

Genomic signatures in pediatric advanced stage Burkitt lymphoma/leukemia in a Chinese population detected by next generation sequencing

Type

Research paper

Keywords

children, Burkitt Lymphoma, whole-exome sequencing, next-generation sequencing

Abstract

Introduction

Burkitt lymphoma/leukemia (BL/BAL) is the most common lymphoma in children, and the sporadic subtype is dominant in Chinese populations. MYC gene translocations are essential for sporadic BL/BAL (sBL/BAL), but other gene mutations also play important roles in sBL/BAL.

Material and methods

Clinical data of ten Chinese children with sBL/BAL were collected, next generation sequencing of tumor tissues was carried out. Cases of BL and diffuse large B cell lymphoma (DLBCL) were collected from a database, bioinformatics analysis was conducted

Results

Nine boys and one girl were enrolled, the average age at diagnosis was 100.10 ± 13.39 m; MYC rearrangements were confirmed. The patients received combination treatment of chemotherapy and rituximab. All patients achieved complete remission and were alive without relapse. Germline causal gene mutations were detected in four (40%) patients by whole-exome sequencing; ID3, BRCA2, ARID1A and SMARCA4 mutations, in addition to MYC mutations, were the most common somatic mutations. Gene functions in etiology were different between the BL and DLBCL datasets. The identified mutated genes were enriched and connected by GO or KEGG pathways, it seemed that the PI3K-Akt signaling pathway played important roles in the etiology of sBL/BAL; the EGFR-TKI resistance pathway was also analyzed.

Conclusions

The different gene mutations might be index to identify BL or DLBCL in case of difficult pathologic samples. Molecular hallmark of sBL/BAL is MYC translocation, whereas additional gene mutations also occur and play roles in the progression of the disease. sBL/BAL might be controlled by curbing the PI3K-Akt signaling pathway; the EGFR-TKI treatment might not be helpful in sBL/BAL.

Genomic signatures in pediatric advanced stage Burkitt lymphoma/leukemia in a Chinese population detected by next generation sequencing

Abstract

Introduction: Burkitt lymphoma/leukemia (BL/BAL) is the most common lymphoma in children, and the sporadic subtype is dominant in Chinese populations. **Material and methods:** Clinical data of ten Chinese children with sBL/BAL were collected, next generation sequencing of tumor tissues was carried out. Cases of BL and diffuse large B cell lymphoma (DLBCL) were also collected from database, bioinformatics analysis was conducted. **Results:** Nine boys and one girl were enrolled in the study, the average age at diagnosis was 100.10 ± 13.39 m; *MYC* rearrangements were confirmed. The patients received combination treatment of chemotherapy and rituximab. All patients achieved complete remission and were alive without relapse. Germline causal gene mutations were detected in four (40%) patients by whole-exome sequencing; *ID3*, *BRCA2*, *ARID1A* and *SMARCA4* mutations, in addition to *MYC* mutations, were the most common somatic mutations. Gene functions in etiology were different between the BL and DLBCL datasets. The identified mutated genes were enriched and connected by GO or KEGG pathways, it seemed that the *PI3K-Akt* signaling pathway played important roles in the etiology of sBL/BAL; the *EGFR-TKI* resistance pathway was also analyzed. **Conclusion:** The different gene mutations might be index to identify BL or DLBCL in case of difficult pathologic samples. Molecular hallmark of sBL/BAL is *MYC* translocation, whereas additional gene mutations also occur and play roles in the progression of the disease. sBL/BAL might be controlled by curbing the *PI3K-Akt* signaling pathway; the *EGFR-TKI* treatment might not be helpful in sBL/BAL. **Keywords:** Burkitt lymphoma; children; next-generation sequencing; whole-exome sequencing.

Background

Burkitt lymphoma/leukemia (BL/BAL) is the most common subtype of non-Hodgkin lymphoma (NHL) in children and adolescents. It occurs in 30–50% of pediatric NHL cases. There are three reported variants of BL/BAL: endemic, sporadic and immunodeficiency-related BL/BAL [1]. Sporadic BL/BAL (sBL/BAL) is familiar throughout the world and it is more common in China relative to the endemic and immunodeficiency-associated BL/BAL types [2].

The Murphy staging system and the revised International Pediatric Non-Hodgkin Lymphoma Staging System (IPNHLSS) have been proposed for pediatric BL cases, and they are classified into localized-stages (stage I or II) or advanced-stages (stage III or IV) [3]. With intensive combination chemotherapy, the prognosis of sBL/BAL

in children and adolescents has improved dramatically. Currently, 5-year event-free survival (EFS) of local and advanced stages of BL/BAL has reached 100% and 85–90%, respectively [1, 2].

It is well known that t(8;14)(q24;q32) or its variants play a key role in BL/BAL. A translocation of the MYC gene, which is located at band 8q24, is detectable in over 95% of cases. Epstein–Barr virus (EBV) infection is also common in BL/BAL. The MYC translocation is the essential driver for overexpression of the MYC gene, and activation of the MYC gene leads to cell cycle progression, inhibition of differentiation and the promotion of cell proliferation and/or genomic instability [4]. Additional chromosomal abnormalities, recurrent abnormally expressed transcripts and/or gene mutations are also detected in BL/BAL patients and have roles in the initiation, progression and aggressiveness of the disease [5]. For instance, somatic mutations of TCF3 can activate the PI3K/MAPK/MTOR pathway, in part by increasing a tonic form of B cell receptor signaling, and patients with germline mutations in the SH2 domain protein 1A gene (SH2D1A) suffer from an increased risk of BL/BAL [6].

Next-generation sequencing (NGS) studies have provided valuable insight into the landscape of genomic alterations in malignancies by using whole-exome sequencing (WES) and/or RNA sequencing (RNAseq). These approaches are helpful for researchers to explore the etiology, pathogenesis and mechanisms of such diseases [7]. Bioinformatics analysis has also been widely used in cancer research.

So far, WES sequencing or mRNAseq of sBL/BAL among Chinese children has not been reported. In this study, we performed WES and/or RNAseq for ten sBL/BAL patients, and we investigated the role of molecular alterations in BL/BAL and diffuse large B cell lymphoma (DLBCL) using large databases. Our objective was to reveal the signaling pathways involved in its pathogenesis and the relationship between clinical characteristics and gene mutations.

Methods

Patients

Ten patients with newly diagnosed sBL/BAL at the Children's Hospital of

Chongqing Medical University (CHCMU) between February 2018 and October 2019 were enrolled in the study. The diagnosis was in accordance with the World Health Organization criteria of 2016, and patients were staged with the revised IPNHLSS [3]. Patients who were ≥ 18 yr at diagnosis or diagnosed with Burkitt-like lymphoma with 11q aberration, secondary lymphoma, or had human immunodeficiency virus infection were excluded; patients received chemotherapy before hospitalization were also excluded from the study. During this period, one patient classified into stage II was admitted, but his parents refused this study. The patients received chemotherapy in accordance with modified non-Hodgkin Lymphoma 1995, Berlin-Frankfurt-Münster, (BFM-95) protocol [8, 9]. Intrathecal injections (iT) were administered as the protocol required and cranial radiotherapy was carried out for the patients with central nervous system (CNS) involvement. The details of the risk groups, course of treatment and drugs used in the modified BFM-95 protocol are listed in the Supplementary material (S1-S4).

The Ethics Administration Office of CHCMU granted ethics approval for this research. Informed consent was obtained from the patients or their guardians. Clinical data, laboratory findings and the outcomes of the enrolled patients were collected from the medical record system and analyzed retrospectively.

Pathologic diagnosis

Pathologic diagnoses of the BL patients were confirmed by lymph node (LN) biopsies. Immunohistochemical (IHC) staining was applied, including for: MYC, Ki-67, CD20, BCL-2, BCL-6, PAX-5, CD10, CD7, CD3, and MUM-1; fluorescence in situ hybridization (FISH) of *MYC*, *MLL*, *BCL-2*, *BCL-6* and *EBER* was also performed [10].

Diagnosis and classification of the BAL patients were confirmed by bone marrow (BM) samples. BM samples were subjected to FAB typing, flow cytometry (FCM), cytogenetic analysis of the chromosomal karyotype, and FISH of *MYC*, *ETV6-RUNX1*, *MLL*, *BCR-ABL1*, *TCF3-PBX1* and *PDGFRB*. In addition, 43 fusion genes were assayed by a multiplex RT-PCR as described previously [11].

DNA, RNA isolation and sequencing

Tumor DNA samples from the BL or BAL patients were obtained from formalin-fixed, **paraffin-embedded** specimens or BM samples at diagnosis, and germline samples were collected from the oral mucosa of the patients and their parents' peripheral blood (PB). Genomic DNA was extracted using a QIAamp DNA Minikit (QIAGEN, China). The genomic DNA was enriched and sequence acquisition was carried out (Agilent SureSelect Human All Exon V6). Then, PCR amplifications of the whole exome were sequenced (Illumina HiSeq PE 150 bp).

BM samples at diagnosis were collected from BAL patients, and total RNA was extracted using a QIAamp RNA Blood Mini Kit (Qiagen, Cat.52304). An enriched and captured mRNA library was constructed and amplified by using a KAPA mRNA HyperPrep Kit (KAPA/Roche, Cat.KK8581). PCR amplifications of mRNAseqs were sequenced using PE150 (Illumina HiSeq ×10).

The original whole exome sequencing (WES) data were read using Illumina pipeline Software (version 1.3.4), and the data were analyzed referring to databases (dbSNP, 1000 Genomes Project, ClinVar, ESP6500, ExAc, Ensembl, HGMD, **UCSC**). Mutated genotypes were determined using the software GATK, LRT, Mutation Taster and SAMtools.

All discovered variants were divided into the following four categories according to prior literature reports [12] and software analysis: 1) Pathogenic genotypes that were confirmed by literature reports; 2) Likely pathogenic genotypes that were reported in the literature and/or that affected the protein by functional prediction; 3) Indefinite variants; and 4) Single nucleotide polymorphisms (SNPs) or single nucleotide variants (SNVs). Pathogenic genotypes and likely pathogenic genotypes were recorded as causal gene mutations. All causal gene mutations in the tumor samples were confirmed by Sanger sequencing. Samples in the control group were cross-checked and detected by Sanger sequencing, and somatic or germline causal gene mutations were identified.

PPI network construction and GO, KEGG pathway enrichment analysis

The detectable mutations by WES sequencing were collected, and R-package wordcloud2 was used to visualize the frequency of mutations in these cases. We used

STRING (STRING, <http://string-db.org>, RRID: SCR_005223) to find protein to protein interactions (PPI) and to visualize the interactions using Cytoscape (version 3.7.1). Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses were performed using R package clusterProfiler and the Biological Networks Gene Ontology tool (BiNGO, RRID:SCR_005736).

Identification of DEGs between the BL and DLBCL datasets

The gene chip datasets GSE4475, GSE10172, GSE43677, and GSE48435 were downloaded from the Gene Expression Omnibus (GEO) database (<https://www.ncbi.nlm.nih.gov/geo/>). We selected the BL and DLBCL datasets for downstream analysis. After quality control (QC), we obtained 41 BL samples and 178 DLBCL samples. Differential expression analysis used the R package limma. The database of essential genes (DEGs) was screened at a statistical significance Benjamini and Hochberg false discovery rate-adjusted p -value cutoff of 0.05 and an absolute value of fold change greater than 1. A volcano plot showing the DEGs was constructed by using the ggplot2 package in R.

WGCNA analysis of the BL and DLBCL dataset

We used the R package weighted gene coexpression network analysis (WGCNA) to construct coexpression modules, and 219 samples were used to calculate the *Pearson's* Correlation Matrices. A power of 6 was selected. An unsigned hybrid coexpression network was then calculated using standard settings. We selected 5000 genes to construct a topological heatmap. We calculated *Pearson's* correlations between the module eigengenes and the traits data to identify module-trait relationships. At last, we selected blue module (related to BL) and red module (related to DLBCL) genes to construct a gene regulatory network and did the GO enrichment analysis.

Statistical Analysis

Events were defined as any of the following situations: refractory or relapsed disease, disease progression, death or diagnosis with a secondary malignancy. With follow-up to April of 2020, data on clinical features, laboratory findings, WES sequencing or RNAseq, treatment responses, treatment-related mortality (TRM) and

the event-free survival (EFS) rates of the patients were collected and analyzed.

The EFS was calculated from the date of diagnosis to the last follow-up, lost to follow-up or to the first event. SPSS 19.0 (IBM Corp., Armonk, NY) software was applied for statistical analysis. Survival curves were calculated with the Kaplan-Meier method. Proportional differences between patient groups were analyzed by *Pearson's* chi-squared (χ^2) tests or *Fisher's* exact tests. *P* values <0.05 were regarded as significantly different.

Results

Clinical, laboratory and pathologic characteristics

Nine boys and one girl were enrolled in this study. The six BL patients were classified as stage III, and the four BAL patients were regarded as stage IV. Age at diagnosis was 34–172 m (median value: 105 m; average value: 100.10±13.39 m). Two BL patients and one BAL patient presented with symptom B (fever), and there was no history of malignancy in their families. Average values of white blood cells (WBC), platelets (PLT), hemoglobin (Hb) and lactate dehydrogenase (LDH) were 13.90±6.92×10⁹/L, 315.20±54.39×10⁹/L, 101.90±6.94 g/L and 2896.21±1328.56 U/L, respectively. Organ involvement was evaluated by physical exams, ultrasonography, and computed tomography (CT) or positron emission tomography (PET-CT). CNS involvement and bulky disease were found in one and six patients, respectively.

The four BAL patients were diagnosed with ALL-L₃ by FAB morphology. Mature B-ALL phenotypes were visible and overexpression of CD20 was detectable. Restricted expression of Kappa or Lambda chains were detected in two patients each. t(8;14) and/or *MYC* rearrangements were confirmed by chromosomal karyotype and/or FISH, but 43 common fusion genes in acute leukemia as literature reports [11] were undetectable. The six BL patients were confirmed by LN biopsy. Overexpression of CD20, Ki-67 and *MYC* were detected by IHC staining. FISH of *MYC* and EBER were positive while FISH of *BCL-2*, *BCL-6* and *MLL* were negative.

Treatment and prognosis

Six BL patients (five patients were classified in the R3 group and one in the R2 group) and four BAL (R4 group) patients received chemotherapy in accordance with

the modified BFM-95 protocol. The treatment effects were evaluated after one or two courses of chemotherapy for the BAL and BL patients, respectively, and all of them achieved complete remission (CR). Chemotherapy was completed after 5-6 m. After that, the patient with CNS involvement received cranial radiotherapy (total dose of 18 Gy, divided into 10 times). With follow-up to May of 2020, all 10 patients were alive without events, the TRM rate was 0%, and the EFS rate was 100%.

Results of NGS sequencing

WES sequencing was conducted for the ten patients. Germline causal gene mutations were detected in four (40%) patients, and in addition, several somatic causal gene mutations were identified in the ten patients.

The data obtained by WES sequencing were analyzed using R-package wordcloud2 to visualize the frequency of candidate gene mutations, construct PPI networks, and conduct GO enrichment and KEGG pathway analysis.

(1) The identified gene mutations were shown by tag cloud using word size according to the gene frequency (Fig-1a), which demonstrated that *ID3*, *BRCA2*, *ARID1A* and *SMARCA4* mutations, in addition to the *MYC* mutations, were the most common mutations; other mutations included: *MGA*, *AKAP9*, *KMT2C*, *OBSCN*, *LAMA5*, *DDX3X*, etc. The identified gene mutations were also subjected to PPI network analysis (Fig-1b), which showed the proteins that were modified by *MYC*, *HRAS*, *TP53* and *NOTCH1*. The figure indicates that these identified gene mutations play important roles in the development of BL/BAL.

(2) The identified genes were enriched by GO analysis (Fig-2a). *P* value was defined as 0.05, and the top gene functions in the etiology were: leukocyte differentiation and regulation of hemopoiesis in biological processes (BP), nuclear chromosomal parts and chromatin in cellular components (CC), and chromatin binding and transcription coregulator activity in molecular function (MF). The genes were also enriched and connected by KEGG pathways (Fig-2b), and it seemed that the PI3K-Akt signaling pathway played a key role in the etiology of BL/BAL.

RNAseq was conducted and the gene transcripts were analyzed for the four BAL patients. The results are listed in the Supplementary material.

Identification of DEGs and WGCNA analysis

For further understanding of the development of BL/BAL, the upregulated genes were selected in the BL and DLBCL datasets. We identified the DEGs and performed WGCNA analysis comparing the BL and DLBCL datasets. The genes expression analysis for the BL and DLBCL microarray data were different as shown by the volcano plot (Fig-3). The upregulated genes in the BL and DLBCL datasets were enriched by GO and KEGG pathway analysis (Fig-4 and Fig-5) where the *P* value was also defined as 0.05. The functions of the DEGs were completely different between the BL and DLBCL datasets.

The Topological Overlap Matrix (TOM) of the coexpressed genes in different modules in the top 5000 genes was shown in a heatmap (Fig-6), and eigengene adjacency heatmap for the different modules was also presented for the module and trait relationships. The coexpression network of the significant genes related to BL (Fig-7a) or DLBCL (Fig-7b) was constructed and GO enrichment of the significant genes related to BL (Fig-8a) and DLBCL (Fig-8b) was conducted. The two datasets also had different coexpression networks and GO enrichment genes.

Discussion

BL/BAL is an aggressive non-Hodgkin lymphoma derived from germinal center B cells, and BAL is regarded as the leukemic **phase** of BL [1]. The peak age of onset of sBL/BAL is 11 yr, and boys are affected much more frequently than girls. The enrolled ten patients included nine boys and one girl. The average age of these patients was 100.10 ± 13.39 m, and thus the patients in this study are similar to that reported in the previous literature [1-2, 10].

The prognosis of sBL/BAL was poor in past decades, but with short, intensive chemotherapy, the survival rate has improved steadily. Overexpression of CD20 and the restricted expression of Kappa or Lambda chains are remarkable in BL/BAL, and with combination treatment of intensive chemotherapy and a specific CD20 monoclonal antibody (rituximab), the survival rate of pediatric sBL/BAL has exceeded 90%. The EFS of the ten patients in the study was 100%, and the treatment, diagnosis, and results of these patients was consistent with the literature [13-14].

The molecular hallmark of BL/BAL is a translocation of the oncogenic *MYC*, and similar translocations are also expressed in other types of NHL [10]. Although a translocation of oncogenic *MYC* is detectable in these subtypes of NHL, their clinical manifestations and prognosis are diverse, and diversity of sBL/BAL also exists between children and adults [1, 4, 15]. This suggests that pediatric sBL/BAL can be distinguished from other types of NHL by gene expression profiling, and differentiation of the gene expression profiling potentially reflects distinct pathogenetic mechanisms.

Datasets of BL and DLBCL were collected, and differential expression analysis was performed. The differences were statistically significant and they revealed different gene expression profiles between BL and DLBCL, which indicates the pathogenetic mechanisms of BL and DLBCL are distinct.

Data of WES sequencing from the ten pediatric sBL/BAL of Chinese populations were analyzed. *ID3*, *BRCA2*, *ARID1A* and *SMARCA4* mutations, in addition to *MYC* mutations, were common (Fig-1). By literature review [10,15-18], NGS sequencing analysis has revealed the importance of the B-cell receptor signaling pathway in the pathogenesis of BL. Mutations of the transcription factor *TCF3* or its negative regulator *ID3* have been reported in around 70% of sBL cases. These mutations activate B-cell receptor signaling, which sustains BL cell survival by engaging the *PI3K* pathway. Other recurrent mutations in *CCND3*, *TP53*, *RHOA*, *SMARCA4*, and *ARID1A* occur in 5–40% of sBL/BAL cases. The findings of our study are similar to those reported in the literature [10,15-20]. In our patients, *ID3* was the 2nd most common recurrent gene mutation after *MYC* mutation. Other genes commonly altered in our patients were *BRCA2*, *ARID1A* and *SMARCA4*. Further studies need to validate our findings.

Both the number of mutations overall and the proportion of cases with mutations in *TCF3* or *ID3* are lower in endemic than in sporadic BL[19,20]. An inverse correlation between EBV infection and the number of mutations has been observed, suggesting that these mutations may take the place of the virus in the activation of B-cell receptor signaling. It has been predicted that these identified gene mutations

play important roles in the development of BL/BAL. Genetically susceptible individuals, such as those with germline *SH2D1A* mutations, are at a greatly increased risk of developing BL. It is interesting that germline causal gene mutations were detected in four of our ten patients, but larger samples and multiple centers are needed to verify their exact detection rate.

These mutated genes were enriched and connected by GO or KEGG pathways (Fig-2), and it seemed that the *PI3K-Akt* signaling pathway has a key role in the etiology of BL/BAL. Similar findings were also found in the datasets and reported in previous studies [16, 17]. BL/BAL may be inhibited by activation of this signaling pathway [18].

Epidermal growth factor receptor (*EGFR*) is a tyrosine kinase. *EGFR* gene mutations and overexpression of its protein are associated with cancer growth [21]. Tyrosine kinase inhibitors (*TKI*) against *EGFR* (*EGFR-TKI*) are used to treat cancer patients with *EGFR* mutations, such as those with lung adenocarcinoma, etc. [22]. However, an *EGFR-TKI* resistance pathway was found by KEGG enrichment in our BL/BAL patients, which reveals that *EGFR-TKI* treatment is invalid for BL/BAL. Further research is necessary to investigate this finding and its implications.

Conclusion

sBL/BAL is highly aggressive but curable subtype of lymphoma. With combination treatment of chemotherapy and rituximab, the survival rate of sBL/BAL has improved steadily although severe infection, included enterobacteriaceae producing extended-spectrum-lactamases, were common[23], but additional research into its pathogenetic mechanisms is necessary. The molecular hallmark of sBL/BAL is *MYC* translocation, but additional chromosomal abnormalities and gene mutations also occur and play roles in the progression of the disease.

In this study, NGS sequencing was used for pediatric sBL/BAL patients from a Chinese population. Other recurrent mutations in addition to *MYC* mutations were detected, and possible signaling pathways were also demonstrated. The limitations including a single centre study, small sample size and short follow up, make it necessary to validate our findings with larger multi-centric studies.

Trial registration

The study has been registered retrospectively at the Chinese Clinical Trial Registry (ChiCTR1900025690 and ChiCTR-IPR-14005706).

Abbreviations

Bone marrow: BM; biological processes: BP; Burkitt lymphoma: BL; Burkitt leukemia: BAL; cellular components: CC; central nervous system: CNS; Children's Hospital of Chongqing Medical University: CHCMU; complete remission: CR; computed tomography: CT; database of essential genes: DEGs; diffuse large B cell lymphoma: DLBCL; Epidermal growth factor receptor: *EGFR*; Epstein Barr virus: EBV; event-free survival: EFS; flow cytometry: FCM; fluorescence in situ hybridization: FISH; Gene Expression Omnibus: GEO; Gene Ontology: GO; hemoglobin: Hb; International Pediatric Non-Hodgkin Lymphoma Staging System: IPNHLSS; immunohistochemical: IHC; intrathecal injections: iT; kinase inhibitors: TKI; Kyoto Encyclopedia of Genes and Genomes: KEGG; lactate dehydrogenase: LDH; lymph node: LN; molecular function: MF; next-generation sequencing: NGS; peripheral blood: PB; platelet: PLT; positron emission tomography: PET-CT; protein to protein interactions; PPI; quality control: QC; RNA sequencing: RNAseq; sBL/BAL; Single nucleotide polymorphisms: SNPs; single nucleotide variants: SNVs; sporadic BL/BAL; SH2 domain protein 1A: *SH2D1A*; Topological Overlap Matrix: TOM; treatment-related mortality: TRM; weighted gene coexpression network analysis: WGCNA; white blood cell: WBC; whole-exome sequencing: WES.

Declarations

Ethics approval and consent to participate

The Ethics Administration Office of CHCMU granted ethics approval for this research (No.2018-75). Informed consent was obtained from the patients or their guardians. The study has been registered at the Chinese Clinical Trial Registry (ChiCTR1900025690 and ChiCTR-IPR-14005706).

Consent for publication

Not applicable.

Availability of data and material

Most data generated or analysed during this study are included in this published article and its supplementary information files. The raw data during the current study are not publicly available due to unfinished study but are available from the corresponding author on reasonable request.

Competing interests

The authors declare that there are no competing interests associated with this manuscript.

Funding

The study was supported by the National Natural Science Foundation of China (Project No. 81900162) and the Chongqing Science and Technology Commission of China PR (Project No. cstc2018jsyj-jsyjX0015).

Authors' contributions

JW X and J Z conceived and designed the study. L S and NG Y prepared the figures and tables. XY L analyzed and interpreted the data. XY L and JW X drafted the manuscript. JW X revised the manuscript. All authors read and approved the initial manuscript.

Acknowledgements

We thanks the patients and their parents for the help in the study.

References

- 1 Gopal S., Gross T G. (2018) How I treat Burkitt lymphoma in children, adolescents, and young adults in sub-Saharan Africa. *Blood*, **19**, 254-263. <https://doi.org/10.1182/blood-2018-04-844472>
- 2 Xiao J., Du S., Dai G., Gao G., Yang D., Zhao H. (2017) Efficacy and tolerability of chemotherapy in Chinese patients with AIDS-related Burkitt lymphoma and diffuse large B-cell lymphoma: An observational study. *Sci Rep*, **15**, 1905. <https://doi.org/10.1038/s41598-017-02086-4>
- 3 Rosolen A., Perkins SL., Pinkerton CR., Guillerman RP., Sandlund JT., Patte C. et al.(2015)Revised International Pediatric Non-Hodgkin Lymphoma Staging System. *J Clin Oncol*, **33**, 2112-2118. <https://doi.org/10.1200/JCO.2014.59.7203>
- 4 Alsharif R, Dunleavy K. (2019) Burkitt Lymphoma and Other High-Grade B-Cell Lymphomas with or without MYC, BCL2, and/or BCL6 Rearrangements. *Hematol Oncol Clin North Am.* **33**, 587-596. <https://doi.org/10.1016/j.hoc.2019.04.001>
- 5 Bialopiotrowicz E., Noyszewska-Kania M2, Kachamakova-Trojanowska N., Loboda A., Cybulska M., Grochowska A., et al. (2020) Serine Biosynthesis Pathway Supports MYC-miR-494-EZH2 Feed-Forward Circuit Necessary to Maintain Metabolic and

- Epigenetic Reprogramming of Burkitt Lymphoma Cells. *Cancers (Basel)*. **12**, pii: E580. <https://doi.org/10.3390/cancers12030580>
- 6 Rohde M., Bonn BR., Zimmermann M., Lange J., Möricke A., Klapper W. et al. (2017) Relevance of ID3-TCF3-CCND3 pathway mutations in pediatric aggressive B-cell lymphoma treated according to the non-Hodgkin Lymphoma Berlin-Frankfurt-Münster protocols. *Haematologica*. 102. 1091-1098. <https://doi.org/10.3324/haematol.2016.156885>
 - 7 Kamps R., Brandão RD., Bosch BJ., Paulussen AD., Xanthoulea S., Blok MJ., et al. (2017) Next-Generation Sequencing in Oncology: Genetic Diagnosis, Risk Prediction and Cancer Classification. *Int J Mol Sci*. **18**. pii: E308. <https://doi.org/10.3390/ijms18020308>
 - 8 Li C, Cai X, Chen X, Liang Y, Xia Z, Wang H. Long-term outcomes of modified BFM-95 regimen in adults with newly diagnosed standard-risk acute lymphoblastic leukemia: a retrospective single-center study. *Int J Hematol*. **110**, 2019:458–465. <https://doi.org/10.1007/s12185-019-02703-0>.
 - 9 Mottl H., Bajciová V., Nemec J., Al Shemmari S., Al Awadi S. High survival rate in childhood non-Hodgkin lymphoma without CNS involvement: results of BFM 95 study in Kuwait. *Pediatr Hematol Oncol*. **20**, 2003:103-110. <https://doi.org/10.1080/0880010390158603>.
 - 10 Christopher W., Eli W., Gru A A. (2018) Updates in the Pathology of Precursor Lymphoid Neoplasms in the Revised Fourth Edition of the WHO Classification of Tumors of Hematopoietic and Lymphoid Tissues. *Curr Hematol Malig Rep*. **13**. 275-288. <https://doi.org/10.1007/s11899-018-0456-8>
 - 11 Liao X., Guo Y., Shen Y., Xiao J. (2020) Recurrent Gastrointestinal Hemorrhage in Children with Philadelphia-Positive B-Cell Acute Lymphoblastic Leukemia Treated with Dasatinib: Case Reports. *Case Rep Hematol*. **10**. 2020:5678210. <https://doi.org/10.1155/2020/5678210>
 - 12 Li MM., Datto M., Duncavage EJ., Kulkarni S., Lindeman NI., Roy S., et al. (2017) Standards and Guidelines for Validating Next-Generation Sequencing Bioinformatics Pipelines: A Joint Recommendation of the Association for Molecular Pathology and the College of American Pathologists. *J Mol Diagn*, **19**. 4-23. <https://doi.org/10.1016/j.jmoldx.2017.11.003>
 - 13 Samochatova EV, Maschan AA, Shelikhova LN, Myakova NV, Belogurova MB, Khlebnikova OP, et al. Therapy of advanced-stage mature B-cell lymphoma and leukemia in children and adolescents with rituximab and reduced intensity induction chemotherapy (B-NHL 2004M protocol): the results of a multicenter study. *J Pediatr Hematol Oncol*, **36**, 2014:395-401. <https://doi.org/10.1097/MPH.0b013e31829d4900>
 - 14 Liu ZL, Liu PP, Bi XW, Lei DX, Wang Y, Li ZM, et al. Trends in Survival of Patients With Stage I/II Burkitt Lymphoma in the United States: A SEER Database Analysis. *Cancer Med*, **8**, 2019:874-881. <https://doi.org/10.1002/cam4.1870>.
 - 15 Bouska A, Bi C, Lone W, Zhang W, Kedwaii A, Heavican T, et al. Adult high-grade B-cell lymphoma with Burkitt lymphoma signature: genomic features and potential therapeutic targets. *Blood*, **130**, 2017:1819-1831. <https://doi.org/10.1182/blood-2017-02-767335>.
 - 16 Gehringer F, Weissinger SE, Möller P, Wirth T, Ushmorov A. Physiological levels of the PTEN-PI3K-AKT axis activity are required for maintenance of Burkitt lymphoma. *Leukemia*. **34**, 2020:857-871. <https://doi.org/10.1038/s41375-019-0628-0>
 - 17 Grande BM, Gerhard DS, Jiang A, Griner NB, Abramson JS, Alexander TB, et al.

- Genome-wide discovery of somatic coding and noncoding mutations in pediatric endemic and sporadic Burkitt lymphoma. *Blood*, **133**, 2019:1313-1324. <https://doi.org/10.1182/blood-2018-09-871418>.
- 18 Giulino-Roth L, van Besien H J, Dalton T, Totonchy JE, Rodina A, Taldone T, et al. Inhibition of Hsp90 Suppresses PI3K/AKT/mTOR Signaling and Has Antitumor Activity in Burkitt Lymphoma. *Mol cancer ther*, **16**, 2017:1779-1790. <https://doi.org/10.1158/1535-7163.MCT-16-0848>.
 - 19 Richter J, Schlesner M, Hoffmann S, Kreuz M, Leich E, Burkhardt B, et al. Recurrent mutation of the ID3 gene in Burkitt lymphoma identified by integrated genome, exome and transcriptome sequencing. *Nat Genet* **44**, 2012:1316–1320. <https://doi.org/10.1038/ng.2469>
 - 20 Love C, Sun Z, Jima D, Li G, Zhang J, Miles R, et al. The genetic landscape of mutations in Burkitt lymphoma. *Nat Genet*, **44**, 2012:1321-1325. <https://doi.org/10.1038/ng.2468>
 - 21 Naruse I, Ohmori T, Ao Y, Fukumoto H, Kuroki T, Mori M, et al. Antitumor activity of the selective epidermal growth factor receptort-yrosine kinase inhibitor (egfr-tki) iressa® (zd1839) in an egfr-expressing multidrug-resistant cell line in vitro and in vivo. *Int J Cancer*, **98**, 2002:310-315. <https://doi.org/10.1002/ijc.10173>.
 - 22 Adrian G. Sacher, Pasi A. J, Geoffrey R. Management of acquired resistance to epidermal growth factor receptor kinase inhibitors in patients with advanced non-small cell lung cancer. *Cancer*, **17**, 2014:2289-2298. <https://doi.org/10.1002/cncr.28723>.
 - 23 Agata B, Jędrzej P, Angelika S, Guzek, A, Piotr M, Katarzyna J, et al. Frequency of infections caused by esbl-producing bacteria in pediatric ward- single center five-year observation. *Arch Med Sci*. **15**, 2019: 688-693. <https://doi.org/10.5114/aoms.2017.72407>.

Fig-1 Identified genes and PPI in the BL/BAL patients

Fig-2 GO and KEGG enrichment for identified genes in the BL/BAL patients

Fig-3 Differential expression analysis for the BL and DLBCL microarray data

Fig-4 GO and KEGG enrichment for upregulated genes in the BL dataset

Fig-5 GO and KEGG enrichment for upregulated genes in the DLBCL dataset

Fig-6 Heatmap showing TOM of coexpressed genes in different modules in the top 5000 genes

Fig-7 The coexpression network of the significant genes related to BL or DLBCL

Fig-8 GO enrichment of the significant genes related to BL or DLBCL

Supplementary Tables

S1 Risk group of modified BFM-95 protocol

| Risk group | Definition |
|------------|---|
| R1 | Stage I or stage II, tumor was removed by surgery. |
| R2 | Stage I or stage II, tumor cannot be completely removed by surgery; stage III with value of LDH $\leq 2N$. |
| R3 | Stage III with value of LDH $> 2N$; stage IV or BAL without CNS involvement, value of LDH $\leq 2N$. |
| R4 | Stage IV or BAL with value of LDH $> 2N$; CNS involvement; Patients cannot achieve complete remission after 2 courses of chemotherapy. |

S2 Treatment planning of modified BFM-95 protocol

| Risk group | course | | | | | |
|------------|---------------------|-----------------|-----|-----------------|-----|-----|
| R1 | Ar ^{&} | Br | | | | |
| R2 | V+Ar | Br [#] | Ar | Br [#] | | |
| R3 | V+Ar | BBr | CCr | AAr | BBr | |
| R4 | V+Ar | BBr | CCr | AAr | BBr | CCr |

&: r=rituximab; #: VP-16 was omitted. Cranial radiotherapy is settled for patients with CNS involvement after chemotherapy; Evaluation is arranged after 1 or 2 courses of chemo for BAL or BL respectively.

S3 Schedule of modified BFM-95 protocol

| Course | Drugs | Dosage | Date of administration |
|---------------|---------------------------|---|------------------------|
| Induction (V) | Cyclophosphamide | 200mg/m ² , iv, 2h | Day 1-2 |
| | Vincristine | 1.5mg/m ² (max 2mg) | Day 1 |
| | Dexamethasone | 10mg/m ² , po or iv | Day 1-5 |
| | iT | Based on age | Day 1 |
| Ar/AAr | Ifosfamide | 800mg/m ² , iv, 1h | Day 1-5 |
| | Vincristine | 1.5mg/m ² (max 2mg) | Day 1 |
| | Methotrexate [#] | 1g or 5g/m ² , iv, 4h or 24h | Day 1 |
| | Etoposide | 100mg/m ² iv, 3h | Day 4-5 |
| | Cytarabine | 150mg/m ² , iv, 1h, q12h | Day 4-5 |
| | Dexamethasone | 10mg/m ² , po or iv | Day 1-5 |
| | Rituximab | 375mg/m ² iv >3h | Day 1 |
| | iT | Based on age | Day 1 |
| Br/BBr | Cyclophosphamide | 200mg/m ² , iv, 1h | Day 1-5 |
| | Etoposide | 60mg/m ² , iv, 2h | Day 1-3 |
| | Methotrexate [#] | 1g or 5g/m ² , iv, 4h or 24h | Day 1 |
| | Vincristine | 1.5mg/m ² (max 2mg) | Day 1 |
| | Etoposide | 100mg/m ² iv, 3h | Day 4-5 |
| | Doxorubicin | 25mg/m ² iv, 1h | Day 4-5 |
| | Dexamethasone | 10mg/m ² , po or iv | Day 1-5 |
| | Rituximab | 375mg/m ² iv >3h | Day 1 |

| | iT | Based on age | Day 1 |
|----|---------------|----------------------------------|---------|
| CC | Vindesine | 3mg/m ² (max 5mg), iv | Day 1 |
| | Dexamethasone | 20mg/m ² , po or iv | Day 1-5 |
| | Etoposide | 100mg/m ² , iv, 3h | Day 4-5 |
| | Cytarabine | 3g/m ² , iv, 3h, q12h | Day 4-5 |
| | Rituximab | 375mg/m ² iv >3h | Day 1 |
| | iT | Based on age | Day 5 |

#: course of A or AA, dosage of methotrexate is 1g/m² iv 4h or 5g/m² iv 24h respectively, administration of tetrahydrofolate begins at 42h after methotrexate, serum level of methotrexate is monitored at 48h, and tetrahydrofolate is withdrawn till serum level of methotrexate $\leq 1 \mu\text{mol/ml}$.

S4 Schedule of intrathecal injection

| Age | Methotrexate | Cytarabine | Dexamethasone |
|-------------------|--------------|------------|---------------|
| <12m | 6 mg | 15 mg | 2.5 mg |
| 12-36m | 9 mg | 25 mg | 2.5 mg |
| $\geq 36\text{m}$ | 12 mg | 30 mg | 5mg |

Fig-1 Identified genes and PPI in the BL/BAL patients



Fig-1a

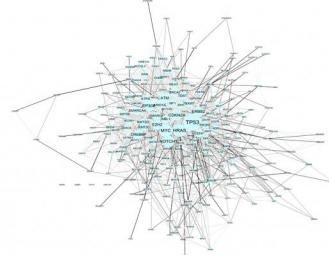


Fig-1b

Fig-2 GO and KEGG enrichment for identified genes in the BL/BAL patients

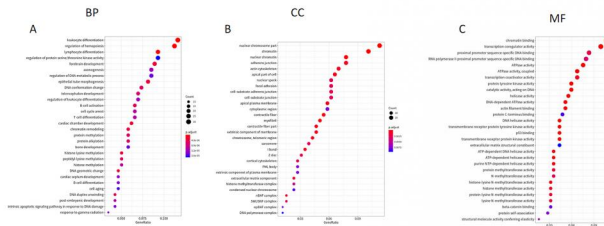


Fig-2a

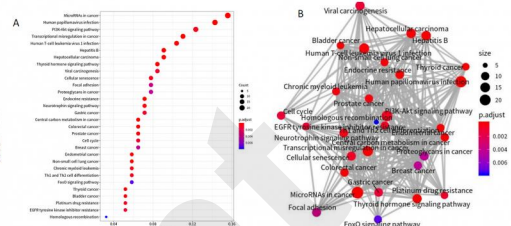


Fig-2b

Fig-3 Differential expression analysis for the BL and DLBCL microarray data

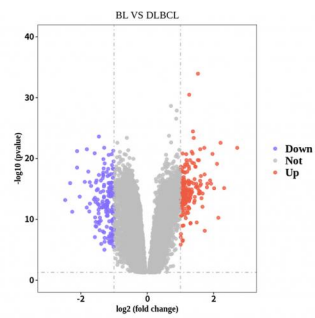


Fig-4 GO and KEGG enrichment for upregulated genes in the BL dataset

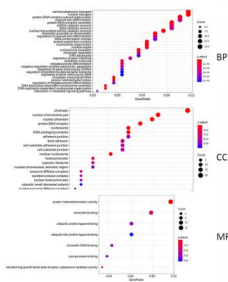


Fig-4a

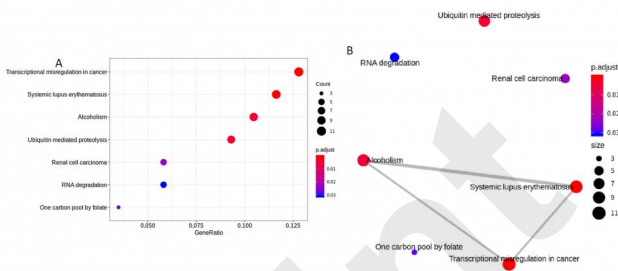


Fig-4b

Fig-7 The coexpression network of the significant genes related to BL or DLBCL



Fig-8 GO enrichment of the significant genes related to BL or DLBCL

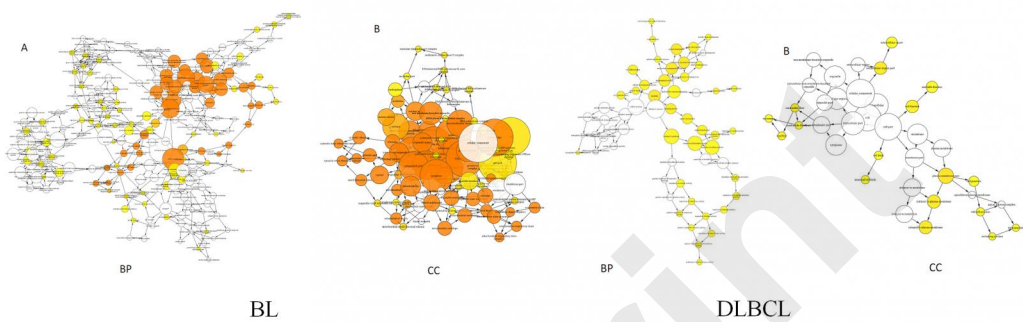


Fig-5 GO and KEGG enrichment for upregulated genes in the DLBCL dataset

Fig-5b

Fig-6 Heatmap showing TOM of coexpressed genes in different modules in the top 5000 genes