

Screening and Analysis of lncRNA Biomarkers and the Relationship with the Corresponding Regulatory miRNA for Progressive Acute Lymphoblastic Leukemia

Jianzhi Deng¹, Yuehan Zhou^{2,*}, Xiaohui Cheng¹ and Mingjun Xu¹

¹Guangxi key Laboratory of Embedded Technology and Intelligent System, Guilin University of Technology, Guilin 541004, China

²College of Pharmacy, Guilin Medical University, Guilin 541004, China

*Corresponding author

Keywords: TARGET, Acute lymphoblastic leukaemia, ROC, Medical big data analysis, C20orf166-AS1, Progressive biomarker.

Abstract. In this study, we aim to find out the lncRNA biomarker of progressive acute lymphoblastic leukaemia (ALL). After the comparison of the phase II and phase III gene expression downloaded from Therapeutically Applicable Research to Generate Effective Treatment (TARGET) online database, 112 differentially expressed lncRNAs and 73 differentially expressed miRNAs were screened by edge R package. The regulatory of lncRNA-miRNA was found based on the screened differentially expressed gene by using the mircode database. C20orf166-AS1 was analyzed as a new progressive biomarker by comparing the gene expression in different tumors by cancer cell line encyclopedia website. And the impact on ALL progression was proven by the receiver operating characteristic (ROC) curve and kmplot survival curve.

Introduction

Acute lymphoblastic leukaemia (ALL) is a complicated hematological malignant disease. With the progression of this disease, some gene expression would be out of regulation. It had already known that there was regulatory between some lncRNAs and miRNAs. In this paper, we aimed to find out the differentially expressed genes (DEGs) in ALL progression and whether the DEGs were in a regulative relationship. The differentially expressed genes which downloaded from Therapeutically Applicable Research to Generate Effective Treatment (TARGET) database website were screened by edgeR package [1-3]. Regulatory of the differentially expressed lncRNAs (DElncRNAs) and differentially expressed miRNAs (DEmiRNAs) were found by mircode database. And the DElncRNA in the regulatory network were considered to be lncRNA biomarkers of progressive ALL. It was analyzed by cancer cell line encyclopedia (CCLE) and analyzed by survival R package [4]. The receiver operating characteristic (ROC) curve, the area under the ROC curve (AUC) and kmplot survival curve were analyzed based on the selected lncRNA biomarker [5, 6].

Materials and Methods

lncRNA Matrix and miRNA Matrix

In order to find out the relationship between the typical lncRNA expression and the ALL progression, differentially expressed lncRNAs and differentially expressed miRNAs were screened based on the comparison of ALL phase II and III RNA-seq matrix that downloaded from the online TARGET database. The two phase gene expression data of ALL patients' sample were downloaded from TARGET () that was saved as txt files. The lncRNA expression files were merged into one matrix, and so does the miRNA expression matrix. The relative clinical data was extracted from the clinical data excel file. Information of the downloaded gene data are shown in Table I.

Table 1. Information of the downloaded gene data

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

DEGs Screening and Regulatory of lncRNA-miRNA Analysis

DElncRNAs, and DEmiRNAs were screened by using the edgeR package and R software according to the threshold values: adjusted P-value < 0.01 and $|\log\text{FoldChange}| > 4$ [7]. Based on the DElncRNAs and DEmiRNAs, the regulatory of lncRNA-miRNA was selected by mircode database (www.mircode.org) [8, 9]. The DElncRNAs which were found in mircode database would be regarded as biomarkers for progressive ALL in the further research [1, 10].

Expression and Progression Analysis for the Biomarkers

Firstly, the expression of biomarker found in the regulatory network was analyzed by CCLE (<https://portals.broadinstitute.org/ccle>). And then, it was analyzed by the survival R package in a 5 years ROC curve AUC analysis and Kmplot survival analysis [11-13].

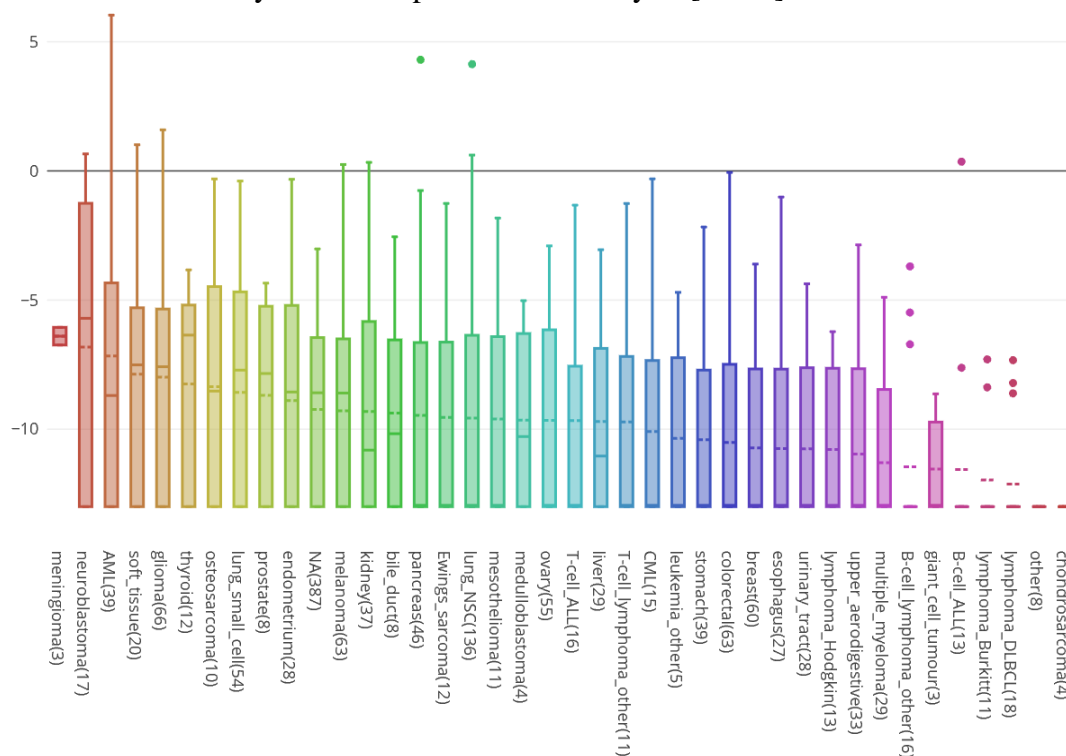


Figure 1. C20orf166-AS1 expression in tumors

Results

Compared the lncRNA and miRNA expression in phase III and phase II, 112 DElncRNAs and 73 DEmiRNAs were screened out. Among them, 21 of the DElncRNAs were high-expressed while the rest DElncRNAs were low-expressed. Besides, there were 17 high-regulated DEmiRNAs and 56 low-expressed DEmiRNAs. The top 20 significant high- or low-expressed DElncRNAs and DEmiRNAs were listed in table II.

Table 2. Differentially expressed gene

gene	expression	Gene symbol
lncRNA	high-	RP11-704M14.1, RP5-991C6.4, CTC-493L21.2, CTC-493L21.1, C20orf166-AS1, KIAA0087, RP11-99A1.2, AC013402.5, CTD-2034I4.1, PLCB1-IT1, RP11-374A22.1, AC132807.1, CACNA1C-IT3, HOXA11-AS, CTD-2034I4.2, ARHGAP22-IT1, RP11-328K2.1, MIR646HG, RP11-121P10.1, C1orf132
	low-	RP11-217E13.1, RP11-470P21.2, CTD-2337I7.1, RP3-326I13.1, D21S2088E, RP11-615I2.1, AF131216.5, RP11-96H17.1, LINC00681, RP11-368L12.1, ENOX1-AS1, RP1-297M16.2, RP11-396N11.1, RP11-717H13.1, LINC00839, RP11-359E19.2, RP11-252P19.1, AP005530.2, LINC01289, CTC-457E21.1
miRNA	high-	hsa-mir-513b, hsa-mir-513a-2, hsa-mir-513a-1, hsa-mir-513c, hsa-mir-214, hsa-mir-137, hsa-mir-1-1, hsa-mir-1-2, hsa-mir-135a-2, hsa-mir-4518, hsa-mir-876, hsa-mir-135a-1, hsa-mir-548a-3, hsa-mir-199a-1, hsa-mir-196a-2, hsa-mir-199a-2, hsa-mir-450a-2
	low-	hsa-mir-92a-2, hsa-mir-124-1, hsa-mir-1266, hsa-mir-892a, hsa-mir-182, hsa-mir-217, hsa-mir-10b, hsa-mir-99b, hsa-mir-183, hsa-mir-891a, hsa-mir-891b, hsa-mir-105-2, hsa-mir-383, hsa-mir-124-2, hsa-mir-4491, hsa-mir-1247, hsa-mir-767, hsa-mir-892b, hsa-mir-4465, hsa-mir-5685

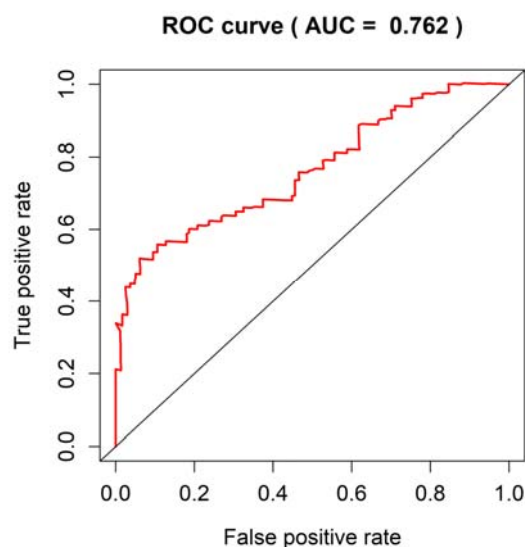


Figure 2. The ROC curve of C20orf166-AS1

In mircode database, 15 miRNAs were found in a relative regulatory with the DElncRNAs. As one of the most significant dys-regulated lncRNAs, C20orf166-AS1, was regulated with the dys-regulated miRNAs, hsa-miR-372, hsa-miR-183 and hsa-miR-22 in this study. By analyzed in CCLE, C20orf166-AS1 was down-regulated in all kinds of ALL than other tumors.

Hsa-miR-372, hsa-miR-183 and hsa-miR-22 were 3 miRNAs that had already known in cancer research. Hsa-miR-183 and has-miR-22 were involved in ovarian cancer and acute myeloid leukaemia [14-16]. Has-miR-372 was found in the prognostic of Colorectal Cancer, hepatocellular carcinoma and lung cancer [17-19].

In a 5-year-survival analysis, the ROC curve and AUC analysis was done based on C20orf166-AS1. In the ROC curve drawn in Figure II, the AUC value is 0.762. The true positive rate was higher than the false positive rate in total. The clinical survival data of C20orf166-AS1 were merged and drawn a kmplot curve by using kaplan-meier method. The KM survival curve ($p=2.514e-06$) in Figure III showed that it was at a higher risk in phase III than phase II. In 5-year-survival time, the survival curve was about 30% higher risk in the high expression group.

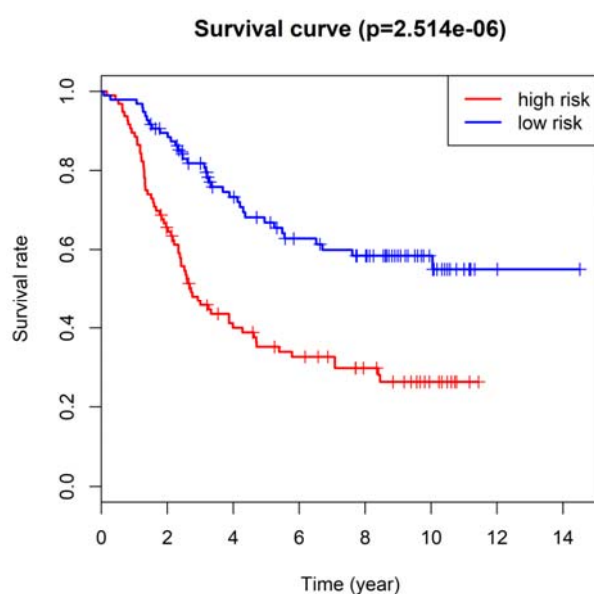


Figure 3. The survival curve of C20orf166-AS1

Discussion

As a complicated hematological malignant disease, the cause of ALL is still unclear, and it progresses rapidly without effective treatment. From previous studies, it is known that the lncRNAs and miRNAs were in a close regulatory network in some case. And the regulatory can be used to slow down or even cure the disease. In this study, in order to find out the lncRNA biomarkers for progressive ALL and its regulatory of lncRNA-miRNA, first of all, the lncRNAs and miRNAs expression of the ALL patients were downloaded from the TAGERT database. Compared patients at the phase III and phase II, DElncRNAs and DEmiRNAs were screened secondly. By using the mircode database, the regulatory of DElncRNA-DEmiRNA was found. The dys-regulated lncRNA, C20orf166-AS1 was regulated with the dys-regulated miRNAs, hsa-mir-372, hsa-mir-183 and hsa-mir-22.

In the further research, C20orf166-AS1 was low expressed in ALL than other tumors by analysed in CCLE. After that, the survival R package was used to do the ROC curve and AUC analysis and kmplot survival curve analysis in a 5 years survival time. The expression of C20orf166-AS1 in a high level was higher risk than the low level ones that shown in Figure 3 ($p=2.514e-06$). And the ROC curve with a high AUC ($AUC=0.762$) provide a strong proof for our research.

From the study above, it can be known that C20orf166-AS1 would be a progressive biomarker of ALL. And it might be used to cure the disease by regulate the relative regulatory miRNAs, hsa-miR-372, hsa-miR-183 and hsa-miR-22.

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] Jian-zhi DENG, X.-h.C.a.Y.-h.Z., *Analysis of Competing endogenous RNA network and prediction of prognosis in acute lymphoblastic leukemia patients of Phase II and III*, in *2019 International Conference on Modeling, Simulation and Big Data Analysis*. 2019: wuhan.
- [2] Li, X., et al., *Gene silencing of MIR22 in acute lymphoblastic leukaemia involves histone modifications independent of promoter DNA methylation*. Br J Haematol, 2010. **148**(1): p. 69-79.
- [3] Newton, J.G., et al., *CD36-positive B-lymphoblasts Predict Poor Outcome in Children With B-lymphoblastic Leukemia*. Pediatr Dev Pathol, 2017. **20**(3): p. 224-231.
- [4] Barretina, J., et al., *Addendum: The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity*. Nature, 2019. **565**(7738): p. E5-E6.
- [5] Blanche, P., J.F. Dartigues, and H. Jacqmin-Gadda, *Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks*. Stat Med, 2013. **32**(30): p. 5381-97.
- [6] Liao, P., H. Wu, and T. Yu, *ROC Curve Analysis in the Presence of Imperfect Reference Standards*. Stat Biosci, 2017. **9**(1): p. 91-104.
- [7] Law, C.W., et al., *RNA-seq analysis is easy as 1-2-3 with limma, Glimma and edgeR*. F1000Res, 2016. **5**.
- [8] Jeggari, A., D.S. Marks, and E. Larsson, *miRcode: a map of putative microRNA target sites in the long non-coding transcriptome*. Bioinformatics, 2012. **28**(15): p. 2062-3.
- [9] Song, W., D.L. Miao, and L. Chen, *Comprehensive analysis of long noncoding RNA-associated competing endogenous RNA network in cholangiocarcinoma*. Biochem Biophys Res Commun, 2018. **506**(4): p. 1004-1012.
- [10] 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*, in *AMMSO2019*. 2019: guilin.
- [11] Ziegler, A., S. Lange, and R. Bender, *[Survival analysis: properties and Kaplan-Meier method]*. Dtsch Med Wochenschr, 2007. **132 Suppl 1**: p. e36-8.
- [12] Barretina, J., et al., *The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity*. Nature, 2012. **483**(7391): p. 603-7.
- [13] Gribskov, M. and N.L. Robinson, *Use of receiver operating characteristic (ROC) analysis to evaluate sequence matching*. Comput Chem, 1996. **20**(1): p. 25-33.
- [14] Jiang, X., et al., *miR-22 has a potent anti-tumour role with therapeutic potential in acute myeloid leukaemia*. Nature Communications, 2016. **7**: p. 11452.
- [15] Zheng, Z., et al., *miR-183-5p Inhibits Occurrence and Progression of Acute Myeloid Leukemia*

via Targeting Erbin. Mol Ther, 2019. **27**(3): p. 542-558.

[16] Jiang, X., et al., *miR-22 has a potent anti-tumour role with therapeutic potential in acute myeloid leukaemia*. Nat Commun, 2016. **7**: p. 11452.

[17] Wang, Q., et al., *MiR-372-3p promotes cell growth and metastasis by targeting FGF9 in lung squamous cell carcinoma*. Cancer Med, 2017. **6**(6): p. 1323-1330.

[18] Wu, G., et al., *Low mir-372 expression correlates with poor prognosis and tumor metastasis in hepatocellular carcinoma*. BMC Cancer, 2015. **15**: p. 182.

[19] Yu, J., et al., *Serum miR-372 is a Diagnostic and Prognostic Biomarker in Patients with Early Colorectal Cancer*. Anticancer Agents Med Chem, 2016. **16**(4): p. 424-31.