L., Wei, L.J., and Dodson, T.B.: ‘Kaplan-Meier analysis of dental implant survival: a strategy for estimating survival with clustered observations’, *J Dent Res*, 2001, 80, (11), pp. 2016-2020