B cell receptor (BCR) diversity and differential CDR3s usage as potential immune indictors of diffuse large B cell lymphoma (DLBCL) Running title: Vital roles of BCR traits in DLBCL

Wenhua Jiang
The Second Hospital of Tianjin Medical University
Shiyong Zhou
Sino-US Center of Lymphoma and Leukemia, Tianjin Medical University Cancer Institute and Hospital, National Clinical Research Center for Cancer, Key Laboratory of Cancer Prevention and Therapy, Tianji
Jian Li
Tianjin University Hospital
Guoqing Zhu
State Key Laboratory of Experimental Hematology National Clinical Research Center for Blood Diseases Institute of Hematology & Blood Diseases Hospital, Chinese Academy of Medical Sciences & Peking Uni
Mingyou Gao
National Clinical Research Center for Cancer, Tianjin Medical University Cancer Institute and Hospital
Kuo Zhao
National Clinical Research Center for Cancer, Tianjin Medical University Cancer Institute and Hospital
Limeng Zhang
National Clinical Research Center for Cancer, Tianjin Medical University Cancer Institute and Hospital
Xiaojing Xie
National Clinical Research Center for Cancer, Tianjin Medical University Cancer Institute and Hospital
Ning Zhao
National Clinical Research Center for Cancer, Tianjin Medical University Cancer Institute and Hospital
Caijuan Tian
Tianjin Marvel Medical Laboratory, Tianjin Marvelbio Technology Co., Ltd.
Zhenzhen Zhang
Tianjin Yunquan Intelligent Technology Co., Ltd
Yanpeng Zhao
Tianjin Yunquan Intelligent Technology Co., Ltd
Yixian Guo
Tianjin Yunquan Intelligent Technology Co., Ltd
Yunlong Cui

National Clinical Research Center for Cancer, Key Laboratory of Cancer Prevention and Therapy of Tianjin, Tianjin's Clinical Research Center for Cancer, Tianjin Medical University Cancer Institute and

Pengfei Liu (✉ liupengfeitj@163.com)

Tianjin Academy of Traditional Chinese Medicine Affiliated Hospital

---

Research Article

Keywords: Diffuse large B cell lymphoma, B cell receptor repertoire, whole-exome sequencing, immune indicator, CDR3

Posted Date: June 7th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1668735/v1

Abstract

The aim of the present study was to characterize DLBCL of BCR distributions, and to screen specific BCR features as potential indicators related to DLBCL prognosis. A total of 13 patients with DLBCL and 5 healthy individuals were enrolled in the present study. In total, two high throughput sequencing methods, BCR repertoire and WES, were used to reveal BCR distributions and gene mutations, respectively. A series of comprehensive analyses were conducted to identify potential complementarity determining region 3 (CDR3) indicators of DLBCL and to analyze their association with clinical characteristics. The relatively higher BCR diversity in patients with DLBCL was associated with shorter progression-free survival (PFS). In addition, the top 10 differential CDR3s were screened as potential DLBCL indicators, and among these, four CDR3s [immunoglobulin heavy chain IGHV1-8/IGHJ6, IGHV3-74/IGHJ6, IGHV4-39/IGHJ3 and IGHV4-34/IGHJ4] exhibited a close relationship with PFS in patients with DLBCL. In further analysis, IGHV4-34/IGHJ4 and IGHV4-39/IGHJ3 were identified to be directly related to tumor metastasis, and IGHV4-39/IGHJ4 was related to hepatitis B virus (HBV) infection. WES data were used to analyze the gene mutations in patients with DLBCL, indicating that two metastasis pathways ('focal adhesion' and 'ECM-receptor interaction') were involved in characteristic changes of BCR. In conclusion, BCR diversity and usage of 10 specific CDR3s were identified as potential DLBCL indicators in the present study. Clinically, inflammation and HBV infection were the two main factors associated with changes of BCR characteristics. In terms of the molecular mechanism, BCR distributions were closely associated with metastasis-related pathways.

Introduction

Diffuse large B cell lymphoma (DLBCL) is the most common subtype of B cell non-Hodgkin lymphoma with high invasiveness and diffuse growth (1). Although DLBCL is a highly curable disease, there remains approximately one-third of patients with poor prognosis (2). In trials in the last two decades, the regimen of rituximab, cyclophosphamide, doxorubicin, vincristine and prednisone has been commonly used as the standard-of-care therapy for patients with DLBCL (3–5). However, due to a lack of specific indicators, it is difficult to detect DLBCL early, as well as to accurately evaluate the prognosis. Although the mutations of *TP53* and *MYC* have been demonstrated to be effective indicators for DLBCL prognosis, the prognostic value of these remains controversial with conflicting research results (6–9).

Since DLBCL is a B cell originated lymphoma, changes of B cell composition and abundance in the microenvironment are a direct reflection of DLBCL progress. Resulting from the DNA recombination of different B cell receptor (BCR) variable (V), diversity (D) and joining (J) genes, complementarity determining region 3 (CDR3) is the most various fragment in BCR, so that each CDR3 with a unique sequence represents one type of B cells (10). With the development of next-generation sequencing technology, systemic exploration of all CDR3s in the tumor microenvironment has become possible. BCR repertoire is a high-throughput sequencing technology specifically for CDR3s utilizing specific primers (11), which can be used to determine most characteristics of B cells in the DLBCL microenvironment.

Therefore, it may be possible to screen indicators related to DLBCL progress using BCR repertoire in order to provide novel guidance for prognosis prediction of DLBCL.

Whole-exome sequencing (WES) is an appropriate strategy for detecting mutations of coding sequences in tumor samples (12), and is commonly used in clinical examination with low cost and high efficiency (13). Similar to *KRAS* in pancreatic cancer (14), oncogenes with high mutation frequency can be the main cause of cancer. Therefore, WES is an effective tool to identify the possible molecular mechanism of DLBCL.

In the present study, BCR repertoire, WES and clinical characteristics were combined to conduct a series of comprehensive analyses in DLBCL, in order to identify specific BCR characteristics related to DLBCL prognosis, and to explore the molecular mechanism.

Materials And Methods

Sample collection. All samples in this study are from the second hospital of Tianjin Medical University (Tianjin, China) from 2018 to 2019. Inclusion criteria: a. age 18–59 years, unlimited gender; b. Clinical diagnosis was DLBCL; c. No history of important organ diseases such as heart, lung, liver and kidney; d. All methods were carried out in accordance with relevant guidelines and regulations.. Exclusion criteria: a. Sjogren's syndrome, ankylosing spondylitis, arthritis, herpetic dermatitis, alcoholic cirrhosis, chronic hepatitis, etc; b. Suffering from heart, lung, liver and other important organ diseases; c. Patients cannot understand the purpose of the study or disagree with the requirements of the study. Formalin fixed paraffin embedded(FFPE) samples of cancer tissues and adjacent tissues were collected from 13 patients with DLBCL. Blood samples were collected from 5 healthy individuals.. All subjects provided written informed consent.

DNA isolation and library construction. Total genomic DNA was extracted from all samples using the QIAamp Circulating Nucleic Acid kit (QIAGEN Ltd) according to the manufacturer's instructions.

BCR repertoire library preparation for the samples from 5 healthy individuals, 12 cancer tissues (one of the 13 cancer tissues was not used due to the bad quality of the library) and 13 paracancerous tissues from patients with DLBCL was performed according to the manufacturer's protocol. The specific primers of the V region and J region(Table) were designed to amplify the BCR CDR3 using the multiplex PCR kit (Qiagen). The reaction conditions were as follows: 1 cycle at 95˚C for 15 min, followed by 35 cycles of denaturation at 94˚C for 30 sec, annealing at 60˚C for 90 sec and extension at 72˚C for 30 sec. Using agarose gel electrophoresis, 100–200 bp fragments were collected and purified with the Qiaquick gel purification kit (Qiagen). The ends of purified DNA were repaired with the KAPA end repair kit (Kapa Biosystems; Roche Diagnostics), and subsequently purified with the QIAquick PCR purification kit (Qiagen). Next, the KAPA A-tailing kit and KAPA adapter ligation kit (Kapa Biosystems; Roche Diagnostics) were used to add A tails and adapters, respectively, to the aforementioned purified products. Finally, via PCR amplification of the aforementioned products using universal primers and the

subsequent purification using Agencourt AMPure XP beads (Agencourt Biosciences Co.), the library for the BCR repertoire was constructed.

WES library preparation for the 12 cancer tissue samples and 12 paracancerous tissue samples from patients with DLBCL (one of the 13 enrolled patients did not accept WES) was performed according to the manufacturer's protocol. The extracted DNA was randomly sheared into 150–220 bp fragments using the Covaris M220 instrument (Covaris, Inc.). Subsequently, these DNA fragments were used to construct the library with the KAPA hyper DNA library prep kit (Kapa Biosystems; Roche Diagnostics). Following size selection using Agencourt AMPure XP beads (Agencourt Biosciences Co.), the aforementioned DNA product was used to conduct the PCR amplification. Subsequently, exon trapping was performed using SeqCap EZ Prime Choice Probes (Roche Diagnostics), which captured a total of 1.1 Mb from 1,000 known cancer-related genes. Finally, the WES library was built.

High throughput sequencing. After quantification, the BCR repertoire library and WES library were subjected to high throughput sequencing using the Illumina HiSeq 2500 platform and the Illumina HiSeq Xten sequencer (Illumina, Inc.), respectively.

Data and statistical analysis. The Kruskal Wallis test (Package agricolae; version 1.3-3) (https://CRAN.R-project.org/package=agricolae) was used to compare three groups, post hoc test is using the Dunn's test (Package FSA; version 0.9.1)(https://github.com/droglenc/FSA), ggplot2 (version 3.2.1) is used for drawing(https://ggplot2.tidyverse.org).Based on WES data and clinical characteristics of each patient with DLBCL, a landscape map was drawn using GenVisR (version 1.16.1) (https://bioconductor.org/packages/GenVisR) to display the somatic mutations. Survival analysis the patients were divided into two groups according to the median. The log rank test in survival (version 2.44) (https://CRANR-project.org/package=survival>) was used and the survival curve was drawn.The Spearman test was used to conduct correlation analysis using Hmisc (version 4.3) (https://CRAN.R-project.org/package=Hmisc) and corrplot (version 0.84) (https://github.com/taiyun/corrplot) was used to draw the plot. The comparison of CDR3s and clinical characteristics was conducted using a unpaired t-test. $P < 0.05$ (two-sided) was considered to indicate a statistically significant difference. GO(Gene Ontology) and KEGG(Kyoto Encyclopedia of Genes and Genomes) enrichment were analyzed by string (https://string-db.org/) And David (https://david.ncifcrf.gov/).

Results

Differences in BCR characteristics between healthy individuals and patients with DLBCL. To systematically characterize the BCR distribution in the DLBCL microenvironment, BCR repertoire was used to display BCR features in the present study. According to the different sample sources, the samples were divided into three groups: Group N, blood samples from the normal; group T, cancer tissues from patients with DLBCL; and group C, paracancerous tissues from patients with DLBCL.

V and J distributions of each sample were first displayed (Fig. 1A and B). To identify the representative V and J types in groups N, T and C, the average proportions of each V and J type in these three groups were

compared (Fig. 1C and D). For V segment usages, the usage of 18 out of 48 V segments markedly differed among groups N, T and C ($P < 0.05$; Fig. 1F), there were 14 difference V segments between group C and group N, and 16 difference V segments between group N and group T. For J segment usages, no significant difference was found in these three groups, expect for immunoglobulin heavy chain IGHJ3 ($P < 0.05$; Fig. 1D, F and G).

The average proportion of each V-J pairing usage in groups N, T and C was further calculated (Fig. 2A) to identify the top 10 most different V-J type CDR3s as the candidate immune indicators in DLBCL based on the P-value (Fig. 2B and C). Among these 10 CDR3s, eight exhibited significant differences between groups T and N, seven exhibited significant differences between groups C and N, and there were two CDR3 differences between T and C (Fig. 2B and C).

The distribution of CDR3s with different amino acid (aa) lengths is another important performance of BCR characteristics. Figure 3A shows significant distribution differences between groups T and N mainly at lengths of 22–32 aa, as well as between groups C and N at lengths of 22,28,30,31 and 34 aa. However, no significant differences were observed in the aa length distribution of CDR3 between groups T and C, except at lengths of 7 aa. Furthermore, in the region of 22–30 aa length, the N group had a markedly higher proportion of CDR3s than groups T and C at the same aa length (Fig. 3B). These results revealed a similar length distribution of CDR3s in the T and C groups, which differed from group N.

In summary, compared with patients with DLBCL, healthy individuals exhibited distinct differences in the aa length distribution and usage of specific CDR3s of BCRs. However, there were no significant differences between cancer and paracancerous tissues in patients with DLBCL. These results suggested that the proportion of CDR3s between 22 and 30 aa, and the percentage of the aforementioned 10 specific CDR3s could potentially be used as immune indicators for the diagnosis of DLBCL.

Prognostic significance of BCRs in DLBCL. Considering the important role of BCRs in DLBCL, the present study further explored the relationship between CDR3s and prognosis of patients with DLBCL. First, the association between BCR diversity and prognosis was analyzed in patients with DLBCL. Shannon index was calculated according to BCR sequence data. Taking the median Shannon index as the cut-off value, 13 patients with DLBCL were divided into two groups The high diversity group exhibited significantly shorter progression-free survival (PFS) than the low diversity group ($P = 0.018$; Fig. 4A). These results indicated that low diversity of BCRs in patients with DLBCL predicted longer PFS.

The present study also comprehensively analyzed the association between DLBCL prognosis and 10 specific CDR3 usages. The PFS was significantly prolonged in patients with DLBCL with the lower proportion of IGHV1-8/IGHJ6 ($P = 0.018$), IGHV3-74/IGHJ6 ($P = 0.018$), IGHV4-39/IGHJ3 ($P = 0.018$) and IGH V4-34/IGHJ4 ($P = 0.018$)(Fig. 4B-E). Additionally, in patients with DLBCL with hepatitis B virus (HBV) infection, the proportion of IGHV4-39/IGHJ4 was significantly lower ($P = 0.016$; Fig. 4F). In patients with DLBCL with metastases, the proportions of both IGHV4-34/IGHJ4 and IGHV4-39/IGHJ3 were relatively high ($P = 0.014$ and 0.049, respectively; Fig. 4G and H). These findings indicated four PFS-related CDR3s

(IGHV1-8/IGHJ6, IGHV3-74/IGHJ6, IGHV4-39/IGHJ3 and IGHV4-34/IGHJ4), one HBV infection-related CDR3 (IGHV4-39/IGHJ4) and two metastasis-related CDR3s (IGHV4-34/IGHJ4 and IGHV4-39/IGHJ3).

Taken together, BCR repertoires could be used to predict prognosis in patients with DLBCL. The Shannon index (overall BCR diversity) and the proportions of four specific CDR3s (IGHV1-8/IGHJ6, IGHV3-74/IGHJ6, IGHV4-39/IGHJ3 and IGHV4-34/IGHJ4) were identified as prognosis-related indicators, which were negatively associated with PFS. In addition, three CDR3s were demonstrated to be associated with HBV infection and metastasis in patients with DLBCL: the IGHV4-39/IGHJ4 usage was relatively low in HBV-infected patients with DLBCL. Wihle in patients with DLBCL with metastasis, the proportions of IGHV4-34/IGHJ4 and IGHV4-39/IGHJ3 were relatively high. Therefore, these BCR repertoire sequences could be used as immune indicators related to the prognosis of patients with DLBCL.

Association between BCR features and clinical characteristics in DLBCL. Since clinical characteristics serve important roles in disease diagnosis and prognosis, an overall correlation landscape between BCR features (Shannon index and usage of the 10 aforementioned CDR3s) and 60 clinical characteristics was established to analyze their mutual relationship in DLBCL (Fig. 5).

A total of 22 clinical biochemical indexes were identified to be significantly associated with the BCR diversity or usage of specific CDR3s. Among them, basophil rate (BASO-R), reticulocyte hemoglobin content, mean corpuscular hemoglobin and white blood cell count (WHC) were all positively associated with the Shannon index of BCR diversity. Interestingly, each of these four clinical characteristics was also found to be positively associated with at least four specific CDR3s. In addition, hepatitis B surface antibody (HBsAb) and mean platelet volume were two clinical characteristics that were negatively associated with the proportions of some specific CDR3s. These findings provided a novel clue regarding the pathogenesis of DLBCL, suggesting that BCR changes in patients with DLBCL could affect certain clinical characteristics, and may be related to the prognosis of DLBCL. Furthermore, these results also provided an easy method to preliminarily deduce the BCR features of patients with DLBCL through clinical routine testing.

Somatic mutation landscape of 13 patients with DLBCL. Since gene mutation is a key factor in cancer progress, the WES data were used to present a somatic mutation landscape in 13 patients with DLBCL (Fig. 6).

Among all mutant genes from patients with DLBCL, ryanodine receptor 1 (*RYR1*) was identified to have the highest mutation frequency, with mutations of *RYR1* being observed in 12 out of 13 patients with DLBCL. As a calcium release channel protein, *RYR1* has been reported to be involved in the enrichment of breast cancer stem cells by regulating Ca^{2+} release (15), and it may affect the progression of DLBCL in a similar way. In the present study, among the 13 patients with DLBCL, T12, who was the only patient with DLBCL without hepatitis A virus (HAV) or HBV infection, showed the lowest number of mutations (4 out of the 20 mutated genes analyzed). It was hypothesized that the overall mutation frequency in patients with

DLBCL was related to infection of HAV or HBV. Additionally, there were no obvious connections between mutation frequency and other factors, such as cancer stage, age and sex.

Enrichment analysis of CDR3 usage-related mutation genes. To explore the molecular mechanism of DLBCL progression, the BCR repertoire was combined with WES data of patients with DLBCL to screen related genes and pathways. First, all differentially mutated genes between groups of samples with high and low proportions of each CDR3 ($P < 0.05$) were counted. A total of 501 differential mutant genes were observed. According to the frequency of occurrence of each enriched gene for the 10 CDR3s, a number of the high frequency genes were related to cancer (Fig. 7).

The annotation of each gene with > 60% occurrence is listed in Table , showing that 10 out of 17 genes were related to cancer. Enrichment analysis revealed 57 Gene Ontology (GO) terms and 14 Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were significantly related to 10 CDR3 usages ($P < 0.05$; Fig. 8A and B). Among them, the 'acute myeloid leukemia pathway' was directly related to DLBCL; 'ECM-receptor interaction' and 'focal adhesion' were two pathways closely associated with tumor metastasis; and the 'PI3K-Akt signaling pathway' and 'pathways in cancer' served important roles in cancer. Second, to investigate the molecular mechanism of BCR diversity in DLBCL, 123 genes with different mutation status between patients with high and low Shannon indexes were screened. These genes were significantly enriched in only three pathways ($P < 0.05$), 'regulation of actin cytoskeleton', 'focal adhesion' and 'estrogen signaling pathway', suggesting that the changes of BCR diversity in DLBCL involved these three pathways (Fig. 8C and D). Furthermore, to explore the mechanism of each of the 10 selected CDR3s, GO and KEGG enrichment analysis was performed based on differential mutant genes in different CDR3 usage groups of patients with DLBCL. The results demonstrated that these 10 CDR3s were enriched in 59 GO terms, and 5 of them were enriched in 14 pathways (Fig. 8E and F).

IGHV3-74/IGHJ6 was the V-J type involved in most gene mutations, and enriched in 40 GO terms and 8 pathways. Among them, 4 specific pathways ('small cell lung cancer', 'focal adhesion', 'ECM-receptor interaction' and 'AMPK signaling pathway') were closely associated with tumorigenesis and progression. Collectively, these results indicated that some specific BCR characteristics in DLBCL were closely related to gene mutations and regulatory pathways.

Discussion

DLBCL is a highly heterogeneous lymphoid neoplasm with B cell lesion, which is a serious threat to human health. In the present study, BCR repertoire and WES were used to evaluate the BCR features and gene mutations of each patient with DLBCL, conducting a series of comprehensive analyses of clinical characteristics. The results suggested that the BCR diversity and 10 specific CDR3 usages were closely related to the prognosis of patients with DLBCL.

Previous studies have demonstrated that diverse BCRs could improve the ability to recognize a mass of foreign antigens and protect against disease in humans (16). In general, the diversity of immune cells is extremely high in healthy individuals but markedly decreased in patients with cancer (17, 18). In the

present study, the same reduction of BCR diversity was also observed in patients with DLBCL. However, the patients with DLBCL with higher BCR diversity exhibited markedly shorter PFS.A possible reason for this is that DLBCL is derived from the carcinogenesis of B cells, and the diversity of BCRs mainly reflects the abundance of cancerous B cells with uncontrolled proliferation. Therefore, high BCR diversity indicated no effective reduction of cancerous B cells in the patients with DLBCL after treatment.

Early diagnosis and accurate prognosis evaluation are two key strategies to improve the survival of patients with cancer. Cancer biomarkers provide an effective way for early cancer detection and prognosis evaluation. A previous study indicated that the immunosignature could be a potential approach in the accurate diagnosis of cancer (19). In the present study, 10 V-J types of CDR3s were screened as the candidate immune biomarkers for DLBCL diagnosis and prognosis (Fig. 2C). BCR characteristics can directly reflect the situation of B-cell neoplasms in patients with DLBCL, and BCR signaling serves an important role in DLBCL (20). Therefore, specific BCR CDR3 proportions were proposed as DLBCL prognosis indicators, different from other tumor biomarkers. The results demonstrated that 10 selected CDR3s showed significantly different usages among the N, T and C groups, and four of them were negatively associated with PFS in the present study. These results provided a novel idea for the diagnosis and prognosis of DLBCL through the BCR repertoire.

It should also be noted that there were almost all differential CDR3s came from N vs. T and N vs. C, with the exception of IGHV3-74/IGHJ6 and IGHV3-11/IGHJ5. These two types of CDR3 usages displayed significant differences between cancer tissues and paracancerous tissues, indicating that IGHV3-74/IGHJ6 and IGHV3-11/IGHJ5 usage could be used as two specific indicators to distinguish cancer tissues from normal tissues in DLBCL. In the present study, IGHV3-74/IGHJ6 usage was also demonstrated to be closely related to PFS. Therefore, the risk of DLBCL recurrence and metastasis can be evaluated based on the two indicators.

Since clinical characteristics are the direct reflection of the condition of the patient, a comprehensive correlation analysis of 60 clinical indicators with Shannon index of BCR diversity and 10 differential CDR3s was performed. Among them, 22 clinical indicators were significantly related to the Shannon index and differential CDR3 usage. BASO-R and WHC were two typical inflammatory indicators, which were positively correlated with both the Shannon index and usage of some specific CDR3s. This result suggested that BCR characteristics in the DLBCL microenvironment are closely related to inflammation, which is often associated with the development and progression of cancer (21). In addition, HBV infection has been reported as a key factor for poor prognosis of DLBCL (22, 23). Advances in HCC genome studies have revealed the importance of HBV integration mutations. Integration of HBV DNA may promote HCC by promoting the continuous expression of isoHBV proteins. Variation or shortening of HBV surface proteins (mediated by integrated HBV DNA) is associated with estrogen receptor (ER) stress responses and can increase the risk of HCC. At the same time, these forms of mutations predispose the corresponding liver cells to hyperplasia and promote their expansion (24–26). In the present study, the correlation analysis indicated a negative correlation between five CDR3s and HBsAb, as well as a positive correlation between three CDR3s and hepatitis Be antigen. Furthermore, IGHV4-39/IGHJ4 usage was

markedly reduced in HBV-infected patients with DLBCL. Therefore, HBV infection may be involved in the regulation of BCRs in DLBCL. Thus, based on this comprehensive correlation analysis, inflammation and HBV infection were identified to be associated with BCR characteristics in DLBCL.

To explore the molecular mechanism of BCR changes in DLBCL, differential mutant genes were used to conduct GO and KEGG enrichment analysis. The ‘focal adhesion-related pathway’ was found to be involved in BCR diversity changes in patients with DLBCL. Additionally, two regulatory pathways associated with tumor metastasis, ‘focal adhesion’ and ‘ECM-receptor interaction’, were enriched in the 10 most differential CDR3s. Specifically, the usages of IGHV4-34/IGHJ4 and IGHV4-39/IGHJ3 were increased in patients with DLBCL with metastasis. Therefore, BCR changes in DLBCL exhibited a close association with metastasis, which is consistent with a previous study reporting that tumor-educated B cells could promote breast cancer lymph node metastasis (27). Furthermore, based on the enrichment analysis of selected CDR3s, it was revealed that IGHV3-74/IGHJ6 was the specific CDR3 involved in most tumor-related pathways, including ‘small cell lung cancer’, ‘focal adhesion’, ‘ECM-receptor interaction’ and ‘AMPK signaling pathway’. Overall, changes of BCR characteristics were mainly related to the tumor metastasis process in DLBCL.

There were still a number of limitations of the present study. First, the number of patients with DLBCL was too small. Only 13 patients with DLBCL and 5 healthy individuals were enrolled in the present study, and thus, each data from these samples had too much influence on the final results. T7 was a special sample with extremely high usages of IGHV3-48/IGHJ6 (47.08%) and IGHV3-7/IGHJ6 (46.75%), showing far different composition from other cancer tissue samples in the present study. Especially in the CDR3 aa length distribution, the high usage of IGHV3-48/IGHJ6 and IGHV3-7/IGHJ6 in T7 made the average usage of 21 aa long CDR3s in the T group extremely high. Therefore, further research needs more samples to identify whether T7 is a particular or common case. Second, only 288 CDR3s (48 V types x 6 J types) were used to describe the entire BCR characteristics in patients with DLBCL. In the present study, the D region of CDR3s was not included due to the complex calculation and limited sequencing depth, which could not display various distributions of CDR3. In addition to the known CDR3 types, the novel CDR3s of tumor B cells also need to be considered, which will provide a direct connection to DLBCL. Third, molecular and cellular experiments should be added to verify the results. Therefore, although a number of meaningful results were obtained in the present study, further research is still required.

In conclusion, the BCR repertoire was used to display BCR distribution in DLBCL and combined with clinical characteristics and WES data to comprehensively analyze the prognostic relationship and related pathways. The present findings indicated that some specific CDR3 usages were also identified as potential immune markers, exhibiting a close association with PFS, metastasis, HBV infection and specific clinical characteristics. Based on the enrichment analysis, BCR distribution was closely related to tumor metastasis pathways. These results provided novel insights into potential DLBCL biomarkers using BCR repertoire analysis.

Declarations

Acknowledgements

Not applicable.

Funding

The present study was supported by the National Natural Science Foundation of China**(grant number 81702098)**.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request, except private information of participants.

Sequencing data https://dataview.ncbi.nlm.nih.gov/object/PRJNA781779?reviewer=l63mfmnth2b71ada274etmncfq

Authors' contributions

Yunlong Cui and Pengfei Liu conceived and designed the study. Wenhua Jiang, Shiyong Zhou and Jian Li provided study materials or patients. Guoqing Zhu, Mingyou Gao, Kuo Zhao, Limeng Zhang, Xiaojing Xie and Ning Zhao collected and assembled data. Caijuan Tian, Zhenzhen Zhang, Yanpeng Zhao and Yixian Guo analyzed and interpreted the data. All authors wrote the manuscript. All authors have read and approved the final manuscript.

Ethics approval and consent to participate

Informed consent was obtained from all subjects and/or their legal guardian(s). The present study was approved by the Ethic Committee of Second Hospital of Tianjin Medical University (Tianjin, China). (Approval number: KY2017K025

Patient consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Alt FW, Yancopoulos GD, Blackwell TK, et al. Ordered rearrangement of immunoglobulin heavy chain variable region segments. EMBO J 3: 1209–1219, 1984.
2. Bartha Á, Győrffy B. Comprehensive Outline of Whole Exome Sequencing Data Analysis Tools Available in Clinical Oncology. Cancers (Basel) 11(11): 1725, 2019.

3. Chen Y, Xu Y, Zhao M, et al. High-throughput T cell receptor sequencing reveals distinct repertoire between tumor and adjacent non-tumor tissues in HBV-associated HCC. Oncoimmunology 5(10): e1219010, 2019.
4. Ennishi D, Mottok A, Ben-Neriah S, et al. Genetic profiling of MYC and BCL2 in diffuse large B-cell lymphoma determines cell-of-origin-specific clinical impact. Blood 129(20): 2760–2770, 2017.
5. Feugier P, Van Hoof A, Sebban C, et al. Long-term results of the R-CHOP study in the treatment of elderly patients with diffuse large B-cell lymphoma: a study by the Groupe d'Etude des Lymphomes de l'Adulte. J Clin Oncol 23(18): 4117–4126, 2005.
6. Gu Y, Liu Y, Fu L, et al. Tumor-educated B cells selectively promote breast cancer lymph node metastasis by HSPA4-targeting IgG. Nat Med 25(2): 312–322, 2019.
7. Jia Z, He J, Cen L, et al. P53 deletion is independently associated with increased age and decreased survival in a cohort of Chinese patients with diffuse large B-cell lymphoma. Leuk Lymphoma 53(11): 2182–2185, 2012.
8. Lu H, Chen I, Shimoda LA, et al. Chemotherapy-induced Ca2 + release stimulates breast cancer stem cell enrichment. Cell Rep 18(8): 1946–1957, 2017.
9. Landsburg DJ, Falkiewicz MK, Petrich AM, et al. Sole rearrangement but not amplification of MYC is associated with a poor prognosis in patients with diffuse large B cell lymphoma and B cell lymphoma unclassifiable. Br J Haematol 175(4): 631–640, 2016.
10. Li S, Young KH, Medeiros LJ. Diffuse large B-cell lymphoma. Pathology 50(1): 74–87, 2018.
11. Liu YY, Yang QF, Yang JS, et al. Characteristics and prognostic significance of profiling the peripheral blood T-cell receptor repertoire in patients with advanced lung cancer. Int J Cancer 145(5): 1423–1431, 2019.
12. Meléndez B, Van Campenhout C, Rorive S, et al. Methods of measurement for tumor mutational burden in tumor tissue. Transl Lung Cancer Res 7(6):661–667, 2018.
13. Ollila J, Vihinen M. B cells. Int J Biochem Cell Biol 37(3): 518–523, 2005.
14. Pfreundschuh M, Schubert J, Ziepert M, et al. Six versus eight cycles of bi-weekly CHOP-14 with or without rituximab in elderly patients with aggressive CD20 + B-cell lymphomas: a randomised controlled trial (RICOVER-60). Lancet Oncol 9(2): 105–116, 2008.
15. Pfreundschuh M, Trümper L, Osterborg A, et al. CHOP-like chemotherapy plus rituximab versus CHOP-like chemotherapy alone in young patients with good-prognosis diffuse large-B-cell lymphoma: a randomised controlled trial by the MabThera International Trial (MInT) Group. Lancet Oncol 7(5): 379–391, 2006.
16. Raphael BJ, Hruban RH, Aguirre AJ, et al. Integrated genomic characterization of pancreatic ductal adenocarcinoma. Cancer Cell 32(2): 185–203, 2017.
17. Rutherford SC, Leonard JP. DLBCL cell of origin: what role should it play in care today? Oncology (Williston Park) 32(9): 445–449, 2018.

18. Rong X, Wang H, Ma J, et al. Chronic hepatitis B virus infection is associated with a poorer prognosis in diffuse large B-cell lymphoma: a meta-analysis and systemic review. J Cancer 10(15): 3450–3458, 2019.
19. Ren W, Ye X, Su H, et al. Genetic landscape of hepatitis B virus-associated diffuse large B-cell lymphoma. Blood 131(24): 2670–2681, 2018.
20. Singh N, Baby D, Rajguru JP, et al. Inflammation and cancer. Ann Afr Med 18(3): 121–126, 2019.
21. Stafford P, Cichacz Z, Woodbury NW, et al. Immunosignature system for diagnosis of cancer. Proc Natl Acad Sci U S A 111(30): E3072-80, 2014.
22. Xu-Monette ZY, Wu L, Visco C, et al. Mutational profile and prognostic significance of TP53 in diffuse large B-cell lymphoma patients treated with R-CHOP: report from an International DLBCL Rituximab-CHOP Consortium Program Study. Blood 120(19): 3986–3996, 2012.
23. Ye B, Smerin D, Gao Q, et al. High-throughput sequencing of the immune repertoire in oncology: Applications for clinical diagnosis, monitoring, and immunotherapies. Cancer Lett 416: 42–56, 2018.
24. Wang, H.C.; Wu, H.C.; Chen, C.F.; Fausto, N.; Lei, H.Y.; Su, I.J. Different types of ground glass hepatocytes inchronic hepatitis B virus infection contain specific pre-S mutants that may induce endoplasmic reticulumstress. Am. J. Pathol. 2003, 163, 2441–2449.
25. Fan, Y.F.; Lu, C.C.; Chen, W.C.; Yao, W.J.; Wang, H.C.; Chang, T.T.; Lei, H.Y.; Shiau, A.L.; Su, I.J. Prevalenceand significance of hepatitis B virus (HBV) pre-s mutants in serum and liver at different replicative stages ofchronic HBV infection. Hepatology 2001, 33, 277–286.
26. Wang, L.H.; Huang, W.; Lai, M.D.; Su, I.J. Aberrant cyclin a expression and centrosome overduplication induced by hepatitis B virus pre-S2 mutants and its implication in hepatocarcinogenesis. Carcinogenesis 2012,33, 466–472.
27. Young RM, Shaffer AL 3rd, Phelan JD, et al. B-cell receptor signaling in diffuse large B-cell lymphoma. Semin Hematol 52(2): 77–85, 2015.

Tables

Table The primers for V-region family and J-region family of BCR

V or J gene/family	Sequence (5'-3')
IGH J	CTGAGGAGACGGTGACCRKKGT
IGHV1-18-NEW	AGAGTCACCATGACCACAGAC
IGHV1-2/1-46-NEW	AGAGTCACCAKKACCAGGGAC
IGHV1-24-NEW	AGAGTCACCATGACCGAGGAC
IGHV1-3/1-45-NEW	AGAGTCACCATTACYAGGGAC
IGHV1-69/1-f-NEW	AGAGTCACGATWACCRCGGAC
IGHV1-8-NEW	AGAGTCACCATGACCAGGAAC
IGHV2-70/26/5	ACCAGGCTCACCATYWCCAAGG
IGHV3-NEW	GGCCGATTCACCATCTCMAG
IGHV4-NEW	CGAGTCACCATRTCMGTAGAC
IGHV5-51-NEW	CAGCCGACAAGTCCATCAGC
IGHV6-1	AGTCGAATAACCATCAACCCAG
IGHV7-NEW	GACGGTTTGTCTTCTCCTTG

BCR: B cell antigen receptors.

K, Y, W, M R: degenerate bases

K: G or T

Y: C or T

W: A or T

M: A or C

R: A or G

Table The mutant genes significantly related to the top 10 CDR3s usages

Gene	Frequency	Function
CCDC141	7/10	Idiopathic hypogonadotropic hypogonadism
AVPR1B	6/10	Social behaviors, such as aggression, depression, anxiety
EPHB6	6/10	Drug resistance in breast cancer
FTSJ3	6/10	A potential regulator of breast cancer progression
GAA	6/10	Pompe disease
GLI1	6/10	An oncogene that mediates hedgehog signaling pathway
KIAA1522	6/10	Non-small cell lung cancer; hepatocellular cancer; esophageal cancer
MAGED2	6/10	antenatal Bartter's syndrome
MLC1	6/10	leukodystrophies
NYNRIN	6/10	Wilms tumor
ODF3L2	6/10	Unknown
PBXIP1	6/10	Tumor cell growth and migration in astrocytoma
PRICKLE1	6/10	Facilitation of cancer cell dissemination by interaction with mTORC2
SCN4A	6/10	Congenita paramyotonia; myotonia
TACC1	6/10	Interaction with FGFR1 to affect tumor cell
WDR6	6/10	Specific expression in breast cancer
ZNFX1	6/10	Antisense RNA (ZFAS1) functioned as an oncogenic lnRNA in multiple cancers

Figures

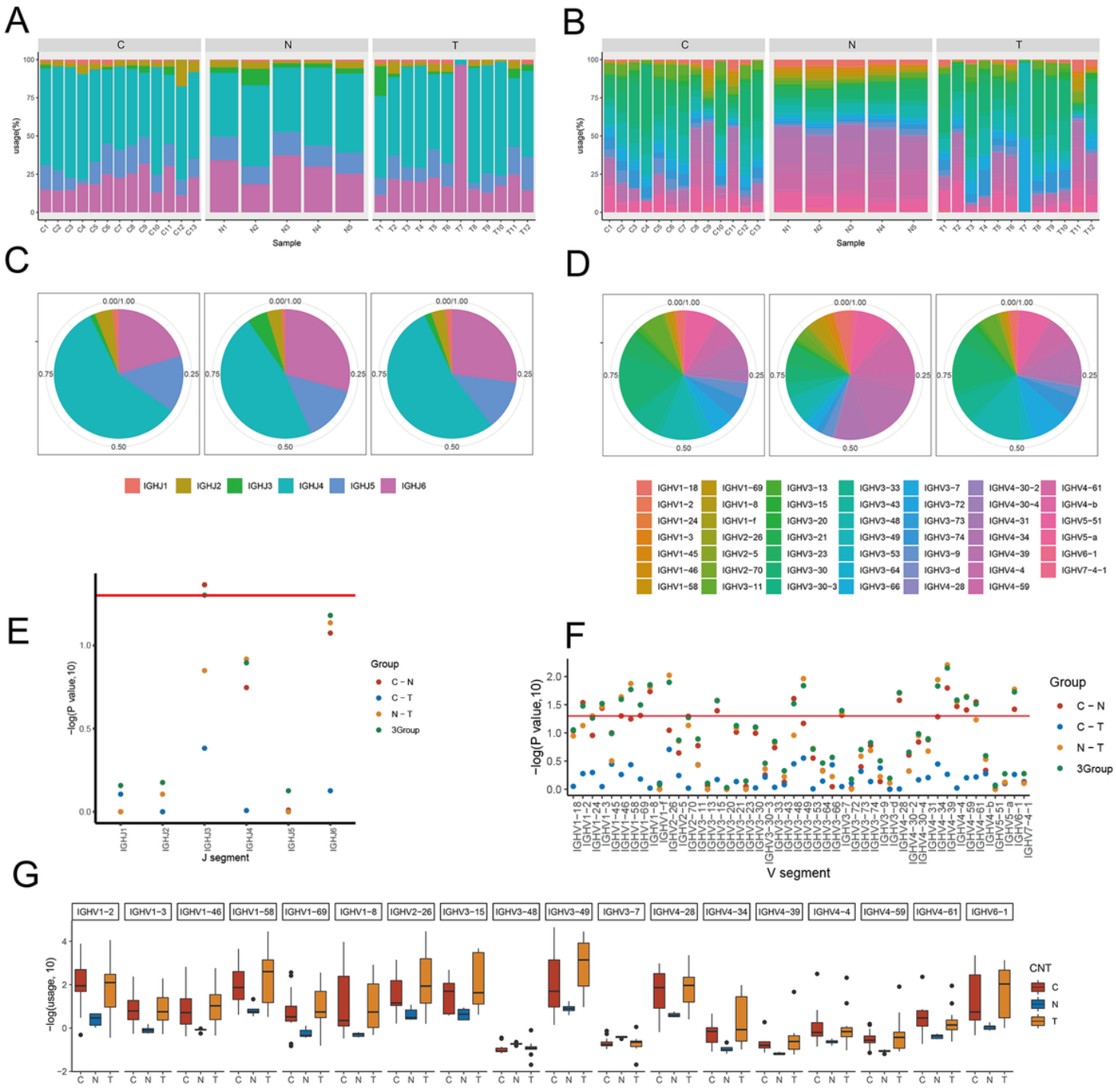

Figure 1

V and J usage

(A-B)V and J distributions for each sample,(C-D)Proportion of V and J in groups N, T and C, (E)Comparison of 6 J segment usage in N,T and C groups,The ordinate is - log (pvalue, 10), the P values of the three groups are the results of Kruskal Wallis test,the p value between the two groups was the result of Dunn's test, (F)Comparison of 48 V segment usage in N,T and C groups,The ordinate is - log (pvalue, 10), the P values of the three groups are the results of Kruskal Wallis test,the p value between the

two groups was the result of Dunn's test, (G)The box plot shows 18 V segment usage with significant differences among the three groups.

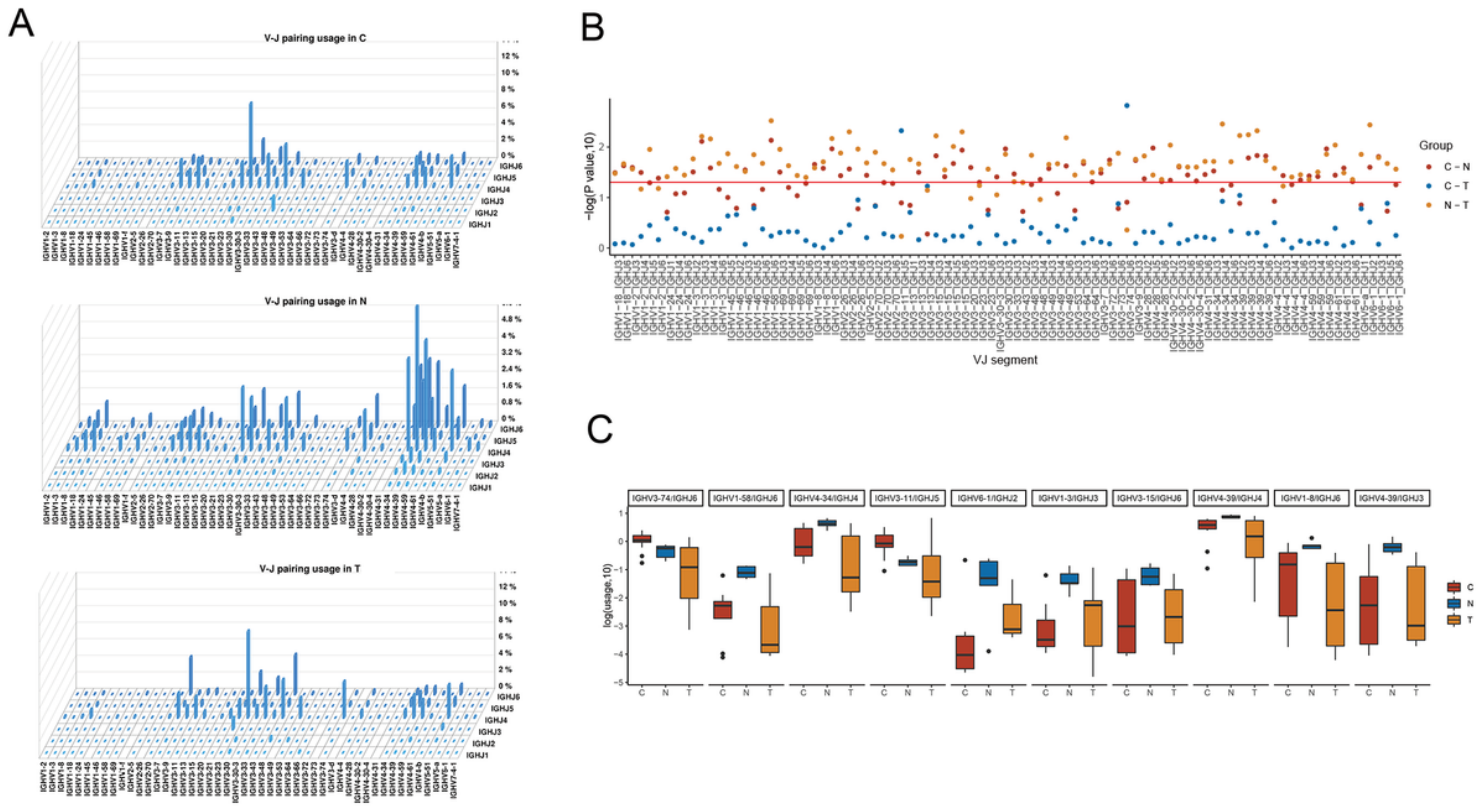

Figure 2

V/J pairing usage

(A)Proportion of V-J pairing in groups N, T and C,(B)Comparison of V-J segment usage in N,T and C groups,The ordinate is - log (pvalue, 10),the P values of the three groups are the results of Kruskal Wallis test,the p value between the two groups was the result of Dunn's test,(C)The box diagram shows top 10 V-J segment usage with significant differences among the three groups.

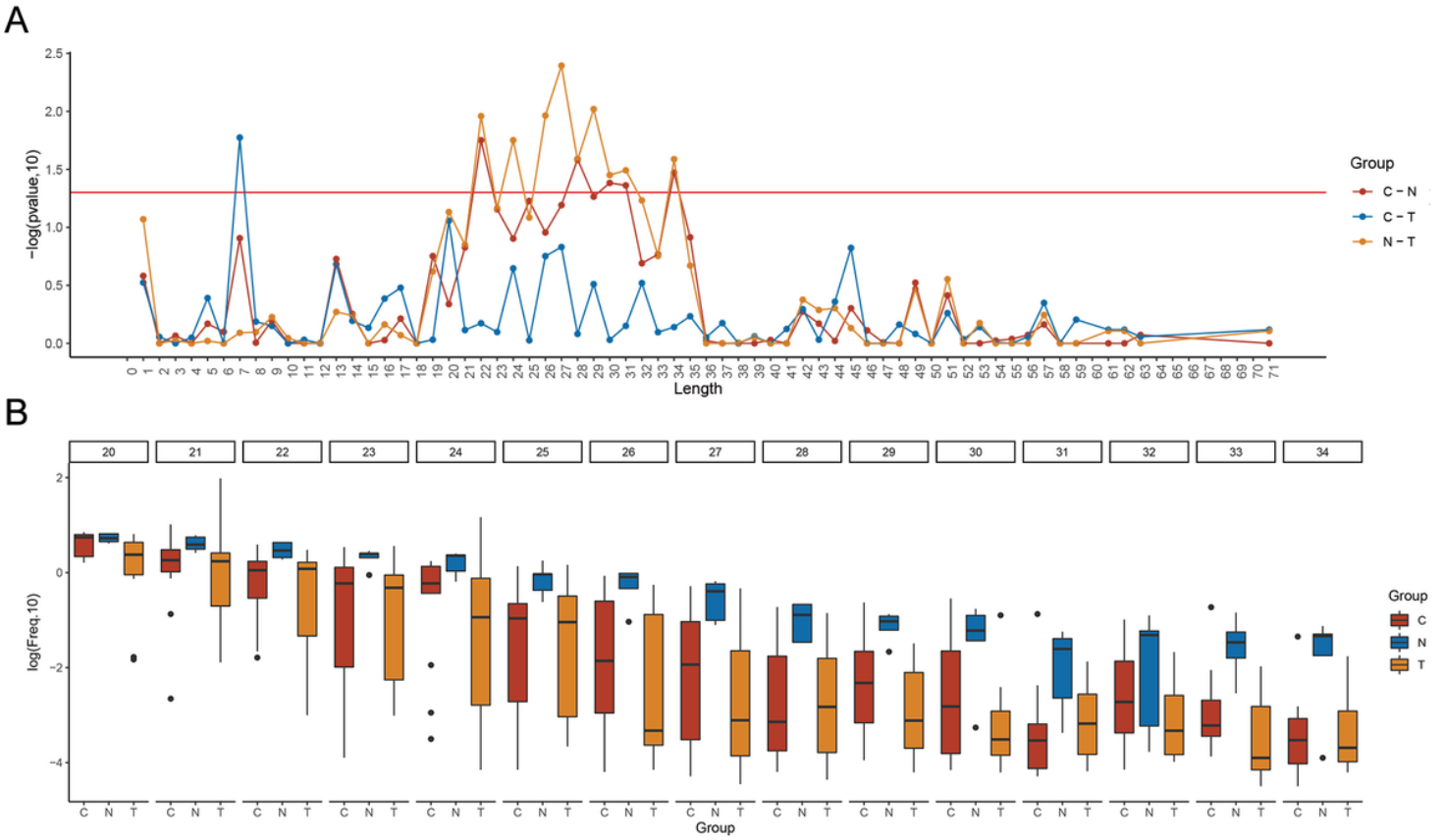

Figure 3

The aa length distribution of three groups

(A)Comparison of CDR3 with different amino acid lengths in N,Tand C, The ordinate is - log (pvalue, 10), the P values of the three groups are the results of Kruskal Wallis test, the p value between the two groups was the result of Dunn's test.(B)The box diagram shows the comparison of CDR3 with amino acid length between 20-34 in groups N, T and C.

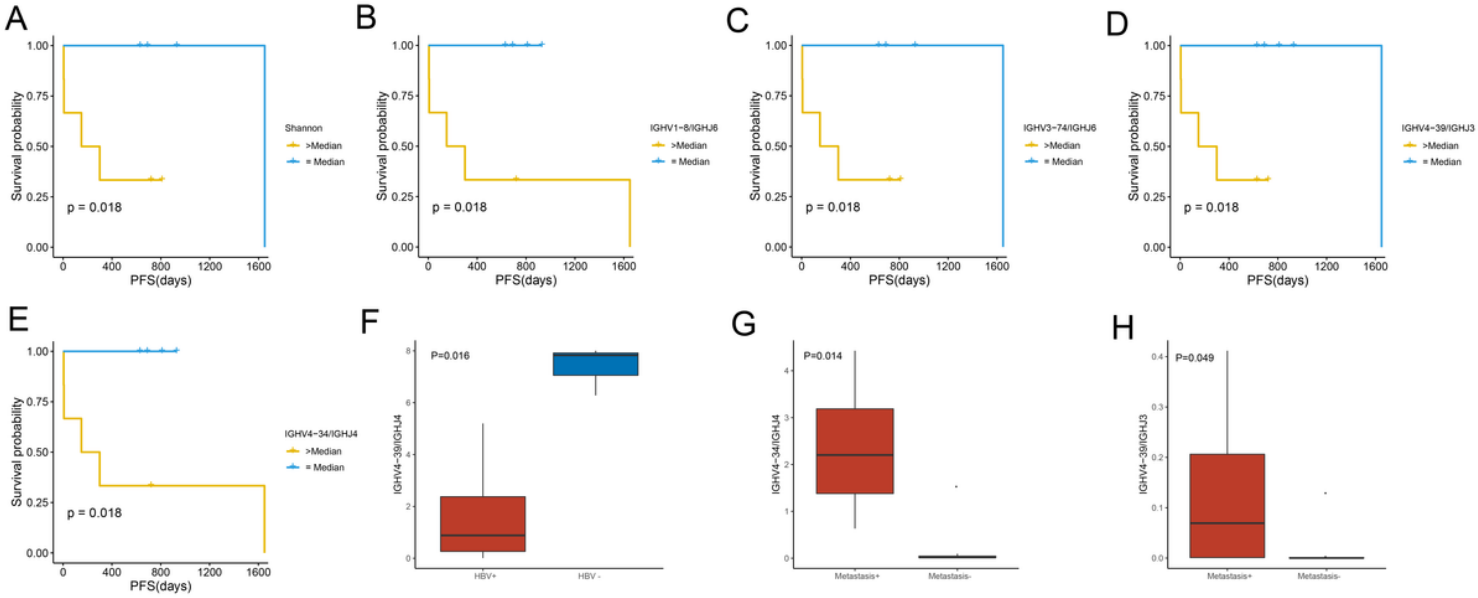

Figure 4

Prognosis analyses of BCR characteristics in DLBCL patients

(A)DLBCL patients were divided into two groups according to the median Shannon index, and the progression free survival curve was drawn,P in the lower left corner was significant,(B-E)According to IGHV1-8/IGHJ6，IGHV3-74/IGHJ，IGHV4-39/IGHJ3，IGH V4-34/IGHJ4,DLCBL was divided into two groups to compare progression free survival curve,P in the lower left corner was significant,(F)Ighv4-39 / ighj4 were compared between HBV positive and HBV negative groups,(G,H)Comparison of IGHV4-34 / IGH J4 and IGHV4-39 / IGHJ3 between metastatic and non metastatic groups in DLBCL patients.

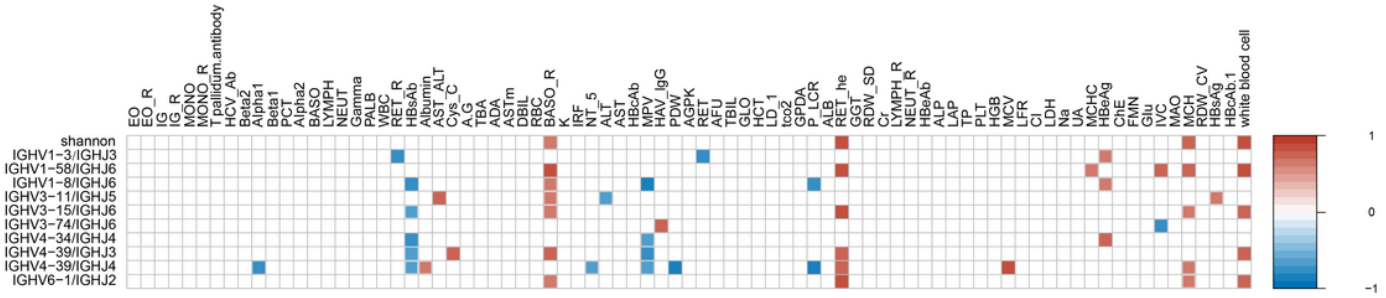

Figure 5

Correlation analysis between BCR features and clinical characteristics

Correlation between Shannon index, 60 clinical features and 10 CDR3 in patients with DLBCL patients .The ordinate is CDR3, and the abscissa is clinical data and Shannon index. Different colors represent correlation coefficients, and the blank part represents no significant difference.

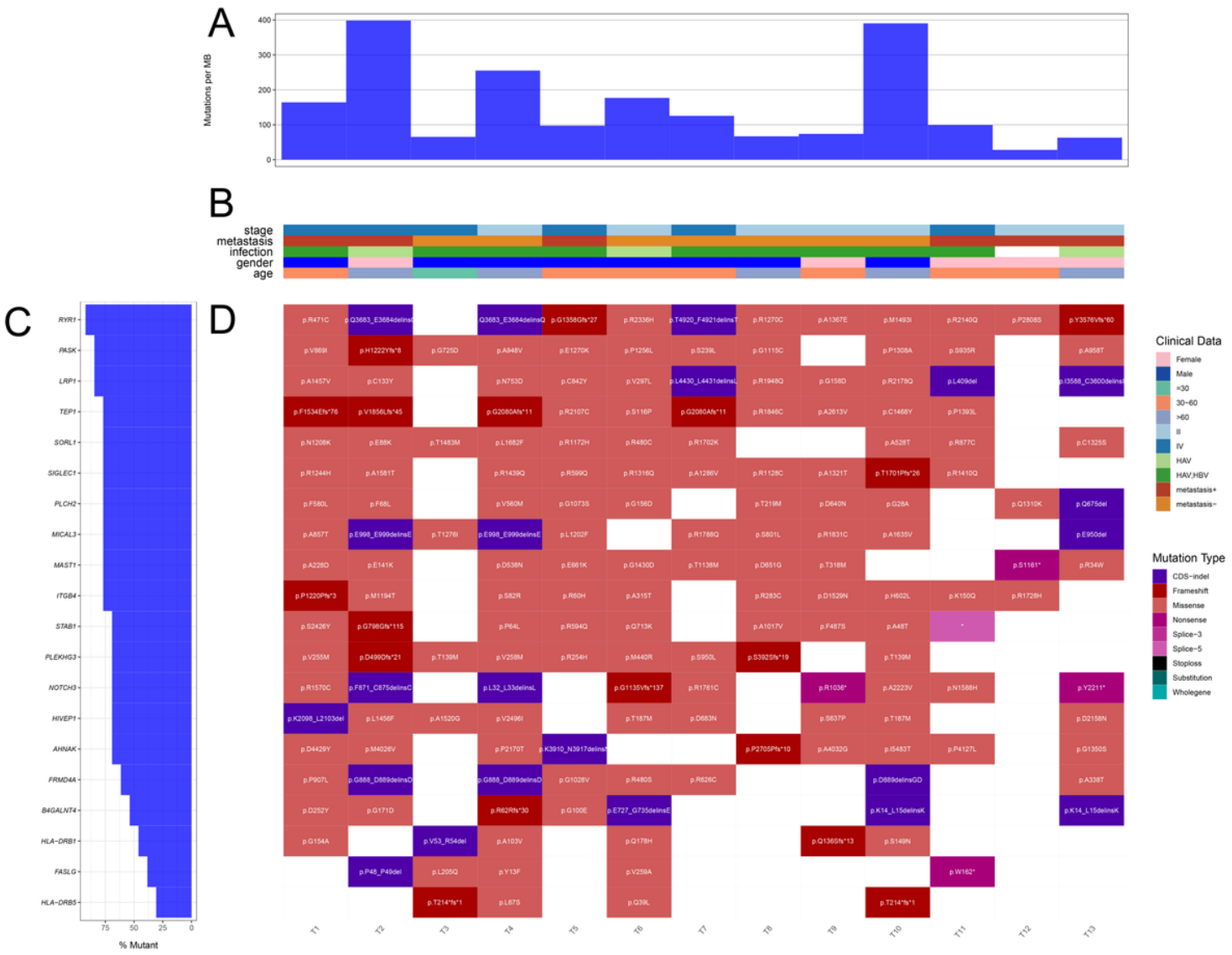

Figure 6

Somatic mutation landscape of 13 DLBCL patients

WES data of 13 DLBCL patients,Figure A shows Mutations per MB.Figure B shows the age, gender, infection,metastasis and stage of 13 patients,different colors represent different clinical information. Please refer to the legend clincial data .Figure C shows the genes with high mutation frequency.Figure D non mutated site information,different colors are different mutation types.

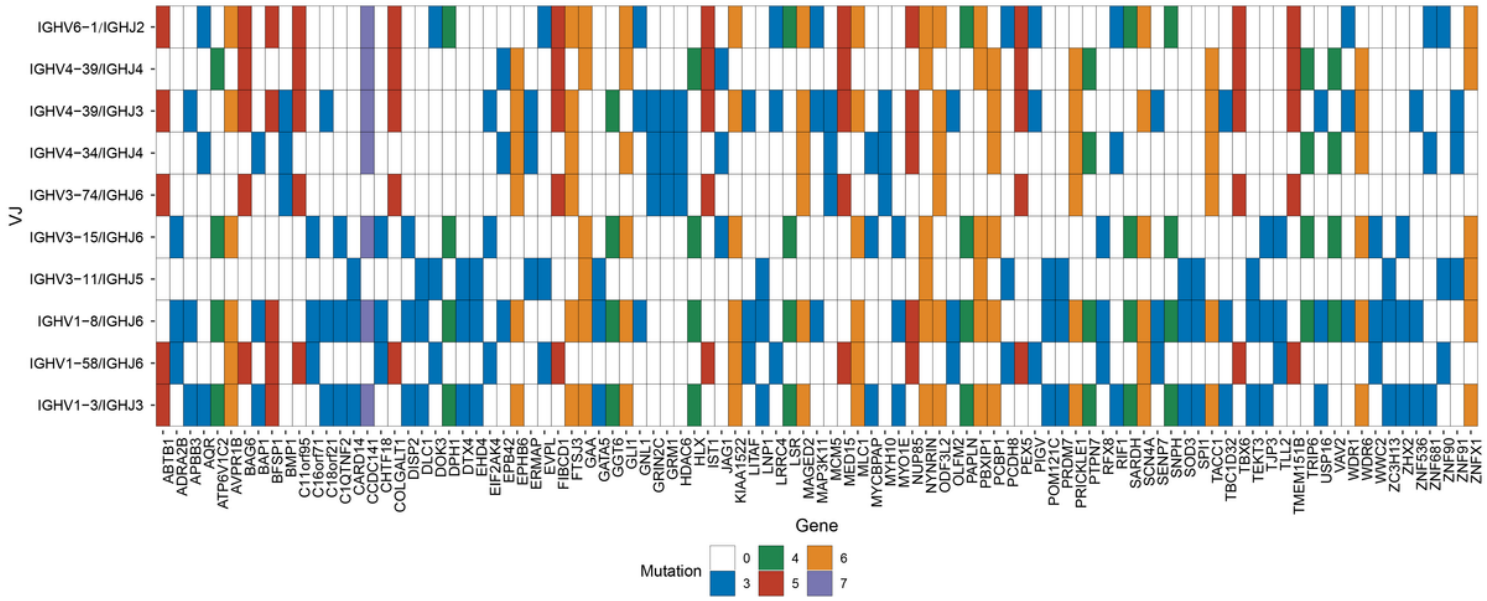

Figure 7

The exhibition of CDR3 usages related mutation genes

The ordinate is CDR3 and the abscissa is gene. Different colors are the number of mutant genes.

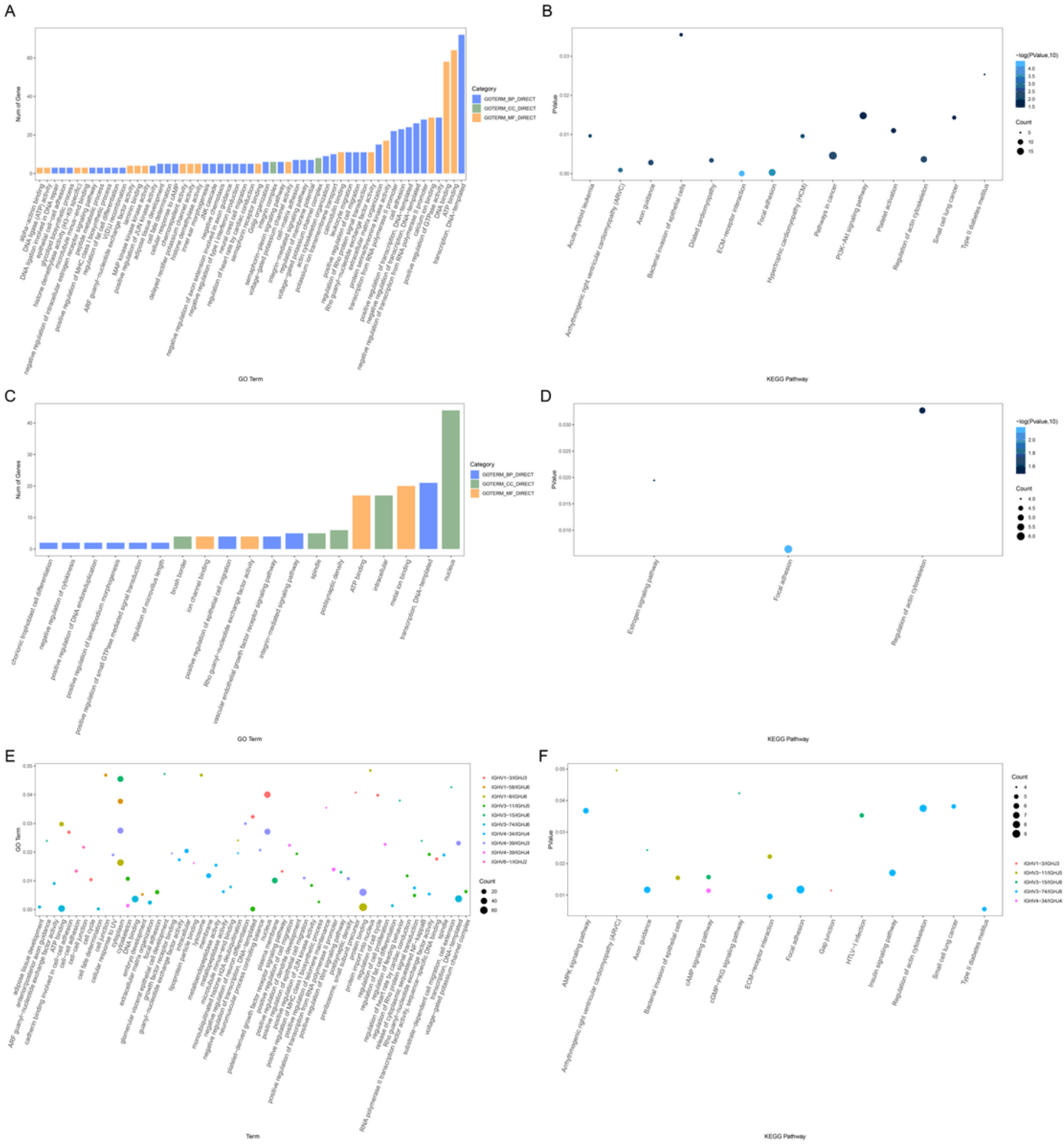

Figure 8

Gene enrichment analyses of 13 DLBCL patients

GO analysis based on the gene mutations related to 10 selected CDR3s usages, Shannon indexes (C) and each CDR3 usage (E),In figure A,C, the ordinate is the number of enriched genes and the abscissa is go term.The color of figure E represents CDR3, the point size represents the number of enriched genes, and

the abscissa is go term,the ordinate is pvalue. KEGG analysis based on the gene mutations related to 10 selected CDR3s usages (B), Shannon indexes (D) and each CDR3 usage (F). In Figure B, the ordinate is pvalue, the abscissa is KEGG pathway, the dot color is - log (pvalue, 10), and the dot size is the number of genes enriched.In Figure D, the ordinate is pvalue, the abscissa is KEGG pathway, the far point color is - log (pvalue, 10), and the dot size is the number of enriched genes.In Figure F, the ordinate is pvalue, the abscissa is KEGG pathway, different colors are CDR3, and the dot size is the number of genes enriched.