Hindawi
Journal of Oncology
Volume 2022, Article ID 5896022, 13 pages
https://doi.org/10.1155/2022/5896022

Research Article

PROM1 and *CTGF* Expression in Childhood *MLL*-Rearrangement Acute Lymphoblastic Leukemia

Lu-lu Wang, Xue Tang, Guichi Zhou, Shilin Liu, Ying Wang, Fen Chen, Tonghui Li, Feiqiu Wen, Sixi Liu, and Huirong Mai

Department of Hematology and Oncology, Shenzhen Children's Hospital, Shenzhen 518038, China

Correspondence should be addressed to Huirong Mai; maihuirong@163.com

Received 24 March 2022; Revised 28 August 2022; Accepted 20 September 2022; Published 14 October 2022

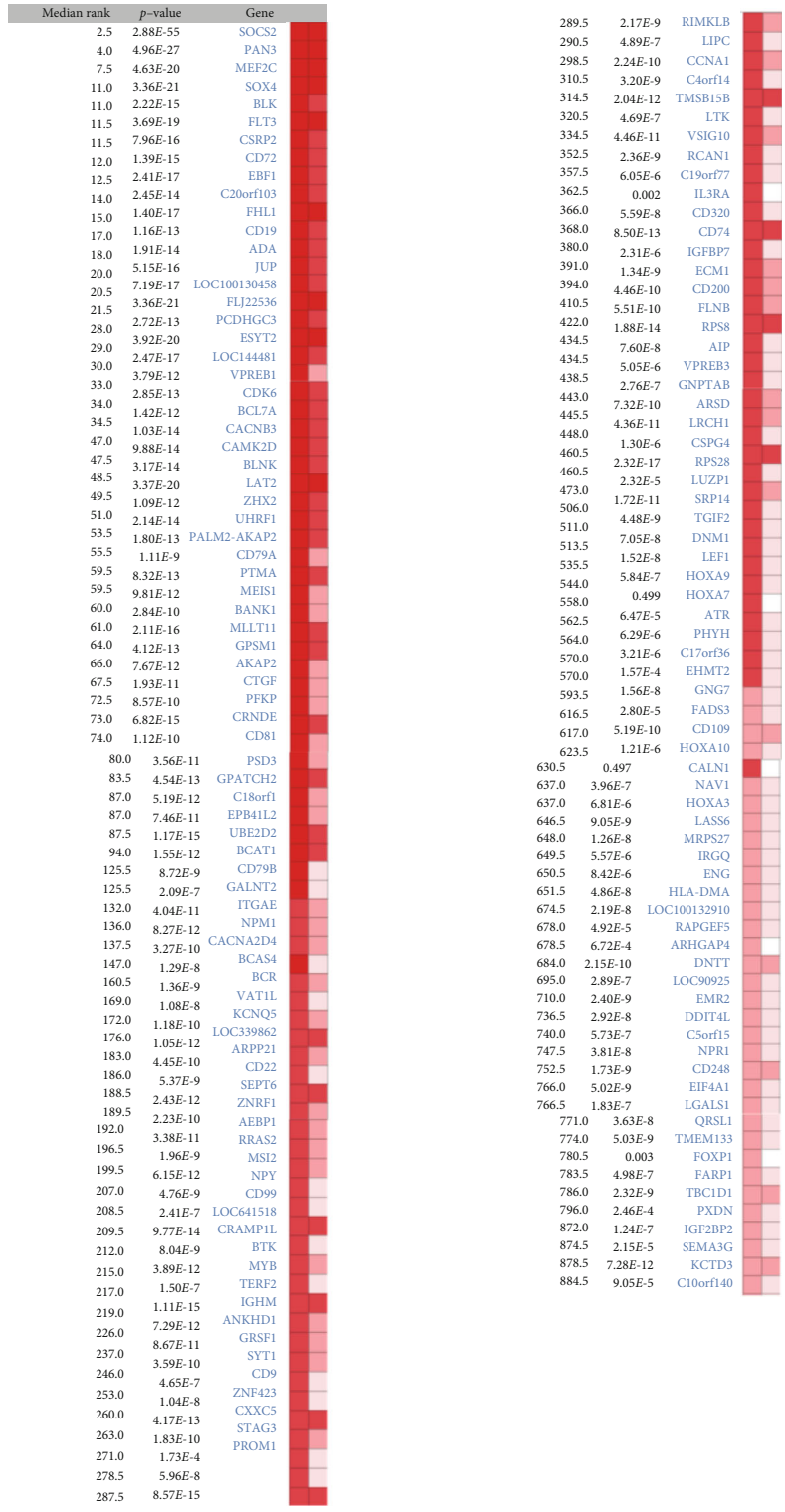
Academic Editor: Zhen-Jian Zhuo

The prognosis of over 90% of infant acute lymphoblastic leukemia (ALL) remains poor because of harboring the mixed-lineage leukemia gene (*MLL*) fusion. To give insight into the critical coexpressed genes related to the *MLL*-rearrangement (*MLL*-R) gene in childhood acute lymphoblastic leukemia, we integrated different bioinformatic methods. First, the gene expression data of *MLL*-R ALL and normal samples from GSE13159 and GSE13164 were analyzed using "compare" function in the Oncomine database. The top 150 overexpressed and 150 underexpressed genes were identified by the Oncomine website. Then, we employed the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) to define functional genes for the 300 DEGs. The Cytoscape identified two important networks for overexpressed genes, including 35 functional genes, among which *PROM1*, *FLT3*, *CTGF*, *LGALS1*, *IGFBP7*, *ZNRF1*, and *RUNX2* were considered as the key genes because of their high expression in *MLL*-R ALL compared to the expression in other subclassification of leukemia in the MILE dataset. Further analysis of GSE68720, GSE19475, and Therapeutically Applicable Research to Generate Effective Treatments (TARGET) ALL (phase I) database confirmed the robust expression of 7 key genes in *MLL*-R compared to MLL-germline (*MLL*-G) childhood ALL. Kaplan-Meier analysis indicated that childhood ALL patients with high *PROM1* and *CTGF* expression had significantly poor overall survival. These findings suggest that *PROM1* and *CTGF* represent two potential therapeutic targets for childhood *MLL*-R ALL.

1. Introduction

Acute lymphoblastic leukemia (ALL) is the most common form of childhood malignancies. It is a heterogeneous hematologic disease characterized by clonal proliferation of immature lymphoid progenitor cells both in bone marrow and extramedullary sites [1]. Thanks to the development of risk-directed chemotherapy and targeted therapy against the gene mutations/fusion, the 5-year survival rate of ALL exceeds 90% [2, 3]. However, the prognosis of over 90% of infant ALL and 35–50% of childhood acute myeloid leukemia remains poor because of harboring the mixed-lineage leukemia gene (*MLL*) fusion [4–8]. For infant *MLL*-rearrangement (*MLL*-R) ALL, the 5-year event-free survival is extremely low, ranging from 20 to 40% [6]. *MLL*-R ALL has unique clinical and biologic features, including the pro-B phenotype, prenatal origin, rapid onset, early relapse, and hyperleukocytosis.

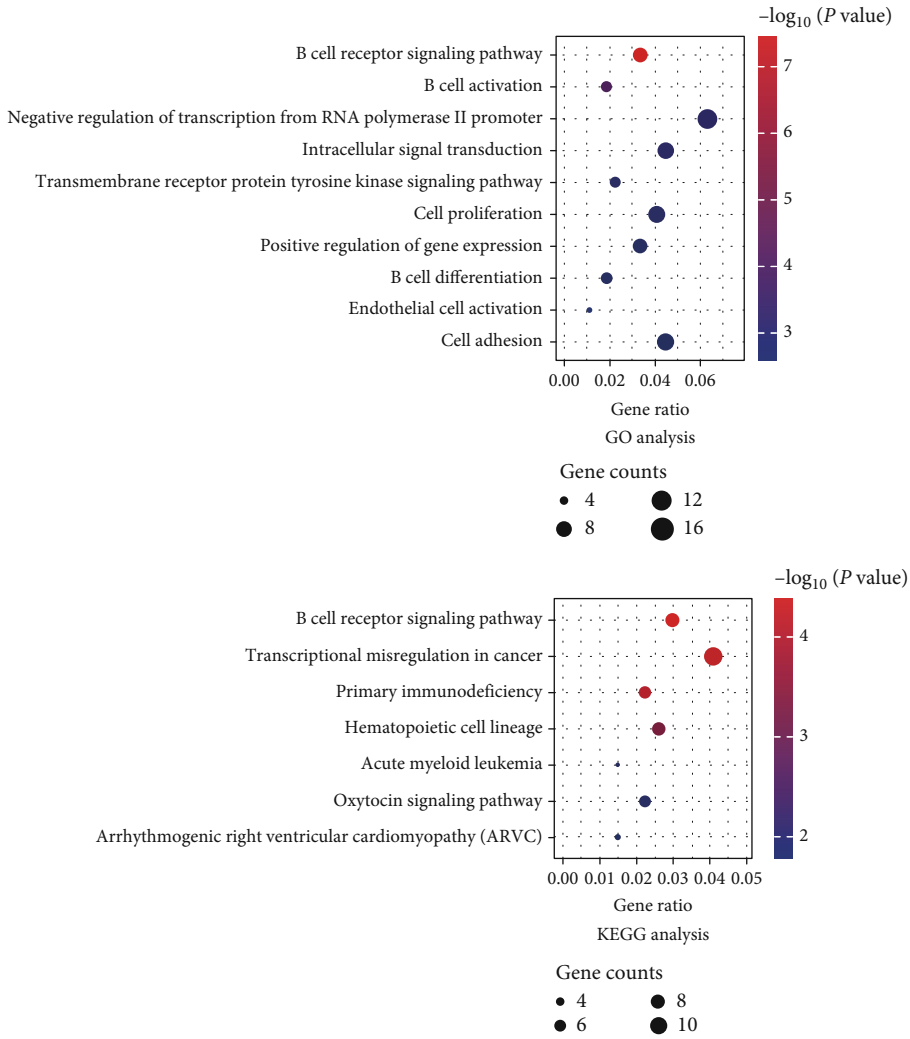
The MLL gene located in chromosome 11q23 fuses to generate chimeric genes with over 80 partners at the C-terminus and forms 135 different MLL rearrangements, of which the most common ones are AF4, AF9, AF17, ELL, and ENL [9]. These fusions are responsible for the gene expression alternation on histone methylation and transcriptional elongation. *MLL*-R activates target genes via H3K79 methylation by DOT1L, stimulation of elongation through P-TEFb, and suppression of the polycomb function [10]. However, as the breakthrough of genome-wide sequencing, a group of MLL target genes was distinguished. It has been reported that MLL fusion genes act as a global regulator by



Median rank	p–value	Gene
2.5	2.88E-55	SOCS2
4.0	4.96E-27	PAN3
7.5	4.63E-20	MEF2C
11.0	3.36E-21	SOX4
11.0	2.22E-15	BLK
11.5	3.69E-19	FLT3
11.5	7.96E-16	CSRP2
12.0	1.39E-15	CD72
12.5	2.41E-17	EBF1
14.0	2.45E-14	C20orf103
15.0	1.40E-17	FHL1
17.0	1.16E-13	CD19
18.0	1.91E-14	ADA
20.0	5.15E-16	JUP
20.5	7.19E-17	LOC100130458
21.5	3.36E-21	FLJ22536
28.0	2.72E-13	PCDHGC3
29.0	3.92E-20	ESYT2
30.0	2.47E-17	LOC144481
33.0	3.79E-12	VPREB1
34.0	2.85E-13	CDK6
34.5	1.42E-12	BCL7A
47.0	1.03E-14	CACNB3
47.5	9.88E-14	CAMK2D
48.5	3.17E-14	BLNK
49.5	3.37E-20	LAT2
51.0	1.09E-12	ZHX2
53.5	2.14E-14	UHRF1
55.5	1.80E-13	PALM2-AKAP2
59.5	1.11E-9	CD79A
59.5	8.32E-13	PTMA
60.0	9.81E-12	MEIS1
61.0	2.84E-10	BANK1
64.0	2.11E-16	MLLT11
66.0	4.12E-13	GPSM1
67.5	7.67E-12	AKAP2
72.5	1.93E-11	CTGF
73.0	8.57E-10	PFKP
74.0	6.82E-15	CRNDE
80.0	1.12E-10	CD81
83.5	3.56E-11	PSD3
87.0	4.54E-13	GPATCH2
87.0	5.19E-12	C18orf1
87.5	7.46E-11	EPB41L2
94.0	1.17E-15	UBE2D2
125.5	1.55E-12	BCAT1
125.5	8.72E-9	CD79B
132.0	2.09E-7	GALNT2
136.0	4.04E-11	ITGAE
137.5	8.27E-12	NPM1
147.0	3.27E-10	CACNA2D4
160.5	1.29E-8	BCAS4
169.0	1.36E-9	BCR
172.0	1.08E-8	VAT1L
176.0	1.18E-10	KCNQ5
183.0	1.05E-12	LOC339862
186.0	4.45E-10	ARPP21
188.5	5.37E-9	CD22
189.5	2.43E-12	SEPT6
192.0	2.23E-10	ZNRF1
196.5	3.38E-11	AEBP1
199.5	1.96E-9	RRAS2
207.0	6.15E-12	MSI2
208.5	4.76E-9	NPY
209.5	2.41E-7	CD99
212.0	9.77E-14	LOC641518
215.0	8.04E-9	CRAMP1L
217.0	3.89E-12	BTK
219.0	1.50E-7	MYB
226.0	1.11E-15	TERF2
237.0	7.29E-12	IGHM
246.0	8.67E-11	ANKHD1
253.0	3.59E-10	GRSF1
260.0	4.65E-7	SYT1
263.0	1.04E-8	CD9
271.0	4.17E-13	ZNF423
278.5	1.83E-10	CXXC5
287.5	1.73E-4	STAG3
289.5	5.96E-8	PROM1
290.5	8.57E-15	
298.5	2.17E-9	RIMKLB
310.5	4.89E-7	LIPC
314.5	2.24E-10	CCNA1
320.5	3.20E-9	C4orf14
334.5	2.04E-12	TMSB15B
352.5	4.69E-7	LTK
357.5	4.46E-11	VSIG10
362.5	2.36E-9	RCAN1
366.0	6.05E-6	C19orf77
368.0	0.002	IL3RA
380.0	5.59E-8	CD320
391.0	8.50E-13	CD74
394.0	2.31E-6	IGFBP7
410.5	1.34E-9	ECM1
422.0	4.46E-10	CD200
434.5	5.51E-10	FLNB
434.5	1.88E-14	RPS8
438.5	7.60E-8	AIP
443.0	5.05E-6	VPREB3
445.5	2.76E-7	GNPTAB
448.0	7.32E-10	ARSD
460.5	4.36E-11	LRCH1
460.5	1.30E-6	CSPG4
473.0	2.32E-17	RPS28
506.0	2.32E-5	LUZP1
511.0	1.72E-11	SRP14
513.5	4.48E-9	TGIF2
535.5	7.05E-8	DNM1
544.0	1.52E-8	LEF1
558.0	5.84E-7	HOXA9
562.5	0.499	HOXA7
564.0	6.47E-5	ATR
570.0	6.29E-6	PHYH
570.0	3.21E-6	C17orf36
593.5	1.57E-4	EHMT2
616.5	1.56E-8	GNG7
617.0	2.80E-5	FADS3
623.5	5.19E-10	CD109
630.5	1.21E-6	HOXA10
637.0	0.497	CALN1
637.0	3.96E-7	NAV1
646.5	6.81E-6	HOXA3
648.0	9.05E-9	LASS6
649.5	1.26E-8	MRPS27
650.5	5.57E-6	IRGQ
651.5	8.42E-6	ENG
674.5	4.86E-8	HLA-DMA
678.0	2.19E-8	LOC100132910
678.5	4.92E-5	RAPGEF5
684.0	6.72E-4	ARHGAP4
695.0	2.15E-10	DNTT
710.0	2.89E-7	LOC90925
736.5	2.40E-9	EMR2
740.0	2.92E-8	DDIT4L
747.5	5.73E-7	C5orf15
752.5	3.81E-8	NPR1
766.0	1.73E-9	CD248
766.5	5.02E-9	EIF4A1
771.0	1.83E-7	LGALS1
774.0	3.63E-8	QRSL1
780.5	5.03E-9	TMEM133
783.5	0.003	FOXP1
786.0	4.98E-7	FARP1
796.0	2.32E-9	TBC1D1
872.0	2.46E-4	PXDN
874.5	1.24E-7	IGF2BP2
878.5	2.15E-5	SEMA3G
884.5	7.28E-12	KCTD3
	9.05E-5	C10orf140

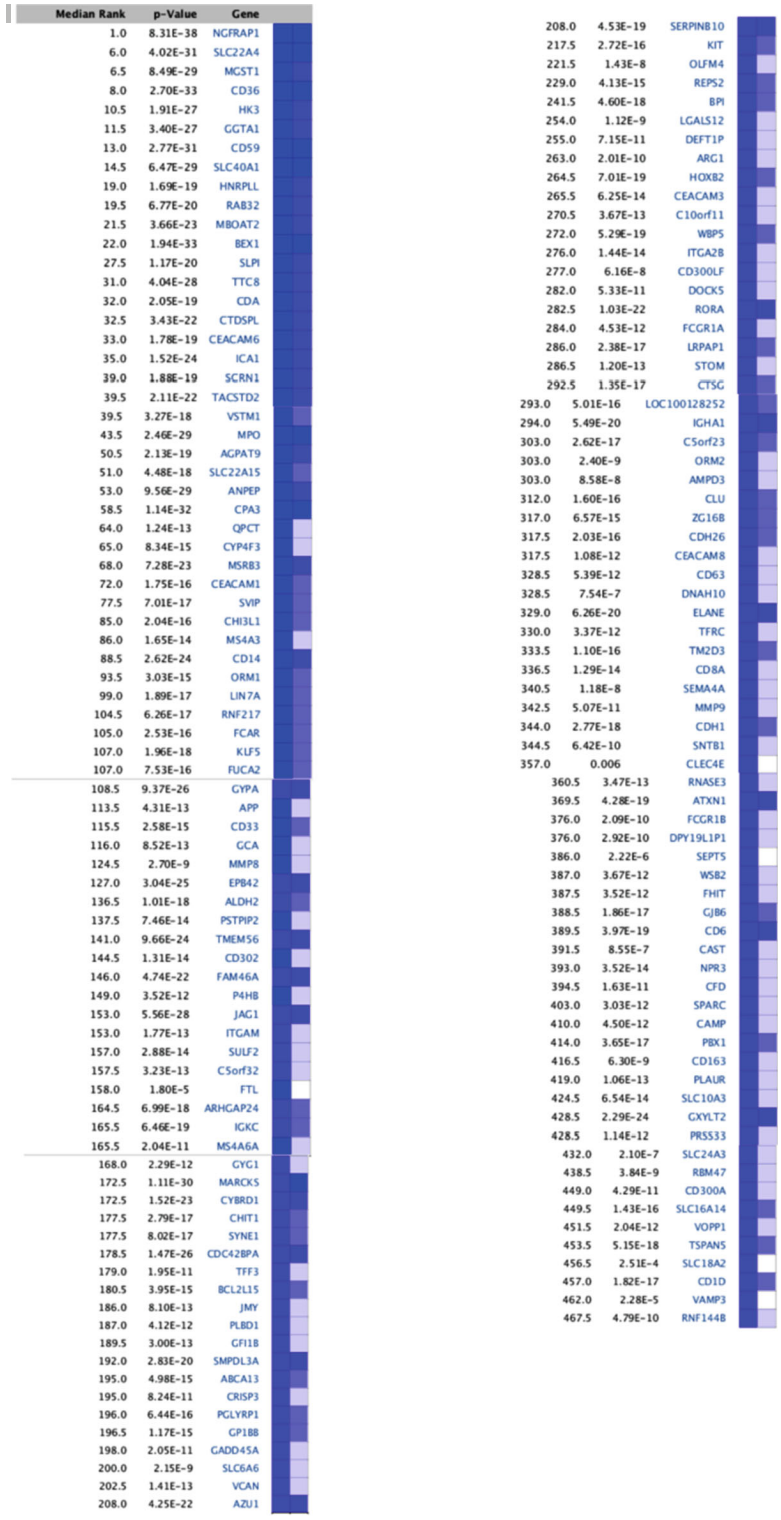
(a)

Figure 1: Continued.



(b)

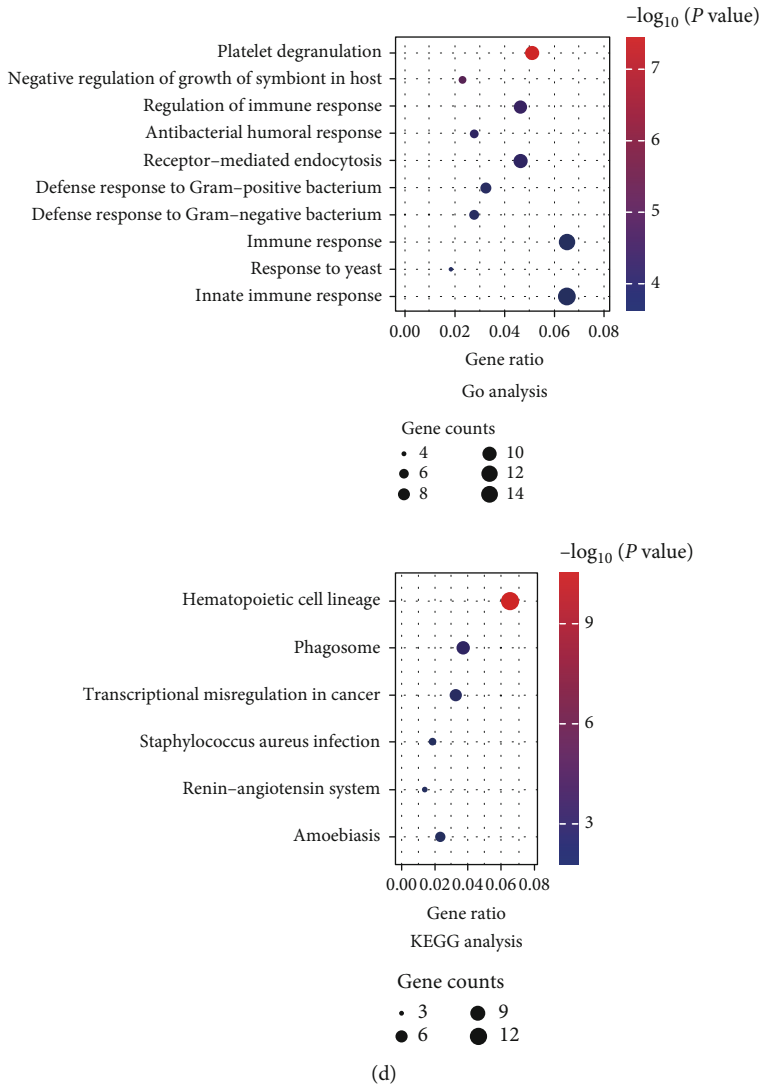
Figure 1: Continued.



Median Rank	p-Value	Gene
1.0	8.31E-38	NGFRAP1
6.0	4.02E-31	SLC22A4
6.5	8.49E-29	MGST1
8.0	2.70E-33	CD36
10.5	1.91E-27	HK3
11.5	3.40E-27	GGTA1
13.0	2.77E-31	CD59
14.5	6.47E-29	SLC40A1
19.0	1.69E-19	HNRPLL
19.5	6.77E-20	RAB32
21.5	3.66E-23	MBOAT2
22.0	1.94E-33	BEX1
27.5	1.17E-20	SLPI
31.0	4.04E-28	TTC8
32.0	2.05E-19	CDA
32.5	3.43E-22	CTDSPL
33.0	1.78E-19	CEACAM6
35.0	1.52E-24	ICA1
39.0	1.88E-19	SCRN1
39.5	2.11E-22	TACSTD2
39.5	3.27E-18	VSTM1
43.5	2.46E-29	MPO
50.5	2.13E-19	AGPAT9
51.0	4.48E-18	SLC22A15
53.0	9.56E-29	ANPEP
58.5	1.14E-32	CPA3
64.0	1.24E-13	QPCT
65.0	8.34E-15	CYP4F3
68.0	7.28E-23	MSRB3
72.0	1.75E-16	CEACAM1
77.5	7.01E-17	SVIP
85.0	2.04E-16	CHI3L1
86.0	1.65E-14	MS4A3
88.5	2.62E-24	CD14
93.5	3.03E-15	ORM1
99.0	1.89E-17	LIN7A
104.5	6.26E-17	RNF217
105.0	2.53E-16	FCAR
107.0	1.96E-18	KLF5
107.0	7.53E-16	FUCA2
108.5	9.37E-26	GYPA
113.5	4.31E-13	APP
115.5	2.58E-15	CD33
116.0	8.52E-13	GCA
124.5	2.70E-9	MMP8
127.0	3.04E-25	EPB42
136.5	1.01E-18	ALDH2
137.5	7.46E-14	PSTPIP2
141.0	9.66E-24	TMEM56
144.5	1.31E-14	CD302
146.0	4.74E-22	FAM46A
149.0	3.52E-12	P4HB
153.0	5.56E-28	JAG1
153.0	1.77E-13	ITGAM
157.0	2.88E-14	SULF2
157.5	3.23E-13	C5orf32
158.0	1.80E-5	FTL
164.5	6.99E-18	ARHGAP24
165.5	6.46E-19	IGKC
165.5	2.04E-11	MS4A6A
168.0	2.29E-12	GYG1
172.5	1.11E-30	MARCKS
172.5	1.52E-23	CYBRD1
177.5	2.79E-17	CHIT1
177.5	8.02E-17	SYNE1
178.5	1.47E-26	CDC42BPA
179.0	1.95E-11	TFF3
180.5	3.95E-15	BCL2L15
186.0	8.10E-13	JMY
187.0	4.12E-12	PLBD1
189.5	3.00E-13	GFI1B
192.0	2.83E-20	SMPDL3A
195.0	4.98E-15	ABCA13
195.0	8.24E-11	CRISP3
196.0	6.44E-16	PGLYRP1
196.5	1.17E-15	GP1BB
198.0	2.05E-11	GADD45A
200.0	2.15E-9	SLC6A6
202.5	1.41E-13	VCAN
208.0	4.25E-22	AZU1
208.0	4.53E-19	SERPINB10
217.5	2.72E-16	KIT
221.5	1.43E-8	OLFM4
229.0	4.13E-15	REPS2
241.5	4.60E-18	BPI
254.0	1.12E-9	LGALS12
255.0	7.15E-11	DEFT1P
263.0	2.01E-10	ARG1
264.5	7.01E-19	HOXB2
265.5	6.25E-14	CEACAM3
270.5	3.67E-13	C10orf11
272.0	5.29E-19	WBP5
276.0	1.44E-14	ITGA2B
277.0	6.16E-8	CD300LF
282.0	5.33E-11	DOCK5
282.5	1.03E-22	RORA
284.0	4.53E-12	FCGR1A
286.0	2.38E-17	LRPAP1
286.5	1.20E-13	STOM
292.5	1.35E-17	CTSG
293.0	5.01E-16	LOC100128252
294.0	5.49E-20	IGHA1
303.0	2.62E-17	C5orf23
303.0	2.40E-9	ORM2
303.0	8.58E-8	AMPD3
312.0	1.60E-16	CLU
317.0	6.57E-15	ZG16B
317.5	2.03E-16	CDH26
317.5	1.08E-12	CEACAM8
328.5	5.39E-12	CD63
328.5	7.54E-7	DNAH10
329.0	6.26E-20	ELANE
330.0	3.37E-12	TFRC
333.5	1.10E-16	TM2D3
336.5	1.29E-14	CD8A
340.5	1.18E-8	SEMA4A
342.5	5.07E-11	MMP9
344.0	2.77E-18	CDH1
344.5	6.42E-10	SNTB1
357.0	0.006	CLEC4E
360.5	3.47E-13	RNASE3
369.5	4.28E-19	ATXN1
376.0	2.09E-10	FCGR1B
376.0	2.92E-10	DPY19L1P1
386.0	2.22E-6	SEPT5
387.0	3.67E-12	WSB2
387.5	3.52E-12	FHIT
388.5	1.86E-17	GJB6
389.5	3.97E-19	CD6
391.5	8.55E-7	CAST
393.0	3.52E-14	NPR3
394.5	1.63E-11	CFD
403.0	3.03E-12	SPARC
410.0	4.50E-12	CAMP
414.0	3.65E-17	PBX1
416.5	6.30E-9	CD163
419.0	1.06E-13	PLAUR
424.5	6.54E-14	SLC10A3
428.5	2.29E-24	GXYLT2
428.5	1.14E-12	PRSS33
432.0	2.10E-7	SLC24A3
438.5	3.84E-9	RBM47
449.0	4.29E-11	CD300A
449.5	1.43E-16	SLC16A14
451.5	2.04E-12	VOPP1
453.5	5.15E-18	TSPAN5
456.5	2.51E-4	SLC18A2
457.0	1.82E-17	CD1D
462.0	2.28E-5	VAMP3
467.5	4.79E-10	RNF144B

(c)

Figure 1: Continued.



(d)

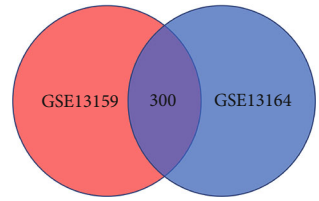
Figure 1: The expression profiles of *MLL*-R pro-B ALL in the MILE study. (a) Top 150 overexpressed DEGs were shown in the table. (b) GO and KEGG enrichment analyses for top 150 overexpressed DEGs. (c) Top 150 underexpressed DEGs were shown in the table. (d) GO and KEGG enrichment analyses for top 150 underexpressed DEGs.

targeting more than 5000 genomic elements [11]. By far, the association between coexpressed genes and the MLL fusion gene has not been comprehensively investigated.

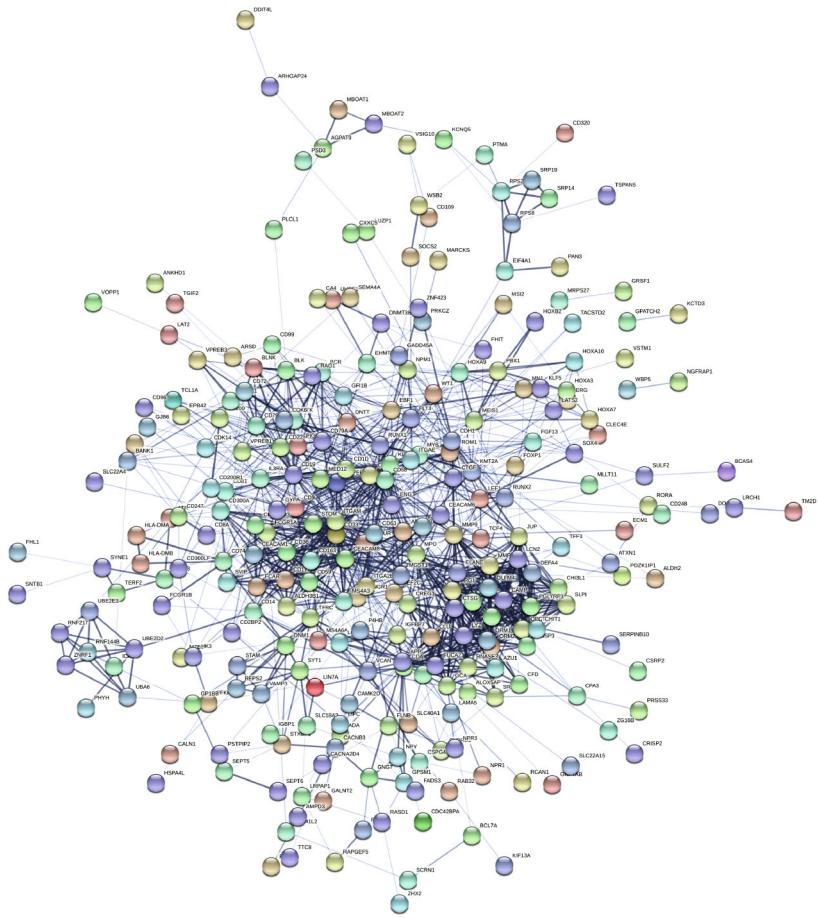
To better understand the whole-genome alteration of leukemia, a retrospective study named Microarray Innovations in LEukemia (MILE) was carried out in 11 laboratories across three continents and included 3334 patients with leukemia [12, 13]. Blood or bone marrow samples of acute and chronic leukemia patients were hybridized to the microarray analysis. On the Gene Expression Omnibus (GEO) website, the MILE study fell into two stages, GSE13159 and GSE13164. In this study, we explored the GSE13159 and GSE13164 datasets on the Oncomine website and defined the top 300 differentiated expressed genes (DEGs) of *MLL*-R pro-B ALL vs. normal samples. Then, we performed Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis for selected DEGs. Moreover, we investigated their protein-protein interaction (PPI) network based on the STRING website and selected functional genes by using Cytoscape software. The 7 key gene expression pattern and their relationship with clinical traits were searched on BloodSpot, and the UCSC Xena website was also constructed. Finally, two GEO datasets, including GSE68720 and GSE19475, studying the infant MLL-R and *MLL*-germline (*MLL*-G) ALL were employed to confirm the key genes. Exploring new genes and pathways associated with *MLL*-R ALL may help to identify potential molecular mechanisms, diagnostic markers, and therapeutic targets for *MLL*-R ALL.

2. Materials and Methods

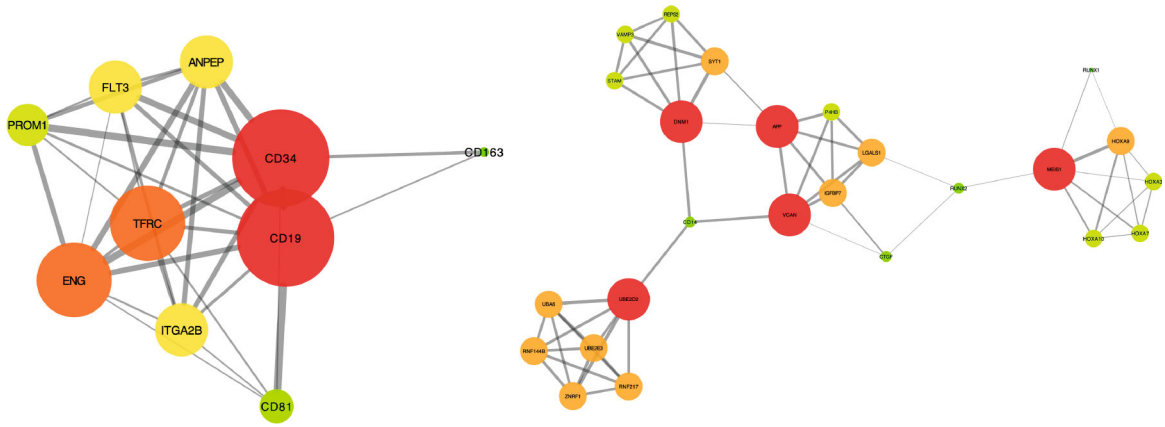
2.1. Oncomine Analysis. Oncomine is an integrated data-mining platform that analyzes previously published or open-access cancer microarray data. Using the keywords "acute lymphoblastic leukemia" and "Cancer vs. Normal



(a)



(b)



(c)

Figure 2: Continued.

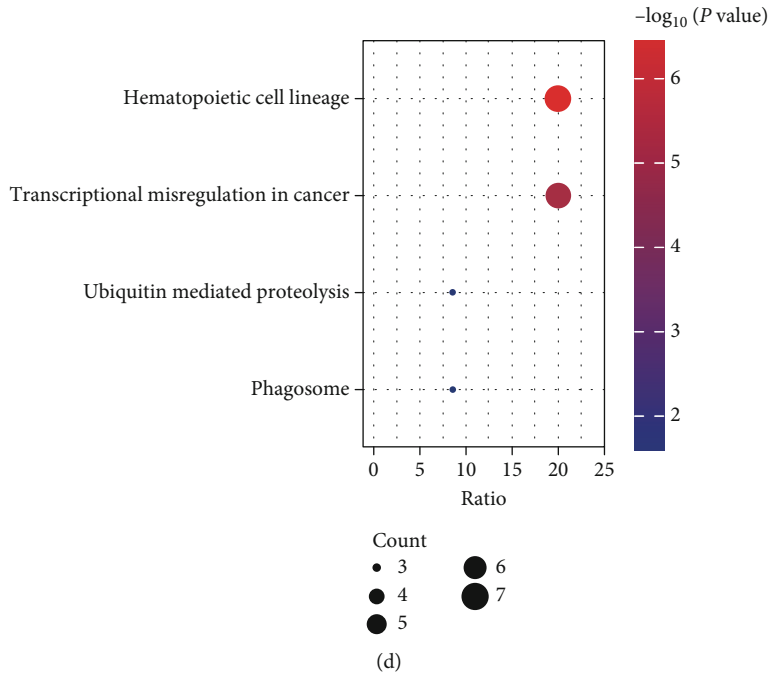


Figure 2: PPI network of DEGs by STRING. (a) The Venn diagram showed the top 300 DEGs. (b) The PPI network was constructed by STRING based on the top 300 DEGs. (c) The functional genes of overexpressed DEGs found by MCODE made up 2 critical subnetworks. (d) KEGG pathway analysis for functional genes.

Analysis," two studies were identified in the Oncomine database (https://www.oncomine.org) with the ID GSE13159 and GSE13164. Gene expression in pro-B ALL vs. normal was analyzed by the "compare" function in the Oncomine database.

According to the description of MILE, all of the pro-B ALL patients harbored MLL fusion in GSE13159 and GSE13164. The result orders genes by median rank across the two analyses and displays the corresponding p values. The overexpressed and underexpressed genes with rank orders above 150 and $p < 0.05$ were selected for further analysis.

2.2. GO and KEGG Enrichment Analyses. The top 150 over- and underexpressed genes were taken into DAVID website separately, analyzed by GO and KEGG enrichment ($p < 0.05$).

2.3. Protein-Protein Interaction Network. The 300 DEGs were taken into Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) with the maximum number of interactors = 0 and a confidence score ≥ 0.4 as the cutoff criteria. Then, to understand the function of the overexpressed gene, the biofunctional modules in the top 150 overexpressed genes were explored using a plug-in MCODE in Cytoscape with a node score cutoff of 0.2, degree cutoff of 2, and k-Core of 2. The top two gene modules with the highest MCODE scores were selected from the network. Then, the genes were taken into DAVID, as demonstrated above. KEGG enrichment analyses were carried out with the significance threshold $p < 0.05$.

2.4. BloodSpot Website Analysis. BloodSpot is a database of mRNA expression in healthy and malignant hematopoiesis and includes data from both humans and mice [14]. The functional gene names were input into the search bar as a query. Gene expression data of the MILE study were identified on the BloodSpot website.

2.5. UCSC Xena Analysis. The gene expression, MLL status, and minimal residual disease (MRD) monitor were verified and analyzed in TARGET ALL (phase I) using the UCSC Xena browser.

2.6. Data Collection and Gene Expression Analysis in the GEO Dataset. Microarray expression data of GSE68720 and GSE19475 were downloaded from the GEO database. To explore the relationship between infant *MLL*-R ALL and *MLL*-G ALL, cel files of 17 *MLL*-G ALL samples and 80 *MLL*-R ALL samples from GSE68720 and 14 *MLL*-G ALL and 58 *MLL*-R ALL samples from GSE19475 were selected. The robust multiarray average in R was applied to explore the gene expression data in the cel files, including background correction, normalization, and summarization. All of the above operations were run with scripts in the R 3.6.3 version. The ggplot2 package in R was used to show the heat map of key genes.

2.7. Kaplan-Meier Analysis. Gene expression was obtained from UCSC Xena website, and the clinical survival information of TARGET ALL (phase I) was downloaded from the official TARGET database website. The ggplot2 of R software was used to plot the Kaplan-Meier survival curve. The TARGET ALL (phase I) project is obtained from

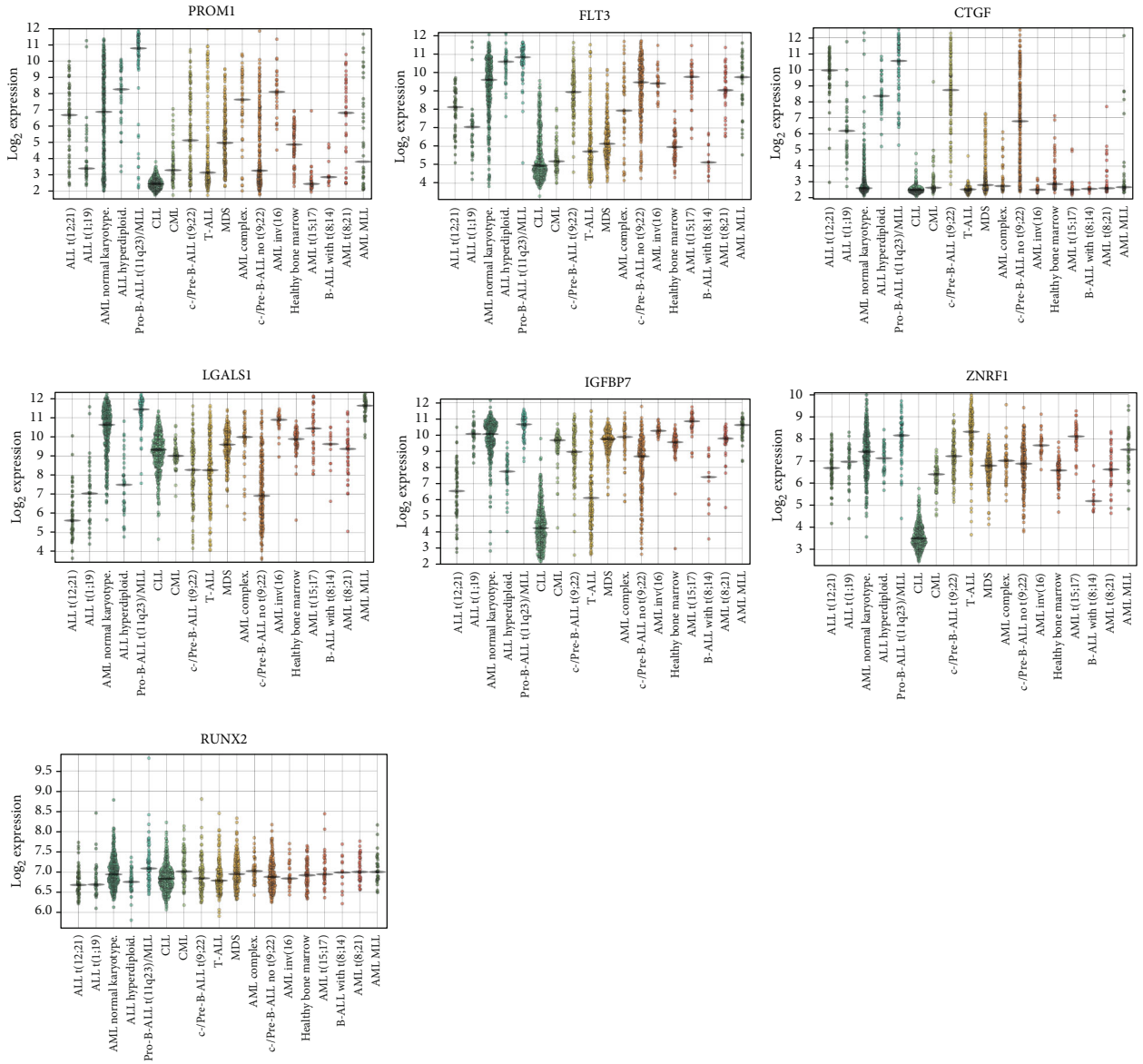


Figure 3: Key gene expression on the BloodSpot website. The box plot showed the expression of *PROM1*, *FLT3*, *CTGF*, *LGALS1*, *IGFBP7*, *ZNRF1*, and *RUNX2* in different subclassifications of ALL on the BloodSpot website.

patients enrolled on biology studies and clinical trials managed through the COG, POG 9906 (clinical trial for patients with newly diagnosed ALL between March 2000 and April 2003 that were defined as high risk for relapse). Patient samples for full characterization were chosen based on the following criteria: the disease onset at >9 years of age, did not have white blood cell count > 50000/μL, did not express the BCR/ABL fusion gene, were not known to be hypodiploid (DNA index > 0.95), and achieved remission (fewer than 5% blasts) following the standard two rounds of induction therapy. The primary patient samples were collected at diagnosis, and gene expression was analyzed following the protocol of Human Genome U133 Plus 2.0 Array (Affymetrix).

2.8. Statistical Analysis. Student t-test of variance was used for comparing the statistical differences of gene expression of samples in GSE19475 and GSE68720. All the analyzes were two sided and $p < 0.05$ was considered to be significant.

3. Results

3.1. Identification of the Top DEGs in MLL-r ALL. The gene expression data of *MLL*-R ALL and normal samples from GSE13159 and GSE13164 were analyzed using the "compare" function in the Oncomine database. The median rank of the overexpressed and underexpressed genes with rank orders above 150 was identified as the genes and selected for further analyses (Figures 1(a) and 1(c)).

Based on the result from the DAVID online analysis tool, the KEGG pathway and GO analysis were carried out to better understand the biological function of the key DEGs in *MLL*-R ALL. The GO enrichment analysis result showed that the overexpressed genes were mainly enriched in

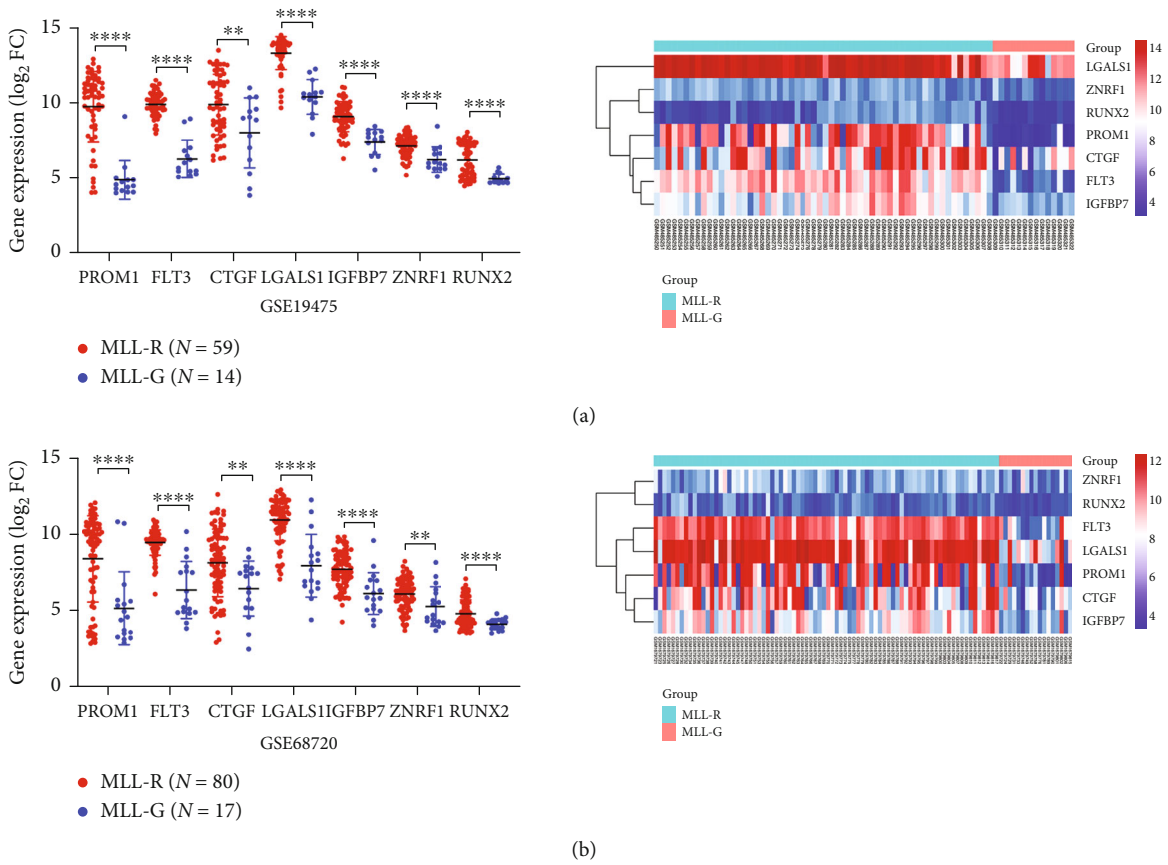


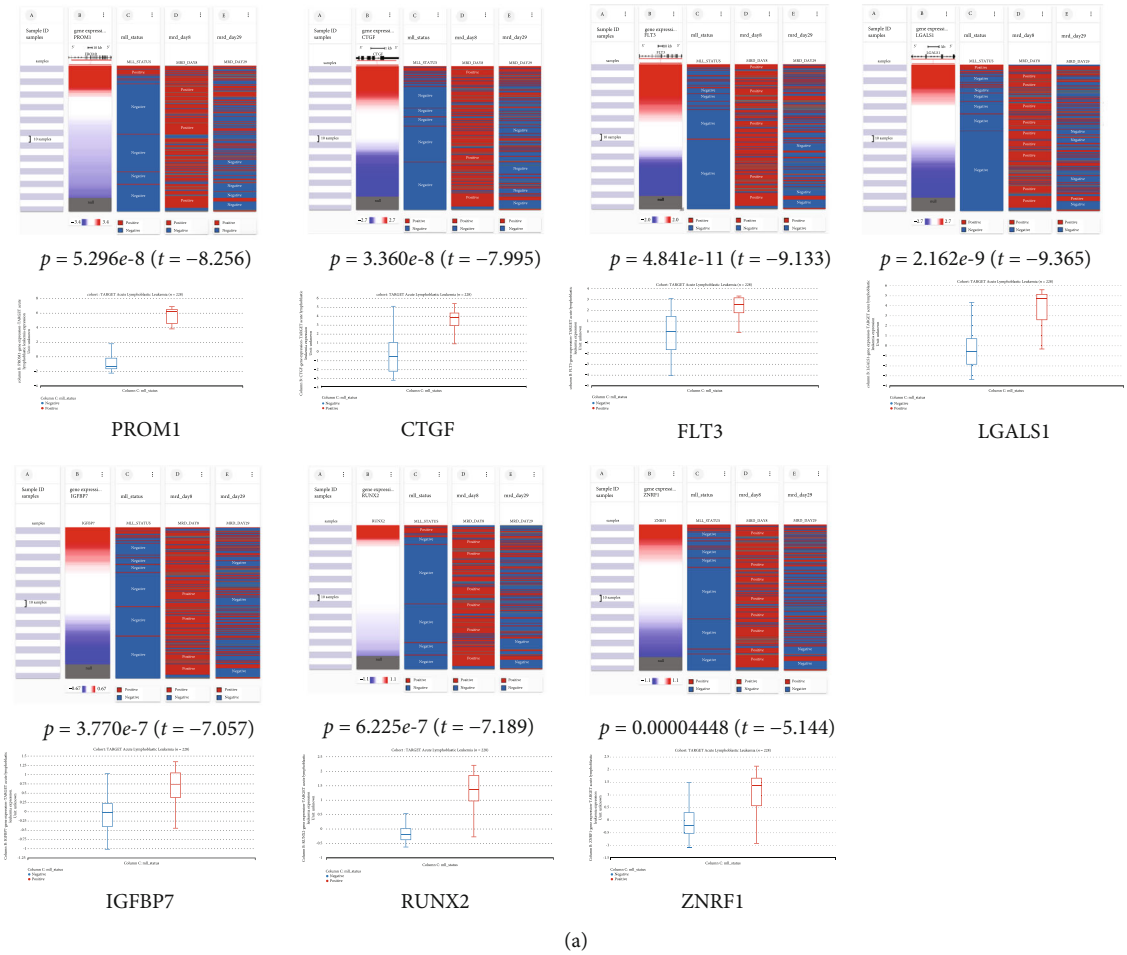
Figure 4: Key gene expression in *MLL*-R compared to *MLL*-G in GSE19475 and GSE68720. (a) Scatter plot and heat map of 7 key gene expression in GSE19475 according to the value of |logFC|. (b) Scatter plot and heat map of 7 key gene expression in GSE68720 according to the value of |logFC|.

biological processes, including the B cell receptor signaling pathway, B cell activation, and negative regulation of transcription from the RNA polymerase II promoter, while KEGG pathway analysis showed that the result was significantly enriched in the B cell receptor signaling pathway, transcriptional misregulation in cancer, and primary immunodeficiency (Figure 1(b)). As for underexpressed genes, GO enrichement analysis demonstrated that they were mainly enriched in platelet degranulation pathway. KEGG pathway analysis showed that the underexpressed genes were mainly enriched in the hematopoietic cell lineage pathway (Figure 1(d)).

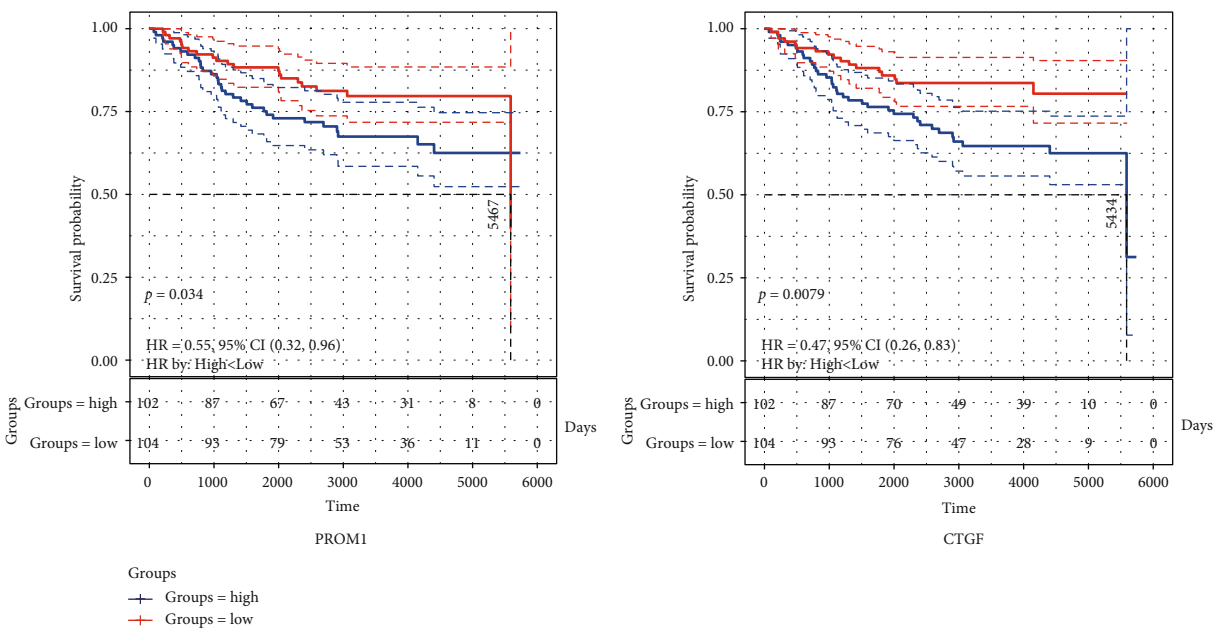
A functional gene usually refers to what is significant in regulation and biological processes and closely interacts with other genes in a network. A total of 300 DEGs, including 150 overexpressed and 150 underexpressed genes, were shown in the overlap of the Venn diagram (Figure 2(a)). To further investigate the function of the DEGs in the GSE13159 and GSE13164 at the protein level, the STRING was employed to screen for functional genes. The PPI network consisted of 295 nodes and 1378 edges (Figure 2(b)). Afterwards, the interactive relationship of overexpressed genes was analyzed separately in Cytoscape. The MCODE, a plug-in in Cytoscape, was employed to calculate the *k*-Core of each gene. The top two significant modules in MCODE with high scores were selected from the PPI network, including module A (MCODE score = 7.556 with 10 nodes) and module B (MCODE score = 4.75 with 25 nodes) (Figure 2(c)). These genes were involved in 4 important KEGG pathways, including the hematopoietic cell lineage, transcriptional misregulation in cancer, ubiquitin-mediated proteolysis, and phagosome (Figure 2(d)).

3.2. Validation of Key Genes in MLL-R ALL. To demonstrate the role of 35 functional genes in ALL subclassifications, we used the BloodSpot website to check their expression in different subclassifications of leukemia. As shown in Figure 3, *PROM1*, *FLT3*, *CTGF*, *LGALS1*, *IGFBP7*, *ZNRF1*, and *RUNX2* were found highly expressed in the *MLL*-R pro-B ALL compared to the other subclassification of leukemia.

To further verify the identified 7 key genes in *MLL*-R ALL, we detected the expression of *PROM1*, *FLT3*, *CTGF*, *LGALS1*, *IGFBP7*, *ZNRF1*, and *RUNX2* between *MLL*-R ALL and *MLL*-G ALL in GSE68720 and GSE19475 datasets by using the R software. In both GSE68720 and GSE19475 datasets, the 7 key genes were significantly overexpressed in *MLL*-R compared to the *MLL*-G ALL samples, especially for *PROM1*. The heat map of the 7 key genes were shown in Figures 4(a) and 4(b). Further analysis in UCSC Xena demonstrated that high expression of these genes was



(a)



(b)

Figure 5: Kaplan-Meier analysis of *PROM1* and *CTGF* expression in childhood ALL. (a) The relationship of key gene expression with the MLL status in the TARGET ALL (phase I) dataset. (b) Survival cure comparing patients with high (blue) vs. low (red) *PROM1* expression was plotted using a log-rank test (HR = 0.61 (0.43–0.87), $P = 0.034$), as well as the *CTGF* (HR = 0.47 (0.26–0.83), $P = 0.0079$).

significantly associated with the MLL status in the TARGET ALL (phase I) database, presenting a high correlation with the status of MLL fusion (Figure 5(a)). These results demonstrated that 7 key genes have extremely high expression in *MLL*-R ALL and maybe the critical targets for MLL fusion.

3.3. Survival Analysis of PROM1 and CTGF in Childhood ALL. To delineate the prognostic value of potential key genes, the overall survival analyses of 7 key gene expression were detected in the TARGET ALL (phase I). The result showed that a high expression level of *PROM1* and *CTGF* was associated with inferior overall survival of ALL (Figure 5(b)).

4. Discussion

Although studies have demonstrated numerous fusion partner proteins, the target genes of *MLL*-fusion and the molecular mechanism involved in target genes were poorly understood. In the past decade, genomic analyses have revolutionized our understanding of the coexpression network in *MLL*-R ALL. *HOX* cluster genes and its cofactor *MEIS1* were the most well-known target genes for the *MLL* fusion gene [15]. Both *HOXA* genes and *MEIS1* are highly expressed in the stem cells and early progenitor cells. MLL drives the proliferation and self-renewal of immature hematopoietic cells by upregulating posterior *HOX* genes and their cofactor *MEIS1* [16, 17]. Coincidentally, in this study, we examined the Oncomine website and investigated DEGs related to *MLL*-R ALL in the MILE study. Using PPI analysis, the critical pathway of functional genes was found involved in the hematopoietic cell lineage and transcriptional misregulation in cancer, including *HOXA10*, *MEIS1*, *FLT3*, *CD14*, *PROM1*, *RUNX2*, and *RUNX1* (data not shown), indicating the dominant roles of *HOXA* and *MEIS1* in *MLL*-R ALL.

Posttranslational modifications of PROM1 play a critical role in *MLL*-R ALL [18, 19]. It was reported that AF4 recruited and activated DOT1L at the H3K79me2/3 locus of the *PROM1* promoter, which is required for the growth of *MLL*-AF4 B-cell ALL cells [20–22]. CD133 is a kind of transmembrane glycoprotein encoded by the *PROM1* gene. It is associated with cancer stem cells in diverse human tumors, including brain, liver, stomach, endometrium, ovary, and colorectum and gliomas and medulloblastoma [23]. Recent studies demonstrated that CD19/CD133 tandem CAR T induces robust cytotoxicity against CD19+ CD133+ and CD19− CD133+ B-cell lines, suggesting CD133 a promising target *MLL*-R ALL immunotherapy [24]. However, this study was challenged by "on-target off-tumor" myeloablative and life-threatening toxicity, because the CD133 was expressed in the hematopoietic stem and progenitor cells [25].

CTGF, CCN2 as the official name, is an extracellular matrix- (ECM) associated protein of 36–38 kDa and a member of the CCN family of proteins. It plays a great role in cell adhesion, proliferation, migration, and differentiation and improves the development of numerous tumor metastases [26–29]. Interestingly, elevated *CTGF* expression is also a feature of precursor B-cell ALL [30–33]. By analyzing COG trial P9906, high expression of *BMPR1B*, *CTGF*, *TTYH2*, *IGJ*, *NT5E (CD73)*, *CDC42EP3*, and *TSPAN7* was found to be associated with poor outcomes in precursor-B ALL patients [34]. Ruling out the possibility of structure alternation, amplification, or base mutation, Welch et al. demonstrated that the *CTGF* locus is hypomethylated in pediatric pre-B ALL [35]. Anti-CTGF monoclonal antibody attenuated tumor growth of precursor-B ALL from pediatric patients propagated in mice [36]. Here in this study, *PROM1* and *CTGF* were overexpressed in *MLL*-R compared to *MLL*-G patients and those with high *PROM1* and *CTGF* expression had significantly poor OS (Figure 5(b)). Further in vitro, in vivo, and clinical studies are warranted to delineate the role of *PROM1* and *CTGF* in *MLL*-R ALL.

In conclusion, we first demonstrated the top DEGs of GSE13159 and GSE13164 by using the Oncomine website. After integrated analyses, we identified from the 300 DEG genes that *PROM1*, *FLT3*, *CTGF*, *LGALS1*, *IGFBP7*, *ZNRF1*, and *RUNX2* were the key genes, as they were highly expressed in *MLL*-R ALL compared to *MLL*-G ALL. Further investigation demonstrated that *PROM1* and *CTGF* were the poor prognostic markers for childhood *MLL*-R ALL. Thus, we provide an insight into ALL that *PROM1* and *CTGF* may be the novel potential target genes for the MLL fusion gene in childhood *MLL*-R ALL.

Abbreviations

ALL: Acute lymphoblastic leukemia
DEG: Differentiated expressed gene
GEO: Gene Expression Omnibus
GO: Gene Ontology
HR: Hazard ratio
KEGG: Kyoto Encyclopedia of Genes and Genomes
MILE: Microarray Innovations in LEukemia
MLL: Mixed-lineage leukemia gene
MLL-G: Mixed-lineage leukemia gene-germline
MLL-R: Mixed-lineage leukemia gene-rearrangement
OS: Overall survival
PPI: Protein-protein interaction
STRING: Search Tool for the Retrieval of Interacting Genes/Proteins
TARGET: Therapeutically Applicable Research to Generate Effective Treatments.

Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest

The authors have no conflicts of interest relevant to the article to disclose.

Authors' Contributions

Lu-lu Wang and Xue Tang contributed equally to this work and share first authorship.

Acknowledgments

We appreciate the GEO datasets (GSE13159, GSE13164, GSE68720, and GSE19475) for the free use. This work was supported by the Sanming Project of Medicine in Shenzhen (SZSM201512033), Guangdong Medical Science and Technology Research Project (A2020101), Shenzhen Fund for Guangdong Provincial High-Level Clinical Key Specialties (SZGSP012), Shenzhen Key Medical Discipline Construction Fund (SZXK034), and Shenzhen Healthcare Research Project (SZLY2018015).

References

[1] G. M. Dores, S. S. Devesa, R. E. Curtis, M. S. Linet, and L. M. Morton, "Acute leukemia incidence and patient survival among children and adults in the United States, 2001-2007," *Blood*, vol. 119, no. 1, pp. 34–43, 2012.

[2] C.-H. Pui, D. Campana, D. Pei et al., "Treating childhood acute lymphoblastic leukemia without cranial irradiation," *The New England Journal of Medicine*, vol. 360, no. 26, pp. 2730–2741, 2009.

[3] S. P. Hunger and C. G. Mullighan, "Acute lymphoblastic leukemia in children," *The New England Journal of Medicine*, vol. 373, no. 16, pp. 1541–1552, 2015.

[4] R. Pieters, M. Schrappe, P. De Lorenzo et al., "A treatment protocol for infants younger than 1 year with acute lymphoblastic leukaemia (Interfant-99): an observational study and a multicentre randomised trial," *The Lancet*, vol. 370, no. 9583, pp. 240–250, 2007.

[5] M. H. van der Linden, M. G. Valsecchi, P. De Lorenzo et al., "Outcome of congenital acute lymphoblastic leukemia treated on the Interfant-99 protocol," *Blood*, vol. 114, no. 18, pp. 3764–3768, 2009.

[6] J. M. Hilden, P. A. Dinndorf, S. O. Meerbaum et al., "Analysis of prognostic factors of acute lymphoblastic leukemia in infants: report on CCG 1953 from the Children's Oncology Group," *Blood*, vol. 108, no. 2, pp. 441–451, 2006.

[7] C. Meyer, J. Hofmann, T. Burmeister et al., "The _MLL_ recombinome of acute leukemias in 2013," *Leukemia*, vol. 27, no. 11, pp. 2165–2176, 2013.

[8] C. Bueno, R. Montes, P. Catalina, R. Rodríguez, and P. Menendez, "Insights into the cellular origin and etiology of the infant pro-B acute lymphoblastic leukemia with MLL-AF4 rearrangement," *Leukemia*, vol. 25, no. 3, pp. 400–410, 2011.

[9] C. Meyer, T. Burmeister, D. Gröger et al., "The _MLL_ recombinome of acute leukemias in 2017," *Leukemia*, vol. 32, no. 2, pp. 273–284, 2018.

[10] A. C. Winters and K. M. Bernt, "MLL-rearranged leukemias-an update on science and clinical approaches," *Frontiers in Pediatrics*, vol. 5, p. 4, 2017.

[11] M. G. Guenther, R. G. Jenner, B. Chevalier et al., "Global and Hox-specific roles for the MLL1 methyltransferase," *Proceedings of the National Academy of Sciences*, vol. 102, no. 24, pp. 8603–8608, 2005.

[12] A. Kohlmann, T. J. Kipps, L. Z. Rassenti et al., "An international standardization programme towards the application of gene expression profiling in routine leukaemia diagnostics: the Microarray Innovations in LEukemia study prephase," *British Journal of Haematology*, vol. 142, no. 5, pp. 802–807, 2008.

[13] T. Haferlach, A. Kohlmann, L. Wieczorek et al., "Clinical utility of microarray-based gene expression profiling in the diagnosis and subclassification of leukemia: report from the International Microarray Innovations in Leukemia Study Group," *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, vol. 28, no. 15, pp. 2529–2537, 2010.

[14] F. O. Bagger, S. Kinalis, and N. Rapin, "BloodSpot: a database of healthy and malignant haematopoiesis updated with purified and single cell mRNA sequencing profiles," *Nucleic Acids Research*, vol. 47, no. D1, pp. D881–D885, 2019.

[15] Z. Li, R. T. Luo, S. Mi et al., "Consistent deregulation of gene expression between human and murine MLL rearrangement leukemias," *Cancer Research*, vol. 69, no. 3, pp. 1109–1116, 2009.

[16] J. Faber, A. V. Krivtsov, M. C. Stubbs et al., "HOXA9 is required for survival in human MLL-rearranged acute leukemias," *Blood*, vol. 113, no. 11, pp. 2375–2385, 2009.

[17] A. R. Kumar, Q. Li, W. A. Hudson et al., "A role for MEIS1 in MLL-fusion gene leukemia," *Blood*, vol. 113, no. 8, pp. 1756–1758, 2009.

[18] G.-Y. Liou, "CD133 as a regulator of cancer metastasis through the cancer stem cells," *The International Journal of Biochemistry & Cell Biology*, vol. 106, pp. 1–7, 2019.

[19] C. Wuchter, R. Ratei, G. Spahn et al., "Impact of CD133 (AC133) and CD90 expression analysis for acute leukemia immunophenotyping," *Haematologica*, vol. 86, no. 2, pp. 154–161, 2001.

[20] A. V. Krivtsov, Z. Feng, M. E. Lemieux et al., "H3K79 methylation profiles define murine and human MLL-AF4 leukemias," *Cancer Cell*, vol. 14, no. 5, pp. 355–368, 2008.

[21] A. B. Mak, A. M. L. Nixon, and J. Moffat, "The mixed lineage leukemia (MLL) fusion-associated gene AF4 promotes CD133 transcription," *Cancer Research*, vol. 72, no. 8, pp. 1929–1934, 2012.

[22] L. Godfrey, N. T. Crump, S. O'Byrne et al., "H3K79me2/3 controls enhancer-promoter interactions and activation of the pan-cancer stem cell marker _PROM1_ /CD133 in MLL-AF4 leukemia cells," *Leukemia*, vol. 35, no. 1, pp. 90–106, 2021.

[23] J. Neuzil, M. Stantic, R. Zobalova et al., "Tumour-initiating cells vs. cancer 'stem' cells and CD133: what's in the name?," *Biochemical and Biophysical Research Communications*, vol. 355, no. 4, pp. 855–859, 2007.

[24] D. Li, Y. Hu, Z. Jin et al., "TanCAR T cells targeting CD19 and CD133 efficiently eliminate MLL leukemic cells," *Leukemia*, vol. 32, no. 9, pp. 2012–2016, 2018.

[25] C. Bueno, T. Velasco-Hernandez, F. Gutiérrez-Agüera et al., "CD133-directed CAR T-cells for MLL leukemia: on-target, off-tumor myeloablative toxicity," *Leukemia*, vol. 33, pp. 2090–2125, 2019.

[26] Y. Kang, P. M. Siegel, W. Shu et al., "A multigenic program mediating breast cancer metastasis to bone," *Cancer Cell*, vol. 3, no. 6, pp. 537–549, 2003.

[27] D. Xie, D. Yin, H.-J. Wang et al., "Levels of expression of CYR61 and CTGF are prognostic for tumor progression and survival of individuals with gliomas," *Clinical Cancer Research*, vol. 10, no. 6, pp. 2072–2081, 2004.

[28] A. Koliopanos, H. Friess, F. F. di Mola et al., "Connective tissue growth factor gene expression alters tumor progression in esophageal cancer," *World Journal of Surgery*, vol. 26, no. 4, pp. 420–427, 2002.

[29] K. L. Bennewith, X. Huang, C. M. Ham et al., "The role of tumor cell-derived connective tissue growth factor (CTGF/CCN2) in pancreatic tumor growth," *Cancer Research*, vol. 69, no. 3, pp. 775–784, 2009.

[30] P. Vorwerk, H. Wex, B. Hohmann, Y. Oh, R. G. Rosenfeld, and U. Mittler, "CTGF (IGFBP-rP2) is specifically expressed in malignant lymphoblasts of patients with acute lymphoblastic leukaemia (ALL)," *British Journal of Cancer*, vol. 83, no. 6, pp. 756–760, 2000.

[31] Y. Tesfai, J. Ford, K. W. Carter et al., "Interactions between acute lymphoblastic leukemia and bone marrow stromal cells influence response to therapy," *Leukemia Research*, vol. 36, no. 3, pp. 299–306, 2012.

[32] J. M. Boag, A. H. Beesley, M. J. Firth et al., "High expression of connective tissue growth factor in pre-B acute lymphoblastic leukaemia," *British Journal of Haematology*, vol. 138, no. 6, pp. 740–748, 2007.

[33] O. Sala-Torra, H. M. Gundacker, D. L. Stirewalt et al., "Connective tissue growth factor (CTGF) expression and outcome in adult patients with acute lymphoblastic leukemia," *Blood*, vol. 109, no. 7, pp. 3080–3083, 2007.

[34] H. Kang, I.-M. Chen, C. S. Wilson et al., "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*, vol. 115, no. 7, pp. 1394–1405, 2010.

[35] M. D. Welch, W. K. Greene, and U. R. Kees, "Hypomethylation of the CTGF gene locus is a common feature of paediatric pre-B acute lymphoblastic leukaemia," *British Journal of Haematology*, vol. 162, no. 4, pp. 537–541, 2013.

[36] H. Lu, K. Kojima, V. L. Battula et al., "Targeting connective tissue growth factor (CTGF) in acute lymphoblastic leukemia preclinical models: anti-CTGF monoclonal antibody attenuates leukemia growth," *Annals of Hematology*, vol. 93, no. 3, pp. 485–492, 2014.