

Identification of prognostic risk factors of acute lymphoblastic leukemia based on mRNA expression profiling

C. LI^{1,2,*}, L. KUANG^{3,4,‡}, B. ZHU^{1,2}, J. CHEN^{1,2}, X. WANG^{1,2}, X. HUANG^{1,2,*}

¹Department of Hematology, Sichuan Academy of Medical Sciences & Sichuan Provincial People's Hospital, Chengdu, Sichuan, People's Republic of China; ²Department of Hematology, Affiliated Medical School of University of Electronic Science and Technology, Chengdu, Sichuan, People's Republic of China; ³Nursing School, Zunyi Medical University, Zunyi, Guizhou, People's Republic of China; ⁴Nephrology Division, Sichuan Academy of Medical Sciences & Sichuan Provincial People's Hospital, Chengdu, Sichuan, People's Republic of China

*Correspondence: huanghhhs@qq.com

‡Contributed equally to this work.

Received July 19, 2016 / Accepted December 12, 2016

We aim to identify prognosis risk factors in acute lymphoblastic leukemia (ALL). mRNA microarray data of adult ALL patients were downloaded from TCGA database, whose mRNAs were isolated from bone marrow aspirate fluid mononuclear cells. Then the differentially expressed genes (DEGs) between good and poor prognosis samples were screened. Following that, the sample dependency network was constructed based on the Pearson connection coefficients of DEGs in the samples. The prognosis-related genes were collected using logistic regression analysis. A classifier for predict the prognosis of ALL patients was established, which was validated in another independent dataset GSE13280 including 173 ALL samples. A total of 578 down-regulated and 637 up-regulated DEGs for worse prognosis were identified. A sample dependency network was established, comprising 100 samples combined by 246 lines. 13 prognosis-related genes were selected to constructed the prognosis classification model, which had an overall precision of 82.7% on distinguishing prognosis status of ALL patients. Total 4 genes were found as the prognosis risk factors in predicting the prognosis of ALL samples, including *ALPK1*, *ACTN4*, *CALR*, and *ZNF695*. *ALPK1*, *ACTN4*, *CALR*, and *ZNF695* were identified as the potential prognosis risk factors in adult ALL.

Key words: acute lymphoblastic leukemia, genes, prognosis, bioinformatics analysis

Acute lymphoblastic leukemia (ALL) is the most common childhood cancer and a major cause of illness and death in adults [1]. The causation of ALL is the cooperation of various oncogenic lesions in T and B progenitor cells, includes the lesions that impact the capacity of unlimited self-renewal and lesions that lead to stage-specific development arrest [2, 3]. With modified therapies based on patient pharmacogenomics and pharmacodynamics, the 5-year survival rate of ALL has increased significantly over the past 10 to 15 years, and it is about 87.5% for patients younger than 15 years and 61.1% for adolescent patients [4, 5]. Even though most patients can be cured, innovative approaches are still needed to gain further improvement in prognosis while reducing adverse effects.

Recent genome-wide study on leukaemic cell DNA profiles have identified certain genetic and sequence mutations that influence the responsiveness to treatment and mutations that serve as the therapeutic targets and prognosis markers for ALL

patients. For instance, the co-inheritance of *IKZF1*, *ARID5B*, *CEBPE*, and *CDKN2A* are found to be the influence factor for childhood ALL, and it may impose an effect on the response to treatment [6, 7]. *IKZF1* alternation is associated with the worse prognosis in Down syndrome ALL and BCR-ABL1 positive and negative ALL [8-10]. The methylation of p21 and p57 were found to be associated with dismal outcome in ALL [11, 12]. Deletion of *CDKN2* is a predictor for poor prognosis in the prognosis of adult B-lineage ALL patients [13]. *Ezh2* controls the survival signaling in early T cell precursor ALL [14]. High expression of *CREB* is associated with a lower overall survival in adult ALL patients [15].

Genomic classifier identified via microarray data analysis can provide prognostic and predictive insight into the clinical features of diseases [16]. A three-gene classifier has been established to stratify early-stage non-small-cell lung cancer patients with significantly different prognosis [17]. Another

three-gene expression signature showed the ability to predict the survival of squamous cell carcinoma of lung in early stage [18]. Zhu *et al* found a predictive gene signature for adjuvant chemotherapy in resected lung cancer [19]. Expression profiling of leukemic blasts can accurately identify the known prognostic subtypes of ALL [20].

In this study, the microarray data of ALL patients were downloaded to identify the prognosis risk factors, the accuracy and precision of identified gene signature was then verified on another independent dataset. Survival analysis was also performed to test the prognostic value of the identified genes.

Material and methods

Microarray data. A total of 173 microarray data of bone marrow aspirate fluid mononuclear cells from adult acute lymphocytic leukemia patients were downloaded from TCGA database (<http://cancergenome.nih.gov/>). These data were divided into two groups according to the survival status of the patients: good prognosis group (survival time longer than 3 months, N=59) and poor prognosis group (survival time shorter than 3 months, N=114). Next, the level 3 data were processed based on the following platform: TCGA_LAML_exp_HiSeqV2_PANCAN-2014-08-22. After the log transformation, the data were normalized using the Z score method [21].

Identification of DEGs. The differentially expressed genes (DEGs) between good and poor expression group were identified using Limma method [22], the threshold was $p < 0.05$ and $\log |\text{FC}(\text{fold change})| > 1$. To test the classification effects of DEGs on good and poor prognosis samples, hierarchical clustering analysis was conducted using Cluster v3.0 and visualized by TreeView [23].

Dependency network construction. To explore the expression differences and connections between good and poor prognosis samples, the Pearson correlation coefficients between them were calculated. The threshold for significant correlation was $|p| < 0.05$. Next, these correlations were visualized via cytoscape software as the sample dependency network [24].

Enrichment analysis. To investigate the underlying mechanism of the prognosis in ALL, the function and pathway enrichment of the DEGs were conducted using molecule annotation system V3.0. FDR (false discovery rate, the adjust P) < 0.05 was the cut-off for significant terms. Genes in the identified significant terms were conducted logistic regression analysis, and genes with $p < 0.05$ were recognized as the prognosis-related genes.

Prognosis classifier construction. Prognosis classifier was constructed using the prognosis-related genes using random forest algorithm under following parameters: the number of trees in the forest, estimator=1000; criterion, the entropy gain algorithm [25]. The classifier was trained using 172 samples, and verified using the other 1 sample to test its precision. To further explore the prediction effect of the prognosis-related

genes on the prognosis of ALL patients, microarray data GSE13280 were downloaded from GEO database, which includes 44 samples of bone marrow aspirate fluid mononuclear cells from ALL patients. Total 22 responsive samples (patients responded to early treatment of IR-induced DNA damage) and 22 resistant samples (patients did not response to IR-induced DNA damage) were eligible.

Survival analysis. The selected prognosis-related genes were used to conduct survival analysis on the downloaded RNA seq data of 179 ALL patients from TCGA. Firstly, self-regression method was applied to identify the combination values of the 13 prognosis-related genes. In self-regression process, genes were randomly removed in the iterations of the overall genes, and the contribution value (coefficient) was obtained by comparing the original p values of genes. Next, the top 5 genes with higher coefficients were applied in combination with the prognosis status of samples downloaded to conduct survival analysis.

Results

Classification effect of DEGs. To identify the genes may associated with prognosis of the samples, DEGs between good and poor expression ALL samples were screened using Limma method. A total of 578 down-regulated and 637 up-regulated genes were identified (Figure 1). Next, clustering analysis were conducted to explore the distinguish effect of the DEGs to all samples with different prognosis status. The results revealed that all 173 samples were clustered into 3 types (displayed as orange, blue and purple in Figure 2), and samples with similar prognosis status were in the same cluster. However, some good prognosis samples exhibited similar gene expression levels as the poor prognosis samples, indicating a high risk of recurrence.

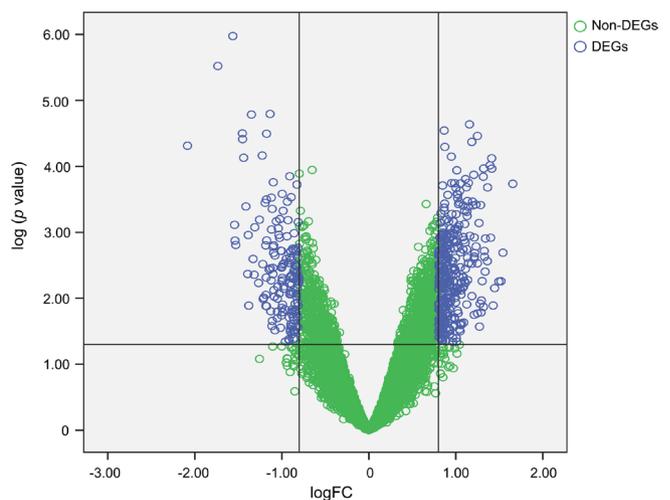


Figure 1. Distribution of DEGs (differentially expressed genes). X-axis represents $\log |\text{FC}(\text{fold change})|$ of gene expressions, y-axis represents $\log p$ values. DEGs are marked in blue, while non-DEGs are marked in green.

Table 1. Significant functional terms of up-regulated genes

GO Term	Count	<i>p</i> value	FDR
GO:0002474 antigen processing and presentation via MHC class I	29	9.15E-58	1.42E-56
GO:0044419 interspecies interaction between organisms	35	7.23E-51	7.88E-50
GO:0019882 antigen processing and presentation	29	2.23E-49	2.21E-48
GO:0006355 regulation of transcription, DNA-dependent	58	2.26E-45	1.45E-44
GO:0006350 transcription	50	2.50E-34	1.36E-33
GO:0006955 immune response	33	5.11E-32	2.53E-31
GO:0055114 oxidation reduction	28	7.46E-30	3.39E-29
GO:0006468 protein amino acid phosphorylation	18	1.98E-15	5.52E-15
GO:0015031 protein transport	18	1.29E-14	3.51E-14
GO:0007165 signal transduction	32	7.44E-13	1.89E-12

Sample dependency network. When comparing patients with different prognosis, not only the gene expression differences, but also the phenotypes of patients should be considered. Normally, patients with similar phenotypes exhibit similar gene expression patterns, that is, positive connection. However, patients with different phenotypes

are not always exhibiting different gene expressions. To find out the connections between patient prognosis and gene expressions, Pearson correlation coefficient between samples at gene expression levels were calculated. In the dependency network of significantly connected samples (Figure 3), there were 100 nodes (samples) and 246 lines (connections). Most samples with similar prognosis exhibited positive connections, while gene expression levels of different prognosis samples differed remarkably. However, there were also exceptions, with different prognosis samples showing positive connections (eg. connection of TCGA-AB-2845-03 with TCGA-AB-2946-03 and TCGA-AB-3000-03), and prognosis-similar samples showing negative connections (eg. connection of TCGA-AB-2805-03 with TCGA-AB-2897-03 and TCGA-AB-2979-03). Therefore, individual differences might be another influencing factor to prognosis in ALL.

Enrichment analysis. The function and pathways changes caused by differential gene expressions are normal ways interrupting disease development. To explore the deeper influence of DEGs to patients' prognosis, function and pathway enrichment analysis was conducted. Significantly enriched terms of up-regulated genes included antigen processing and presentation via MHC class I, immune response, chronic myeloid leukemia and natural killer cell mediated cytotoxicity (Table 1 and Table 2), indicating that the chronic disease

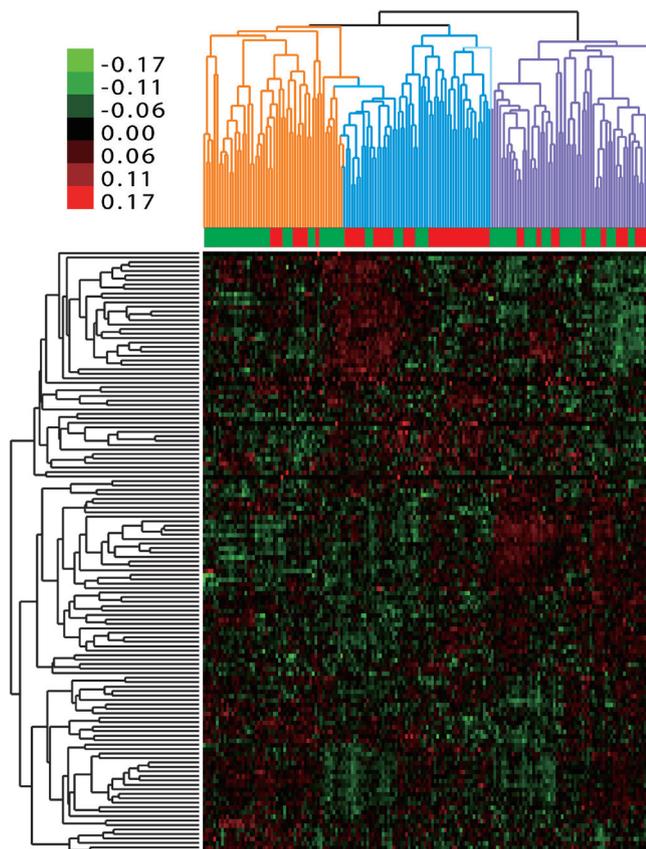


Figure 2. Sample clustering results. Green represents good prognosis samples, and red represents poor prognosis samples. The color changing from green to red represents the change from down-regulation to up-regulation of DEGs. Orange, blue and purple are the clustered 3 types of samples, with poor prognosis samples dominated in blue types.

Table 2. Significant pathway terms of up-regulated genes

Pathway	Count	<i>p</i> value	FDR
Focal adhesion	10	9.48E-06	5.05E-05
Chronic myeloid leukemia	6	3.90E-05	1.27E-04
Phosphatidylinositol signaling system	6	4.20E-05	1.32E-04
Leukocyte transendothelial migration	7	7.18E-05	1.95E-04
Prostate cancer	6	9.58E-05	2.38E-04
Regulation of actin cytoskeleton	9	9.60E-05	2.38E-04
Axon guidance	7	1.13E-04	2.72E-04
Alzheimer's disease	8	1.26E-04	2.84E-04
Citrate cycle (TCA cycle)	4	1.58E-04	3.47E-04
Natural killer cell mediated cytotoxicity	7	1.63E-04	3.48E-04

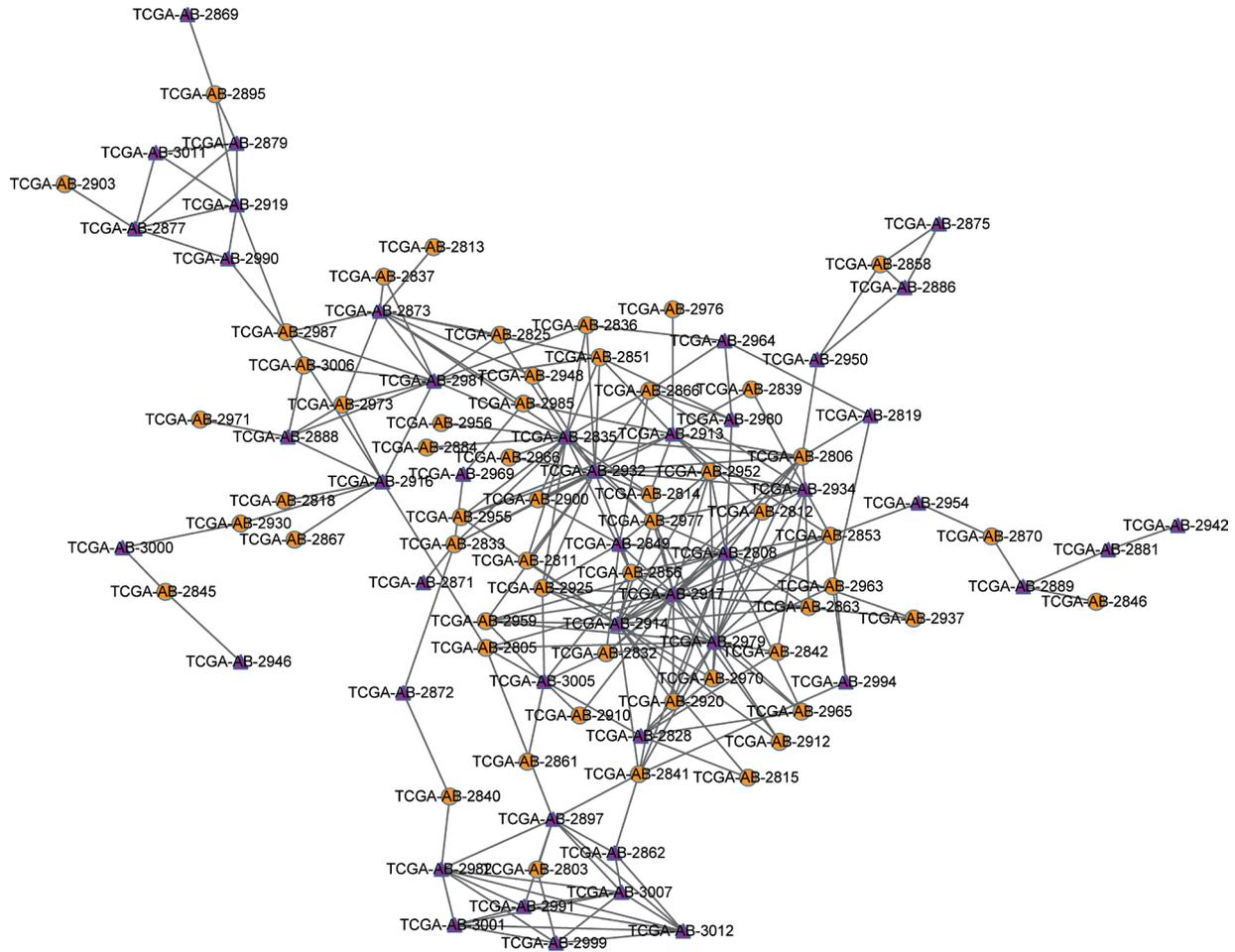


Figure 3. Sample dependency network. Good prognosis samples are expressed in triangle, while poor prognosis samples are expressed in circular. The positive connections are marked in solid line, while the negative connections are marked in dotted lines.

processing and higher natural killer cell-mediated cytotoxicity immune response may be connected with good prognosis in ALL. Significantly enriched terms of down-regulated genes included ubiquitin mediated proteolysis, cell cycle, and wnt signaling pathway (Table 3 and Table 4), suggesting good

prognosis patients may exert slower cell proliferation and good cell differentiation.

Prognosis-related genes and prognosis classifier. Genes in the significant function and pathway terms were conducted logistic regression analysis, to identify the DEGs associated with

Table 3. Significant functional terms of down-regulated genes

GO Term	Count	p value	FDR
GO:0006350 transcription	63	1.29E-55	3.10E-54
GO:0006355 regulation of transcription, DNA-dependent	60	4.50E-54	9.45E-53
GO:0019941 modification-dependent protein catabolism	25	3.88E-29	4.28E-28
GO:0007165 signal transduction	37	4.99E-20	3.99E-19
GO:0000122 negative regulation of transcription	12	1.55E-16	1.04E-15
GO:0006468 protein amino acid phosphorylation	17	5.76E-16	3.72E-15
GO:0006508 proteolysis	17	4.94E-13	2.52E-12
GO:0015031 protein transport	15	8.62E-13	4.26E-12
GO:0007049 cell cycle	15	5.00E-12	2.27E-11
GO:0007018 microtubule-based movement	8	5.38E-12	2.38E-11

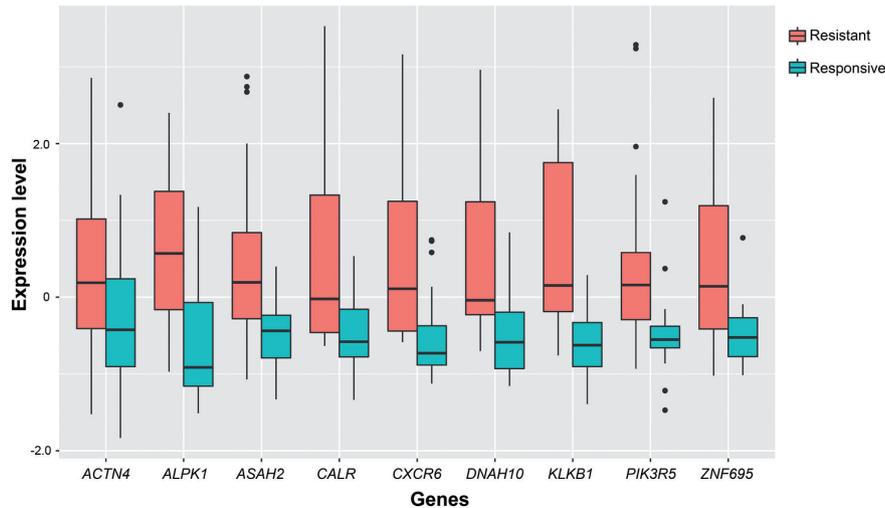


Figure 4. Boxplot of 9 prognosis-related genes in GSE13280. Blue bars represent treatment responsive samples, and red bars represent resistant samples. X-axis shows the prognosis-related gene, y-axis shows their expression levels.

prognosis. 13 prognosis-related genes were identified: *ALPK1*, *UBE2J1*, *ZNF695*, *ACTN4*, *CLK3*, *KLKB1*, *CALR*, *MED12*, *ASAH2*, *PIK3R5*, *FBXL5*, *CXCR6*, and *DNAH10* (Table 5).

The 13 prognosis-related genes were utilized to construct a prognosis classification model. 172 samples were used as the training data, while the rest 1 sample was used as the prediction data. The training result of the classifier showed an overall precision of 82.7%, and the classification precision for good prognosis samples was 71.2%, for poor prognosis samples was 88.6% (Table 6), suggesting a better classification effect on poor prognosis samples.

Among all 13 genes, 9 genes were found differentially expressed in 22 responsive and 22 resistant ALL samples downloaded from GEO database GSE13280 (Figure 4): *KLKB1*, *ALPK1*, *ACTN4*, *CXCR6*, *ASAH2*, *DNAH10*, *ZNF695*, *PIK3R5*, and *CALR*. The prognosis classifier was also tested using these 9 genes. The training result of the classifier showed an overall precision of 81.82% to predict the prognosis. Moreover, the classification precision of the classifier for responsive samples was 86.4%, for resistant samples was 77.3% (Table 7), suggesting a better classification effect on treatment sensible samples.

Survival analysis. To testify the portability and robustness of the prognosis classifier, the classification effect was also tested on the RNA-seq samples from 179 ALL patients downloaded from TCGA database. Firstly, 5 genes with the higher coefficients among 13 prognosis-related genes were selected, and they were *ALPK1*, *ZNF695*, *ACTN4*, *CALR*, and *FBXL5*. Then, the survival analysis based on the expressions of these 5 genes were conducted. Overall analysis result is shown in

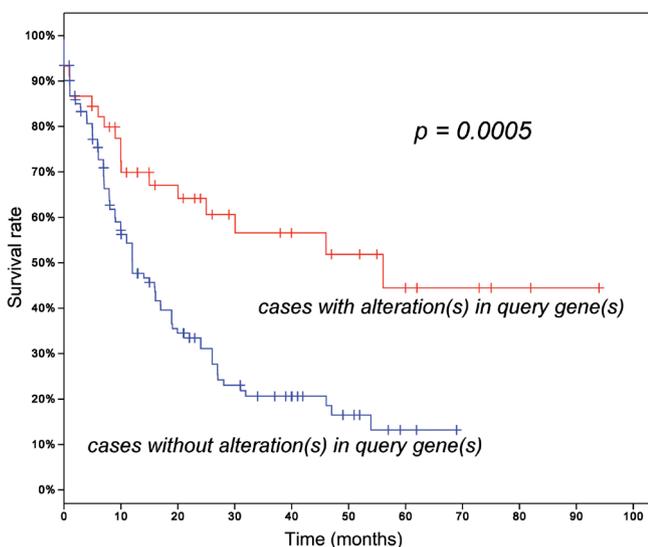


Figure 5. Overall survival analysis. The survival curve of patients without any alterations of the 5 genes is marked in red, and the survival curve of patients with expression alterations of the 5 genes is marked in blue. X-axis shows survival time, y-axis shows patient percentage.

Table 4. Significant pathway terms of down-regulated genes

Pathway	Count	<i>p</i> value	FDR
Ubiquitin mediated proteolysis	13	9.80E-12	1.23E-09
Small cell lung cancer	7	1.66E-06	2.47E-05
Wnt signaling pathway	8	8.00E-06	8.39E-05
Cell cycle	7	1.44E-05	1.35E-04
Prostate cancer	6	2.59E-05	1.98E-04
Aminosugars metabolism	4	3.76E-05	2.63E-04
Systemic lupus erythematosus	7	4.93E-05	3.10E-04
Melanogenesis	6	6.65E-05	3.57E-04
T cell receptor signaling pathway	6	8.64E-05	4.19E-04
Hedgehog signaling pathway	4	5.42E-04	0.001821

Figure 5, and patients with different survival times could be distinguished ($p = 0.000534$).

Discussion

Using the microarray data of bone marrow aspirate fluid mononuclear cells from adult ALL patients, 578 down-regulated and 637 up-regulated genes between good and poor prognosis samples were identified. From the clustering analysis results, we noticed that DEGs cannot be used as the prognosis specific biomarkers, so further analysis was conducted to identify more specific genes in ALL. Based on the Pearson connection coefficients of these DEGs, a sample dependency network comprised of 100 samples was also established. Normally, samples with similar prognosis exhibited positive connections, however, in the sample dependency network, there was also few prognosis-similar samples showed negative connections, suggesting the expression differences are the influential factor for prognosis. Thus, prognosis specific genes are more likely to be the predictive factors for ALL.

Chronic myeloid leukemia and natural killer cell mediated cytotoxicity were the significantly enriched pathways of up-regulated gene, indicating that the enhanced chronic progress and cytotoxicity are favorable factors for prognosis in ALL. Similarly, the enriched ubiquitin mediated proteolysis, cell cycle, and Wnt signaling pathways suggested lower proliferation speed and better differentiation in good prognosis ALL patients.

Totally, 13 prognosis-related genes were selected based on the significantly enriched function and pathway terms. The verification on the established classifier showed an overall precision of 82.7%, a precision of 71.2% to good prognosis samples, and a precision of 88.6% to poor prognosis, indicating a better distinguishing effect on poor prognosis samples. By further verification on another two sets of microarray profiles, four genes showed the close connection with the prognosis status in ALL patients, and they were *ALPK1*, *ZNF695*, *ACTN4*, and *CALR*.

ALPK1 (α -kinase 1) belongs to α -kinase family, which exerts crucial roles in the sorting and polarization of protein in epithelial cells. *ALPK1* was reported to be involved in epithelial cell polarity and exocytic vesicular transport towards the apical plasma membrane [26]. *ALPK1* has been implicated in female cancers such as breast and ovarian cancers [27], and recognized as one of the critical genes to inhibit differentiation of bi-potential tumor-initiating cells in human breast cancer [28]. Despite the lack of direct evidence, the finding in this study warranted the prognostic roles of *ALPK1* in ALL patients.

ACTN4 is a gene located on human chromosome 19q, and actinin-4 protein could be used to access the invasion and movement capacities of cancer cells [29]. It is reported that patients with the amplifications of *ACTN4* have worse outcomes than those without amplification [30]. Copy number increase of *ACTN4* is an indicator for the unfavorable outcome in patients with salivary gland carcinoma [31] and an predictor for

Table 5. Prognosis-related genes selected via logistic regression

Gene	B	SE	Wals coefficient	Df	(p value)	Exp (B)
ALPK1	0.825	0.335	6.059	1	0.014	2.281
UBE2J1	0.979	0.304	10.397	1	0.001	2.662
ZNF695	1.077	0.286	14.197	1	0.000	2.937
ACTN4	1.075	0.327	10.787	1	0.001	2.930
CLK3	-1.329	0.336	15.654	1	0.000	0.265
KLKB1	0.552	0.268	4.236	1	0.040	1.737
CALR	-1.137	0.329	11.963	1	0.001	0.321
MED12	-1.19	0.344	11.992	1	0.001	0.304
ASAH2	0.785	0.341	5.294	1	0.021	2.192
PIK3R5	1.788	0.42	18.155	1	0.000	5.975
FBXL5	-1.997	0.465	18.464	1	0.000	0.136
CXCR6	-1.201	0.331	13.154	1	0.000	0.301
DNAH10	0.725	0.301	5.783	1	0.016	2.064

B: regression coefficient; SE: standard error; Df: degree of freedom.

Table 6. Classification result in TCGA dataset

	Class	Predict		
		Good	Poor	Percent
Observe	Good	42	17	71.2
	Poor	13	101	88.6
Overall percent				82.7

Table 7. Classification result in GSE13280 microarray dataset

	Class	Predict		
		Responsive	Resistant	Percent
Observe	Responsive	19	3	86.4
	Resistant	5	17	77.3
Overall percent				81.82

chemoradiotherapy efficacy in patients with locally advanced pancreatic cancer [32]. Although ALL is not solid tumor, the finding in this study indicted that *ACTN4* may play a similar role in predicting prognosis in ALL.

CALR is a gene encoding calreticulin protein, which a multi-functional Ca^{2+} -binding protein chaperone mostly localized in endoplasmic reticulum [33]. *CALR* mutations have been recognized as an important diagnostic marker for myeloproliferative neoplasms [33].

It is reported that *ZNF695* (zinc finger 965) methylation predicts a response to definitive chemoradiotherapy in esophageal squamous cell carcinoma [34]. The findings in this study show a potential correlation between *ZNF695* and the prognosis of ALL patients.

In conclusion, *ALPK1*, *ACTN4*, *CALR*, and *ZNF695* were identified as the potential prognosis risk factors in acute lymphoblastic leukemia, which may promote to understand

the mechanism of ALL and to develop novel therapeutic strategies. Further researches should be performed to clarify their exact roles in ALL.

References

- [1] HARRISON CJ, JOHANSSON B. Acute lymphoblastic leukemia. p 198–251. In: S. Heim and F. Mitelman (Eds.), *Cancer Cytogenetics: Chromosomal and Molecular Genetic Aberrations of Tumor Cells*, 4th edition. Wiley-Blackwell, 2015, pp. 648. ISBN 9781118795538. <https://doi.org/10.1002/9781118795569.ch10>
- [2] PUI C-H, ROBISON LL, LOOK AT. Acute lymphoblastic leukaemia. *Lancet* 2008; 371: 1030–1043. [https://doi.org/10.1016/S0140-6736\(08\)60457-2](https://doi.org/10.1016/S0140-6736(08)60457-2)
- [3] MARGOLIN J, RABIN KR, STEUBER CP, POPLACK DG. Acute lymphoblastic leukemia. p 518–565. In: PA Pizzo and DG Poplack (Eds.), *Principles and Practice of Pediatric Oncology*, 6th edition. Lippincott Williams & Wilkins, Philadelphia, 2011, pp 1531. ISBN 9781605476827.
- [4] PULTE D, GONDOS A, BRENNER H. Trends in 5- and 10-year survival after diagnosis with childhood hematologic malignancies in the United States, 1990–2004. *J Natl Cancer Inst* 2008; 100: 1301–1309. <https://doi.org/10.1093/jnci/djn276>
- [5] PULTE D, GONDOS A, BRENNER H. Improvement in survival in younger patients with acute lymphoblastic leukemia from the 1980s to the early 21st century. *Blood* 2009; 113: 1408–1411. <https://doi.org/10.1182/blood-2008-06-164863>
- [6] YANG JJ, CHENG C, YANG W, PEI D, CAO X et al. Genome-wide interrogation of germline genetic variation associated with treatment response in childhood acute lymphoblastic leukemia. *JAMA* 2009; 301: 393–403. <https://doi.org/10.1001/jama.2009.7>
- [7] SHERBORNE AL, HOSKING FJ, PRASAD RB, KUMAR R, KOEHLER R et al. Variation in CDKN2A at 9p21.3 influences childhood acute lymphoblastic leukemia risk. *Nat Genet* 2010; 42: 492–494. <https://doi.org/10.1038/ng.585>
- [8] BUITENKAMP T, PIETERS R, GALLIMORE N, VAN DER VEER A, MEIJERINK J et al. Outcome in children with Down's syndrome and acute lymphoblastic leukemia: role of IKZF1 deletions and CRLF2 aberrations. *Leukemia* 2012; 26: 2204–2211. <https://doi.org/10.1038/leu.2012.84>
- [9] MULLIGHAN CG, SU X, ZHANG J, RADTKE I, PHILLIPS LA et al. Deletion of IKZF1 and prognosis in acute lymphoblastic leukemia. *N Engl J Med* 2009; 360: 470–480. <https://doi.org/10.1056/NEJMoa0808253>
- [10] DEN BOER ML, VAN SLEGTENHORST M, DE MENEZES RX, CHEOK MH, BUIJS-GLADDINES JG et al. A subtype of childhood acute lymphoblastic leukaemia with poor treatment outcome: a genome-wide classification study. *Lancet Oncol* 2009; 10: 125–134. [https://doi.org/10.1016/S1470-2045\(08\)70339-5](https://doi.org/10.1016/S1470-2045(08)70339-5)
- [11] SHEN L, TOYOTA M, KONDO Y, OBATA T, DANIEL S et al. Aberrant DNA methylation of p57KIP2 identifies a cell-cycle regulatory pathway with prognostic impact in adult acute lymphocytic leukemia. *Blood* 2003; 101: 4131–4136. <https://doi.org/10.1182/blood-2002-08-2466>
- [12] ROMAN-GOMEZ J, CASTILLEJO JA, JIMENEZ A, GONZALEZ MG, MORENO F et al. 5' CpG island hypermethylation is associated with transcriptional silencing of the p21CIP1/WAF1/SDI1 gene and confers poor prognosis in acute lymphoblastic leukemia. *Blood* 2002; 99: 2291–2296. <https://doi.org/10.1182/blood.V99.7.2291>
- [13] XU N, LI YL, ZHOU X, CAO R, LI H et al. CDKN2 Gene Deletion As Poor Prognosis Predictor Involving in the Progression of Adult B-Lineage Acute Lymphoblastic Leukemia Patients. *J Cancer*. 2015; 6: 1114–1120. <https://doi.org/10.7150/jca.11959>
- [14] DANIS E, YAMAUCHI T, ECHANIQUE K, ZHANG X, HALADYNA JN et al. Ezh2 Controls an Early Hematopoietic Program and Growth and Survival Signaling in Early T Cell Precursor Acute Lymphoblastic Leukemia. *Cell Rep* 2016; 14: 1953–1965. <https://doi.org/10.1016/j.celrep.2016.01.064>
- [15] VAN DER SLIGTE NE, KAMPEN KR, TER ELST A, SCHERPEN FJ, MEEUWSEN-DE BOER TG et al. Essential role for cyclic-AMP responsive element binding protein 1 (CREB) in the survival of acute lymphoblastic leukemia. *Oncotarget* 2015; 6: 14970–14981. <https://doi.org/10.18632/oncotarget.3911>
- [16] DEN RB, YOUSEFI K, TRABULSI EJ, ABDOLLAH F, CHOEURNG V et al. Genomic classifier identifies men with adverse pathology after radical prostatectomy who benefit from adjuvant radiation therapy. *J Clin Oncol* 2015; 33: 944–951. <https://doi.org/10.1200/JCO.2014.59.0026>
- [17] LAU SK, BOUTROS PC, PINTILIE M, BLACKHALL FH, ZHU C-Q et al. Three-gene prognostic classifier for early-stage non-small-cell lung cancer. *J Clin Oncol* 2007; 25: 5562–5569. <https://doi.org/10.1200/JCO.2007.12.0352>
- [18] SKRZYPSKI M, JASSEM E, TARON M, SANCHEZ JJ, MENDEZ P et al. Three-gene expression signature predicts survival in early-stage squamous cell carcinoma of the lung. *Clin Cancer Res* 2008; 14: 4794–4799. <https://doi.org/10.1158/1078-0432.CCR-08-0576>
- [19] ZHU C-Q, DING K, STRUMPF D, WEIR BA, MEYERSON M et al. Prognostic and predictive gene signature for adjuvant chemotherapy in resected non-small-cell lung cancer. *J Clin Oncol* 2010; 28: 4417–4424. <https://doi.org/10.1200/JCO.2009.26.4325>
- [20] ROSS ME, ZHOU X, SONG G, SHURTLEFF SA, GIRTMAN K et al. Classification of pediatric acute lymphoblastic leukemia by gene expression profiling. *Blood* 2003; 102: 2951–2959. <https://doi.org/10.1182/blood-2003-01-0338>
- [21] CHEADLE C, VAWTER MP, FREED WJ, BECKER KG. Analysis of microarray data using Z score transformation. *J Mol Diagn* 2003; 5: 73–81. [https://doi.org/10.1016/S1525-1578\(10\)60455-2](https://doi.org/10.1016/S1525-1578(10)60455-2)
- [22] PAIC F, IGWE JC, NORI R, KRONENBERG MS, FRANCESCHETTI T et al. Identification of differentially expressed genes between osteoblasts and osteocytes. *Bone* 2009; 45: 682–692. <https://doi.org/10.1016/j.bone.2009.06.010>
- [23] DROY-DUPR L, BOSSARD C, VOLTEAU C, BEZIEAU S, LABOISSE CL et al. Hierarchical clustering identifies a subgroup of colonic adenocarcinomas expressing crypt-like differentiation markers, associated with MSS status and better

- prognosis. *Virchows Arch* 2015; 466: 383–391. <https://doi.org/10.1007/s00428-015-1724-9>
- [24] Shi X, BARNES RO, CHEN L, SHAJAHAN-HAQ AN, HILAKIVI-CLARKE L et al. BMRF-Net: a software tool for identification of protein interaction subnetworks by a bagging Markov random field-based method. *Bioinformatics* 2015; 31: 2412–2414. <https://doi.org/10.1093/bioinformatics/btv137>
- [25] FERREIRA LC, DOSCIATTI MM, NIEVOLA JC, PARAISO EC. Using a Genetic Algorithm Approach to Study the Impact of Imbalanced Corpora in Sentiment Analysis. p 163–168. In: I. Russell and W. Eberle (Eds.), *Proceedings of FLAIRS Conference 2015*, AAAI Press, Palo Alto, California, 2015, pp 656. ISBN 9781577357308.
- [26] HEINE M, CRAMM-BEHRENS CI, ANSARI A, CHU H-P, RYAZANOV AG et al. α -kinase 1, a new component in apical protein transport. *J Biol Chem* 2005; 280: 25637–25643. <https://doi.org/10.1074/jbc.M502265200>
- [27] CHEN H-M, SCHMEICHEL KL, MIAN IS, LELIEVRE S, PETERSEN OW et al. AZU-1: a candidate breast tumor suppressor and biomarker for tumor progression. *Mol Biol Cell* 2000; 11: 1357–1367. <https://doi.org/10.1091/mbc.11.4.1357>
- [28] STRIETZ J, STEPPUTTIS SS, PRECA BT, VANNIER C, KIM MM et al. ERN1 and ALPK1 inhibit differentiation of bi-potential tumor-initiating cells in human breast cancer. *Oncotarget* 2016; 7: 83278–83293. <https://doi.org/10.18632/oncotarget.13086>
- [29] HONDA K, YAMADA T, ENDO R, INO Y, GOTOH M et al. Actinin-4, a novel actin-bundling protein associated with cell motility and cancer invasion. *J Cell Biol* 1998; 140: 1383–1393. <https://doi.org/10.1083/jcb.140.6.1383>
- [30] YAMAMOTO S, TSUDA H, HONDA K, ONOZATO K, TAKANO M et al. Actinin-4 gene amplification in ovarian cancer: a candidate oncogene associated with poor patient prognosis and tumor chemoresistance. *Mod Pathol* 2009; 22: 499–507. <https://doi.org/10.1038/modpathol.2008.234>
- [31] WATABE Y, MORI T, YOSHIMOTO S, NOMURA T, SHIBAHARA T et al. Copy number increase of ACTN4 is a prognostic indicator in salivary gland carcinoma. *Cancer Med* 2014; 3: 613–622. <https://doi.org/10.1002/cam4.214>
- [32] WATANABE T, UENO H, WATABE Y, HIRAOKA N, MORI-ZANE C et al. ACTN4 copy number increase as a predictive biomarker for chemoradiotherapy of locally advanced pancreatic cancer. *Br J Cancer* 2015; 112: 704–713. <https://doi.org/10.1038/bjc.2014.623>
- [33] NANGALIA J, MASSIE CE, BAXTER EJ, NICE FL, GUNDEM G et al. Somatic CALR mutations in myeloproliferative neoplasms with nonmutated JAK2. *N Engl J Med* 2013; 369: 2391–2405. <https://doi.org/10.1056/NEJMoa1312542>
- [34] TAKAHASHI T, YAMAHSITA S, MATSUDA Y, KISHINO T, NAKAJIMA T et al. ZNF695 methylation predicts a response of esophageal squamous cell carcinoma to definitive chemoradiotherapy. *J Cancer Res Clin Oncol* 2015; 141: 453–463. <https://doi.org/10.1007/s00432-014-1841-x>