ORIGINAL ARTICLE

Molecular Genetics & Genomic Medicine   WILEY
Open Access

Development of an immune-related prognostic model for pediatric acute lymphoblastic leukemia patients

Xi Quan [ID] | Nan Zhang | Ying Chen | Hanqing Zeng | Jianchuan Deng

Department of Hematology, The Second Affiliated Hospital, Chongqing Medical University, Chongqing, P.R. China

Correspondence
Jianchuan Deng, Department of Hematology, The Second Affiliated Hospital, Chongqing Medical University, 76 Linjiang Road, Chongqing 400010, P.R. China.
Email: dengjccq@hospital.cqmu.edu.cn

Abstract

Background: Acute lymphoblastic leukemia (ALL) is the most common hematological malignancy in pediatrics, and immune-related genes (IRGs) play crucial role in its development. Our study aimed to identify prognostic immune biomarkers of pediatric ALL and construct a risk assessment model.

Methods: Pediatric ALL patients' gene expression data were downloaded from Therapeutically Applicable Research to Generate Effective Treatments (TARGET) database. We screened differentially expressed IRGs (DEIRGs) between the relapse and non-relapse groups. Cox regression analysis was used to identify optimal prognostic genes, then, a risk model was constructed, and its accuracy was verified in different cohorts.

Results: We screened 130 DEIRGs from 251 pediatric ALL samples. The top three pathways that DEIRGs may influence tumor progression are NABA matrisome-associated, chemotaxis, and antimicrobial humoral response. A set of 84 prognostic DEIRGs was identified by using univariate Cox analysis. Then, Lasso regression and multivariate Cox regression analysis screened four optimal genes (*PRDX2*, *S100A10*, *RORB*, and *SDC1*), which were used to construct the prognostic risk model. The risk score was calculated and the survival analysis results showed that high-risk score was associated with poor overall survival (OS) ($p = 3.195 \times 10^{-7}$). The time-dependent survival receiver operating characteristic curves showed good prediction accuracy (Area Under Curves for 3-year, 5-year OS were 0.892 and 0.89, respectively). And the predictive performance of our risk model was successfully verified in testing cohort and entire cohort.

Conclusions: Our prognostic risk model can effectively divide pediatric ALL patients into high-risk and low-risk groups, which may help predict clinical prognosis and optimize individualized treatment.

KEYWORDS
immune-related genes, pediatric acute lymphoblastic leukemia, prognosis, survival analysis

1 | INTRODUCTION

Acute lymphoblastic leukemia (ALL) is the most frequent malignancy and the leading cause of cancer-related deaths in pediatrics. The cure rate has exceeded 80% in last decade, owing to improved supportive care and optimized treatment regimens (Brassesco et al., 2018; Kato & Manabe, 2018). However, a significant number of patients still suffer from drug resistance or relapse (Tasian & Hunger, 2017), resulting in treatment failure. In addition, treatment may have to be discontinued because of its high toxicity (Santiago, Vairy, Sinnett, Krajinovic, & Bittencourt, 2017). Taking these factors into account, new biomarkers and precise treatment regimens will be a priority for these patients.

A large number of studies have focused on the development and application of biomarkers in ALL. For example, some researchers have suggested a tumor suppressor role of TLE1 in T-ALL (Brassesco et al., 2018). In addition, mTOR inhibitors have been used in combination with chemotherapy regiments for the treatment of relapse ALL (Santiago et al., 2017), and bcl-2 inhibitors have also been used in the treatment for all subtypes of pediatric ALL (Jones et al., 2016). Recent studies have shown that gene expression in patients with recurrent leukemia after transplantation is highly enriched in immune-related processes (Toffalori et al., 2019). It has also been mentioned that this is related to the escape of tumor cells from the control of allogeneic immune response (Zeiser & Vago, 2019). These results suggest that immune-related biomarkers may be significant signatures for predicting the prognosis of ALL.

With the development of bioinformatics, the immune-related genes (IRGs)-based prognostic signatures have been developed in patients diagnosed with renal papillary cell carcinoma (Wang et al., 2019), colorectal cancer (Bai, Zhang, Xiang, Zhong, & Xiong, 2020), and lung adenocarcinoma (Song et al., 2019), which can predict survival outcomes. However, the prognostic value of IRGs-based signatures in pediatric ALL patients is still unknown.

The purpose of this study was to investigate the clinical significance of IRGs on the prognosis of pediatric ALL and its biological function. In this paper, we comprehensively analyzed the expression profile data and the clinical information of pediatric ALL patients. A prognostic model based on IRGs was developed and validated in public dataset, which may be helpful in predicting prognosis and optimizing individualized treatment.

2 | MATERIALS AND METHODS

2.1 | Gene expression datasets

The transcriptomic data and corresponding clinical information of 251 pediatric ALL patients were downloaded from the Therapeutically Applicable Research to Generate Effective Treatments (TARGET) portal (https://ocg.cancer.gov/programs/target) (Kang et al., 2010). And the 2,498 IRG sets were obtained from the ImmPort database (https://www.immport.org/home) (Bhattacharya et al., 2014). The expression data were preprocessed by the following steps: (a) removing samples with no clinical data; (b) removing samples that expression data and clinical information did not match; (c) preserving only the expression profiles of IRGs. Form this, 185 patients with complete gene expression profiles and clinical information were utilized to further analyze the model. The data downloaded from the TARGET and ImmPort databases is publicly available and accessible, therefore, this study does not require ethics committee approval.

2.2 | Identification of DEIRGs

The pediatric ALL samples were divided into relapse group and non-relapse group. And the treatment regimens before the study endpoint events were chemotherapy treatments. The differentially expressed IRGs (DEIRGs) were screened by using edge R package (Robinson, McCarthy, & Smyth, 2010) in R3.6.2 software. The FDR < 0.05 and |log2 fold-change [FC]| > 1.5 were cutoff values. Then, the gene expression values were visualized by pheatmap package (Li, Zhang, Rui, Sun, & Guo, 2018). Enrichment analysis was performed to predict the biological functions of the DEIRGs by using Metascape (http://metascape.org/), an online bioinformatics pipeline (Zhou et al., 2019).

2.3 | Construction of the risk score prognostic model

The 185 samples were randomly divided into a training cohort ($n = 93$) and a testing cohort ($n = 92$). The training cohort was used to build the risk score prognostic model, the testing and entire TARGET cohorts were used to test the model. First, univariate Cox analysis was used to identify possible prognostic DEIRGs (PDEIRGs), and $p < .05$ was considered significant. Then, the Lasso regression was applied to select potential risk genes and eliminate genes that would overfit the model. Finally, we used multivariate Cox regression analysis to construct a prognostic risk model.

2.4 | Risk score calculation and Model validation

To evaluate the contribution of each gene to prognosis, the multivariate Cox regression analysis was performed. Then, we obtained a computational formula that weight the
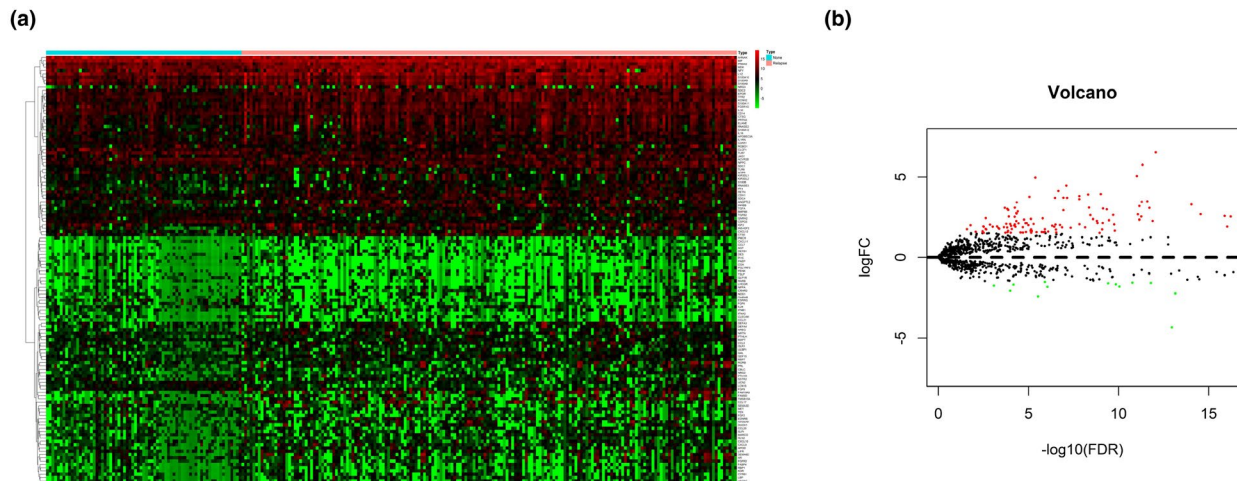
FIGURE 1 Differentially expressed immune-related genes. (a) Heatmap of DEIRGs; the green spectrum means low gene expression while the red means high gene expression. (b) Volcano plot of DEIRGs; the green dots indicate downregulated IRGs, the red dots indicate upregulated IRGs, and the black dots represent IRGs that were not significantly differentially expressed. DEIRGs, differentially expressed IRGs; IRGs, immune-related genes

expression values of selected genes with the regression coefficients as follows:

$$\text{Risk score (patient)} = \sum_{i=1}^{n} \text{coefficient (gene } i\text{) expression value of (gene } i\text{)}$$

The risk model was used to measure the prognostic risk of each pediatric ALL patient.

We substituted the expression profile data into the model to calculate the risk score of each sample from the training cohort and entire TARGET cohort. Then, Kaplan–Meier survival analysis, receiver operating characteristic (ROC) analysis, risk score distribution, survival status, and risk gene expression of the training cohort, entire TARGET cohort were performed to verify our risk score prognostic model. Multivariate Cox regression analysis was used to assess the independent prognostic ability of the model.

2.5 | Statistical analyses

Statistical analyses were performed using R software (https://www.r-project.org/) and Perl (https://www.activestate.com/products/perl/). Univariate Cox regression analysis was used to identify factors affecting the survival of pediatric ALL patients. Lasso regression was used to evaluate univariate analysis of the link between PDEIRGS. Multivariate Cox regression analysis was used to identify prognostic factors. Kaplan–Meier curves and log-rank tests were used to analyze the survival data. An Area Under Curve (AUC) > 0.60 was regarded as acceptable for predictions, and an AUC > 0.75 was deemed to have excellent predictive value (Cho et al., 2019; Han et al., 2018).

3 | RESULTS

3.1 | DEIRGs screening based on the pediatric ALL samples

The mRNA expression data of 2,498 IRGs in pediatric ALL ($n = 251$) from TARGET database was examined. After screening the expression data by using edge R package, a total of 130 DEIRGs were obtained from with relapse group ($n = 180$) and without group ($n = 71$). The results showed that 10.8% (14/130) of DEIRGs were downregulated in relapse group while 89.2% (116/130) of DEIRGs was significantly upregulated (FDR < 0.05, |log2 fold-change [FC]| > 1.5) (Figure 1a,b).

Then, we conducted enrichment analysis to identify the possible biological functions of DEIRGs. Data showed that the top three signaling pathways affected by DEIRGs were NABA matrisome-associated, chemotaxis, and antimicrobial humoral response [Figure 2a,b]. All these three signaling pathways were reported to be associated with tumor progression (Chen, Lin, Wu, Her, & Hui, 2009; Naba et al., 2012; Shields et al., 2007), providing evidence for further study on the mechanism of pediatric ALL progression.

3.2 | Identification of prognostic DEIRGs

The 185 samples were randomly divided into a training cohort ($n = 93$) and a testing cohort ($n = 92$), see Table 1. To identify possible prognostic DEIRGs, we performed a univariate Cox regression analysis of the expression of each DEIRG in the training cohort. As a result, 84 PDEIRGs were found to be significantly associated with the overall survival (OS) of pediatric ALL patients ($p < .05$).
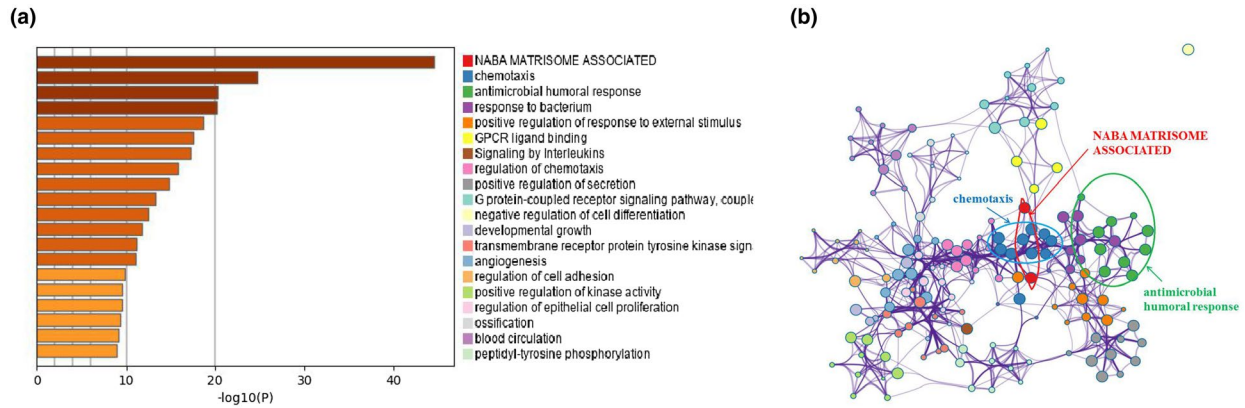
(a)

(b)

FIGURE 2 Biological functions of DEIRGs. (a) Significantly enriched pathways of the DEIRGs. (b) Network of enriched pathways. Each node represents an enriched GO term and node size represents the number of gene in the pathways. DEIRGs, differentially expressed IRGs

TABLE 1 Clinical information of pediatric ALL patients in the training and validation cohorts

	Training cohort (n = 93)	Testing cohort (n = 92)	Entire TARGET cohort (n = 185)
Sex			
Male	47 (50.5%)	44 (47.8%)	91 (49.2%)
Female	46 (49.5%)	48 (52.2%)	94 (50.8%)
Age at diag (years)			
<10	64 (68.8%)	60 (65.2%)	124 (67%)
≥10	29 (31.2%)	32 (34.8%)	61 (33%)
WBC at diag (×10⁹/L)			
<50	60 (64.5%)	57 (62%)	117 (63.2%)
≥50	33 (35.5%)	35 (38%)	68 (36.8%)
CNS status at diag			
CNS1	73 (78.5%)	76 (82.6%)	149 (80.5%)
CNS2	19 (20.4%)	14 (15.2%)	33 (17.8%)
CNS3	1 (1.1%)	2 (2.2%)	3 (1.7%)
First event			
Relapse	57 (61.3%)	64 (69.6%)	121 (65.4%)
None	36 (38.7%)	28 (30.4%)	64 (34.6%)
Vital status			
Dead	39 (41.9%)	43 (46.7%)	82 (44.3%)
Alive	54 (58.1%)	49 (53.3%)	103 (55.7%)

Abbreviations: ALL, acute lymphoblastic leukemia; CNS, central nervous system; CNS1: no lymphoblasts in CSF; CNS2: present lymphoblasts in CSF, WBC count of the CSF < 5 cells/μl; CNS3: present lymphoblasts in CSF or a cranial nerve palsy, WBC count of the CSF ≥ 5 cells/μl; diag, diagnosis; TARGET, Therapeutically Applicable Research to Generate Effective Treatments; WBC, white blood cell count.

3.3 | Screening prognostic genes for constructing risk model

We further analyzed and screened PDEIRGs for constructing cox regression hazard model. First, to avoid model overfitting, we used Lasso regression to remove PDEIRGs that are highly correlated to each other. Therefore, we obtained

seven candidate PDEIRGs (Figure 3a,b). Then, multivariate Cox proportional risk regression analysis was performed (with forward selection and backward selection). Finally, we obtained four optimal PDEIRGs (risk genes) to incorporate into the prognostic risk model: *PRDX2*, *S100A10*, *RORB,* and *SDC1*. These four genes were identified as high-risk genes (predicting a poor prognosis) in terms of the OS of patients (Figure 4).

3.4 | Construction of prognostic risk model in training cohort

Based on the results of multivariate Cox regression analysis, we constructed a model to assess the significance of risk genes in predicting prognosis in pediatric ALL patients. The computational formula was as follows:

$$\text{Training cohort risk score} = (0.1615 \times \text{expression of PRDX2})$$
$$+ (0.3387 \times \text{expression of S100A10})$$
$$+ (0.0903 \times \text{expression of RORB}) + (0.1940 \times \text{expression of SDC1}).$$

We calculated the risk score of each patient in the training cohort using the risk model, and patients were sorted into a high-risk group ($n = 46$) and a low-risk group ($n = 47$). To investigate the difference in prognosis between the high-risk and low-risk groups, we created a Kaplan–Meier curve based on the log-rank test. The prognosis was better in the low-risk group than in the high-risk group ($p = 3.195 \times 10^{-7}$) (Figure 5a). The OS rates at 3 years and 5 years for the high-risk group in the training cohort were 46.3% and 33.2%, respectively, while the corresponding rates for the low-risk group were 91.5% and 86.9%, respectively. Then, we tested the predictive accuracy of the model for 3-year and 5-year OS through the time-dependent ROC curves. The AUC values for the prognostic model were 0.892 at 3 years and 0.89 at 5 years (Figure 5b,c). We then sorted and analyzed the distribution of patients' risk scores in the training cohort (Figure 6a). The survival status of each patient in the training cohort is marked on the dot plot in Figure 6b. The heatmap we completed showed the expression of risk genes in both risk groups (Figure 6c). In the high-risk group of the training cohort, four high-risk genes (*PRDX2*, *S100A10*, *RORB,* and *SDC1*) were upregulated. In the low-risk group, the expression of these risk genes was downregulated.

3.5 | Verification of the performance of the prognostic model

To validate the predictive ability of the prognostic risk model, we used it to analyze the testing cohort (the remaining 92 patients from the 185 total) and the entire TARGET cohort. First, the risk score for each patient in the testing cohort and the entire TARGET cohort was calculated according to the coefficient value of the four risk genes. Patients were divided into high-risk and low-risk groups with the median risk score of the training cohort utilized as the cutoff value. In the testing cohort, 52 patients were divided as high risk and 40 were divided as low risk. In the entire TARGET cohort, 98 patients were classified as high risk and 87 were classified as low risk.

Then, Kaplan–Meier survival analysis was performed for both the testing cohort and the entire TARGET cohort. Patients of high risk were with poor OS compared with those of low risk in both the testing cohort ($p = 1.427 \times 10^{-3}$) and the entire TARGET cohort ($p = 3.255 \times 10^{-9}$) (Figure 7a,d). In the testing cohort, the OS rates at 3 years and 5 years for the high-risk group were 59.1% and 45.2%, respectively, while the corresponding rates for the low-risk group were 87.5% and 77.3%, respectively. In the entire TARGET cohort, the OS rates at 3 years and 5 years for the high-risk group were 54.3% and 40.7%, respectively, while the corresponding rates for the low-risk group were 89.7% and 82.5%, respectively. To evaluate the accuracy in prognosis prediction of our four-gene model, we performed
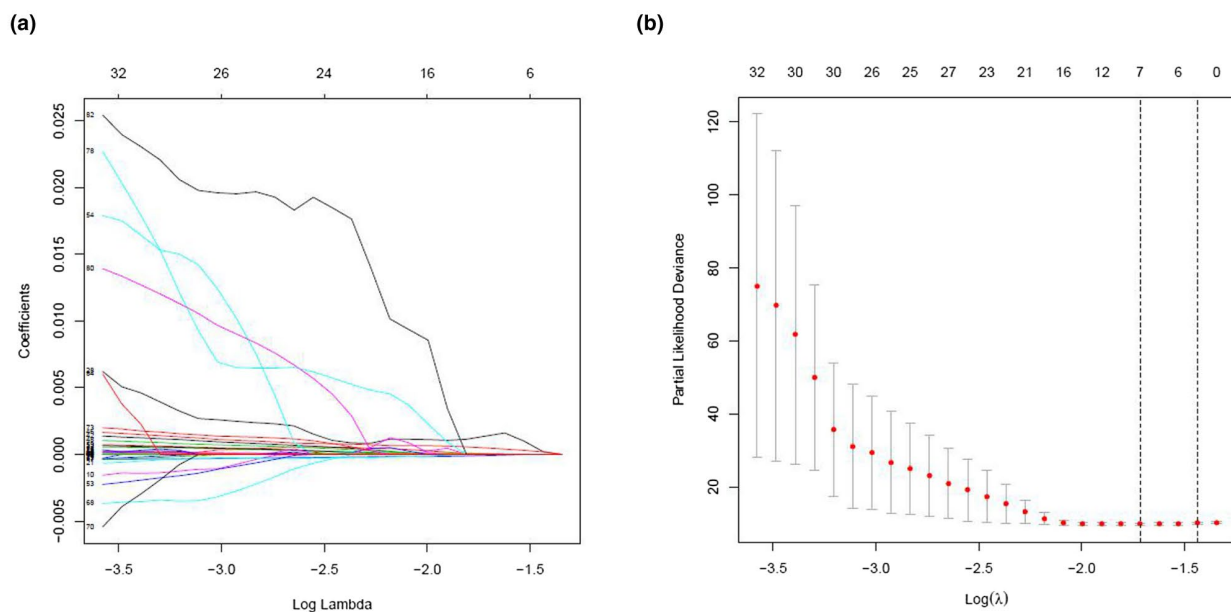
(a)

(b)
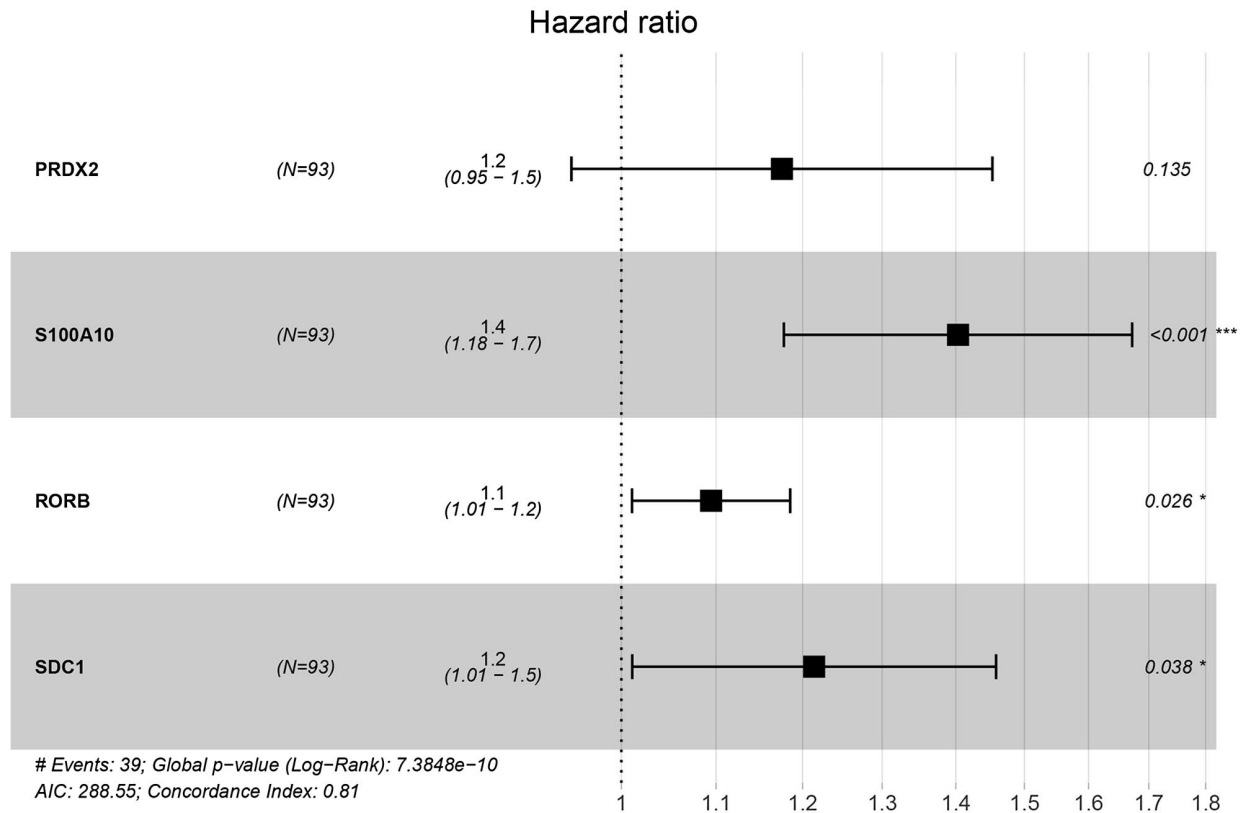

FIGURE 3 (a and b) PDEIRGs screened through Lasso regression

Hazard ratio

PRDX2	(N=93)	1.2 (0.95 – 1.5)		0.135
S100A10	(N=93)	1.4 (1.18 – 1.7)		<0.001 ***
RORB	(N=93)	1.1 (1.01 – 1.2)		0.026 *
SDC1	(N=93)	1.2 (1.01 – 1.5)		0.038 *

Events: 39; Global p−value (Log−Rank): 7.3848e−10
AIC: 288.55; Concordance Index: 0.81

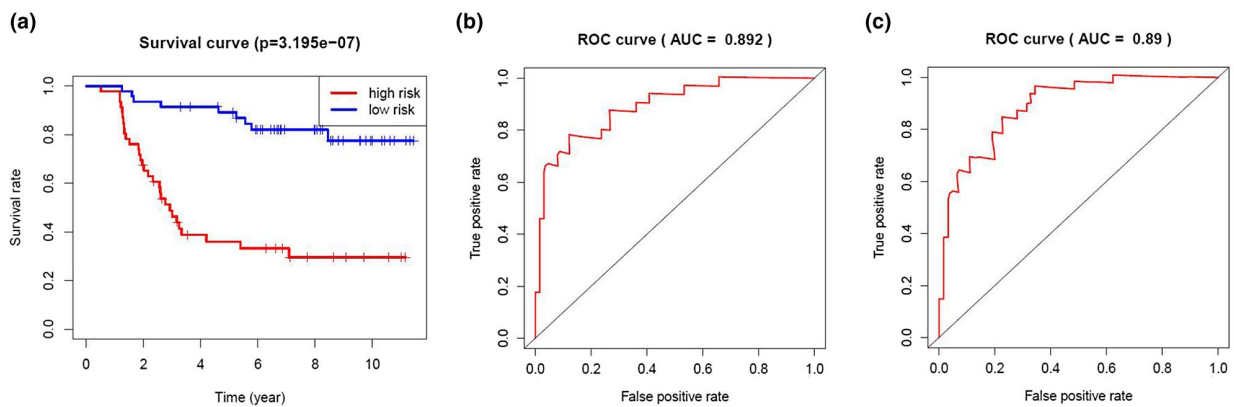**FIGURE 4** Risk genes of the prognostic risk model. *, $p < .05$; ***, $p < .001$

FIGURE 5 Prognosis analysis of training cohort. (a) Kaplan–Meier curve analysis of the high-risk and low-risk groups. Time-dependent ROC curve analysis for the predictive accuracy of the risk model for 3-year (b) and 5-year OS (c). OS, overall survival; ROC, receiver operating characteristic

time-dependent ROC curve analysis. In the testing cohort, the AUCs at 3 and 5 years were 0.814 and 0.751, respectively (Figure 7b,c). In the entire TARGET cohort, the AUCs at 3 and 5 years were 0.852 and 0.819, respectively (Figure 7e,f).

The risk score distribution, survival status, and risk gene expression in the testing cohort and the entire TARGET cohort are shown in Figure 8a–f. Similar to the results in the training cohort, risk gene levels were lower in the low-risk group than in the high-risk group. These results suggested

that our prognostic risk model can accurately predict the prognosis of pediatric ALL patients.

The univariate and multivariate Cox analysis of risk score generated by our model and clinical parameters in entire TARGET cohort is shown in Table 2. The univariate analysis indicated that the variables of age, minimal residual disease (MRD) status at day 29 of induction therapy, and risk score were associated with the prognosis of pediatric ALL patients. And in the multivariate analysis, the risk score can serve as an independent prognostic factor for OS
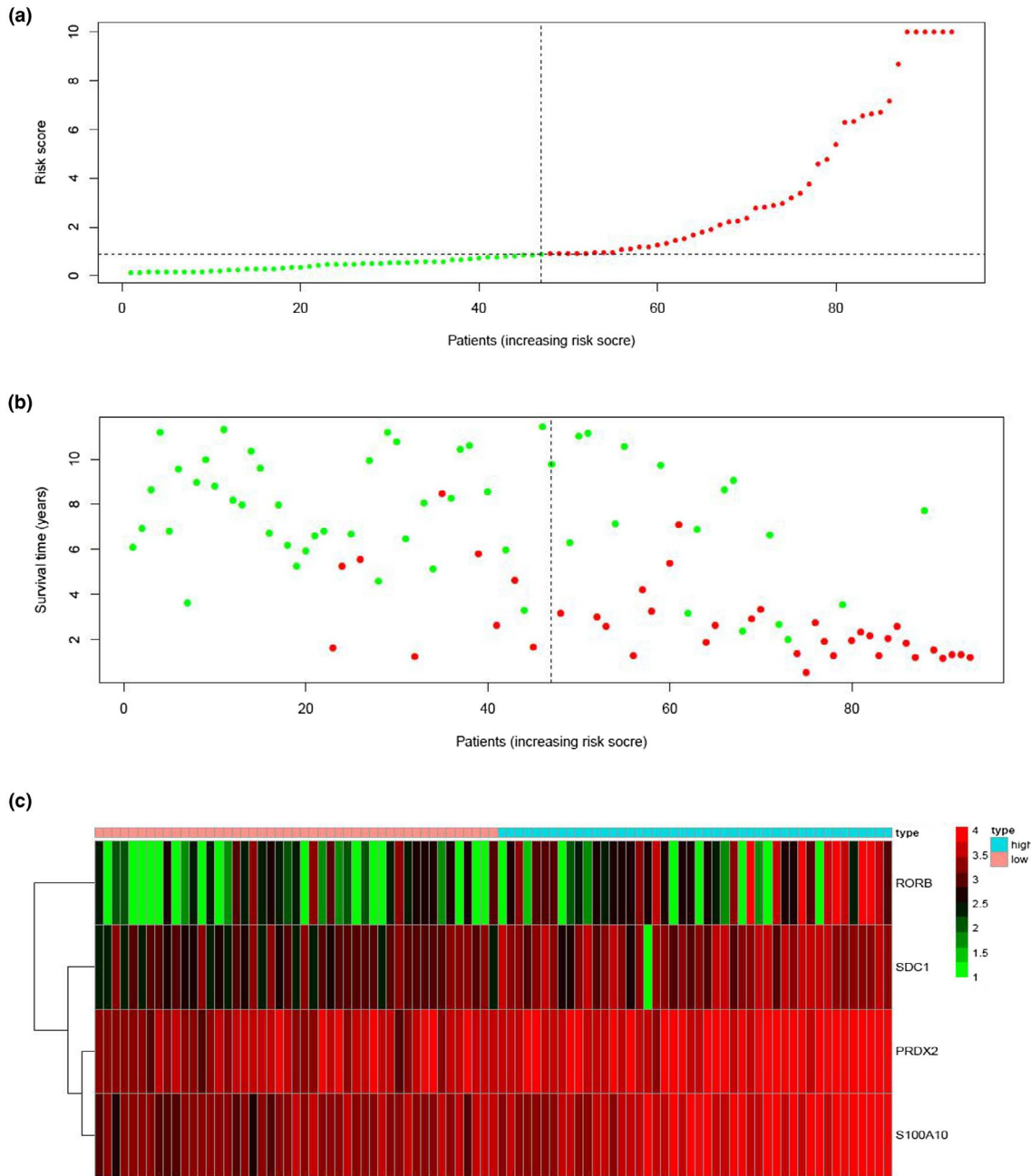
FIGURE 6 Prognosis analysis of training cohort. (a) Risk score distribution of patients based on the prognostic risk model. (b) Survival status of patients in different groups. (c) Heatmap of expression profiles of risk genes

in the entire TARGET cohort ($p < .05$). These results suggested that our prognostic risk model can be independently used to predict the prognosis of pediatric ALL patients. In addition, the variables of age, MRD status at day 29 also had important prognostic value in the multivariate analysis ($p < .05$).

4 | DISCUSSION

Although cure rate of pediatric ALL have improved recently, some patients still suffer from relapse and refractory. With the

development of second-generation sequencing, researchers expect to improve clinical outcomes through more accurate risk stratification and molecular-targeted therapies (Tasian & Hunger, 2017). Studies have shown that gene expression profiles of patients with relapse leukemia are highly enriched in immune-related processes (Toffalori et al., 2019). In addition, some tumor relapses are associated with cancer cells mimicking the IRGs of healthy cells (van der Bruggen et al., 1991; Knuth, Danowski, Oettgen, & Old, 1984; Old, 1981; Sahin et al., 1995; Schreiber, Old, & Smyth, 2011). Therefore, immune-related biomarkers may be an important indicator of prognosis in pediatric ALL.
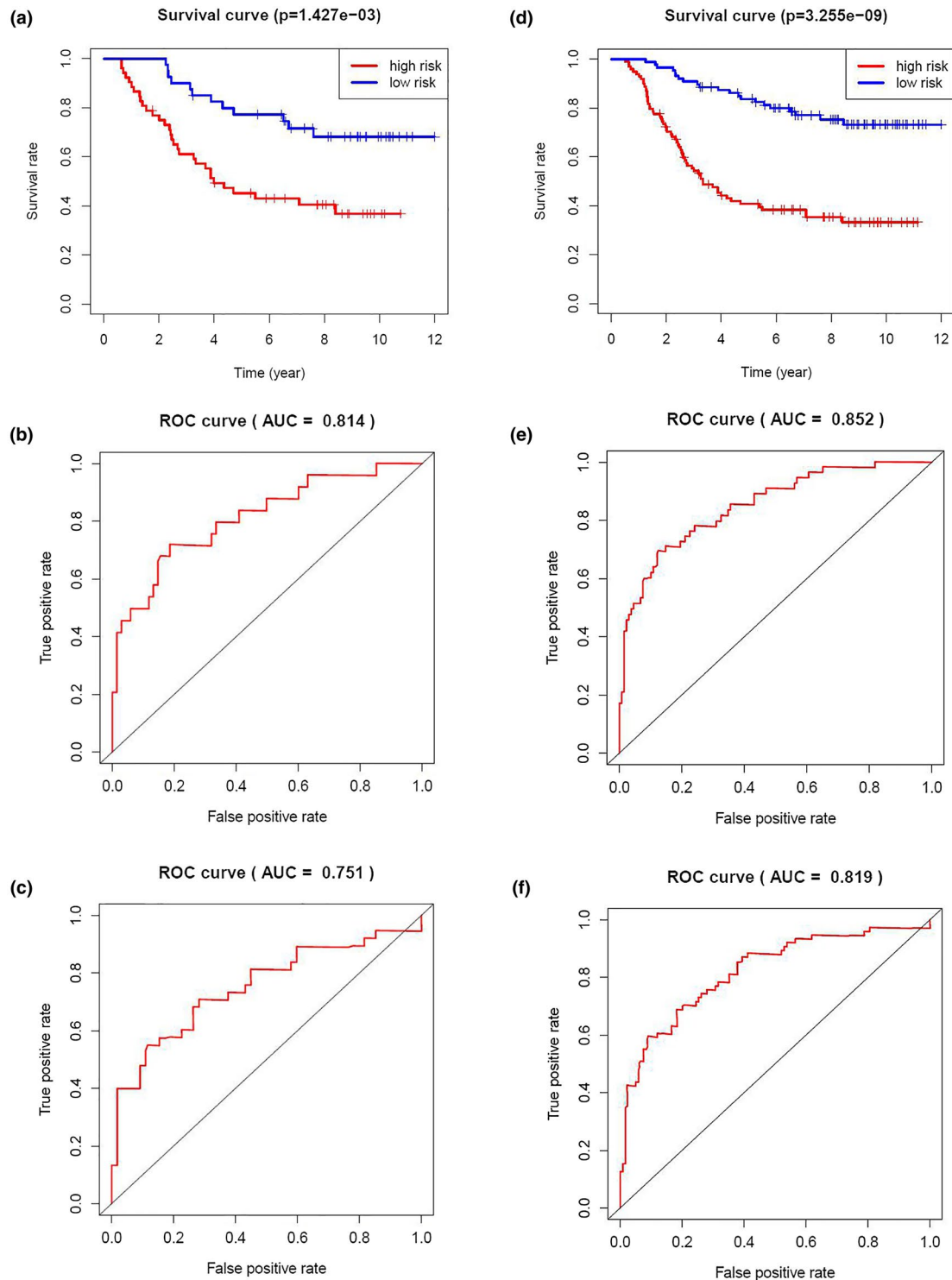
FIGURE 7 Prognosis analysis of testing cohort and entire TARGET cohort. Kaplan–Meier curve analysis of the high-risk and low-risk groups ((a) for testing cohort, (d) for entire TARGET cohort). Time-dependent ROC curve analysis for the predictive accuracy of the risk model for 3-year ((b) for testing cohort, (e) for entire TARGET cohort) and 5-year OS ((c) for testing cohort, (f) for entire TARGET cohort). ROC, receiver operating characteristic; OS, overall survival; TARGET, Therapeutically Applicable Research to Generate Effective Treatments

We analyzed the differential immune gene expression between the relapse and non-relapse groups of 251 pediatric ALL patients and screened 130 DEIRGs. Pathway enrichment analysis was performed to explore the potential biological mechanisms of them. And the top three pathways were NABA matrisome-associated, chemotaxis, and antimicrobial humoral response, which were reported to be involved in tumor development (Chen et al., 2009; Naba et al., 2012; Shields et al., 2007). Based on comprehensive analysis, we identified four optimal genes (*PRDX2*, *S100A10*, *RORB*, and
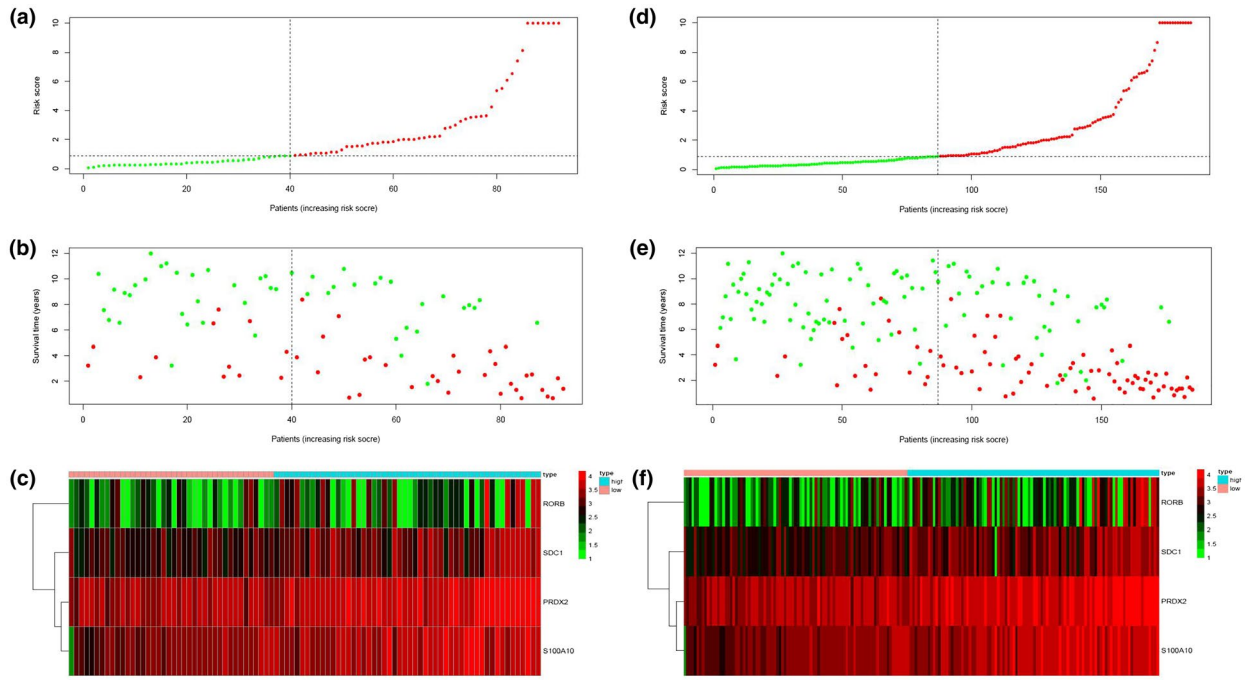
FIGURE 8 Prognosis analysis of testing cohort and entire TARGET cohort. Risk score distribution of patients based on the prognostic risk model ((a) for testing cohort, (d) for entire TARGET cohort). Survival status of patients in different groups ((b) for testing cohort, (e) for entire TARGET cohort). Heatmap of expression profiles of risk genes ((c) for testing cohort, (f) for entire TARGET cohort). TARGET, Therapeutically Applicable Research to Generate Effective Treatments

TABLE 2 Univariate and multivariate Cox regression analyses of the entire TARGET cohort

	Univariate analysis		Multivariate analysis	
Variables	**HR (95% CI)**	**_p_ value**	**HR (95% CI)**	**_p_ value**
Risk score (from risk model)				
High versus low	1.25 (1.19–1.32)	6.07E-19	1.29 (1.22–1.36)	4.29E-20
Age at diagnosis				
≥10 versus <10 years old	2.15 (1.38–3.34)	6.72E-04	2.24 (1.43–3.51)	4.27E-04
Gender				
Male versus female	1.40 (0.90–2.18)	0.131	1.69 (1.07–2.65)	2.37E-02
WBC at diagnosis				
> 50 versus ≤50 × 10^9/L	0.93 (0.59–1.47)	0.751	0.98 (0.61–1.57)	0.927
CNS status at diagnosis				
CNS3/CNS2 versus CNS1	1.05 (0.60–1.84)	0.871	1.05 (0.59–1.87)	0.87
MRD day 29				
≥10^{-4} versus < 10^{-4}	1.80 (1.15–2.79)	9.46E-03	2.06 (1.30–3.25)	2.09E-03

Note: MRD day 29, minimal residual disease status in bone marrow, by flow cytometry, at day 29 of induction therapy.

Abbreviations: CI, confidence interval; CNS, central nervous system; CNS1: no lymphoblasts in CSF; CNS2: present lymphoblasts in CSF, WBC count of the CSF < 5 cells/μl; CNS3: present lymphoblasts in CSF or a cranial nerve palsy, WBC count of the CSF ≥ 5 cells/μl; HR, hazard ratio; MRD, minimal residual disease; TARGET, Therapeutically Applicable Research to Generate Effective Treatments; WBC, white blood cell count.

SDC1) and used them to conduct a prognostic risk model for pediatric ALL patients. The model was able to classify pediatric ALL patients into two subgroups with statistically different survival outcomes, which were validated in both the testing cohort and the entire TARGET cohort. In addition, we verified and analyzed the risk score distribution, survival status, and risk gene expression of testing cohort and entire TARGET cohort. We came to the conclusion that low-risk group had lower levels of the risk gene than high-risk group, which is similar to that of the training cohort. These results suggest that the model may represent the risk status of pediatric ALL patients and provide reliable prognostic value for them. And the multivariate Cox regression analysis confirmed that our model could independently predict the prognosis of pediatric ALL patients.

We identified four optimal signatures from IRGs: *PRDX2*, *S100A10*, *RORB,* and *SDC1*. *PRDX2* can regulate oxidative and metabolic stress, whose carcinogenic role in several solid cancers has been reported (Kim et al., 2000; Stresing et al., 2013). *PRDX2* has also been shown to induce the growth of lymphoma cells (Trzeciecka et al., 2016). And *S100A10* can promote the invasion and metastasis of cancer by increasing the production of fibrinolytic enzyme (Choi, Fogg, Yoon, & Waisman, 2003; Madureira et al., 2016; O'Connell, Madureira, Berman, Liwski, & Waisman, 2011; Zhang, Fogg, & Waisman, 2004). *RORB* regulates Wnt pathway activity, which may be correlated with tumorigenesis and tumor stages (Wen et al., 2017). In addition, *SDC1* has been reported to play an important role in the malignant progression of tumors (Li et al., 2019). At present, no reports concerning these genes were published in ALL, so the role of them in pediatric ALL needs further investigation.

Many researches focused on the relapse and prognosis of leukemia. Cristina Toffalori et al. found that the gene expression profile of patients with recurrence was highly enriched in immune-related processes by analyzing the genome of patients with acute myeloid leukemia transplantation, and frequent new genomic changes in patients who relapsed after transplantation were observed (Toffalori et al., 2019), which was consistent with Miguel et al's report (Waterhouse et al., 2011). Joanna et al. suggested that the most striking characteristics were pathways leading to drug-resistant phenotypes in ALL relapsed patients with high-resolution genomic techniques, which can be targeted to prevent or treat relapse (Pierro, Hogan, Bhatla, & Carroll, 2017). Plenty of studies on prognosis of leukemia, nevertheless, no immune gene-related prognostic research of pediatric ALL has been carried out. Therefore, we focused on the pediatric ALL sample data from public dataset TARGET, which includes comprehensive clinical information and sequencing data. We used multiple algorithms (including univariate Cox, multivariate Cox, and Lasso regression) at the genome-wide level to construct a risk model for predicting the prognosis of pediatric ALL patients.

And the model was successfully verified in testing cohort and entire TARGET cohort. Therefore, the research data are comprehensive and research method is reliable. Our predictive model can represent the risk status and provide reliable prognostic value for the whole cohort and subgroups of pediatric ALL patients. Still, our survey has some limitations. We used retrospective data that were not validated in prospective clinical trials. In addition, further studies are needed on the mechanism by which IRGs affect pediatric ALL prognosis.

5 | CONCLUSIONS

In conclusion, we identified and verified four risk signatures based on IRGs. Then based them a risk model for pediatric ALL patients was developed, which can classify patients into high-risk and low-risk groups. These findings may provide insights for predicting clinical outcomes and individualized treatment based on risk scores.

CONFLICT OF INTERESTS
The authors declare that they have no competing interests.

AUTHOR CONTRIBUTIONS
XQ and NZ designed the study and downloaded the data. XQ performed the data analyses. XQ, YC, and HZ wrote the manuscript. JD revised the manuscript. All authors read and approved the final manuscript.

DATA AVAILABILITY STATEMENT
Expression data can be found in the Therapeutically Applicable Research to Generate Effective Treatments (TARGET) portal (https://ocg.cancer.gov/programs/target) and immune-related gene sets can be accessed from the ImmPort database (https://www.immport.org/home).

ORCID
Xi Quan https://orcid.org/0000-0003-1829-2649

REFERENCES
Bai, J., Zhang, X., Xiang, Z. X., Zhong, P. Y., & Xiong, B. (2020). Identification of prognostic immune-related signature predicting the overall survival for colorectal cancer. *European Review for Medical and Pharmacological Sciences*, *24*(3), 1134–1141. https://doi.org/10.26355/eurrev_202002_20164

Bhattacharya, S., Andorf, S., Gomes, L., Dunn, P., Schaefer, H., Pontius, J., … Butte, A. J. (2014). ImmPort: Disseminating data to the public for the future of immunology. *Immunologic Research*, *58*(2–3), 234–239. https://doi.org/10.1007/s12026-014-8516-1

Brassesco, M. S., Pezuk, J. A., Cortez, M. A., Bezerra, S. K., Scrideli, C. A., & Tone, L. G. (2018). TLE1 as an indicator of adverse prognosis in pediatric acute lymphoblastic leukemia. *Leukemia Research*, *74*, 42–46. https://doi.org/10.1016/j.leukres.2018.09.010

Chen, J. Y., Lin, W. J., Wu, J. L., Her, G. M., & Hui, C. F. (2009). Epinecidin-1 peptide induces apoptosis which enhances antitumor effects in human leukemia U937 cells. *Peptides*, *30*(12), 2365–2373. https://doi.org/10.1016/j.peptides.2009.08.019

Cho, S. H., Pak, K., Jeong, D. C., Han, M. E., Oh, S. O., & Kim, Y. H. (2019). The AP2M1 gene expression is a promising biomarker for predicting survival of patients with hepatocellular carcinoma. *Journal of Cellular Biochemistry*, *120*(3), 4140–4146. https://doi.org/10.1002/jcb.27699

Choi, K. S., Fogg, D. K., Yoon, C. S., & Waisman, D. M. (2003). p11 regulates extracellular plasmin production and invasiveness of HT1080 fibrosarcoma cells. *The FASEB Journal*, *17*(2), 235–246. https://doi.org/10.1096/fj.02-0697com

Han, M. E., Kim, J. Y., Kim, G. H., Park, S. Y., Kim, Y. H., & Oh, S. O. (2018). SAC3D1: A novel prognostic marker in hepatocellular carcinoma. *Scientific Reports*, *8*(1), 15608. https://doi.org/10.1038/s41598-018-34129-9

Jones, L., Carol, H., Evans, K., Richmond, J., Houghton, P. J., Smith, M. A., & Lock, R. B. (2016). A review of new agents evaluated against pediatric acute lymphoblastic leukemia by the Pediatric Preclinical Testing Program. *Leukemia*, *30*(11), 2133–2141. https://doi.org/10.1038/leu.2016.192

Kang, H., Chen, I.-M., Wilson, C. S., Bedrick, E. J., Harvey, R. C., Atlas, S. R., … Willman, C. L. (2010). Gene expression classifiers for relapse-free survival and minimal residual disease improve risk classification and outcome prediction in pediatric B-precursor acute lymphoblastic leukemia. *Blood*, *115*(7), 1394–1405. https://doi.org/10.1182/blood-2009-05-218560

Kato, M., & Manabe, A. (2018). Treatment and biology of pediatric acute lymphoblastic leukemia. *Pediatrics International*, *60*(1), 4–12. https://doi.org/10.1111/ped.13457

Kim, H., Lee, T. H., Park, E. S., Suh, J. M., Park, S. J., Chung, H. K., … Shong, M. (2000). Role of peroxiredoxins in regulating intracellular hydrogen peroxide and hydrogen peroxide-induced apoptosis in thyroid cells. *Journal of Biological Chemistry*, *275*(24), 18266–18270. https://doi.org/10.1074/jbc.275.24.18266

Knuth, A., Danowski, B., Oettgen, H. F., & Old, L. J. (1984). T-cell-mediated cytotoxicity against autologous malignant melanoma: Analysis with interleukin 2-dependent T-cell cultures. *Proceedings of the National Academy of Sciences of the United States of America*, *81*(11), 3511–3515. https://doi.org/10.1073/pnas.81.11.3511

Li, G. M., Zhang, C. L., Rui, R. P., Sun, B., & Guo, W. (2018). Bioinformatics analysis of common differential genes of coronary artery disease and ischemic cardiomyopathy. *European Review for Medical and Pharmacological Sciences*, *22*(11), 3553–3569. https://doi.org/10.26355/eurrev_201806_15182

Li, K., Li, L., Wu, X., Yu, J., Ma, H., Zhang, R., … Wang, W. (2019). Loss of SDC1 expression is associated with poor prognosis of colorectal cancer patients in Northern China. *Disease Markers*, *2019*, 1–7. https://doi.org/10.1155/2019/3768708

Madureira, P. A., Bharadwaj, A. G., Bydoun, M., Garant, K., O'Connell, P., Lee, P., & Waisman, D. M. (2016). Cell surface protease activation during RAS transformation: Critical role of the plasminogen receptor, S100A10. *Oncotarget*, *7*(30), 47720–47737. https://doi.org/10.18632/oncotarget.10279

Naba, A., Clauser, K. R., Hoersch, S., Liu, H., Carr, S. A., & Hynes, R. O. (2012). The matrisome: In silico definition and in vivo characterization by proteomics of normal and tumor extracellular matrices. *Molecular & Cellular Proteomics: MCP*, *11*(4), M111–M14647. https://doi.org/10.1074/mcp.M111.014647

O'Connell, P. A., Madureira, P. A., Berman, J. N., Liwski, R. S., & Waisman, D. M. (2011). Regulation of S100A10 by the PML-RAR-alpha oncoprotein. *Blood*, *117*(15), 4095–4105. https://doi.org/10.1182/blood-2010-07-298851

Old, L. J. (1981) Cancer immunology: the search for specificity–G. H. A. Clowes Memorial lecture. *Cancer Research*, *41*(2), 361–375.

Pierro, J., Hogan, L. E., Bhatla, T., & Carroll, W. L. (2017). New targeted therapies for relapsed pediatric acute lymphoblastic leukemia. *Expert Review of Anticancer Therapy*, *17*(8), 725–736. https://doi.org/10.1080/14737140.2017.1347507

Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, *26*(1), 139–140. https://doi.org/10.1093/bioinformatics/btp616

Sahin, U., Tureci, O., Schmitt, H., Cochlovius, B., Johannes, T., Schmits, R., … Pfreundschuh, M. (1995). Human neoplasms elicit multiple specific immune responses in the autologous host. *Proceedings of the National Academy of Sciences of the United States of America*, *92*(25), 11810–11813. https://doi.org/10.1073/pnas.92.25.11810

Santiago, R., Vairy, S., Sinnett, D., Krajinovic, M., & Bittencourt, H. (2017). Novel therapy for childhood acute lymphoblastic leukemia. *Expert Opinion on Pharmacotherapy*, *18*(11), 1081–1099. https://doi.org/10.1080/14656566.2017.1340938

Schreiber, R. D., Old, L. J., & Smyth, M. J. (2011). Cancer immunoediting: Integrating immunity's roles in cancer suppression and promotion. *Science*, *331*(6024), 1565–1570. https://doi.org/10.1126/science.1203486

Shields, J. D., Fleury, M. E., Yong, C., Tomei, A. A., Randolph, G. J., & Swartz, M. A. (2007). Autologous chemotaxis as a mechanism of tumor cell homing to lymphatics via interstitial flow and autocrine CCR7 signaling. *Cancer Cell*, *11*(6), 526–538. https://doi.org/10.1016/j.ccr.2007.04.020

Song, Q., Shang, J., Yang, Z., Zhang, L., Zhang, C., Chen, J., & Wu, X. (2019). Identification of an immune signature predicting prognosis risk of patients in lung adenocarcinoma. *Journal of Translational Medicine*, *17*(1), 70. https://doi.org/10.1186/s12967-019-1824-4

Stresing, V., Baltziskueta, E., Rubio, N., Blanco, J., Arriba, M. C., Valls, J., … Sierra, A. (2013). Peroxiredoxin 2 specifically regulates the oxidative and metabolic stress response of human metastatic breast cancer cells in lungs. *Oncogene*, *32*(6), 724–735. https://doi.org/10.1038/onc.2012.93

Tasian, S. K., & Hunger, S. P. (2017). Genomic characterization of paediatric acute lymphoblastic leukaemia: An opportunity for precision medicine therapeutics. *British Journal of Haematology*, *176*(6), 867–882. https://doi.org/10.1111/bjh.14474

Toffalori, C., Zito, L., Gambacorta, V., Riba, M., Oliveira, G., Bucci, G., … Vago, L. (2019). Immune signature drives leukemia escape and relapse after hematopoietic cell transplantation. *Nature Medicine*, *25*(4), 603–611. https://doi.org/10.1038/s41591-019-0400-z

Trzeciecka, A., Klossowski, S., Bajor, M., Zagozdzon, R., Gaj, P., Muchowicz, A., … Firczuk, M. (2016). Dimeric peroxiredoxins are druggable targets in human Burkitt lymphoma. *Oncotarget*, *7*(2), 1717–1731. https://doi.org/10.18632/oncotarget.6435

van der Bruggen, P., Traversari, C., Chomez, P., Lurquin, C., De Plaen, E., Van den Eynde, B., … Boon, T. (1991). A gene encoding an antigen recognized by cytolytic T lymphocytes on a human melanoma. *Science*, *254*(5038), 1643–1647. https://doi.org/10.1126/science.1840703

Wang, Z., Song, Q., Yang, Z., Chen, J., Shang, J., & Ju, W. (2019). Construction of immune-related risk signature for renal papillary cell carcinoma. *Cancer Medicine*, *8*(1), 289–304. https://doi.org/10.1002/cam4.1905

Waterhouse, M., Pfeifer, D., Pantic, M., Emmerich, F., Bertz, H., & Finke, J. (2011). Genome-wide profiling in AML patients relapsing after allogeneic hematopoietic cell transplantation. *Biology of Blood and Marrow Transplantation*, *17*(10), 1450–1459. https://doi.org/10.1016/j.bbmt.2011.07.012

Wen, Z., Pan, T., Yang, S., Liu, J., Tao, H., Zhao, Y., … Zhu, Y. (2017). Up-regulated NRIP2 in colorectal cancer initiating cells modulates the Wnt pathway by targeting RORbeta. *Molecular Cancer*, *16*(1), 20. https://doi.org/10.1186/s12943-017-0590-2

Zeiser, R., & Vago, L. (2019). Mechanisms of immune escape after allogeneic hematopoietic cell transplantation. *Blood*, *133*(12), 1290–1297. https://doi.org/10.1182/blood-2018-10-846824

Zhang, L., Fogg, D. K., & Waisman, D. M. (2004). RNA interference-mediated silencing of the S100A10 gene attenuates plasmin generation and invasiveness of Colo 222 colorectal cancer cells. *Journal of Biological Chemistry*, *279*(3), 2053–2062. https://doi.org/10.1074/jbc.M310357200

Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A. H., Tanaseichuk, O., … Chanda, S. K. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nature Communications*, *10*(1), 1523. https://doi.org/10.1038/s41467-019-09234-6