PPARG, GNG12, and CD19 are potential independent predictors of central nerve recurrence in childhood acute lymphoblastic leukemia

Shan Zhang,a,b Yansong Tu,c Hurong Lai,a,b Huijun Chen,a Huaijun Tu,b,d and Jian Lia,b

Objective: To identify biomarkers that can predict the recurrence of the central nervous system (CNS) in children with acute lymphoblastic leukemia (ALL).

Materials and Methods: The transcriptome and clinical data of ALL in children were downloaded from the TARGET database. Transcriptome data were analyzed by bioinformatics method to identify core (hub) genes and establish a risk assessment model. Univariate Cox analysis was performed on each clinical data, and multivariate Cox regression analysis was performed on the obtained results and risk score. The children ALL phase I samples from TARGET database were used for validation.

Results: Univariate multivariate Cox analysis of 10 hub genes identified showed that PPARG (HR = 0.78, 95% CI = 0.67–0.91, p = 0.007), CD19 (HR = 1.15, 95% CI = 1.05–1.26, p = 0.003) and GNG12 (HR = 1.25, 95% CI = 1.04–1.51, p = 0.017) had statistical differences. The risk score was statistically significant in univariate (HR = 3.06, 95% CI = 1.30–7.19, p = 0.011) and multivariate (HR = 1.81, 95% CI = 1.16–2.32, p = 0.046) Cox regression analysis. The survival analysis results of the high and low-risk groups were different when the validation dataset was substituted into the model (p = 0.018). Then, we constructed a Nomogram which had a concordance index of survival prediction of 0.791 (95% CI = 0.779–0.803). In addition, the CNS involvement grading status at first diagnosis CNS3 vs. CNS1 (HR = 0.78, 95% CI = 0.67–0.91, p = 0.046) Cox regression analysis. The survival analysis results of the high and low-risk groups were different when the validation dataset was substituted into the model (p = 0.011) and multivariate (HR = 1.81, 95% CI = 1.16–2.32, p = 0.046) Cox regression analysis. The survival analysis results of the high and low-risk groups were different when the validation dataset was substituted into the model (p = 0.018).

Conclusions: PPARG, GNG12, and CD19 may be predictors of CNS relapse in childhood ALL.
Materials and methods

Clinical data and preprocessing

We downloaded clinical and transcriptome data for Phase I and II of the ALL program from the TARGET database (https://ocg.cancer.gov/programs/target) managed by NCI’s Office of Cancer Genomics and Cancer Therapy Evaluation Program, which is a registry specifically for pediatric tumors. The TARGET database is publicly available and includes clinical and omics information on pediatric cancer patients. The TARGET ALL phase I project was obtained from patients enrolled in biology studies and clinical trials managed through the Children’s Oncology Group (COG), POG 9906 (clinical trial for patients with newly diagnosed ALL between March 2000 and April 2003). The Phase II project is collected from patients enrolled in the COG biology studies and clinical trials, most who experienced an early bone marrow relapse (within 4 years of initial diagnosis) and adequate amount of high-quality nucleic acids for comprehensive genomic profiling.

The data obtained from the database were divided into CNS relapsed group and non-relapsed group based on the presence or absence of CNS relapse. The inclusion and exclusion criteria for the two groups of data were: (1) patients were younger than 18 years old at the time of diagnosis; (2) complete transcriptome data were obtained in both groups; (3) clinical information should include complete relapse-free survival, CNS relapse status, etc.; and (4) relapsed group included patients that had only CNS relapse and no other site relapse at the end of the follow-up; the non-relapsed group refers to patients with no recurrence in any part by the end of follow-up.

Differentially expresses genes (DEGs) and enrichment analysis

The edge R was used to identify DEGs between the two groups (relapsed group vs. non-relapsed group). A false discovery rate (FDR) adjusted p < 0.05 combined with a simultaneous absolute value of ≥2 for logFC was set as the threshold for DEGs identification. The results of the differential analysis were visualized using gplots on the volcanic map and heat map. The differential genes were imported into the Biomart application of the Ensembl website for gene ID conversion. GO function enrichment analysis (annotated from the biological process (BP), cell composition (CC), and molecular function (MF)) and KEGG pathway analysis were performed using the David website [14]. The STRING was used for the analysis of protein interactions encoded by differential genes; the credibility >0.7 was used as a cut-off value to plot the protein–protein interaction network (PPI) [15].

Twelve topological analysis methods, including Degree, Edge Percolated Component (EPC), Maximum neighborhood component (MNC), Density of Maximum Neighborhood Component (DMNC), and Maximal Clique Centrality (MCC), etc., were provided by the CytoHubba plug-in in the Cytoscape software [16]. These methods were used to analyze protein interaction networks. The top 10 scoring genes obtained by various algorithms were counted for frequency, and the top 10 genes with the highest occurrence frequency were regarded as hub genes.

Survival analysis and stepwise backward regression, ROC curve

The top 10 DEGs with the highest occurrence frequency were used as hub genes for univariate Cox regression analysis using Survival R packet. Multivariate Cox regression included statistically significant DEGs, and the stepwise backward regression was used to establish the optimal model. Finally, the risk value was calculated using the following formula: the risk score = (β1 · expression level of PPARγ) + (β2 · expression level of CD19) + (β3 · expression level of GNG12). The ROC curves of 3–5 years were drawn in batches using the survival ROC package. ROC curve at the time point with the maximum AUC was selected to intercept the turning point, which was used as the cut-off value of the risk score. The research subjects were divided into high and low-risk groups based on the cut-off value of the risk score. The Kaplan-Meier survival curves for the cases predicted to have low or high risk were produced to verify whether the two groups showed statistical significance.

The establishment and evaluation of the nomogram

Integrate the factors associated with CNS recurrence and create the nomogram of training data. The predictive power of the model was assessed and quantified by measuring the extent to which the C-index and the baseline time predicted by the nomogram in the standard curve fit the actual recurrence time.

Validation

The data collected by the COG collaboration group in the TARGET database for phase I of the study on childhood ALL were used to verify. The expression levels of the three target genes identified in the study were included in the model, and the risk score was calculated. The high-risk group and the low-risk group were also divided, and survival analysis was conducted to verify the accuracy of the model. Finally, to further validate whether the biomarkers’ prediction was independent of other clinical variables, univariate and multivariate Cox regression, and stratified analyses were conducted. Statistical and analytical software R3.6.0 was used for all data
analysis and mapping. Gene ID conversion through Ensembl (https://grch37.ensembl.org/info/data/biomart/index.html). Enrichment analysis was conducted by the David website (https://david.ncifcrf.gov/tools.jsp). PPI plot was drawn by String website String (https://string-db.org/). Hub gene was determined by Cytoscape 3.8.2. P-value of < 0.05 was considered statistical significance, and all reported p-values were two-sided.

Results

Patient characteristics

Clinical information of the subjects in the relapsed group and non-relapsed group were described by means and percentages. A total of 278 subjects were included in the training dataset, with an average age of 8.7 ± 4.6 years. The time of diagnosis was from 2001 to 2011, and the follow-up period was from 2005 to 2016 (the mean follow-up time was 6.3 years); the male-to-female ratio was 3:7. Based on the presence or absence of CNS relapse, the data were divided into CNS relapsed and non-relapsed groups, with 28 patients in the CNS relapsed group and 250 patients in the non-relapsed group. Due to many missing values of the ETV6-RUNX1 fusion gene and only four positive cases, and since only one case of TCF3-PBX1 and BCR-ABL1 genes were positive, subsequent studies were not included in the analysis (Table 1).

Table 1. Patients' characteristics in training dataset.

| Term                              | Total (n = 278) | Relapsed (n = 28) | Non-relapsed (n = 250) |
|-----------------------------------|----------------|------------------|------------------------|
| Age (years, mean ± SD)            | 8.7 ± 4.6      | 7.4 ± 5.0        | 8.8 ± 4.5              |
| Sex (n, %)                        |                |                  |                        |
| Male                              | 194 (69.8)     | 15 (53.6)        | 179 (71.6)             |
| Female                            | 84 (30.2)      | 13 (46.4)        | 71 (28.4)              |
| Race (n, %)                       |                |                  |                        |
| White                             | 205 (73.7)     | 24 (85.7)        | 181 (72.4)             |
| Othersa                           | 73 (26.3)      | 4 (14.3)         | 69 (27.6)              |
| Ethnicityb (n, %)                 |                |                  |                        |
| Hispanic or Latino                | 49 (18.9)      | 8 (28.6)         | 41 (17.7)              |
| Not                               | 210 (81.1)     | 20 (71.4)        | 190 (82.3)             |
| Relapse-free survival (months)    | 76.9 ± 26.4    | 60.6 ± 32.1      | 78.7 ± 25.1            |
| Survival status (n, %)            |                |                  |                        |
| Alive                             | 261 (93.9)     | 11 (39.3)        | 250 (100)              |
| Dead                              | 17 (6.1)       | 17 (60.7)        | 0 (0)                  |
| WBC count (109/L) (n, %)          |                |                  |                        |
| >100                              | 128 (46.0)     | 12 (42.9)        | 128 (46.0)             |
| ≤10                               | 124 (44.6)     | 13 (46.4)        | 124 (44.6)             |
| Day-29 MRD (n, %)                 |                |                  |                        |
| >0                                | 98 (35.5)      | 12 (42.9)        | 86 (34.7)              |
| 0                                 | 178 (64.5)     | 16 (57.1)        | 162 (65.3)             |
| CNS classificationb (n, %)         |                |                  |                        |
| CNS1                              | 197 (71.1)     | 18 (64.2)        | 179 (71.9)             |
| CNS2                              | 61 (22.0)      | 5 (17.9)         | 56 (22.5)              |
| CNS3                              | 19 (6.9)       | 5 (17.9)         | 14 (5.6)               |
| ETV6 RUNX1b (n, %)                |                |                  |                        |
| Negative                          | 122 (96.8)     | 16 (64.2)        | 106 (99.1)             |
| Positive                          | 4 (3.2)        | 3 (15.8)         | 1 (0.9)                |
| TCF3-PBX1b (n, %)                 |                |                  |                        |
| Negative                          | 199 (99.5)     | 18 (54.7)        | 181 (100)              |
| Positive                          | 1 (0.5)        | 1 (5.3)          | 0 (0)                  |
| BCR-ABL1b (n, %)                  |                |                  |                        |
| Negative                          | 277 (99.6)     | 28 (100)         | 249 (99.6)             |
| Positive                          | 1 (0.4)        | 0 (0)            | 1 (0.4)                |
| TRISOMY 4 10 Statusb              |                |                  |                        |
| Negative                          | 221 (96.5)     | 23 (100)         | 198 (96.1)             |
| Positive                          | 8 (3.5)        | 0                | 8 (3.9)                |
| MLL Statusb (n, %)                |                |                  |                        |
| Negative                          | 227 (95.4)     | 26 (100)         | 201 (94.6)             |
| Positive                          | 11 (4.6)       | 0                | 11 (5.4)               |
| BMA Blasts Day 8                  | 20.3 ± 26.3    | 30.0 ± 31.4      | 19.2 ± 25.6            |
| BMA Blasts Day 15                 | 4.6 ± 8.8      | 7.4 ± 16.8       | 4.3 ± 7.1              |
| BMA Blasts Day 29                 | 0.7 ± 1.2      | 1.1 ± 1.5        | 0.7 ± 1.1              |
| DNA Indexb (n, %)                 |                |                  |                        |
| >1                               | 228 (83.8)     | 20 (76.9)        | 208 (84.6)             |
| 1                                | 44 (16.2)      | 6 (23.1)         | 38 (15.4)              |
| Cell of originb (n, %)            |                |                  |                        |
| B cell                           | 74 (27.1)      | 13 (50.0)        | 61 (24.7)              |
| T cell                           | 199 (72.9)     | 13 (50.0)        | 186 (75.3)             |

*Include black and Native Hawaiian or other Pacific Islander Asian and Unknown.
†The original data in the database is missing.
MRD, Minimal Residual Disease; WBC, white blood cell; CNS, central nervous system; MLL, Mixed Lineage Leukemia; and BMA, bone marrow aspirate.
DEGs analysis

A total of 1230 differential genes were screened, including 685 up-regulated and 545 down-regulated genes. G plots were used to visualize the different analysis results, as shown in Figure 1 and Supplementary Figure 1.

Identification of differentially expressed genes and functional enrichment analysis

The results showed that these differential genes were mainly involved in biological processes such as antigen processing and positive regulation of T cell aggregation, angiogenesis, and cell adhesion, migration, and proliferation of presenting cells. The cell components were mainly concentrated on the cell surface, transport vesicle membrane, endocytic vesicle membrane, and involved in the composition of the MHC class II protein complex. As for molecular function, the results showed that genes were mainly involved in antigen processing and presentation-related pathways, PI3K-Akt signaling pathway, Ras signaling pathway, and cancer-related signaling pathways (p < 0.05) (Figure 2B).

Protein interaction network (PPI) and hub gene identification

A total of 580 proteins and 647 edges were obtained with a cut-off value of credibility > 0.7. The average interaction between each node was 2.23, and the aggregation coefficient was 0.315, as shown in Figure 3.

Frequency statistics were carried out for the top 10 genes scored by various algorithms, and the top 10 genes (NPY, IL1B, IL6, BMP2, PPARG, CD19, GNG13, CD40, GNG7, and GNG12) with the highest occurrence frequency were regarded as hub genes (Supplementary Table 1 and Figure 4A).

PPARG, CD19, and GNG12 were obtained by Cox regression analysis

Univariate Cox regression analysis was performed for the top 10 hub genes using Survival packet (recurrence as
the outcome variable, time without recurrence as time-variable). The results showed that all 10 hub genes were statistically significant, and they were included in multivariate Cox regression analysis (Table 2).

Figure 2. Gene ontology (A) and KEGG (B) pathway analysis of DEGs. The top 10 pathway with the smallest p value were selected for plotting.
The results of multivariate analysis showed that there were three genes with significant differences, namely PPARG, CD19, and GNG12 (Table 2). In the CNS relapsed group, PPARG (logFC = −2.67) was down-regulated expression, while CD19 (logFC = 2.06) and GNG12 (logFC = 2.20) were up-regulated expression. The risk model was obtained according to the regression results as follows:

\[
\text{risk score} = \text{PPARG} \times (-0.19009) + \text{CD19} \times (0.13794) + \text{GNG12} \times (0.22447)
\]

For the obtained risk model, the ROC curves at multiple time points across 3–5 years were done in batches to compare and screen out the time point at which the maximum AUC was obtained. The results showed that the model established in this study has a good predictive ability in predicting the 5-year relapse rate (AUC max = 0.75) (Figure 4B).

We calculated the risk score of the 278 research objects. The true positive and false positive difference maximum points were taken as cut-off of the risk score. Values above this score were defined as high-risk group, and below this score as a low-risk group.
Kaplan–Meier survival curve and log-rank test were adopted to appraise the association of the overall survival (Figure 4C).

In order to verify the accuracy of the model, we used transcriptome and clinical data collected from the COG collaboration group in the Target database for the phase I of childhood ALL for validation. The expression levels of the three target genes identified in the study were included in the model, and the risk values were calculated. The validation dataset was also divided into high- and low-risk groups, and survival analysis (K-M curve) was performed. Our results proved that the established model has a good predictive ability ($p = 0.018$) (Figure 4D).

The risk score can be used as an independent prognostic factor

To judge whether the model’s risk score can be used as an independent predictor of a childhood ALL relapse of CNS, we conducted Cox regression analyses with clinical data (Table 3). The results showed that the risk score was statistically significant in univariate (HR = 3.06, 95%CI = 1.30–7.19, $p = 0.011$) and

### Table 2. Cox regression results of hub gene.

| gene  | Univariate Cox. | Multivariate Cox. |
|-------|-----------------|-------------------|
|       | $\beta$ | HR(95%CI) | p-value | $\beta$ | HR(95%CI) | p-value |
| NPY   | 0.17 | 1.18(1.10–1.28) | <0.01 | −0.19 | 0.78(0.67–0.91) | 0.007 |
| IL1B  | 0.21 | 1.23(1.10–1.39) | <0.01 | 0.14 | 1.15(1.05–1.26) | 0.003 |
| IL6   | 0.26 | 1.29(1.07–1.56) | <0.01 | 0.012 |
| BMP2  | 0.18 | 1.20(1.12–1.28) | <0.01 | 0.01 |
| PPARG | −0.21 | 0.81(0.66–0.93) | 0.039 | 0.01 |
| CD19  | 0.19 | 1.21(1.11–1.31) | <0.01 | 0.23 | 1.25(1.04–1.51) | 0.017 |
| GNG13 | 0.29 | 1.34(1.07–1.68) | 0.012 | 0.01 |
| GNG7  | 0.19 | 1.21(1.10–1.32) | <0.01 | 0.01 |
| GNG12 | 0.35 | 1.42(1.19–1.70) | <0.01 | 0.01 |

Figure 4. Identification and validation of hub genes. (A) Top 10 genes with the highest occurrence frequency of twelve topological analysis methods. (B) ROC curve of the prediction model at multiple time points across 3–5 years. (C) Kaplan-Meier curve of high risk and low-risk group in training dataset. D Kaplan-Meier curve of high risk and low-risk group in validation dataset.
leukemia is still very challenging [17]. Therefore, identifying new potential biomarkers that could predict CNS relapse in children is extremely important. In this study, bioinformatics correlation analysis was used to identify biomarkers, revealing that PPARG, GNG12, and CD19 were independent prognostic factors for childhood ALL with CNS relapse.

PPARG gene is a member of the nuclear receptor family, which encodes peroxisome proliferative activated receptor gamma (PPARγ). PPARγ was first identified in adipocytes, as a transcription factor, the PPARγ receptor binds to its ligand, inhibits tumor proliferation, metastasis, and angiogenesis, and promotes apoptosis and immune regulation by activating signaling pathways of different tumor cells and tumor stem cells [18]. For example, PPARγ can induce apoptosis by inhibiting the PI3K/Akt signaling pathway, PPARγ ligand can induce the apoptosis of leukemia cells, and can be potentially used to treat hematologic malignancies, especially leukemia [19]. In our study, we found a low expression of PPARG in childhood ALL with CNS relapse. Moreover, KEGG analysis showed the activation of the PI3K/Akt signaling pathway, which suggests that PPARG might regulate CNS relapse in ALL children through the PI3K/Akt signaling pathway. Boyd et al. [20] found that the injection of PPARS antibody agonist could induce adipogenesis in bone marrow tissue in mice, thereby saving healthy hematopoietic cell formation during the inhibition of leukemia cell growth. The most common way leukemia cells invade the central nervous system is from the bone marrow of the skull to the subarachnoid space via dural bridging vein

**Table 3. Cox regression results of independent predictors.**

| Factor                        | Univariate Cox |          | Multivariate Cox |          |
|-------------------------------|----------------|----------|------------------|----------|
|                               | HR (95% CI)    | p-value  | HR (95% CI)      | p-value  |
| Age (0,6)                     | 1              |          |                  |          |
| (6,12)                        | 1.72 (0.66–4.48) | 0.265   | 1.20 (0.72–2.02) | 0.490   |
| (12,18)                       | 0.91 (0.32–2.63) | 0.863   |                  |          |
| Sex (female vs. male)         | 2.10 (1.09-4.42) | 0.017   |                  |          |
| Race (others* vs White)       | 0.34 (0.05–2.51) | 0.289   |                  |          |
| Eth (Hispanic vs. Not)        | 1.70 (1.02–2.63) | 0.204   |                  |          |
| WBC count (10³/L)             |                |          |                  |          |
| ≤10                           | 1              |          |                  |          |
| (10,100)                      | 1.19 (0.34–4.25) | 0.780   |                  |          |
| >100                          | 1.11 (0.51–2.42) | 0.803   |                  |          |
| Day-29 MRD (>0 vs. 0)         | 1.70 (1.13-2.57) | 0.360   |                  |          |
| CNS classification            |                |          |                  |          |
| CNS1                          | 0.86 (0.32–2.32) | 0.770   | 0.52 (0.19–2.29) | 0.519   |
| CNS2                          | 3.38 (1.25–9.10) | 0.016   | 5.75 (2.01–16.4) | 0.001   |
| Trisomy (N v P)               | 0.98 (0.36–2.67) | 0.964   |                  |          |
| MLL (N v P)                   | 3.57 (0.50–25.66) | 0.206  |                  |          |
| BMA Blasts Day 8 (≥6 vs.<6)   | 1.17 (0.79–1.71) | 0.436   |                  |          |
| BMA Blasts Day 15 (≥2 vs. <2) | 1.06 (0.64–1.76) | 0.829   |                  |          |
| BMA Blasts Day 29 (>0)        | 0.98 (0.64–1.50) | 0.930   |                  |          |
| DNA index (>1 vs. 1)          | 1.50 (0.61–3.73) | 0.386   |                  |          |
| T Cell V B Cell ALL           | 2.59(1.18–5.68)  | 0.017   | 1.63(1.06–2.49)  | 0.026   |
| Prediction model risk score   | 3.06(1.30–7.19)  | 0.011 | 1.81(1.16–2.32)  | 0.046   |

N: negative; P: positive.

*Include Black and Native Hawaiian or other Pacific Islander Asian and Unknown; WBC, white blood cell; CNS, central nervous system; MLL, Mixed Lineage Leukemia; and BMA, bone marrow aspirate.

**Construction of the Nomogram**

We established a nomogram for the prediction of 3-year and 5-year recurrence rate in training dataset. Six independent risk factors such as age, sex, WBC at diagnosis, CNS classification, cell origin, and predictive model risk score were included in the model. (Figure 5). The points of the factors indicate their corresponding contribution to the survival probability. The total points of each patient provided the estimated 3-year and 5-year recurrence rate. The C-index of our nomogram was 0.791(95% CI = 0.779–0.803). The actual recurrence rate and nomogram-predicted recurrence rate matched well at 3 and 5 years, as shown by the calibration curves(Figure 5 B, C).

We stratified the analysis of the expression level of the CD19 gene according to the source of T cells and B cells. The results showed that this gene differed only in B lymphocytic leukemia (p = 0.018), as shown in Table 4.

**Discussion**

The CNS is a common site for the relapse of ALL. Yet, considering that most drugs cannot pass through the brain blood–brain barrier (BBB), the treatment of CNS multivariate (HR = 1.81, 95%CI = 1.16–2.32, p = 0.046) Cox regression analysis. CNS3 vs. CNS1 (HR = 5.74, 95%CI = 2.01–16.4, p = 0.001) and T cell vs. B cell (HR = 1.63, 95%CI = 1.06–2.49, p = 0.026) were significantly different between the two groups; while no difference was found for patients' gender (HR = 1.20, 95%CI = 0.72–2.02, p = 0.490).

It is also worth noting that in our study, we found a low expression of PPARG in childhood ALL with CNS relapse. Moreover, KEGG analysis showed the activation of the PI3K/Akt signaling pathway, which suggests that PPARG might regulate CNS relapse in ALL children through the PI3K/Akt signaling pathway. Boyd et al. [20] found that the injection of PPARS antibody agonist could induce adipogenesis in bone marrow tissue in mice, thereby saving healthy hematopoietic cell formation during the inhibition of leukemia cell growth. The most common way leukemia cells invade the central nervous system is from the bone marrow of the skull to the subarachnoid space via dural bridging vein...
When the expression of the PPARG gene is low, this induction function of PPARs is inhibited, which might in turn lead to the CNS relapse of ALL in children. The GNG12, a gene encoding Gγ12, a member of the G12 subfamily of heterotrimeric G proteins, is involved in the regulation of expression and distribution of tightly connected endothelial cells [22]. The BBB is mainly composed of tight junctions (TJ) found between endothelial cells, which limit the passage of substances in the brain [23]. Donato et al. found that the heterotrimer G protein has an important role in TJ's formation and regulation, which is regulated through the intercellular signaling pathway and has an important influence on the BBB structure [24]. It has also been found that Gγ12 preferentially interacts with adrenergic receptors in activated central synapses [25]. Adrenergic receptors are involved in the regulation of protein kinase A (PKA) and cAMP [26]. The BBB appears to be highly sensitive to PKA levels and opens when cAMP and PKA levels decrease [27]. This may result in leukemic cells crossing the blood–brain barrier more easily and invading the CNS.

Table 4. Layered analysis of CD19 gene between T cells and B cells.

| Non-relapse vs. relapse | Log FC | P-value   | FDR    |
|-------------------------|--------|-----------|--------|
| B-cells                 | 2.176  | 0.018     | 0.037  |
| T-cells                 | -0.318 | 0.184     | 0.367  |

[21]. [22]. [23]. [24]. [25]. [26]. [27]. [28].
target for drug development [28]. In particular, CAR T cells can penetrate the BBB into the CNS, killing leukemic cells in the cerebrospinal fluid [29]. Jin MY et al. [13] reported two cases of acute leukemia patients with CNSL, which showed positive CD19 expression in tumor cell immunotyping and remission after anti-CD19 CAR T cell therapy. In addition, studies have reported the efficacy of anti-CD19 CAR T cell therapy in patients with ALL-CNSL [30,31]. These treatments further revealed the relationship between CD19 and CNSL and also verified the high relapse risk of CNS in patients with high expression of the CD19 gene in this study. Meanwhile, stratified analysis results show that this difference only appeared in B-ALL, which also verified the reliability of the results of this study. In other words, the high expression of the CD19 gene is not only a marker of B-ALL but may also lead to CNS relapse when its expression level reaches a certain level.

In this study, the cell composition, molecular function, biological process, and KEGG pathway corresponding to the DEGs were enriched and analyzed. These results revealed a related pathway leading to the relapse of CNS, in which calmodulin-binding was related to the molecular composition and signal regulation of the tight junction of the BBB. Some researchers have reported that leukemia cells enter CNS via vascular transport, which involves the interaction between A6 integrin and laminin and can be blocked by specific PI3K inhibitors [32]. Our results further suggest that CNS relapse in childhood ALL is related to the PI3K-Akt signaling pathway. In addition, cell adhesion molecules (CAMs), antigen processing and presentation, and neuroactive ligand–receptor interactions provide upstream and downstream factors and targets for future studies on CNSL recurrence, which should be further explored.

The results of our study confirmed that CNS grade 3 at initial diagnosis was a risk factor for CNS relapse, which was consistent with previous studies [7]. Our data also suggested that T lymphocytic leukemia in children was more likely to lead to CNS relapse than B lymphocytic leukemia, which was also consistent with previous relevant reports [33,34]. This study failed to prove that hyperleukocytosis and male in patients were high-risk factors for CNS relapse, which was inconsistent with previous studies [6,35]. We speculate that it may be due to sample differences and should be further clarified by future studies. Moreover, we established a nomogram for clinical-decision support that could well predict the 3-year and 5-year recurrence rate for CNSL patients by incorporating the three biomarkers and prognostic factors.

There are some limitations in our study. First, prognostic value of the critical genes was only analyzed and validated in the TARGET dataset. Thus, the results need to be validated in other databases. Second, the sample size in the validation dataset was limited and there are many missing clinical data in the training dataset. In addition, our study was based on a purely bioinformatic analysis. Therefore, further experiments are needed to verify the results.

To sum up, this study verified that CNS3 classification and the types of T lymphocytes are risk factors for CNS relapse of childhood ALL. We found that PPARG, GNG12, and CD19 can be used as independent predictors of CNS relapse in childhood ALL, yet, more studies are needed to confirm these findings. Our results may provide a theoretical basis for further studies to elucidate the molecular mechanism of CNS relapse in childhood ALL and provide more therapeutic targets for future clinical interventions.

Acknowledgments
We acknowledge the TARGET database for providing their platforms and contributors for uploading their meaningful datasets.

Disclosure statement
No potential conflict of interest was reported by the authors.

Funding
This work was supported by grant from National Natural Science Foundation of China [81760033]; [82160030].

Data availability statement and ethical approval
The results published here are in whole based upon data generated by the Therapeutically Applicable Research to Generate Effective Treatments (https://ocg.cancer.gov/programs/target) initiative, pHS000218. The patients involved in the database have obtained ethical approval. Users can download relevant data for free for research and publish relevant articles, so there are no ethical issues and other conflicts of interest. The data used for this analysis are available at https://portal.gdc.cancer.gov/projects.

References
[1] Mueller KT, Maude SL, Porter DL, et al. Cellular kinetics of CTL019 in relapsed/refractory B-cell acute lymphoblastic leukemia and chronic lymphocytic leukemia. Blood. 2017;130(21):2317–2325.
[2] Ikonomidou C. Cerebrospinal fluid biomarkers in childhood leukemias. Cancers (Basel). 2021;13:438.
[3] Saeedi A, Baghestani A, Khadem Maboudi A, et al. Determining the significant prognostic factors for the recurrence of pediatric acute lymphoblastic leukemia using a competing risks approach. Iran J Med Sci. 2020;45(4):304–310.
Janani C, Ranjitha Kumari BD. PPAR gamma gene—a review. Diabetes Metab Syndr. 2015;9(1):46–50.

Yousefnia S, Momenzadeh S, Seyed Forootan F, et al. The influence of peroxisome proliferator-activated receptor γ (PPARγ) ligands on cancer cell tumorigenicity. Gene. 2018;649:14–22.

Boyd A, Bhatia M. Acute myeloid leukaemia disrupts endogenous myelo-erythropoiesis by compromising the adipocyte bone marrow niche. Nat Cell Biol. 2017;19(11):1336–1347.

Zhou F, Wen Y, Jin R, et al. New attempts for central nervous infiltration of pediatric acute lymphoblastic leukemia. Cancer Metastasis Rev. 2019;38(4):657–671.

Li J, Jin C, Zou C, et al. GNG12 regulates PD-L1 expression by activating NF-κB signaling in pancreatic ductal adenocarcinoma. FEBS Open Bio. 2020;10(2):278–287.

Langen UH, Ayloo S, Gu C. Development and cell biology of the blood-brain barrier. Annu Rev Cell Dev Biol. 2019;35:591–613.

González-Mariscal L, Raya-Sandino A, González-González L, et al. Relationship between G proteins coupled receptors and tight junctions. Tissue Barriers. 2018;6:e1414015.

Yim YY, Betke KM, McDonald WH, et al. The in vivo specificity of synaptic Gβ and Gγ subunits to the α(2a) adrenergic receptor at CNS synapses. Sci Rep. 2019;9:1718.

Shi Q, Li M, Mika D, et al. Heterologous desensitization of cardiac β-adrenergic signal via hormone-induced βAR/arrestin/PDE4 complexes. Cardiovasc Res. 2017;113(6):656–670.

Fukuda S, Nakagawa S, Tatsumi R, et al. Glucagon-Like peptide-1 strengthens the barrier integrity in primary cultures of Rat brain endothelial cells under basal and hyperglycemia conditions. J Mol Neurosci. 2016;59 (2):211–219.

Penaceta HL, Laetsch TW, John S. CD19 CAR T cells for the treatment of pediatric Pre-B cell acute lymphoblastic leu- kemia. Paediatr Drugs. 2020;22(1):1–11.

Maus MV, Grupp SA, Porter DL, et al. Antibody-modified T cells: CARs take the front seat for hematologic malignancies. Blood. 2014;123(17):2625–2635.

Qi Y, Zhao M, Hu Y, et al. Efficacy and safety of CD19-specific CAR T cell-based therapy in B-cell acute lymphoblastic leukemia patients with CNSL. Blood. 2022;139(23):3376–3386.

Nair S, Wang JB, Tsao ST, et al. Functional improvement of chimeric antigen receptor through intrinsic interleukin-15Ra signaling. Curr Gene Ther. 2019;19(1):40–53.

Yao H, Price TT, Cantelli G, et al. Leukaemia hijacks a neural mechanism to invade the central nervous system. Nature. 2018;560(7676):55–60.

Gupta S, Teachey DT, Chen Z, et al. Sex-based dispar- ities in outcome in pediatric acute lymphoblastic leu- kemia: a children’s oncology group report. Cancer. 2022;128(9):1863–1870.

Schrappe M, Reiter A, Ludwig WD, et al. Improved outcome in childhood acute lymphoblastic leukemia despite reduced use of anthracyclines and cranial radiotherapy: results of trial ALL-BFM 90. German-Austrian-Swiss ALL-BFM study group. Blood. 2000;95(11):3310–3322.

Tevatia MS, Sharma I, Jadhav T, et al. Isolated CNS relapse in acute lymphoblastic leukemia (ALL): an experience from a tertiary care center. J Lab Physicians. 2021;13(2):134–138.