

Genetic Map of Relapsed and Refractory Diffuse Large B-cell Lymphoma: A Systemic Review and Association Analysis

Fan Gao

Peking University Third Hospital

Lei Tian

Peking University Third Hospital

Jing Wang

Peking University Third Hospital

Fei Dong

Peking University Third Hospital

Kai Hu

Peking University Third Hospital

Xiaoyan Ke (✉ xykbysy@163.com)

Peking University Third Hospital

Research article

Keywords: B-cell lymphoma, DLBCL, Hodgkin lymphoma

Posted Date: October 21st, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-93266/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Diffuse large B-cell lymphoma (DLBCL) is the most common histological subtype of non-Hodgkin lymphoma (NHL). In recent years, a deeper understanding of the genetic subtypes of diffuse large B lymphoma has been reached, and these advances have also been applied to research on relapsed and refractory diffuse large B-cell lymphoma (RRDLBCL). We screened 1495 documents, compiled the whole-exome sequencing data of several studies, formed a data set including 92 observations, and performed association analysis on the high-frequency mutations among them. The most common mutations in the data set include TTN (34/92, 37.0%), KMT2D (29/92, 31.5%), TP53 (25/92, 27.2%), IGLL5 (25/92, 27.2%), CREBBP (21 /92, 22.8%), BCL2 (21/92, 22.8%), MYD88 (20/92, 21.7%), and SOCS1 (19/92, 20.7%). Among these, CREBBP, KMT2D, and BCL2 have a strong association with each other, and SOCS1 has a strong association with genes such as ACTB, CIITA, and GNA13. There is also a strong association between SOCS1 and STAT6. Though TP53 and MYD88 lack significant associations with most genes, the association between MYD88 and PIM1 is significant. Through SOM clustering and expression-level analysis of common gene mutations, we believe that RRDLBCL can be divided into four main types: (1) JAK-STAT-related type, including STAT6, SOCS1, ITPKB, CIITA, and B2M. The expression lineage is similar to PMBL and cHL. (2) EZB type: BCL2 and EZHZ are the main types of mutations. Epigenetic mutations such as KMT2D and CREBBP are more common in this type, and are often accompanied by BCL2 mutations. (3) MCD type, including MYD88, CD79B and PIM1. These genes are involved in the BCR signaling pathway and related pathways, and are connected by the common NF- κ B pathway. (4) Undefined type (Sparse Mutation type). These patients are mainly individuals with sparse mutations, including some patients with TP53 mutations (30.3%, 10/33), but who generally lack characteristic mutations. Among the common gene mutations, the expression changes in BCL2, PIM1, STAT6, ITPKB, and GNA13 have more significant prognostic significance. We also reviewed the literature from recent years concerning the previously mentioned common gene mutations.

Introduction

Diffuse large B-cell lymphoma (DLBCL) is the most common histologic subtype of non-Hodgkin lymphoma (NHL). For patients who are refractory to initial treatment or relapse after an initial response, only a small percentage will experience prolonged disease-free survival with salvage chemoimmunotherapy alone [1]. In recent years, our understanding of the classification of diffuse large B lymphomas has deepened, and these advances have also been applied to the study of relapsed and refractory diffuse large B-cell lymphomas (RRDLBCL). Among recent advances, the multi-platform genome analysis performed by Schmitz et al. in 2018, based on gene expression classification of DLBCL, adds genetic classification that may be helpful for understanding the pathogenesis of DLBCL [2]. This milestone study identified four prominent genetic subtypes in DLBCL, termed MCD (based on the co-occurrence of MYD88 L265P and CD79B mutations), BN2 (based on BCL6 fusions and NOTCH2 mutations), N1 (based on NOTCH1 mutations), and EZB (based on EZH2 mutations and BCL2 translocations). Recently, Sha et al. defined a class of molecular high-grade diffuse large B-cell lymphoma (MHG) in their research and concluded that different treatment strategies were required for this group of patients [3].

With the contributions of these researchers, our understanding of the heterogeneity of DLBCL has reached a critical point. Since DLBCL includes various genetic subtypes, it is necessary to understand what role these components play in relapsed and refractory DLBCL. In the study by Schmitz et al., the results of survival analysis showed that MCD and N1 have the worst prognosis in the activated B-cell (ABC) subtype, while EZB has the worst prognosis in the germinal center B-cell (GCB) subtype. As clinicians, we pay particular attention to the composition of each genetic subtype in RRDLBCL. At the same time, whether there are undisclosed subtypes remains unclear. Further findings in this area can guide clinical design of new treatment strategies to improve the prognosis of patients with RRDLBCL.

At present, there are not enough large-scale studies to summarize the genetic characteristics of this particular population of RRDLBCL patients. To address this, we screened 1495 documents, from which we summarized whole-exome sequencing data from 4 studies of RRDLBCL [4–7]. By conducting an association study of whole-exome sequencing results for a total of 92 patients, we observed the most common mutations in RRDLBCL and attempted to cluster them. In addition, for the most high-frequency genes, we also used another data set (GSE10846) [8] to study the correlation between expression levels. Finally, we reviewed the research progress on these genes, and speculated the possible molecular biological mechanism of RRDLBCL.

Methods

Study selection

A systematic literature search for all relevant articles from January 2005 through August 2019 was conducted in MEDLINE. The search strategy combined the Medical Subject Headings and key words with terms for diffuse large B-cell lymphoma (DLBCL) and for whole-exome sequencing (WES). No language restrictions were imposed. Two investigators (T.L. and W.J.) reviewed all potentially relevant articles independently. We include both prospective and retrospective studies and, critically, complete whole-exome sequencing results of RRDLBCL. One reviewer (T.L.) extracted whole-exome sequencing data and clinical information from each eligible study, and another reviewer (W.J.) confirmed the data. In the end, we selected 4 articles from 1495 articles (Table 1) [4–7], and the data came from 5 independent clinical studies. We extracted the following information from eligible studies: first author, year of publication, journal, pathological subtypes of RRDLBCL, and whole-exome sequencing results (mutated genes, mutation types, etc.) corresponding to RRDLBCL patients.

Data processing and analysis

We standardized whole-exome sequencing extracted from different studies (standardization of mutation-type tags) and adopted partial exclusion criteria to avoid serious bias: (1) exclude mutations found in only one data set (highly suspected false positives); (2) exclude genes that affect a number of patients less than or equal to 3; (3) mutation type labels adopted: Truncating SNV, Splice site SNV, Missense SNV, Indel (insert or delete), Synonymous SNV, Multiple types (refers to the presence of multiple types of mutations in the

same patient sample); (4) use consistent gene names. After the above processing, we obtained a data set consisting of 92 observations and 128 variables (genes). We then conducted an association analysis [Apriorii and self-organizing map (SOM) algorithms, both based on the R package] on the data set to analyze the strength of the association between each gene and try to cluster it. For genes with strong associations, we rely on another data set (GSE10846) to analyze the correlation of their expressions. For the specific method, please refer to Supplemental S1. Visualization of relevant results was achieved using the R package and Gephi 0.9.2.

Table 1
List of all selected articles.

Author	Year*	Title	Number*	Included subtypes	Identifier	WES preparation
Mareschal S et al.	2016	Whole-exome sequencing of relapsed/refractory patients expands the repertoire of somatic mutations in diffuse large B-cell lymphoma	14	GCB,ABC□PMBL	GHE	SureSelect Human All Exon V5 kit (Agilent Technologies, Santa Clara, CA)
Greenawalt DM et al.	2017	Comparative analysis of primary versus relapse/refractory DLBCL identifies shifts in mutation spectrum	47	GCB,ABC□unclassified	E	NimbleGen Sequence Capture Human Exome 2.1M Array (Roche NimbleGen, Inc.)/Agilent SureSelect
Park HY et al.	2016	Whole-exome and transcriptome sequencing of refractory diffuse large B-cell lymphoma	6	GCB, non-GCB	F	SureSelect Human All Exon 50M (Agilent Technologies, Santa Clara, CA)
Morin RD et al.	2016	Genetic landscapes of relapsed and refractory diffuse large B-cell lymphomas	25	GCB, ABC, unclassified(NA)	QC-2/CH	Not mentioned; sequenced with the HiSeq 2000 platform (Illumina)
Abbreviation: GCB:germinal center B cell-like; ABC:activated B cell-like; PMBL: primary mediastinal large B cell lymphoma.						
#Year: the year of publication						
*Number: number of RRDLBCL patients						

Results

Association mining

The whole exome sequencing data set of 92 patients processed by the aforementioned method is relatively sparse. In this data set, the most common mutations are in TTN (34/92, 37.0%), KMT2D (29/92, 31.5%), TP53 (25/92, 27.2%), IGLL5 (25/92, 27.2%), CREBBP (21/92, 22.8%), BCL2 (21/92, 22.8%), MYD88 (20/92, 21.7%), and SOCS1 (19/92, 20.7%). These genes are also frequent item sets in association analysis (Figure 1a). After K-means clustering the data set, the data was divided into 2 groups (Figure S1). Cluster 1 consists mainly of patients with SOCS1 mutations. The three most common mutations in the group are SOCS1 (18/24, 75%), TTN (14/24, 58.3%), and IGLL5 (14/24, 58.3%). The most common mutations in cluster 2 are TTN (20/68, 29.4%), TP53 (18/68, 26.4%), and MYD88 (18/68, 26.5%). 18 of 24 patients in cluster 1 had SOCS1 mutations, while only 2 had MYD88 mutations; correspondingly, cluster 2 had 18 cases of MYD88 mutation and only 1 case with a SOCS1 mutation. Notably, only 1 patient has both MYD88 and SOCS1 mutation, only 3 patients have both MYD88 and CREBBP mutation, and the number of patients with TP53, MYD88, and SOCS1 mutation is 0 (Figure 1b). We used the apriori algorithm to observe the association between genes, and found that CREBBP, KMT2D, and BCL2 have a strong correlation with each other, and SOCS1 has a strong association with ACTB, CIITA, GNA13, and other genes (Figure S2a&b). There is a strong correlation between SOCS1 and STAT6. Among patients with SOCS1 mutations, 42.1% (8/19) also have STAT6 mutations (lift = 2.75). TP53 and MYD88 lack a meaningful association with most genes (lift \leq 1 indicates independence) (Figure S2c&d). Among genes with a relatively high incidence ($n \geq 8$), MYD88 is only strongly associated with PIM1, CD79B, HLA-A, USH2A, and UBR4 (lift between 1.59 and 2.27). Figure 1c shows the 300 rules with the highest lift value in the apriori algorithm, displayed based on the Fruchterman-Reingold layout. Genes such as SOCS1, BCL2, KMT2D, and CREBBP have more association rules with higher lift values, so the level of in-degree is higher. Although the incidence of TTN, IGLL5, DMD, and other genes is also high, the strength of association with other genes is relatively weak.

Gene clustering based on SOM self-organizing mapping neural network model

In order to further group the gene mutations in RRDLBCL, we took the lift of genes in the dataset relative to SOCS1, STAT6, KMT2D, CREBBP, PIM1, and MYD88 as variables, and clustered them by SOM algorithm. Figure 2 shows the results of clustering, where the size of the slice reflects the influence of variables on the objects in the cluster. The mean distance of objects to their closest code book vectors is 1.596, which indicates good mapping quality (Figure S3 shows the iterative process). By evaluating the generated SOM clusters, a set of ideal clusters were produced (Figure S4 from left to right and from bottom to top, clusters 1-16 respectively). Evaluation shows that all clusters except cluster 6 are ideal. Through cluster analysis based on a SOM self-organizing network, we found that from the perspective of inter-gene association, there are at least two major gene clusters in the genetic mutations of RRDLBCL. One cluster is a group of mutations including SOCS1 and STAT6, which are strongly associated with the SOCS1 gene and may be related to the JAK-STAT pathway. The other cluster is a group of genes, including KMT2D and CREBBP, related to epigenetic changes and strongly associated with BCL2. High-frequency genes such as MYD88 and TP53 are relatively independent. PIM1 has a broad association with many genes, but it is only strongly associated with MYD88 in high-frequency mutations.

Analysis of correlation between genes by expression array data and grouping of RRDLBCL mutant genes

In order to further analyze the relationship between hot-spot genes in RRDLBCL revealed by whole-exome sequencing, we used another public dataset based on whole human genome expression array of Affymetrix Human Genome U133 Plus 2.0 Array. Based on the framework provided by SOM self-organizing network clustering analysis, we perform linear analysis between the expression data of hot-spot genes. We found that STAT6 and SOCS1 not only showed a strong correlation in association analysis, but their expression levels were also highly correlated ($P < 0.01$). Among genes strongly associated with STAT6 and SOCS1, there are also genes such as ITPKB, CIITA, FAT4, ACTB, for which expression levels are highly correlated with both genes. Figure 3 shows the relationship of some hot-spot genes. The figure shows that the expression levels of ITPKB and CIITA are not only related to STAT6 and SOCS1, but also to each other. These genes constitute the most closely related group of hot-spot genes. In addition, B2M, RYR2, LOXHD1, and PKHD1L1 are closely related to STAT6. Although they are also strongly related to SOCS1, they are not related to SOCS1 at the expression level.

As mentioned, KMT2D, CREBBP and BCL2 are strongly related. However, analysis suggests that BCL2 is not related to KMT2D or CREBBP in terms of expression level. However, the expression levels of CREBBP and STAT6 are linearly related. MYD88 lacks closely related genes; among the high-frequency genes, only PIM1 has a strong correlation, linearly correlating with MYD88 in expression level ($P < 0.01$). PIM1 and CD79B also showed a strong correlation (lift = 3.04, $P < 0.01$). PIM1, CD79B, and PRDM15 constitute a closely related group. Although TP53 is one of the most high-frequency mutations, there is no high-frequency gene either strongly associated with it or showing a meaningful linear correlation in expression level.

Through the above analysis, we believe that there may be several combinations in RRDLBCL that summarize most of the mutation types: (1) mutations represented by STAT6 and SOCS1 that may be related to the JAK-STAT pathway; (2) KMT2D, CREBBP, and BCL2 mutations; (3) MYD88, CD79B, and other mutations that show strong correlation with PIM1. Based on this classification, we classified 92 patients (Supplemental S1) to observe the situation of each sub-category through Figure 4; patients on the left side of the heat map mainly carry SOCS1 and STAT6 mutations, and also include mutations considered to be closely related to SOCS1 and STAT6, such as ITPKB, FAT4, and B2M. In the middle of the heat map, there are mainly KMT2D, CREBBP, and BCL2 mutations. These three mutations are concentrated in some RRDLBCL patients. To the right of the KMT2D-CREBBP-BCL2 patient group on the heat map, there are patients with MYD88 or CD79B mutations, and about 1/4 (6/23) of them also carry PIM1 mutations. However, PIM1 mutations also appear in both JAK-STAT6 and KMT2D-CREBBP-BCL2. There is some overlap between these categories, and we call them complex types of RRDLBCL patients. Complex type 1 shown in the figure has characteristics of KMT2D-CREBBP-BCL2 and JAK-STAT6; complex type 2 has characteristics of KMT2D-CREBBP-BCL2 and PIM1-MYD88-CD79B, and of these, one case has all three types of features.

Relying on data from the same expression array, we predicted expression of the 18 major genes in Figure 3 based on survival status (survival/death). As can be seen from Figure S5, among deceased DLBCL patients,

STAT6, PIM1, ACTB, MYD88, and B2M are generally up-regulated, while LOXHD1, RYR2, PKHD1L1, FAT4, PRDM15, SOCS1, BCL2, CD79B, ITPKB, CIITA, CREBBP, and KMT2D are generally lowered.

Survival analysis based on expression of high-frequency genes

We used data from the same expression array to conduct survival analysis to evaluate the research value of the above genes. Patients were grouped according to expression levels of the corresponding genes (expression upregulation, expression downregulation, significant upregulation, significant downregulation). Results showed that expression levels of BCL2, PIM1, STAT6, ITPKB, and GNA13 had significant prognostic value. Among these, the expression level of BCL2 has relatively significant value in patients receiving the drug combination treatments CHOP ($P < 0.01$) and R-CHOP ($P = 0.1$) (Figure 5a-b). Patients with up-regulated BCL2 expression have a worse prognosis than others. The prognostic value of PIM1 is more prominent in patients receiving a CHOP regimen ($P = 0.02$), but there is no significant prognostic value in patients receiving an R-CHOP regimen (Figure 5c-d). Patients with upregulated PIM1 expression had a worse prognosis in patients receiving CHOP regimen. STAT6, ITPKB, and GNA13 have no significant value in patients treated with a CHOP regimen, but show significant prognostic value in patients treated with an R-CHOP regimen (Figure 5e-j). Patients with up-regulated STAT6 expression had a worse prognosis in patients receiving R-CHOP regimen ($P = 0.1$). Among patients treated with the R-CHOP regimen, patients with down-regulated expression of ITPKB ($P < 0.01$) and GNA13 ($P < 0.01$) have a worse prognosis.

Based on the aforementioned gene classification, we established a scoring system based on expression level and evaluated its prognostic significance. The system includes the following items: (1) STAT6 is significantly up-regulated or ITPKB/GNA13 is significantly down-regulated; (2) BCL2 expression level is significantly up-regulated; (3) PIM or MYD88 expression is significantly up-regulated. Each of these three items earns 1 point. As can be seen from Figure 6, with the accumulation of more types of abnormal expression, the patient's prognosis has a clear tendency to deteriorate. When the above three abnormalities exist at the same time, all of these patients die in about 3 years, and the 5-year survival rate is 0. This suggests that the occurrence of RRDLBCL and drug resistance are the result of the accumulation of multiple mechanisms.

Discussion

MYD88, CD79B, and PIM1: solitary independent MCD subtype

Wild-type MYD88 functions as an adaptor protein that mediates Toll-like receptor (TLR) and/or IL-1 receptor signaling. Mutations in MYD88 activate Toll-like receptor (TLR) signaling, and activating genetic events involving other positive regulators, including TLR2 and TRAF6 [2], have also been reported. CD79A and CD79B form a heterodimer and constitute an important part of the BCR signaling complex. Most mutations in CD79B affect the first tyrosine of the ITAM, resulting in increased cell surface IgM expression and reduced LYN kinase activity, which normally inhibits BCR signaling via negative feedback [9]. Thus, CD79B mutations

may promote lymphomagenesis by increasing cell surface BCR expression and reducing the negative feedback-mediated inhibition of BCR signaling [9].

MYD88 mutations and CD79A or CD79B mutations co-occur in ~10% of ABC DLBCLs. CD79A or CD79B mutations are more prevalent in ABC DLBCLs harboring MYD88(L265P) (34%) than in those without (18%) [10]. This indicates that these mutations have a synergistic effect in the development of ABC DLBCL. In a clinical trial, the BTK inhibitor ibrutinib created a response in 37% of ABC cases. The most significant response rate (80%) was observed in tumors with CD79B and MYD88 (L265P) mutations [11]. Phelan et al. discovered a My-T-BCR super complex: compared with other ABC DLBCL and GCB DLBCL cell lines, the survival rate of MYD88 L265P and CD79A mutant or CD79B mutant ABC DLBCL cell lines is more dependent on this complex [12]. This mechanism can explain the connection between MYD88 and CD79B. Patients with MYD88 mutations account for a large proportion of RRDLBCL. The MCD-DLBCL proposed by Schmitz et al. is named after MYD88 and CD79B. The results of this study also show that PIM1 is an important mutant gene in MCD DLBCL. Our research also found that among the high-frequency genes of RRDLBCL, PIM1 is most closely related to MYD88.

PIM1 is a member of the proto-oncogene serine/threonine-protein kinase (PIM) family. PIM1 was found to be an important effector of the signal transducer and activator of transcription (STAT) 3 and 5 transcription factors [13]. All PIM kinases can act as survival factors, preventing cells from undergoing apoptosis by activating BCL2 and inactivating the pro-apoptotic proteins BAD and ASK1 [14-16]. In DLBCL, PIM1 is a target for aberrant hypermutation, particularly in extra-nodal cases [17]. A 2012 study by Brault et al. showed that PIM expression in DLBCL is related to the activation of the JAK/STAT signaling pathway and proliferative activity. Correlation of nuclear PIM1 expression with disease stage, as well as a modest response to small-molecule inhibitors, suggests that PIM kinases are progression markers rather than primary therapeutic targets in DLBCL [18]. In our research, PIM1 and MYD88 are indeed the most closely related (6/14, 42.8%), but PIM1 does not always appear simultaneously with MYD88. At the expression level, PIM1 also showed significant correlation with CREBBP, TP53, STAT6, and KMT2D ($P < 0.01$). This may indicate that the impact of PIM1 is actually broader than originally supposed. In a 2018 study, PIM1 and MYD88 immunohistochemistry was performed on samples from 57 patients with PCNSL. The results suggest that PIM1 is mainly expressed in the nucleus, while MYD88 staining is largely cytoplasmic, with almost no nuclear staining. Compared with the control group, the positive expression rates of PIM1 and MYD88 were higher in PCNSL, and their expression levels were positively correlated ($r = 0.581$, $P = 2.0 \times 10^{-6}$) [19].

PRDM15 encodes a novel DNA-binding protein that regulates expression of key activators and repressors of the WNT and MAPK-ERK pathways at the chromatin level to safeguard naive pluripotency [20]. A recent study showed that PRDM15 regulates the transcription of key effectors of NOTCH and WNT/PCP pathways to preserve early midline structures in the developing embryo [21]. The correlation analysis in this study suggests that PRDM15 is related to PIM1 and CD79B. However, in most previous studies, PRDM1 is closely related to MCD rather than PRDM15. Nevertheless, the MCD subtypes in RRDLBCL are notable, and undoubtedly a subtype with poor prognosis in DLBCL.

BCL2, KMT2D and CREBBP: the only protagonist of EZB

In a study published in 2019, a class of molecular high-grade diffuse large B-cell lymphoma (MHG) was defined, which had significantly higher mutation frequencies than GCB in KMT2D, BCL2, MYC, or DDX3X. The progression-free survival rate at 36 months after R-CHOP in the MHG group was 37% (95% CI, 24-55%) compared with 72% (95% CI, 68-77%) for others. A 2019 NGS study on DHL and THL showed that the most frequently mutated genes were CREBBP (16/20 cases), followed by BCL2 (12/20), KMT2D (12/20), MYC (9/20), EZH2 (8/20), IGLL5 (8/20), FOXO1 (6/20), and SOCS1 (6/20) [22]. These studies indicate that mutations in KMT2D and CREBBP appear more in the so-called high-grade DLBCL.

KMT2D and CREBBP are both epigenetically regulated genes. KMT2D is a histone monomethyltransferase, which induces both monomethylated and dimethylated H3K4 (H3K4me1 and H3K4me2, respectively), thereby promoting transcription. KMT2D is mutated in about 30% of DLBCL and is one of the most frequent mutations in DLBCL (most are frameshift or nonsense mutations) [2,23]. Integrative genomic analyses indicate that KMT2D affects methylation of lysine 4 on histone H3 (H3K4) and expression of a set of genes, including those in the CD40, JAK-STAT, Toll-like receptor, and B-cell receptor signaling pathways. At the same time, KMT2D may promote malignant outgrowth by perturbing the expression of tumor suppressor genes that control B-cell-activating pathways [24].

CREBBP is a histone acetyltransferase that mediates H3K27 acetylation and is important for gene enhancer activation. Evidence suggests that CREBBP mutations are an early event in lymphoma because mutations in this gene are also found in hematopoietic stem cells [25]. CREBBP-mutant lymphomas have decreased expression of genes involved in germinal center exit, those responsible for plasma cell differentiation, and those associated with antigen presentation by MHC class II, suggesting that CREBBP deficiencies contribute to lymphomagenesis by blocking B-cell differentiation and facilitating immune escape [26]. In our research, we found that KMT2D and CREBBP mutations often appear at the same time, but there is no expression correlation between the two. At present, there are no other studies to prove that there is a straightforward mechanistic relationship between the two genes.

BCL2 abnormal activation is usually driven by genetic abnormalities in BCL2 itself, the most important of which is the t(14; 18) (q32; q21) translocation [27]. In addition, BCL2 gain or amplification is also related to BCL2 overexpression, which occurs almost exclusively in the ABC subtype (~14%). Somatic mutations involving the promoter and coding regions of BCL2 are caused by somatic hypermutation, can be detected in ~35% of DLBCLs, and are significantly enriched in the GCB subtype. The expression level of BCL2 is sufficient to indicate the prognosis, and we have repeated this result in the study. BCL2, KMT2D, and CREBBP are the main mutations in EZB-DLBCL in the system described by Schmitz et al. These three genes affect about 40% of RRDLBCL patients (37/92, 40.2%). The EZB-DLBCL category also includes mutations such as SOCS1 and STAT6. However, we believe that in RRDLBCL, the simultaneous occurrence of SOCS1 and STAT6 with BCL2, KMT2D, and CREBBP is still relatively rare. These two sets of mutations are a manifestation of DLBCL heterogeneity, and it is necessary to distinguish them. We discuss further below.

STAT6 and SOCS1: trunk of the JAK-STAT subtype

The JAK2-STAT signaling pathway is a survival pathway involved in the proliferation and differentiation of B cells. Deregulated JAK2-STAT signaling participates in the pathogenesis of subtypes of lymphoma [28]. In general, activation of the JAK2-STAT signaling pathway appears more in HL and PMBL [29-30]. HL is frequently associated with a 9p24.1 genomic amplification that includes the JAK2 locus, as well as with a cytokine-enriched tumor microenvironment. Thus, activation of the JAK2-STAT signaling pathway may promote tumor growth in HL [31-32]. As for PMBL, it shares molecular features with HL [33]. However, Schmitz states that activators of transcription (JAK-STAT) signaling may have been promoted in 49% of cases by a STAT6 mutation or amplification or by a mutation or deletion targeting SOCS1 [2].

SOCS1 encodes one of eight members of the suppressor-of-cytokine-signaling (SOCS) protein family. These proteins contain an SH2 domain and a "SOCS box" that play pivotal roles in the down-regulation of JAK activity [34]. SOCS1 mutations are found primarily in patients with PMBL or HL. These mutations were first identified in 42% of patients with cHL, and subsequently in a similar proportion of those patients with PMBL or nodular lymphocyte-predominant Hodgkin's lymphoma (NLPHL) [35-37]. In DLBCL, the SOCS1 mutation has previously been shown to be associated with good survival [38]. A 2016 study also showed that the SOCS1 mutations were exclusive to non-recurrent primary DLBCL and were completely absent in cases of relapsing DLBCL, which also supported a good prognosis [39]. However, our research shows that SOCS1 mutation still occurs in a considerable proportion of RRDLBCL, affecting about 20% of patients, which should not be ignored. These patients likely did not show a good prognosis because there are a large number of other mutations that are clearly associated with SOCS1, such as STAT6, ITPKB, and CIITA. This shows that the role of SOCS1 in DLBCL is likely to be more complicated than previously recognized.

STAT6 belongs to the STAT family of both adaptor proteins and transcription factors. STAT family members display a shared protein structure that is instrumental for their activation and functions [40]. This family plays a key role in the proliferation and survival of B lymphocytes and is often dysregulated in lymphomas [41]. In particular, expression and/or activation of STAT6 as well as amplification of the locus encoding STAT6 on chromosome 12 have been detected in more than 50% of PCNSL specimens [42]. In addition to PCNSL, similar to SOCS1, STAT6 mutations are also found in PMBL. An earlier study suggested that 20 of 55 (36%) PMBL cases harbor heterozygous missense mutations in the STAT6 DNA-binding domain, whereas no mutation in the gene was found in 25 diffuse large B-cell lymphoma samples [43]. Some studies have shown that STAT6 is associated with poor prognosis, but the sample size is small in these cases [44].

ITPKB, CIITA, B2M, FAT4, GNA13: branches of JAK-STAT subtype

Some genes with relatively few background studies have been screened in this study. Among them, ITPKB, GNA13, CIITA, B2M and FAT4 are closely related to STAT6 and/or SOCS1. ITPKB encodes for a kinase that converts the second messenger inositol trisphosphate (IP3) to IP4, a soluble antagonist of the AKT-activating PI3K-product IP3 [45]. Studies have shown that ITPKB may be associated with Alzheimer's

disease, Parkinson's disease, and common-variant immunodeficiency diseases [46-48]. In a 2018 study, cHL-encoding genomes were analyzed in tumors and normal cells of individually isolated lymphomas in biopsied tissue from 34 patients. The study found that ITPKB is one of the most frequently mutated genes in cHL (16% of total cases), and other mutated genes include STAT6 (32%), GNA13 (24%), and XPO1 (18%) [49]. In the data we compiled, ITPKB affected 16.3% of patients (15/92), of which only one case was PMBL (GHE0645). 1/3 of these patients also carried GNA13 (5/15, 33.3%), which shows that a segment of RRDLBCL has a mutational gene spectrum highly similar to cHL.

CIITA encodes a protein with an acidic transcriptional activation domain, 4 LRRs (leucine-rich repeats), and a GTP-binding domain. Research into the role of this gene in lymphoma has focused on gray zone lymphoma (GZL) [50]. In a 2019 study of GZL, 139 cases of were classified into cHL-like GZL and LBCL-like GZL. A continuous morphologic and immunophenotypic spectrum was observed within these 2 GZL categories. The majority of cases presented genetic immune escape features with CD274/PDCD1LG2 and/or CIITA structural variants by fluorescence *in situ* hybridization. Cases with LBCL-like morphology had more PD-L1/PD-L2 or CIITA rearrangements than cHL-like cases [51].

Beta-2-microglobulin (B2M) encodes a serum protein found in association with the major histocompatibility complex (MHC) class I heavy chain on the surface of nearly all nucleated cells. B2M is the most commonly altered gene in HRS cells. Studies have found that 7/10 of cHL cases have inactive mutations. The study also indicated that wild-type B2M expression in a cHL cell line restored MHC-I expression [52]. In addition, recent studies have suggested that B2M is also altered frequently in PMBL [53].

There is little published research on FAT atypical cadherin 4 (FAT4). The protein encoded by this gene is a member of the protocadherin family. This gene may play a role in regulating planar cell polarity (PCP). Previous studies have found frequent mutations in FAT4 in spleen marginal lymphomas and extranodal nasal NK/T-cell lymphomas. Because of their potential role in lymphoma formation, there is value in further studying this gene in the context of neoplastic growth [54-55].

The GNA13 mutation affected 13% (12/92) of RRDLBCL patients and showed a clear prognostic significance in the survival analysis of expression data in this study. Previous studies suggest that GNA13 is the most common mutant gene in germinal center (GC)-derived B-cell lymphoma, including nearly a quarter of Burkitt's lymphoma and GC-derived diffuse large B-cell lymphoma cases [56]. GNA13 is involved in Gna13 signaling, which restricts germinal center B cells to the germinal center, suggesting that GNA13 mutations are related to cell migration [57]. Some researchers have modeled the GNA13-deficient state exclusively in GC B cells by crossing the Gna13 conditional knockout mouse strain with the GC-specific AID-Cre transgenic strain. GNA13 deficiency, combined with conditional MYC transgene expression in mouse GC B cells, ultimately promoted lymphomagenesis [58]. In the association analysis conducted in our study, the association between GNA13 and SOCS1 and STAT6 was prominent, but the expression level was only linearly correlated with STAT6 (P=0.06).

Of these genes that are potentially linked to the JAK-STAT pathway, some are related to immune escape (B2M, CIITA), some are related to cell migration (GNA13), and some require further study. We found that the expression levels of ITPKB and GNA13 are related to prognosis, which previous studies have also suggested

[59]. If SOCS1 and STAT6 represent the significance of JAK-STAT pathway in the pathogenesis of DLBCL, then these genes related to JAK-STAT may affect the remission and prognosis of DLBCL on multiple levels. The role of the entire JAK-STAT pathway in RRDLBCL may be more complicated than we imagined, suggesting that we should treat it independently from EZB as represented by BCL2 and CREBBP.

RRDLBCL classification and precision medicine

Our research relies on part of the published research data, and there is an inevitable lack of information. For example, the NOTCH2 mutation was discarded in the process of dimensionality reduction because of a significant false positive in one of the data sources. However, overall, the results of this study are still beneficial for the classification and treatment of RRDLBCL. In terms of classification, the so-called MCD and EZB types cover about 2/3 of RRDLBCL. In EZB classification, we believe that it is meaningful to distinguish mutations related to the JAK-STAT pathway from mutations represented by BCL2 and CREBBP; this refines classification of RRDLBCL and provides guidance for further research. In sum, RRDLBCL is classified into four types, and the pathways and mechanisms involved are shown in Figure 7:

- (1) JAK-STAT related type: including STAT6, SOCS1, ITPKB, CIITA, and B2M. The expression lineage is similar to PMBL and cHL. This type involves many mechanisms, such as cell migration and immune escape. However, in general, there is a close relationship with the JAK-STAT pathway.
- (2) EZB type: BCL2 and EZHZ are the main types of mutations. Epigenetic mutations such as KMT2D and CREBBP are more common in this type, and are often accompanied by BCL2.
- (3) MCD type: including MYD88, CD79B, and PIM1. These genes are involved in the BCR signaling and related pathways, and are connected with the common NF- κ B pathway. This is the most independent category, with a more conservative mutant gene lineage.
- (4) Undefined type (Sparse Mutation type): These patients are mainly individuals with sparse mutations, including some patients with TP53 mutations (30.3%, 10/33), and they generally lack characteristic mutations. The most frequent mutations other than TP53 are found in TTN, IGLL5, and DMD, but these mutations are also very common in other patients. Of the 128 genes we used as variables, the Sparse Mutation type has an average of 11.65 gene mutations per person. In contrast, each of the other types of patients has an average of 13.8 genes mutated.

Two points bear consideration: (1) There are complex types of mutations. That is, a patient may have two or more types of mutations in the above categories 1-3; (2) For RRDLBCL of Undefined type, the relapse refractory mechanism may be difficult to analyze by genetic mutation, which means that there may be other reasons (we have not observed a meaningful mutation pattern in this group of patients). We observe that patients with multiple types of expression patterns have a worse prognosis, but there remains a lack of robust clinical research evidence. The presence of Sparse Mutation types indicates that NGS is by no means the endpoint of RRDLBCL studies. The above classification does not require WES for every patient. In fact, it only needs to detect some hot-spot genes to achieve the classification goal. We suggest that a reasonable

panel should include at least the following genes: (1) meaningful for sub-classification: SOCS1, STAT6, B2M, CIITA, BCL2, KMT2D, CREBBP, EZHZ, MYD88, CD79B, NOTCH2, and TP53; (2) May be helpful for classification and have prognostic significance: PIM1, ITPKB, GNA13, MYC, and CD58 [59].

In terms of treatment, patients may benefit from a more detailed classification system. Currently, salvage immunochemotherapy and autologous transplantation are still the preferred treatment options for RRDLBCL. Currently, salvage immunochemotherapy and autologous hematopoietic stem cell transplantation (ASCT) are still the preferred treatment options for RRDLBCL [60]. However, there are patients who are not suited to receive ASCT, and who find it difficult to achieve remission by general immunochemotherapy. In this context, many optional chimeric antigen receptor T-cell (CAR-T) immunotherapies, other immunotherapies, and targeted therapies make individualized treatment options complicated and subjective. By further classifying RRDLBCL based on mutations, combined with the pathological classification of GCB and ABC, the choice of treatment may become more directional and individualized. The NCT02077166 study applied ibrutinib combined with lenalidomide and rituximab to patients with RRDLBCL. In response-evaluable patients, the overall response rate (ORR) was 44% (complete response [CR], 28%); among these, the ORR was 65% (CR, 41%) in non-GCB and 69% and 56% in relapsed (n=16) and secondary refractory (n=27) disease, respectively [61]. As mentioned, the ibrutinib response rate is comparatively good for MCD DLBCL [11]. Available data from early clinical trials indicate that the single-agent pembrolizumab shows encouraging efficacy in patients with DLBCL subtypes with 9p24.1 chromosome amplification and PD-L1 overexpression, and it is well tolerated [62]. JAK-STAT related types share similar gene mutation profiles with cHL, and there is already evidence that patients with relapsed and refractory cHL can benefit from treatment of anti-human PD-1 monoclonal antibodies [63-64]. Therefore, we anticipate that immune checkpoint inhibitors may also improve the prognosis of these DLBCL patients. In summary, in order to practice precision medicine, a more specific and practical subtype classification is needed. This study proposes some feasible designs for achieving this purpose. At the same time, more in-depth research is required to translate these results into clinical practice.

Abbreviations

DLBCL ☒ Diffuse large B-cell lymphoma

NHL ☒ Non-Hodgkin lymphoma

RRDLBCL ☒ Relapsed and refractory diffuse large B-cell lymphoma

MHG ☒ Molecular high-grade diffuse large B-cell lymphoma

WES ☒ Whole-exome sequencing

SNV ☒ Single nucleotide variants

Indel ☒ Insertion or deletion mutation

GCB ☒ Germinal center B cell-like

ABC☒Activated B cell-like

PMBL: Primary mediastinal large B cell lymphoma

TLR☒Toll-like receptor

ITAM☒Immunoreceptor tyrosine-based activation motif

cHL☒Classical Hodgkin lymphoma

GC☒Germinal center

CAR-T☒Chimeric antigen receptor T-cell

ASCT☒Autologous stem cell transplantation

ORR☒Overall response rate

Declarations

Ethics approval and consent to participate

This research does not involve human and animal experiments.

Consent for publication

Not applicable.

Availability of data and materials

Data and materials supporting the conclusion of this study have been included within the article and the additional files. Some datasets generated during and/or analyses during the current study are available in the NCBI database (GSE10846).

Consent for publication

The authors declare that they have no competing interests.

Funding

Not applicable.

Authors' contributions

Gao Fan conducted data analysis, visualization and article writing. Tian Lei and Wang Jing screened candidate literatures and integrated whole-exome sequencing data. Ke Xiaoyan and Hu Kai supervised methodology and results analysis of the research.

Authors' information

Department of Hematology, Peking University Third Hospital, Beijing.

Gao Fan✉Tian Lei✉Wang Jing✉Dong Fei, Ke Xiaoyan

Department of Adult Lymphoma, Beijing Boren Hospital, Beijing

Hu Kai

Corresponding author

Correspondence to Ke Xiaoyan.

References

1. Gisselbrecht C, Glass B, Mounier N, et.al. Salvage regimens with autologous transplantation for relapsed large B-cell lymphoma in the rituximab era. *J Clin Oncol.* 2010;28(27):4184. Epub 2010 Jul 26.
2. Schmitz R, Wright GW, Huang DW, et al. Genetics and Pathogenesis of Diffuse Large B-Cell Lymphoma. *N Engl J Med.* 2018;378(15):1396–1407.
3. Sha C, Barrans S, Cucco F et al. Molecular High-Grade B-Cell Lymphoma: Defining a Poor-Risk Group That Requires Different Approaches to Therapy. *J Clin Oncol.* 2019 Jan 20;37(3):202-212.
4. Morin RD, Assouline S, Alcaide M, et.al. Genetic Landscapes of Relapsed and Refractory Diffuse Large B-Cell Lymphomas. *Clin Cancer Res.* 2016 May 1;22(9):2290-300.
5. Mareschal S, Dubois S, Viailly PJ, et.al. Whole-exome sequencing of relapsed/refractory patients expands the repertoire of somatic mutations in diffuse large B-cell lymphoma. *Genes Chromosomes Cancer.* 2016 Mar;55(3):251-67.
6. Park HY, Lee SB, Yoo HY, et.al. Whole-exome and transcriptome sequencing of refractory diffuse large B-cell lymphoma. *Oncotarget.* 2016 Dec 27;7(52):86433-86445.
7. Greenawalt DM, Liang WS, Saif S, et.al. Comparative analysis of primary versus relapse/refractory DLBCL identifies shifts in mutation spectrum. *Oncotarget.* 2017 Jun 15;8(59):99237-99244.
8. Lenz G, Wright G, Dave SS, et.al. Stromal gene signatures in large-B-cell lymphomas. *N Engl J Med.* 2008 Nov 27;359(22):2313-23.

9. Davis, R. E. et al. Chronic active B cell-receptor signaling in diffuse large B cell lymphoma. *Nature* 463, 88–92 (2010).
10. Ngo, V. N. et al. Oncogenically active MYD88 mutations in human lymphoma. *Nature* 470, 115–119 (2011).
11. Wilson WH et al. Targeting B cell receptor signaling with ibrutinib in diffuse large B cell lymphoma. *Nat Med* 21, 922–926, (2015).
12. Phelan, J. D. et al. A multiprotein supercomplex controlling oncogenic signaling in lymphoma. *Nature* 560, 387–391 (2018).
13. Yip-Schneider MT, Horie M, Broxmeyer HE. Transcriptional induction of pim-1 protein kinase gene expression by interferon gamma and posttranscriptional effects on costimulation with steel factor. *Blood*. 1995 Jun 15; 85(12):3494-502.
14. Lilly M, Sandholm J, Cooper JJ, Koskinen PJ, Kraft A. The PIM-1 serine kinase prolongs survival and inhibits apoptosis-related mitochondrial dysfunction in part through a bcl-2-dependent pathway. *Oncogene*. 1999 Jul 8; 18(27):4022-31.
15. Aho TL, Sandholm J, Peltola KJ, Mankonen HP, Lilly M, Koskinen PJ. Pim-1 kinase promotes inactivation of the pro-apoptotic Bad protein by phosphorylating it on the Ser112 gatekeeper site. *FEBS Lett*. 2004 Jul 30; 571(1-3):43-9.
16. Gu JJ, Wang Z, Reeves R, Magnuson NS. PIM1 phosphorylates and negatively regulates ASK1-mediated apoptosis. *Oncogene*. 2009 Dec 3; 28(48):4261-71.
17. Deutsch AJ, Aigelsreiter A, Staber PB, et.al. MALT lymphoma and extranodal diffuse large B-cell lymphoma are targeted by aberrant somatic hypermutation. *Blood*. 2007 Apr 15; 109(8):3500-4.
18. L Brault, T Menter, E C Obermann, et.al. PIM kinases are progression markers and emerging therapeutic targets in diffuse large B-cell lymphoma. *Br J Cancer*. 2012 Jul 24; 07(3): 491–500.
19. Zhou Y, Liu W, Xu Z, et.al. Analysis of Genomic Alteration in Primary Central Nervous System Lymphoma and the Expression of Some Related Genes. *Neoplasia*. 2018 Oct;20(10):1059-1069.
20. Mzoughi S, Zhang J, Hequet D, et.al. PRDM15 safeguards naive pluripotency by transcriptionally regulating WNT and MAPK-ERK signaling. *Nat Genet*. 2017 Sep;49(9):1354-1363.
21. Mzoughi S, Di Tullio F, Low DHP, Motofeanu CM, et.al. PRDM15 loss of function links NOTCH and WNT/PCP signaling to patterning defects in holoprosencephaly. *Sci Adv*. 2020 Jan 10;6(2):eaax9852.
22. Evrard SM, Péricart S, Grand D. Targeted next generation sequencing reveals high mutation frequency of CREBBP, BCL2 and KMT2D in high-grade B-cell lymphoma with MYC and BCL2 and/or BCL6 rearrangements. *Haematologica*. 2019 Apr;104(4):e154-e157.
23. Morin, R. D. et al. Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. *Nature* 476, 298–303 (2011).
24. Ortega-Molina A, Boss IW, Canela A, et.al. The histone lysine methyltransferase KMT2D sustains a gene expression program that represses B cell lymphoma development. *Nat Med*. 2015 Oct;21(10):1199-208.
25. Horton, S. J. et al. Early loss of Crebbp confers malignant stem cell properties on lymphoid progenitors. *Nat. Cell Biol*. 19, 1093–1104 (2017).

26. Jiang, Y. et al. CREBBP inactivation promotes the development of HDAC3-dependent lymphomas. *Cancer Discov.* 7, 38–53 (2017).
27. Yi Miao, L. Jeffrey Medeiros, Yong Li, Jianyong Li & Ken H. Young. Genetic alterations and their clinical implications in DLBCL. *Nat Rev Clin Oncol.* 2019 Oct;16(10):634-652.
28. L.M. Scott, M.K. Gandhi et.al. Deregulated JAK/STAT signaling in lymphomagenesis, and its implications for the development of new targeted therapies. *Blood Rev.*, 29 (2015), pp. 405-415.
29. Hartmann S, Martin-Subero JI, Gesk S, et al. Detection of genomic imbalances in microdissected Hodgkin and reed-Sternberg cells of classical Hodgkin's lymphoma by array-based comparative genomic hybridization. *Haematologica.* 2008;93(9):1318–1326.
30. Rosenwald A, Wright G, Leroy K, et al. Molecular diagnosis of primary mediastinal B cell lymphoma identifies a clinically favorable subgroup of diffuse large B cell lymphoma related to Hodgkin lymphoma. *J Exp Med.* 2003;198(6):851–862.
31. Aldinucci D, Gloghini A, et.al. The classical Hodgkin's lymphoma microenvironment and its role in promoting tumour growth and immune escape. *J Pathol.* 2010;221(3):248–263.
32. Navarro A, Diaz T, Martinez A, et al. Regulation of JAK2 by miR-135a: prognostic impact in classic Hodgkin lymphoma. *Blood.* 2009;114(14):2945–2951.
33. Kim SJ, Yoon DH, Kang HJ, et.al. Ruxolitinib shows activity against Hodgkin lymphoma but not primary mediastinal large B-cell lymphoma. *BMC Cancer.* 2019 Nov 10;19(1):1080.
34. Scott LM, Gandhi MK. Deregulated JAK/STAT signaling in lymphomagenesis, and its implications for the development of new targeted therapies. *Blood Rev.* 2015 Nov;29(6):405-15.
35. M.A. Weniger, I. Melzner, C.K. Menz, S. et al. Mutations of the tumor suppressor gene SOCS-1 in classical Hodgkin lymphoma are frequent and associated with nuclear phospho-STAT5 accumulation. *Oncogene*, 25 (2006), pp. 2679-2684
36. A. Mottok, C. Renne, K. Willenbrock, et.al. Somatic hypermutation of SOCS1 in lymphocyte-predominant Hodgkin lymphoma is accompanied by high JAK2 expression and activation of STAT6. *Blood*, 110 (2007), pp. 3387-3390
37. A. Mottok, C. Renne, M. Seifert, E. et al. Inactivating SOCS1 mutations are caused by aberrant somatic hypermutation and restricted to a subset of B-cell lymphoma entities. *Blood*, 114 (2009), pp. 4503-4506.
38. Schif B, Lennerz JK, Kohler CW, Bentink S, Kreuz M, Melzner I, et al. SOCS1 mutation subtypes predict divergent outcomes in diffuse large B-cell lymphoma (DLBCL) patients. *Oncotarget.* 2013;4:35–47.
39. Juskevicius D, Lorber T, Gsponer J, Perrina V, Ruiz C, Stenner-Liewen F, et al. Distinct genetic evolution patterns of relapsing diffuse large B-cell lymphoma revealed by genome-wide copy number aberration and targeted sequencing analysis. *Leukemia.* 2016;30:2385–95.
40. Miloudi H, Leroy K, Jardin F, Sola B STAT6 is a cargo of exportin 1: Biological relevance in primary mediastinal B-cell lymphoma. *Cell Signal.* 2018 Jun;46:76-82.
41. Mondello P, Cuzzocrea S, Arrigo C, et.al. STAT6 activation correlates with cerebrospinal fluid IL-4 and IL-10 and poor prognosis in primary central nervous system lymphoma. *Hematol Oncol.* 2020 Feb;38(1):106-110.

42. Schwindt H, Vater I, Kreuz M, et al. Chromosomal imbalances and partial uniparental disomies in primary central nervous system lymphoma. *Leukemia*. 2009;23(10):1875-1884.
43. Ritz O, Guitter C, Castellano F, et al. Recurrent mutations of the STAT6 DNA binding domain in primary mediastinal B-cell lymphoma. *Blood*. 2009 Aug 6;114(6):1236-42.
44. Yang SH, Lee KS, Kim IS, et al. Long-term survival in primary CNS lymphoma treated by high-dose methotrexate monochemotherapy: role of STAT6 activation as prognostic determinant. *J Neurooncol*. 2009 Mar;92(1):65-71.
45. Westernberg L, Conche C, Huang YH, et al. Non-canonical antagonism of PI3K by the kinase Itpkb delays thymocyte β -selection and renders it Notch-dependent. *eLife*. 2016;5:e10786.
46. Salta E, Sierksma A, Vanden Eynden E, De Strooper B. miR-132 loss de-represses ITPKB and aggravates amyloid and TAU pathology in Alzheimer's brain. *EMBO Mol Med*. 2016 Sep 1;8(9):1005-18.
47. Chew EGY, Tan LCS, Au WL, et al. ITPKB and ZNF184 are associated with Parkinson's disease risk in East Asians. *Neurobiol Aging*. 2020 Feb;86:201.e15-201.e17.
48. Louis AG, Yel L, Cao JN, Agrawal S, Gupta S. Common variable immunodeficiency associated with microdeletion of chromosome 1q42.1-q42.3 and inositol 1,4,5-trisphosphate kinase B (ITPKB) deficiency. *Clin Transl Immunology*. 2016 Jan 22;5(1):e59.
49. Tiacci E, Ladewig E, Schiavoni G, Pervasive mutations of JAK-STAT pathway genes in classical Hodgkin lymphoma. *Blood*. 2018 May 31;131(22):2454-2465.
50. Eberle FC, Salaverria I, Steidl C, et al. Gray zone lymphoma: chromosomal aberrations with immunophenotypic and clinical correlations. *Mod Pathol*. 2011;24:1586-1597.
51. Sarkozy C, Copie-Bergman C, Damotte D, et al. Gray-zone Lymphoma Between cHL and Large B-Cell Lymphoma: A Histopathologic Series From the LYSA. *Am J Surg Pathol*. 2019 Mar;43(3):341-351.
52. Reichel J, Chadburn A, Rubinstein PG, et al. Flow sorting and exome sequencing reveal the oncogenome of primary Hodgkin and Reed-Sternberg cells. *Blood*. 2015 Feb 12;125(7):1061-72.
53. Chapuy B, Stewart C, Dunford AJ, et al. Genomic analyses of PMBL reveal new drivers and mechanisms of sensitivity to PD-1 blockade. *Blood*. 2019 Dec 26;134(26):2369-2382.
54. Parry M, Rose-Zerilli MJ, Gibson J, et al. Whole-exome sequencing identifies novel recurrently mutated genes in patients with splenic marginal zone lymphoma. *PLoS One*. 2013 Dec 13;8(12):e83244.
55. Dobashi A, Tsuyama N, Asaka R, et al. Frequent BCOR aberrations in extranodal NK/T-Cell lymphoma, nasal type. *Genes Chromosomes Cancer*. 2016 May;55(5):460-71.
56. Zhang J, Grubor V, Love CL, et al. Genetic heterogeneity of diffuse large B-cell lymphoma. *Proc Natl Acad Sci U S A*. 2013 Jan 22;110(4):1398-403.
57. Green, J. A. et al. The sphingosine 1-phosphate receptor S1P(2) maintains the homeostasis of germinal center B cells and promotes niche confinement. *Nat. Immunol*. 12, 672–680 (2011).
58. Healy JA, Nugent A, Rempel RE, GNA13 loss in germinal center B cells leads to impaired apoptosis and promotes lymphoma in vivo. *Blood*. 2016 Jun 2;127(22):2723-31.
59. Dubois S, Viailly PJ, Mareschal S, et al. Next-Generation Sequencing in Diffuse Large B-Cell Lymphoma Highlights Molecular Divergence and Therapeutic Opportunities: a LYSA Study. *Clin Cancer Res*. 2016

Jun 15;22(12):2919-28.

60. https://www.nccn.org/professionals/physician_gls/pdf/b-cell.pdf (Accessed on September 20, 2018).
61. Andre Goy, Radhakrishnan Ramchandren, Nilanjan Ghosh. Brutinib Plus Lenalidomide and Rituximab Has Promising Activity in Relapsed/Refractory Non-Germinal Center B-cell-like DLBCL. *Blood*. 2019 Sep 26;134(13):1024-1036.
62. Semira Sheikh, John Kuruvilla, Pembrolizumab for the Treatment of Diffuse Large B-cell Lymphoma. *Expert Opin Biol Ther*. 2019 Nov;19(11):1119-1126.
63. Younes A, Santoro A, Shipp M, et.al. Nivolumab for classical Hodgkin's lymphoma after failure of both autologous stem-cell transplantation and brentuximab vedotin: a multicentre, multicohort, single-arm phase 2 trial. *Lancet Oncol*. 2016 Sep;17(9):1283-94. Epub 2016 Jul 20.
64. Chen R, Zinzani PL, Lee HJ, Pembrolizumab in relapsed or refractory Hodgkin lymphoma: 2-year follow-up of KEYNOTE-087. *Blood*. 2019;134(14):1144. Epub 2019 Aug 13.

Figures

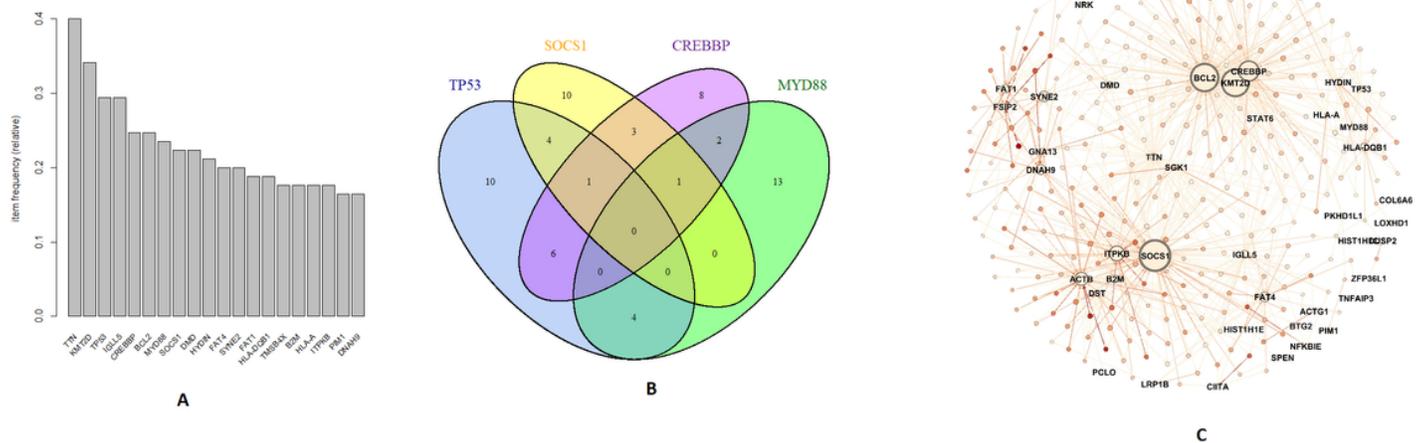


Figure 1

a) Frequent itemsets of exon mutant genes in 92 patients. b) Different patient sets with four mutations of SOCS1, TP53, MYD88, and CREBBP, visualized with a Venn diagram. c) Visualization of 300 rules with Gephi (Fruchterman Reingold layout). The shade depends on the lift value, and label size is proportional to the in-degree, the number of rules the consequent participates in.

Results of SOM: Importance of variables

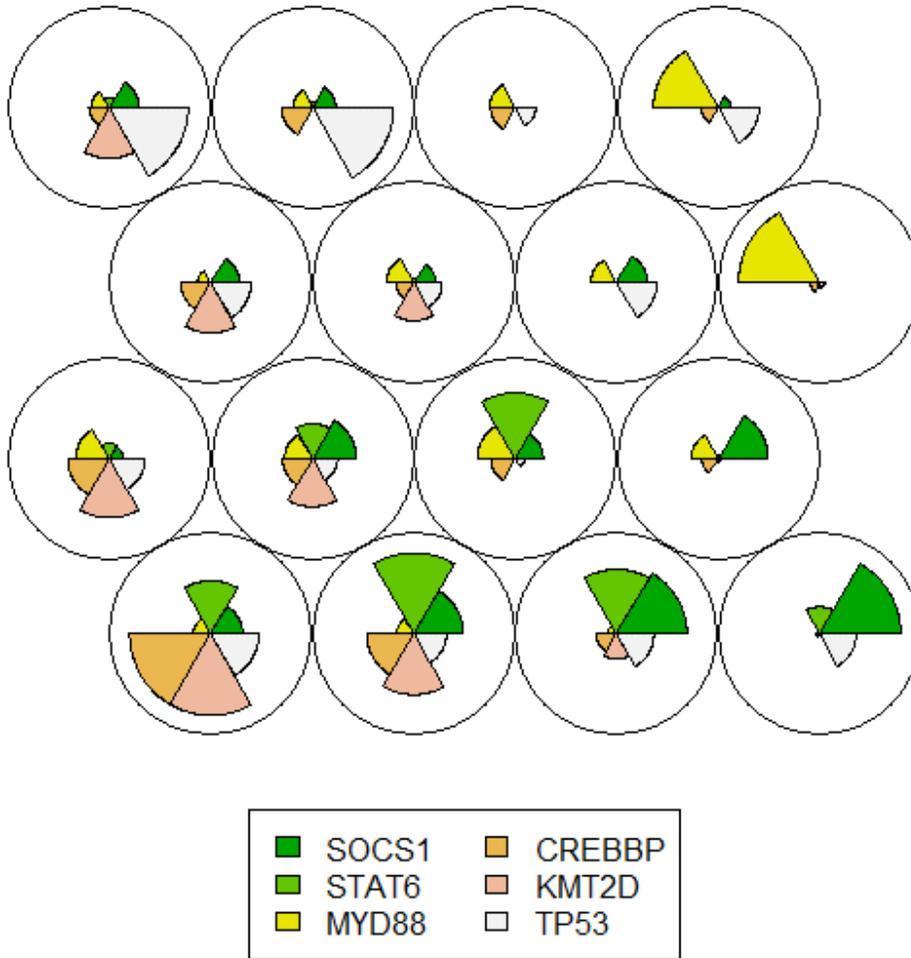


Figure 2

Cluster analysis of SOM self-organizing network for 128 genes. Each circle represents a cluster, and the size of the sector represents the strength of the association of several major genes with genes within the cluster.

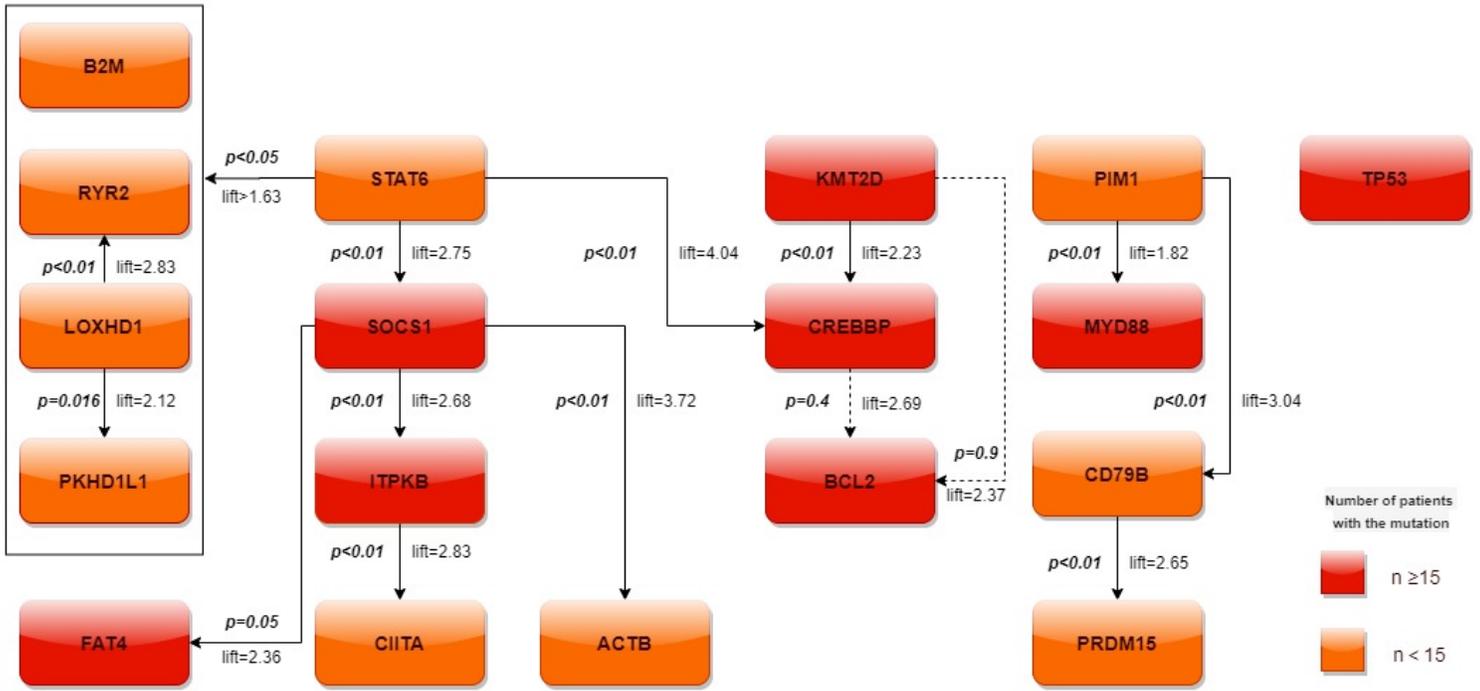


Figure 3

Hot-spot gene relationship diagram based on association analysis and expression analysis. The depth of the color indicates the frequency of gene mutations in the 92-patient exome sequencing dataset. “n” in the label refers to the number of patients carrying a certain gene.

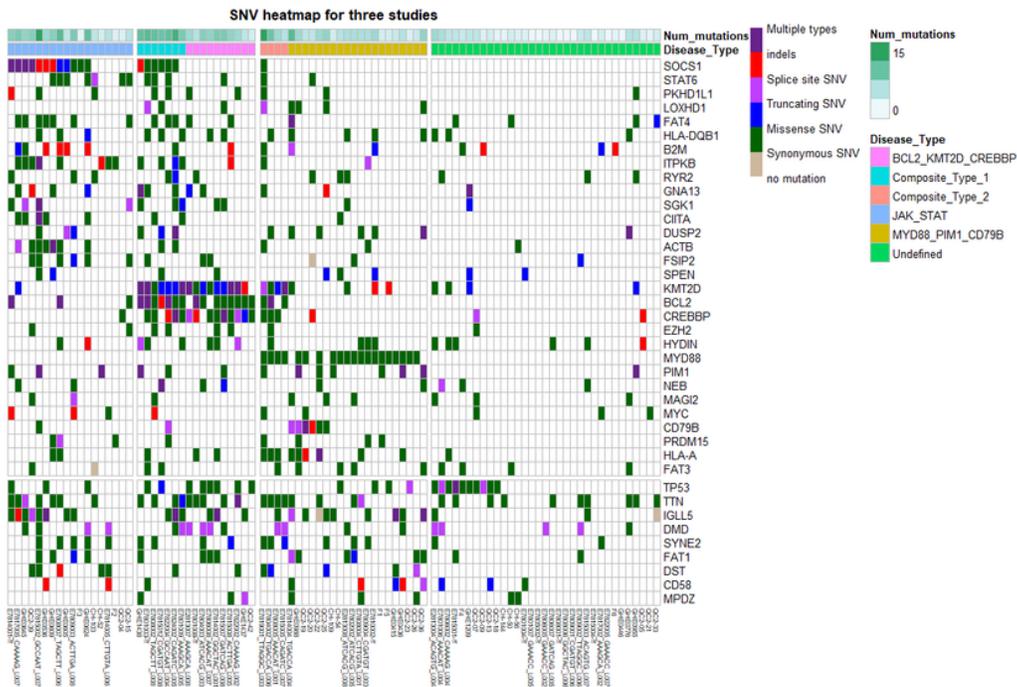


Figure 4

Clustering heat map based on association analysis results. Num_mutations refers to the number of gene mutation types carried by a patient. Complex type 1 shown in the figure has characteristics of KMT2D-CREBBP-BCL2 and JAK-STAT6. Complex type 2 has characteristics of KMT2D-CREBBP-BCL2 and PIM1-MYD88-CD79B.

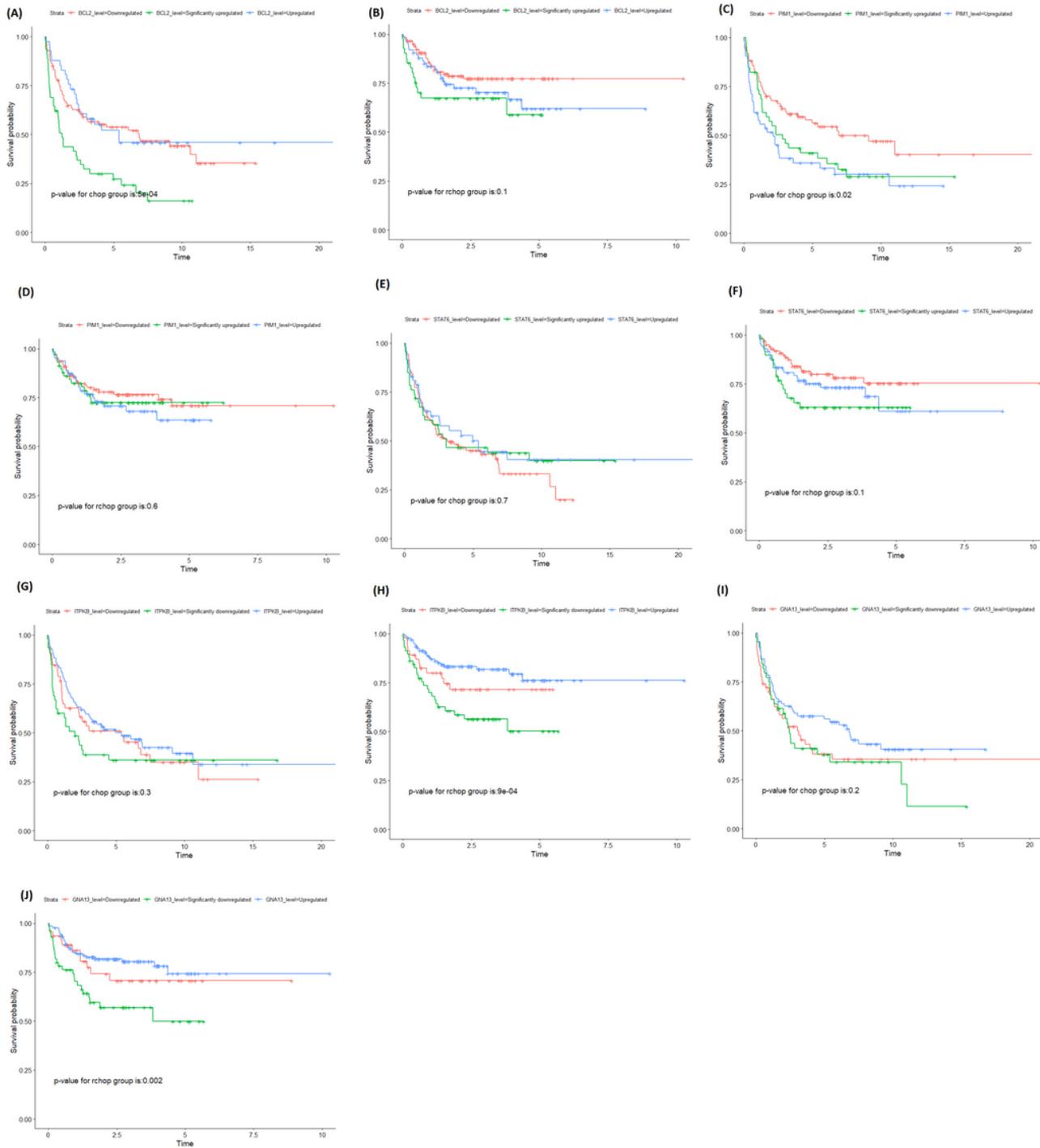


Figure 5

Survival analysis grouped according to several gene expression levels. a. Survival curve in CHOP group according to BCL2 expression level; b. Survival curve in R-CHOP group according to BCL2 expression level; c.

Survival curve in CHOP group according to PIM1 expression level; d. Survival curve in R-CHOP group according to PIM1 expression level; e. Survival curve in CHOP group according to STAT6 expression level; f. Survival curve in R-CHOP group according to STAT6 expression level; g. Survival curve in CHOP group according to ITPKB expression level; h. Survival curve in R-CHOP group according to ITPKB expression level; i. Survival curve in CHOP group according to GNA13 expression level; j. Survival curve in R-CHOP group according to GNA13 expression level.

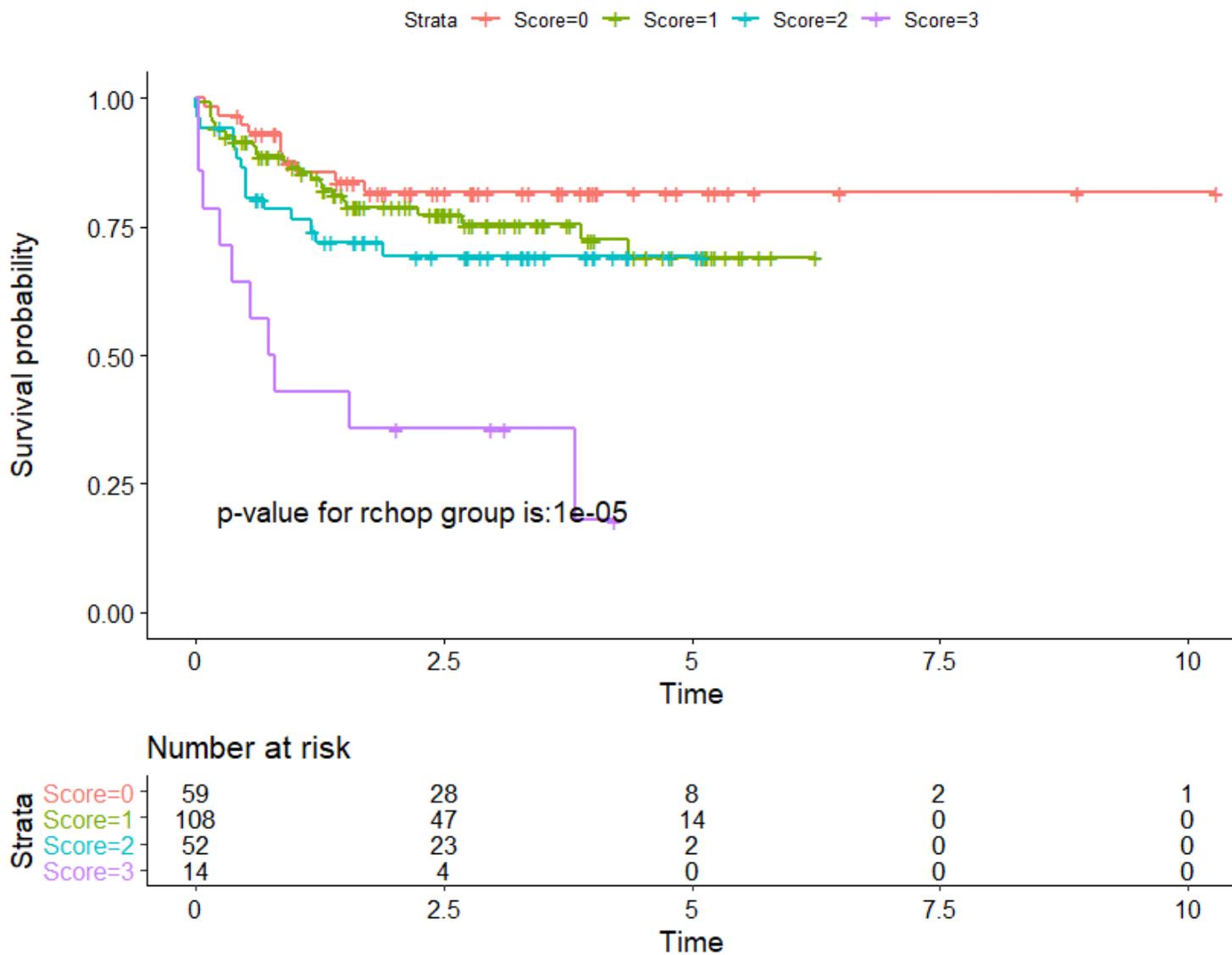


Figure 6

Survival analysis of the cumulative points. Analysis was performed based on the expression of three combinations of genes,

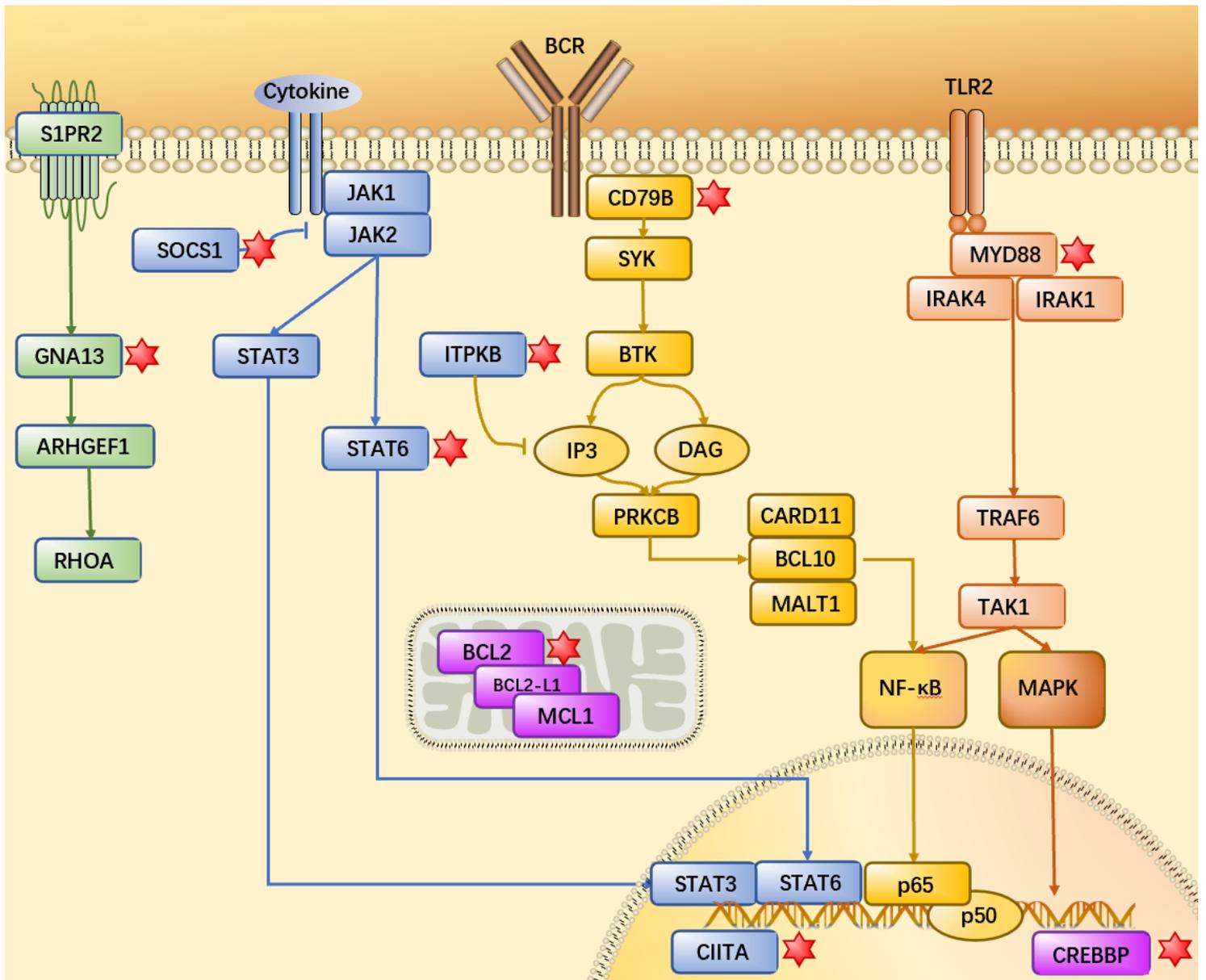


Figure 7

Schematic diagram of the main mutations of RRDLBCL and the pathways and mechanisms involved. Red stars mark some of the highest-frequency mutations in RRDLBCL.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [FigureS1.png](#)
- [figS2.png](#)
- [FigureS3.png](#)
- [FigureS4.png](#)
- [FigureS5.png](#)

- [SupplementaryS1202009.doc](#)
- [SupplenmentaryS2202008.doc](#)
- [TableS1.xlsx](#)