CD1C was identified as a potential biomarker by the comprehensive exploration of tumor mutational burden and immune infiltration in diffuse large B cell lymphoma (#82787)

First submission

Guidance from your Editor

Please submit by **22 Apr 2023** for the benefit of the authors (and your token reward) .

Structure and Criteria
Please read the 'Structure and Criteria' page for general guidance.

Custom checks
Make sure you include the custom checks shown below, in your review.

Raw data check
Review the raw data.

Image check
Check that figures and images have not been inappropriately manipulated.

If this article is published your review will be made public. You can choose whether to sign your review. If uploading a PDF please remove any identifiable information (if you want to remain anonymous).

Files

Download and review all files from the materials page.

13 Figure file(s)
5 Table file(s)
1 Raw data file(s)
1 Other file(s)

Custom checks

Human participant/human tissue checks

- Have you checked the authors ethical approval statement?
- Does the study meet our article requirements?
- Has identifiable info been removed from all files?
- Were the experiments necessary and ethical?

Structure and Criteria

Structure your review

The review form is divided into 5 sections. Please consider these when composing your review:

1. **BASIC REPORTING**
2. **EXPERIMENTAL DESIGN**
3. **VALIDITY OF THE FINDINGS**
4. General comments
5. Confidential notes to the editor

You can also annotate this PDF and upload it as part of your review

When ready submit online.

Editorial Criteria

Use these criteria points to structure your review. The full detailed editorial criteria is on your guidance page.

BASIC REPORTING

- Clear, unambiguous, professional English language used throughout.
- Intro & background to show context. Literature well referenced & relevant.
- Structure conforms to PeerJ standards, discipline norm, or improved for clarity.
- Figures are relevant, high quality, well labelled & described.
- Raw data supplied (see PeerJ policy).

EXPERIMENTAL DESIGN

- Original primary research within Scope of the journal.
- Research question well defined, relevant & meaningful. It is stated how the research fills an identified knowledge gap.
- Rigorous investigation performed to a high technical & ethical standard.
- Methods described with sufficient detail & information to replicate.

VALIDITY OF THE FINDINGS

- Impact and novelty not assessed. *Meaningful* replication encouraged where rationale & benefit to literature is clearly stated.
- All underlying data have been provided; they are robust, statistically sound, & controlled.
- Conclusions are well stated, linked to original research question & limited to supporting results.

Standout reviewing tips

The best reviewers use these techniques

Tip	*Example*
Support criticisms with evidence from the text or from other sources	*Smith et al (J of Methodology, 2005, V3, pp 123) have shown that the analysis you use in Lines 241-250 is not the most appropriate for this situation. Please explain why you used this method.*
Give specific suggestions on how to improve the manuscript	*Your introduction needs more detail. I suggest that you improve the description at lines 57- 86 to provide more justification for your study (specifically, you should expand upon the knowledge gap being filled).*
Comment on language and grammar issues	*The English language should be improved to ensure that an international audience can clearly understand your text. Some examples where the language could be improved include lines 23, 77, 121, 128 – the current phrasing makes comprehension difficult. I suggest you have a colleague who is proficient in English and familiar with the subject matter review your manuscript, or contact a professional editing service.*
Organize by importance of the issues, and number your points	*1. Your most important issue* *2. The next most important item* *3. …* *4. The least important points*
Please provide constructive criticism, and avoid personal opinions	*I thank you for providing the raw data, however your supplemental files need more descriptive metadata identifiers to be useful to future readers. Although your results are compelling, the data analysis should be improved in the following ways: AA, BB, CC*
Comment on strengths (as well as weaknesses) of the manuscript	*I commend the authors for their extensive data set, compiled over many years of detailed fieldwork. In addition, the manuscript is clearly written in professional, unambiguous language. If there is a weakness, it is in the statistical analysis (as I have noted above) which should be improved upon before Acceptance.*

CD1C was identified as a potential biomarker by the comprehensive exploration of tumor mutational burden and immune infiltration in diffuse large B cell lymphoma

Xiaoyu Xiang [Corresp., 1], **Li-Min Gao** [1], **Yuehua Zhang** [1], **Qiqi Zhu** [1], **Sha Zhao** [1], **Weiping Liu** [1], **Yunxia Ye** [1], **Yuan Tang** [1], **Wenyan Zhang** [Corresp. 1]

[1] Department of Pathology, West China Hospital of Sichuan University, Chengdu, Sichuan, China

Corresponding Authors: Xiaoyu Xiang, Wenyan Zhang
Email address: xxy071796@163.com, zhangwenyan@wchscu.cn

Backgrouund: Tumor mutational burden (TMB) is a useful biomarker to predict prognosis. This study was to explore the prognostic value of TMB and the potiential association between TMB and immune infiltration in DLBCL.

Methods: We downloaded the gene expression profile, somatic mutation and clinical data of DLBCL patients from the Cancer Genome Atlas (TCGA) database. TMB was calculated and we classified the samples into high- and low-TMB group in order to identify differentially expressed genes (DEGs). Functional enrichments analyzes were performed to identify the biological functiona of the DEGs. Besides, we utilized the CIBERSORT algorithm to estimate the abundance of 22 immune fractions, and the significant difference were determined by Wilcoxon rank-sum test between high- and low-TMB group. Hub gene had been screened as the prognostic TMB-related immune biomarker by the combination of the Immunology Database and Analysis Portal (ImmPort) database and the univariate Cox analysis from the Gene Expression Omnibus (GEO) database including three DLBCL datasets. Various database application (TIMER, CellMiner, knockTF, GETx) verified the functions of target gene.

Results: SNP occurred more frequently than insertion and deletion, and C > T was the most common of SNV in DLBCL. Survival analysis showed that high-TMB group conferred poor survival outcomes. A total of 62 DEGs were obtained and 13 TMB-related immune genes were identified. Univariate Cox analysis result illustrated that CD1C mutation was associated with lower TMB and manifested a satisfactory clinical prognosis by analysis of large samples from GEO database. In addition, infiltration levels of immune cells in high-TMB group were lower. Using the TIMER database, we further systematically analyzed the relationships between mutants of CD1C and immune infiltration levels. Drug sensitivity showed that there was a significant correlation between CD1C expression level and clinical drug sensitivity from CellMiner database. KnockTF database was used to comprehensively explore the regulation of gene-related transcription factors and signaling pathways. We searched the GETx database to compare the mRNA expression levels of CD1C between lymphoma and normal tissues and the results suggested that there was significant difference between tumor and normal tissues in most studies.

Conclusions: Higher TMB correlated with poor survival outcomes and might inhibit the immune infiltrates in DLBCL. Our results suggest that CD1C is a TMB-related prognostic biomarker.

CD1C was identified as a potential biomarker by the comprehensive exploration of tumor mutational burden and immune infiltration in diffuse large B cell lymphoma

Xiaoyu Xiang[1], Li-Min Gao[1], Yuehua Zhang[1], Qiqi Zhu[1], Sha Zhao[1], Weiping Liu[1], Yunxia Ye[1], Yuan Tang[1], and Wenyan Zhang[1]

[1] Department of Pathology, West China Hospital of Sichuan University, 610000, chengdu, Sichuan, China

Corresponding Author: Xiaoyu Xiang, Yuan Tang, Wenyan Zhang
Guoxue Alley No 37, Chengdu, 610041, Sichuan, China
Email address: xxy071797@163.com, 1202ty@163.com, zhangwenyan@wchscu.cn

ABSTRACT

Backgrouund: Tumor mutational burden (TMB) is a useful biomarker to predict prognosis. This study was to explore the prognostic value of TMB and the potiential association between TMB and immune infiltration in DLBCL.

Methods: We downloaded the gene expression profile, somatic mutation and clinical data of DLBCL patients from the Cancer Genome Atlas (TCGA) database. TMB was calculated and we classified the samples into high- and low-TMB group in order to identify differentially expressed genes (DEGs). Functional enrichments analyzes were performed to identify the biological functiona of the DEGs. Besides, we utilized the CIBERSORT algorithm to estimate the abundance of 22 immune fractions, and the significant difference were determined by Wilcoxon rank-sum test between high- and low-TMB group. Hub gene had been screened as the prognostic TMB-related immune biomarker by the combination of the Immunology Database and Analysis Portal (ImmPort) database and the univariate Cox analysis from the Gene Expression Omnibus (GEO) database including three DLBCL datasets. Various database application (TIMER, CellMiner, knockTF, GETx) verified the functions of target gene.

Results: SNP occurred more frequently than insertion and deletion, and C > T was the most common of SNV in DLBCL. Survival analysis showed that high-TMB group conferred poor survival outcomes. A total of 62 DEGs were obtained and 13 TMB-related immune genes were identified. Univariate Cox analysis result illustrated that CD1C mutation was associated with lower TMB and manifested a satisfactory clinical prognosis by analysis of large samples from GEO database. In addition, infiltration levels of immune cells in high-TMB group were lower. Using the TIMER database, we further systematically analyzed the relationships between mutants of CD1C and immune infiltration levels. Drug sensitivity showed that there was a significant correlation between CD1C expression level and clinical drug sensitivity from CellMiner database. KnockTF database was used to comprehensively explore the regulation of gene-related transcription factors and signaling pathways. We searched the GETx database to compare the mRNA expression levels of CD1C between lymphoma and normal tissues and the results suggested that there was significant difference between tumor and normal tissues in most studies.

Conclusions: Higher TMB correlated with poor survival outcomes and might inhibit the immune infiltrates in DLBCL. Our results suggest that CD1C is a TMB-related prognostic biomarker.

Keywords Diffuse large B cell lymphoma, tumor mutational burden, immune infiltration, CD1C, prognostic biomarker

INTRODUCTION

Diffuse large B-cell lymphomas (DLBCL), comprising approximately 30% of all non-Hodgkin lymphoma (NHL) cases, are the most common NHL types (*Lenz & Staudt, 2010*). And the incidence rate of DLBCL is over 40 per 100,000 individuals in adults ages 70 years and older (*Morton et al., 2006; Smith et al., 2015; Teras et al., 2016*). DLBCL is a heterogeneous subtype of aggressive B-cell neoplasm with varied clinical, immunophenotypic, cytogenetic, and genetic features (*Berglund et al., 2002; Lossos & Morgensztern, 2006; Schmitz et al., 2018; Sehn & Gascoyne, 2015*). About 70% of DLBCL patients are located in the advanced stage and require systemic therapy (*Morin et al., 2011; Sehn & Gascoyne, 2015*). Currently, the standard chemotherapy regimen is R-CHOP (*B. Coiffier et al., 2002; Czuczman et al., 1999; Roschewski, Staudt, & Wilson, 2014; Vose et al., 2001)*, but about 10% to 15% of patients receiving R-CHOP is under a primary refractory disease (i.e., incomplete remission or relapse within 6 months of treatment) *(Chaganti et al., 2016; B. Coiffier et al., 2002; Bertrand Coiffier et al., 2010; S. Li, Young, & Medeiros, 2018*). Another 20%-25% of patients progress 2 years after initial remission *(Chaganti et al., 2016; Crump et al., 2017; Sehn & Gascoyne, 2015*). Therefore, it is urgent to identify new biomarkers for clinical diagnosis and treatment of DLBCL.

In recent years, immunotherapy has become a common treatment for metastatic and invasive tumors (*Finn, 2012; Topalian et al., 2015; Tran, Robbins, & Rosenberg, 2017*). Tumor immunotherapy is a treatment that uses the body's immune system to attack cancer cells *(Conlon et al., 2015; Farkona, Diamandis, & Blasutig, 2016*). Immunotherapy mainly includes tumor vaccines, biotherapy, CAR-T therapy, and immune checkpoint inhibitors (*Efremova, Finotello, Rieder, & Trajanoski, 2017; Khalil, Smith, Brentjens, & Wolchok, 2016*). More and more studies have shown that molecular targets related to the tumor microenvironment invasion may become the key of immunotherapy(*Shain, Dalton, & Tao, 2015*).

Based on advances in gene sequencing and expression profiling, studies have shown that the prognosis of DLBCL patients is related to TME *(Ciavarella et al., 2019; Nicholas, Apollonio, & Ramsay, 2016*). For example, immune checkpoint inhibitors (ICI) have shown significant efficacy in some refractory hematological malignancies *(Thanarajasingam, Thanarajasingam, & Ansell, 2016*). Tumor mutation burden (TMB) is the number of somatic mutations per megabase (Mb) of the genome in a tumor, representing genomic instability (*Chalmers et al., 2017; Chen et al., 2020; Schumacher & Schreiber, 2015*). Tumors with a high mutation burden are more likely to induce neoantigen production, so TMB has become a predictor of the response rate to immune checkpoint inhibitors *(Schumacher & Schreiber, 2015*). Previous studies have shown that TMB is associated with cancer immunotherapy response and cancer prognosis (*Chan et al., 2019; Samstein et al., 2019; X. Wang & Li, 2019*).

In this study, we collected somatic mutation data, transcriptome data, and clinical information from the TCGA database, aiming to study the association between TMB and gene mutation, immune response, and prognosis of DLBCL combined with immune infiltration. We attempted to elucidate the relationships between TMB groups and clinicopathological factors, between TMB groups and different immune-infiltrating cells, and between TMB groups and prognosis. The results of these studies may provide novel biomarkers and potential treatment options for DLBCL.

MATERIALS AND METHODS

Acquisition of somatic mutation data and expression profile from TCGA.

We obtained the somatic mutation profile from the publicly available TCGA database via the GDC data portal (https://portal.gdc.cancer.gov/). Among the four subtypes of files, "Masked Somatic Mutation" data were selected and processed based on VarScan software. Summarizing, analyzing, annotating, and visualizing mutation annotation format (MAF) files used to store detected somatic variation using the "maftools" bioconductor package. In addition, we also downloaded HTseq-FPKM transcription profile from UCSC Xena (https://xena.ucsc.edu/), which were respectively of TCGA-DLBC tumor samples and normal samples of "Cells-EBV-Transformed Lymphocytes" locating in GETx database (https://www.gtexportal.org/home/). Corresponding clinical information was also collected from UCSC Xena, including age, sex, tumor grade, pathological stage, TNM stage, survival time, and OScensor.

Calculation of TMB score of each sample and prognostic analysis.

TMB values for each sample were determined by measuring the total number of nonsynonymous detected per million bases, which could be calculated as (whole counts of gene variants) / (the whole length of exons). In our study, we calculated the mutation frequency of variation/exon length (38 million) per sample based on the "maftools" R package. TCGA-DLBC samples were subdivided into low- (18 cases) and high- (19 cases) TMB groups according to the median data. Then, TMB mutation data were combined with the corresponding survival data by sample ID and Kaplan-Meier (KM) analysis was performed to compare survival differences between the low- and high-TMB groups with log-rank sum test. In addition, the association between TMB and clinical features was further assessed, in which the Wilcoxon rank-sum test was used to calculate the p-value for the two groups, and the Kruskal-Wallis (KW) test was used for three or more groups.

Differentially expressed genes and functional pathways analysis.

According to TMB level, the transcriptome profile was assigned into low- and high-TMB groups by R software. "limma" R package was used to identify the DEGs between the low- and high-TMB groups, and the thresholds were set at p-value < 0.05 and $|\log_2(\text{fold change})| > 1.0$. A heat map was drawn by using the "pheatmap" R package. Then, the entreID of each DEG was generated using the "org.Hs.eg.db" R package and we used the "clusterProfiler", "GOplot", "ggplot2" R packages for Gene Ontology (GO) analysis, and "enrichplot" for Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis. Besides, we performed gene set enrichment analysis (GSEA; https://genepattern.broadinstitute.org/) (*Hung, Yang, Hu, Weng, & DeLisi, 2012*) based on the JAVA8 platform using the TMB group as the phenotype and TCGA-DLBC mRNA expression profile as expression spectrum data file. Then we selected "c2.cp.kegg.v6.2.symbol.gmt" gene set as a reference gene set which it is derived from MsigDB Database (http://software.broadinstitute.org/gsea/msigdb/). The significant enrichment pathway was considered only when p-value < 0.05.

Co-analyses of TMB and immune infiltration.

We evaluated the proportion of immune cells using the deconvolution algorithm CIBERSORT (https://cibersort.stanford.edu/). CIBERSORT (CIBERSORT R Script V1.03) was a general calculation method that can accurately estimate the composition of 22 immune cells in tumor tissues by combining the prior knowledge of the composition spectrum of purified leukocyte subsets with support vector regression (*Newman et al., 2015*). We then identified the differences in the composition of immune cells between low- and high-TMB groups and the number of permutations were set to 1000 as well as p-value < 0.05. The "pheatmap" R package showed the distribution of immune cells between the two groups and the "vioplot" package was used to show the differential immune infiltration by the Wilcoxon rank-sum test. The threshold p-value < 0.05 was the standard to calculate the significance of a single immune cell between the two groups. In addition, we obtained a list of 2483 immune-related genes from the Immunology Database and Analysis Portal (https://www.immport.org/). The "VennDiagram" R package was used to screen the intersecting genes between TMB-DEGs and immune-related genes. Univariate

Cox regression analysis was performed to determine the prognostic TMB-related immune genes using "survival", "survminer" and "forestplot" R packages.

Validating the prognostic TMB-related immune genes in the GEO database.

We systematically searched for the GEO database (https://www.ncbi.nlm.nih.gov/geo/) open clinical annotations of DLBCL gene expression profile data, and obtained the three datasets, including GSE10846, GSE31312, GSE32918. Then proceed with datasets processing, (i) Data downloading, download dataset file of a series matrix; (ii) Background correction and standardization of data, such as quantile standardization; (iii) Using GPL570 and GPL8432 annotation files for ID translation; (iv) The same gene corresponds to multiple probes, and the average value of the probes was calculated as the expression level; (v)Complete expression profile data files and corresponding clinical information of patients, including survival time and survival status, were obtained. As a result, 1133 DLBCL samples were selected, including 470 (GSE31312), 249 (GSE32918), and 414 (GSE10846) samples. The Combat function in the "sva" R package was used to remove the batch effect and integrated three datasets to obtain the expression spectrum. Prognostic TMB-related immune genes were screened to verify whether there was a statistical significance between their expression and prognosis in GSE datasets. We selected 5 TMB-related immune genes with $|\log_2(\text{fold change})| > 1$ and p-value <0.05 to further assess the prognostic value of differential immune genes in patients with low- and high-TMB levels. Kaplan-Meier analysis was conducted via a "for cycle" R script to find the hub immune genes associated with survival outcomes.

Copy number variations and correlated immune cells of prognostic TMB-related immune genes.

Tumor Immune Estimation Resource database (TIMER, https://cistrome.shinyapps.io/timer/), a web server for comprehensive analysis of tumor infiltrating immune cells, was used to estimate the abundance of six types of immune infiltrating cells such as B cells, CD4+ T cells, neutrophils, macrophages, and dendritic cells (*T. Li et al., 2020*). Changes in copy Number Variations (CNV) were observed in prognostic TMB-related immune genes, and the correlations between CNV and immune cells abundance, and between immune cells and survival were further assessed.

Analysis of drug sensitivity of target genes.

The drug sensitivity data used in this study were obtained from the CellMiner database (https://discover.nci.nih.gov/cellminer/home.do) (*Reinhold et al., 2012*). The transcriptome and drug sensitivity data of the same batch of samples were downloaded, and the expression profile of the target gene and drugs verified by the Food and Drug Administration (FDA) were retained by sorting the data. Then, the correlation between target gene expression level and drug sensitivity was extracted and further explored by Spearman correlation analysis. The higher the cor value, the stronger the correlation.

Target gene-related transcription factors signaling pathway and validation of target gene.

KnockTF (http://www.licpathway.net/KnockTF/index.html) was used to explore a combination of the regulation of gene-related transcription factors and log FC >1.0 signaling pathways (*Feng et al., 2020*). Meanwhile, we searched the GETx database (https://www.gtexportal.org/home/) to compare mRNA expression levels of target genes between lymphoma tissues and normal tissues. The difference of CD1C mRNA expression between tumor and normal tissues was verified by wet assay. 22 paraffin samples from 2021.12 to 2022.2 from the Department of Pathology of West China Hospital of Sichuan University were screened, of which 13 cases were confirmed as DLBCL samples and 9 samples of normal lymphoid tissue hyperplasia. Ethics Committee on biomedical Research, West China Hospital of Sichuan University approved this study (IRB:2020-703) and waived informed consent. Depending on the manufacturer's protocol, total RNA was extracted from FFPE samples and gDNA removed using the RNApure FFPE kit

(CW0535, CoWin Bioscience, Beijing, China). HiScript® III All-in-one RT SuperMix was utilized Perfect for qPCR (R333, Vazyme, NanJing, China) reverse transcription and used cDNA as a template for real-time fluorescence quantification. RT-qPCR was performed with the SYBR® Green Premix Ex Taq™ II (Tli RNaseH Plus) (RR820A, TaKaRa, Beijing, China) on a Real-time PCR Detection System (Bio-rad). Independent experiments are performed in triplicate, β actin as an internal control. The following primers (Tsingke Biotechnology Co., Ltd., Beijing, China) were used: CD1C: FP 5'-CACTTGCCCCCGATTTCTCT-3'; RP 5'-ATGGAAAAGTGGTGTCCCCAG-3'. ACTIN: FP 5'-CCGCGAGAAGATGACCCAGA-3'; RP 5'-GATAGCACAGCCTGGATAGCA-3'.

RESULTS

The landscape of mutation profiles in DLBCL.

The research strategy is presented in (Fig. 1). Somatic mutation profiles of 37 DLBCL samples were downloaded from the TCGA database. We used the "maftools" R package to visualize mutation data in VAF format. In general, missense mutation accounted for the largest proportion of mutation types (Fig. 2A), and the occurrence frequency of single nucleotide polymorphism (SNP) was higher than insertion and deletion (Fig. 2B). The most common type of base substitution was C>T (Fig. 2C). The boxplot (Fig. 2D and 2E) showed different mutation types in DLBCL patients, and Fig. 2F showed the top 10 genes with mutation frequency, including PIM1 (22%), IGLV3-1 (38%), IGLL5 (27%), IGHG1 (22%), IGHV2-70 (27%), BTG (27%), IGHM (24%), KMT2D (32%), IGLC2 (24%), CARD11 (22%). The mutation landscape displayed the mutation information of each sample, in which the mutation frequency of IGLV3-1 and KMT2D accounted for 38% and 32%, respectively (Fig. 2G). Heatmap of gene correlations shows gene-to-gene relationships, for example, there is a synergistic effect between MUC16 and FAT4, while SOCS1 and KMT2D are mutually exclusive (Fig. 2H). Meanwhile, the Genecloud plot displayed the frequency of mutations in genes (Fig. S1), and the higher the mutation frequency, the larger the gene name.

TMB correlated with survival outcomes and clinical pathological characteristics.

We calculated the mutation event per million bases as the TMB for DLBCL patients, worked out the optimal cutpoint using surv_cutpoint function in the "survival" R package, and set the parameter minprop = 0.1 to divide patients into low- (18 cases) and high- (19 cases) TMB groups. TMB ranged from 0.14 to 6.92 with a median of 1.9 per MB (Fig. 3A). Kaplan-Meier survival analysis was carried out, and the result showed that the 5-year survival rate of the high-TMB group was lower than the low-TMB group (Fig. 3B). In addition, none of the clinical traits was significantly correlated with TMB level, which may be due to the small samples (Table S1).

Identifying differentially expressed genes based on TMB grouping and functional enrichment analysis of GO, KEGG, and GSEA.

Differentially expressed genes (DEGs) were calculated by R software. A total of 62 DEGs were identified in low- and high-TMB group using "limma" R package by setting the p-value < 0.05 and $|\log_2(FC)| > 1$ (Table 1). The heatmap visualized DEGs between the low- and high-TMB groups (Fig. 4A). The volcano map showed 42 up-regulated genes and 20 down-regulated genes (Fig. 4B). Subsequently, we conducted GO enrichment analysis on DEGs and found that the differential genes were mainly involved in immune-related pathways, such as lymphocyte-mediated immunity, adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains, immunoglobulin mediated immune response, and complement activation, classical pathway (Fig. 4C, Table S2). The enrichment information of the GO pathways is illustrated in Fig. 4D.
CD1C, CCL21, TP63, ORM1, ACTG2, IGHG3, IGHM, TRPM4 and so on are involved in all of the top GO pathways (including Molecular Function, Cellular Component, Biological Process) and were identified as hub genes. KEGG pathways illustrated that the differential genes were mainly enriched in vascular

smooth muscle contraction, hematopoietic cell lineage, carbon metabolism, neuroactive 210 ligand−receptor interaction pathway (Fig. S2, Table S3). In addition, we further selected the GSEA results of the top TMB-related items in Figs. 4E-G, including one carbon pool by folate, rig- Ⅰ like receptor signaling pathway-creative diagnostics, and the tight junction, which were associated with the TMB level (p-value < 0.05).

Differential abundance of immune cells in the low- and high-TMB groups using CIBERSORT.

After DEGs screening, in order to further compare the difference in the degree of immune cells infiltration between low- and high-TMB groups, we calculated the composition ratio of immune cells per sample by the "CIBERSORT" R package. The boxplot in Fig. 5A showed a specific portion of 22 immune cells in each DLBCL sample. We also calculated the proportion of immune cells in the whole DLBCL cohort accounting for the most including B cells naive, CD8+ T cells, M2 Macrophages, M0 Macrophages (Fig. 5B). The heatmap showed the distribution of immune cells between low- and high-TMB groups, and the result displayed that the high-TMB group had a lower immune score (Fig. 5C). In addition, the Wilcoxon rank-sum test demonstrated that monocytes, dendritic cells activated, and dendritic cells resting were lower in the high-TMB group (p-value < 0.05) (Fig. 5D). According to the above analysis results, the high-TMB group inhibited the level of immune cells infiltration in DLBCL samples.

Screening TMB-related immune genes and verifying the prognosis of the screened genes using the GEO database.

The immune-related genes were downloaded from the ImmPort database and intersected with the selected DEGs. 13 TMB-related immune genes were obtained, including CD1C, ORM1, ORM2, CCL21, CR2, IGHG3, IGHM, IGHV1-69, IGHV3-23, IGKV1-5, IGKV1D-8, SSTR2, TRBJ2-1 (Fig .6A). Subsequently, univariate regression analysis was performed on the above genes, and it was found that there was no significant correlation between these genes and prognosis (p-value > 0.05, Fig. S3), possibly due to the small samples in the TCGA-DLBC cohort. Therefore, we expanded the sample size and screened a total of 1133 samples of DLBCL gene expression microarray datasets (GSE31312, GSE10846, GSE32918) from the GEO database, as well as the clinical information of the corresponding samples. After ID translation, data homogenization and standardization, and removal of batch effect, 5 genes including CD1C, CCL21, ORM1, CR2, SSTR2 (the rest of the 13 genes were not included in the expression profile data) were obtained after joint analysis of the three datasets. Kaplan-Meier survival analysis was performed, and the results showed that the level of CD1C expression was significantly correlated with prognosis (p-value < 0.05, Figs .6B-F).

CNV of CD1C, immune cells and survival in DLBCL using TIMER database.

In general, CNV refers to an increase or decrease in the copy number of large segments of the genome that is more than 1kb in length. To verify the CNV of CD1C and the relationship between immune cells content and prognosis, we utilized the TIMER database to obtain CD1C expression between normal and tumor tissues in various cancers (Fig .7A). Especially in DLBCL, CD1C expression was positively correlated with B cells, neutrophils, dendritic cells and negatively correlated with CD8+ T cells, CD4+ T cells, and macrophages, among which, the correlation with B cells was the highest (cor = 0.693, p-value = 1.44E −03, Fig .7B). In addition, high amplification of CD1C was significantly different compared to other CNVs (p-value < 0.01, Fig .7C). As for the relationship between immune cells content and prognosis, high levels of CD8+ T cells and dendritic cells indicate a superiors survival result, while low expression of CD1C may promote better survival (Fig .7D).

Analysis of the relationship between CD1C and drug sensitivity in DLBCL.

According to the correlation analysis of target genes and drug sensitivity in the CellMiner database, it was found that there was a significant correlation between the expression level of CD1C and clinical drug sensitivity, mainly with nelarabine, methylprednisolone, chelerythrine, ribavirin, fluphenazine was

positive (Figs .8A and 8B). Therefore, the lower the expression of CD1C, the more sensitive cells are to these drugs.

Modulated CD1C transcription factors and verified CD1C by GETx database and RT-qPCR.

KonckTF database result showed that the regulation of CD1C may be related to the effects of transcription factors such as CREB1, AHR, and TOX, resulting in the corresponding biological effects (Table 2). CD1C mRNA levels were compared between normal tissues downloaded by GETx and tumor tissues of TCGA-DLBC. The results showed that the expression level of CD1C was significantly different between normal tissue and tumor tissue (p-value < 0.05, Fig. S4). To verify this results in the FFPE samples, RT-qPCR was employed. The expression levels of CD1C in DLBCL tissue and normal lymphoid tissue hyperplasia were significant difference (Fig. 9, p-value = 0.019458). The wet experiment further verified the reliability of bioinformatics consequences.

DISCUSSION

Recent studies have shown that the tumor microenvironment plays an important role in DLBCL initiation, progression and drug responsiveness (*Solimando et al., 2020*). The DLBCL immune microenvironment is composed of immune cells, stromal cells, blood vessels and extracellular matrix (*Leivonen et al., 2019*). Among them, immune cells are the key part of tumor microenvironment, including T cells, macrophages, NK cells and dendritic cells. Different immune cells have distinct prognostic effects on DLBCL, but even in the same cell population, the related research is still full of controversy. Studies have demonstrated that tumor infiltration and activation of CD4+ memory T cells are independent prognostic factors regardless of R-CHOP regimen (*Keane et al., 2013*). Khalifa et al. found that lymphomas with increased CD14 monocyte numbers and loss of HLA-DR expression were more aggressive and more frequently associated with refractory disease or recurrent therapy (*Khalifa, Badawy, Radwan, Shehata, & Bassuoni, 2014*). In DLBCL outcome, the prognosis of TAM is also controversial, which depends on M1/M2 macrophages. Riihijarvi et al. found that in patients treated with R-CHOP, both CD68 TAM and CD68 mRNA levels were associated with poor prognostic factors for OS, but in patients treated with R-CHOP, the prognosis of CD68 was favorable and OS was improved *(Sari et al., 2015*). However, Marchesi et al. (n = 61) (*Marchesi et al., 2015*), Nam et al. (n = 165) *(Nam et al., 2014*), and Wada et al. (n = 101) (*Wada et al., 2012*) found that M2 TAM was an important factor for poor prognosis and an independent predictor of shorter OS and PFS. Assessing the composition of the tumor microenvironment as well as the extent of immune cell infiltration could better distinguish subgroups of patients with poor prognosis, which would allow the implementation of personalized therapies (*Steen et al., 2021*).

Immunotherapy has made substantial progress in DLBCL. The tumor microenvironment (TME) can significantly affect the prognosis of DLBCL (*Camicia, Winkler, & Hassa, 2015*). Transcriptome analysis of the microenvironment of 4,655 DLBCL patients from multiple independent cohorts described four distinct lymphoma microenvironments (LME), i.e., germinal center-like type (GC), mesenchymal type (MS), inflammatory type (IN), and depleted type (DP). They closely relate to different biological aberrations and clinical behaviors (*Kotlov et al., 2021*). This measured tumor immunogenicity score (TIGS) shows consistently improved correlations with immunotherapy ORR in various types of cancer when compared to TMB, a novel biomarker that predicts response to immune checkpoint blockade (ICB) (*S. Wang, He, Wang, Li, & Liu, 2019*). For example, among Non-Small-Cell Lung Cancer patients receiving anti-PD-1/L1 treatment, patients with high TMB were associated with longer Progression Free Survival (PFS) than those with low TMB (*Rizvi et al., 2018*). Birkbak et al. studied TMB in ovarian cancer with BRCA1 and BRCA2 and found that TMB coupled with BRCA1 or BRCA2 mutations could be used as a genomic marker of prognosis and a predictor of treatment response (*Birkbak et al., 2013*). The association between the prognosis of TMB in DLBCL and immunotherapy has not been explored.

Therefore, we conducted this study to investigate the prognostic effect of TMB and its potential association with immune infiltration in DLBCL.

TMB was calculated based on the DLBCL mutation profile. The mutation landscape map showed the genes with a high mutation spectrum and the types of mutations, among which nonsense mutations account for the majority. The relationship between the survival curve and TMB suggested that TMB may not be just an independent prognostic factor for DLBCL. Therefore, we speculated that TMB combined with other prognostic factors may have a better predictive effect. Then according to the TMB groups, we selected DEGs, and GO enrichment analysis showed that the associated with TMB DEGs mainly involved in immune-related pathways, such as lymphocyte-mediated immunity, adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains, immunoglobulin mediated immune response and complement activation, classical pathway, etc. 13 TBM-related immune genes were obtained by the intersection of immune genes in ImmPort database and DEGs, which were CD1C, ORM1, ORM2, CCL21, CR2, IGHG3, IGHM, IGHV1-69, IGHV3-23, IGKV1-5, IGKV1D-8, SSTR2, TRBJ2-1.

In order to investigate the relationship between TMB and immune infiltration, the CIBERSORT algorithm was utilized to prove that the content of neutrophils, and dendritic cells in the low-TMB group was lower than that in the high-TMB group. According to recent studies in DLBCL, a lower proportion of dendritic cells were associated with better outcomes (*Ciavarella et al., 2019*). On the other hand, tumor immunogenicity was high in the high-TMB group, leading to CD4+ T cells infiltration, memory T cells, and M0/M1 macrophages to activate the immune response. It had been observed that the proportion of immune checkpoint positive T cells in the tumor microenvironment in DLBCL was high, resulting in a poor prognosis (*Matias et al., 2020)*.

In order to screen 13 TMB-related immune genes associated with prognosis, 1133 samples were screened from 3 datasets in GEO database, and their expression profiles and clinical information were downloaded. Survival analysis results showed that CD1C expression was associated with prognosis. CD1C was eventually identified as a TMB-related immune gene for DLBCL prognosis, and its function was further explored. CD1C is as known as T-cell surface glycoprotein CD1C. This gene encodes a member of the CD1 family of transmembrane glycoproteins and is associated with β2-microglobulin (*Fairhurst, Wang, Sieling, Modlin, & Braun, 1998*). The CD1 family includes CD1a, CD1b, CD1c, CD1d, and CD1e (*Martin, Calabi, & Milstein, 1986*). As with thymic leukemia (TL), CD1 is expressed in cortical thymus and some lymphomas and resembles MHC Class Ⅰ antigens (*Calabi & Milstein, 1986*). Study on early lung adenocarcinomas with EGFR mutations has found that the main types of tumor-infiltrating T cells are depletion and regulatory T cells, which are associated with an increase in dendritic cells specifically expressing CD1C genes (*He et al., 2021*). Expression analysis of B-cell chronic lymphocytic leukemia showed that CD1 mediated immune deficiency, polarization of cytokine response, altered adhesion, increased intracellular protein delivery, and leukemia cell processing *(Zheng, Venkatapathy, Rao, & Harrington, 2002*). In renal cell carcinoma, CD1C+ dendritic cells predicted progression-free survival (*van Cruijsen et al., 2008*). In the peripheral blood of NSCLC, the percentage of CD1C+ dendritic cells was significantly lower than that of normal donors, suggesting that NSCLC cells may prevent the maturation of DC cells and thus avoid an effective immune response (*Tabarkiewicz, Rybojad, Jablonka, & Rolinski, 2008*). In hepatocellular carcinoma, the authors analyzed the expression of ILT4 in mDCs subsets in hepatocellular carcinoma microenvironment. The results showed that the percentage of CD1C+ decreased significantly in peripheral blood mononuclear cells (PBMC) of HCC patients compared with normal controls, suggesting that the increased ILT4+CD1C+ subsets in tumor tissues may play an important role in the immunosuppression of HCC patients (*L. Wang et al., 2019*). Study has also shown that CD1C+ restricted T cells exhibit potent anti-leukemia activity in mouse models, so this lipid antigen may represent a new target for immunotherapy of hematological malignancies (*Lepore et al., 2015*). At present, no studies have reported the biological behavior of CD1C or CD1C as a biomarker of immunotherapy in DLBCL, but in our study, we explored the relationship between CD1C and prognosis

of DLBCL through bioinformatics analysis, and the results showed that low expression of CD1C predicted better prognosis.

We further performed a series of in-depth analyses of CD1C, and the high amplification of CD1C in B cells and dendritic cells suggested that CD1C mutations inhibit the efficient mediating and maintenance of normal immune responses by antigen-presenting cells. The poor prognosis of patients with higher levels of dendritic cells supported this idea. Drug sensitivity analysis of target genes showed that CD1C was associated with multiple clinical drug sensitivities. Molecular studies had shown that transcription factors CREB1 and TOX drive tumor growth and metastasis and were associated with poor prognosis of DLBCL.

CONCLUSIONS

In conclusion, based on the co-analysis of TMB and immune invasion, our study identified the immune genes associated with prognosis in DLBCL mutations and explored the internal correlation between TMB and immune invasion. CD1C was recognized as a potential marker of DLBCL, which may provide new insights into immunotherapy of DLBCL. CD1C is also a gene which shows an association with the tumor mutational burden of diffuse large B cell lymphoma. This predicts poor survival. According to bioinformatic analysis, CD1C is involved in tumor-related signaling pathways and immune and metabolic processes. Thus, the study offers a novel target to investigate the underlying mechanism for diffuse large B cell lymphoma.

ACKNOWLEDGEMENTS

We gratefully acknowledge contributions from the public databases. All the data included in the study was from open databases. Data generated through this work could be requested from the corresponding author via reasonable request. We sincerely acknowledge the TCGA database (https://portal.gdc.cancer.gov/), GEO database (http://www.ncbi.nlm.nih.gov/geo/), ImmPort database (https://www.immport.org/shared/home), UCSC Xena (https://xena.ucsc.edu/), GETx database (https://www.gtexportal.org/home/) and so on for data collection.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This research was funded by 1·3·5 projects for disciplines of excellence–Clinical Research Incubation Project, West China Hospital, Sichuan University (No: 2019HXFH035).

Grant Disclosures

The following grant information was disclosed by the authors:
1·3·5 projects for disciplines of excellence–Clinical Research Incubation Project, West China Hospital, Sichuan University: 2019HXFH035.

Competing Interests

The authors declare that they have no competing interest.

Author Contributions

- Xiaoyu Xiang conceived and designed the experiments, performed the experiments, analyzed the dat, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.
- Li-Min Gao analyzed the dat, authored or reviewed drafts of the article, and approved the final draft.
- Yuehua Zhang analyzed the dat, and approved the final draft.
- Qiqi Zhu analyzed the dat, authored or reviewed drafts of the article, and approved the final draft.
- Sha Zhao analyzed the dat, authored or reviewed drafts of the article, and approved the final draft.

- Weiping Liu analyzed the dat, authored or reviewed drafts of the article, and approved the final draft.
- Yunxia Ye analyzed the dat, authored or reviewed drafts of the article, and approved the final draft.
- Yuan Tang conceived and designed the experiments, analyzed the dat, authored or reviewed drafts of the article, and approved the final draft.
- Wenyan Zhang conceived and designed the experiments, analyzed the dat, prepared figures and/or tables, authored or reviewed drafts of the article, and approved the final draft.

Supplementary Information

Supporting information for this article can be found online at:
FigureS1: Genecloud plot showed mutation information of genes in DLBCL;
Figure S2: KEGG pathway analysis revealed that these DEGs were involved in immune-related pathways;
Figure S3: Forest map of 13 TMB-related immune genes by univariate Cox analysis;
Figure S4: Comparison of CD1C mRNA expression between tumor and normal tissues by combined analysis of GETx and TCGA-DLBC database;
Table S1: The differences in clinical characteristics between low- and high-TMB groups were obtained from the TCGA cohort;
Table S2: Top GO items for differentially expressed genes;
Table S3: Top KEGG enrichment pathways for differentially expressed genes ordered by p-value (p-value < 0.05).

Data Availability

The following information was supplied regarding data availability:
The raw data is available at figshare: Xiaoyu Xiang (2022): Raw_data.zip. figshare. Dataset. https://figshare.com/articles/dataset/Raw_Data_zip/22140245.

REFERENCES

Berglund M, Enblad G, Flordal E, Lui WO, Backlin C, Thunberg U, Sundström C, Roos G, Allander SV, Erlanson M, Rosenquist R, Larsson C, & Lagercrantz S. (2002). Chromosomal imbalances in diffuse large B-cell lymphoma detected by comparative genomic hybridization. *Modern Pathology* **15(8):** 807-816 DOI: 10.1097/01.Mp.0000024375.04135.2b.

Birkbak NJ, Kochupurakkal B, Izarzugaza JM, Eklund AC, Li Y, Liu J, Szallasi Z, Matulonis UA, Richardson AL, Iglehart JD, & Wang ZC. (2013). Tumor mutation burden forecasts outcome in ovarian cancer with BRCA1 or BRCA2 mutations. *PLoS One* **8(11):** e80023 DOI: 10.1371/journal.pone.0080023.

Calabi F, & Milstein C. (1986). A novel family of human major histocompatibility complex-related genes not mapping to chromosome 6. *Nature* **323(6088):** 540-543 DOI: 10.1038/323540a0.

Camicia R, Winkler HC, & Hassa PO. (2015). Novel drug targets for personalized precision medicine in relapsed/refractory diffuse large B-cell lymphoma: a comprehensive review. *Mol Cancer* **14:** 207 DOI: 10.1186/s12943-015-0474-2.

Chaganti S, Illidge T, Barrington S, McKay P, Linton K, Cwynarski K, McMillan A, Davies A, Stern S, & Peggs K. (2016). Guidelines for the management of diffuse large B-cell lymphoma. *Br J Haematol* **174(1):** 43-56 DOI: 10.1111/bjh.14136.

Chalmers ZR, Connelly CF, Fabrizio D, Gay L, Ali SM, Ennis R, Schrock A, Campbell B, Shlien A, Chmielecki J, Huang F, He Y, Sun J, Tabori U, Kennedy M, Lieber DS, Roels S, White J, Otto GA, Ross JS, Garraway L, Miller VA, Stephens PJ, & Frampton GM. (2017). Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. *Genome Medicine* **9(1):** 34 DOI: 10.1186/s13073-017-0424-2.

Chan TA, Yarchoan M, Jaffee E, Swanton C, Quezada SA, Stenzinger A, & Peters S. (2019). Development of tumor mutation burden as an immunotherapy biomarker: utility for the oncology clinic. *Annals of Oncology* **30(1):** 44-56 DOI: https://doi.org/10.1093/annonc/mdy495.

Chen Y, Wang Y, Luo H, Meng X, Zhu W, Wang D, Zeng H, & Zhang H. (2020). The frequency and inter-relationship of PD-L1 expression and tumour mutational burden across multiple types of advanced solid tumours in China. *Experimental Hematology & Oncology* **9(1):** 17 DOI: 10.1186/s40164-020-00173-3.

Ciavarella S, Vegliante MC, Fabbri M, De Summa S, Melle F, Motta G, De Iuliis V, Opinto G, Enjuanes A, Rega S, Gulino A, Agostinelli C, Scattone A, Tommasi S, Mangia A, Mele F, Simone G, Zito AF, Ingravallo G, Vitolo U, Chiappella A, Tarella C,

Gianni AM, Rambaldi A, Zinzani PL, Casadei B, Derenzini E, Loseto G, Pileri A, Tabanelli V, Fiori S, Rivas-Delgado A, López-Guillermo A, Venesio T, Sapino A, Campo E, Tripodo C, Guarini A, & Pileri SA. (2019). Dissection of DLBCL microenvironment provides a gene expression-based predictor of survival applicable to formalin-fixed paraffin-embedded tissue. *Annals of Oncology* **30(12):** 2015 DOI: https://doi.org/10.1093/annonc/mdz386.

Coiffier B, Lepage E, Briere J, Herbrecht R, Tilly H, Bouabdallah R, Morel P, Van Den Neste E, Salles G, Gaulard P, Reyes F, Lederlin P, & Gisselbrecht C. (2002). CHOP chemotherapy plus rituximab compared with CHOP alone in elderly patients with diffuse large-B-cell lymphoma. *N Engl J Med* **346(4):** 235-242 DOI: 10.1056/NEJMoa011795.

Coiffier B, Thieblemont C, Van Den Neste E, Lepeu G, Plantier I, Castaigne S, Lefort S, Marit G, Macro M, Sebban C, Belhadj K, Bordessoule D, Fermé C, & Tilly H. (2010). Long-term outcome of patients in the LNH-98.5 trial, the first randomized study comparing rituximab-CHOP to standard CHOP chemotherapy in DLBCL patients: a study by the Groupe d'Etudes des Lymphomes de l'Adulte. *Blood* **116(12):** 2040-2045 DOI: 10.1182/blood-2010-03-276246 %J Blood.

Conlon KC, Lugli E, Welles HC, Rosenberg SA, Fojo AT, Morris JC, Fleisher TA, Dubois SP, Perera LP, Stewart DM, Goldman CK, Bryant BR, Decker JM, Chen J, Worthy TA, Figg WD, Sr., Peer CJ, Sneller MC, Lane HC, Yovandich JL, Creekmore SP, Roederer M, & Waldmann TA. (2015). Redistribution, hyperproliferation, activation of natural killer cells and CD8 T cells, and cytokine production during first-in-human clinical trial of recombinant human interleukin-15 in patients with cancer. *Journal of Clinical Oncology* **33(1):** 74-82 DOI: 10.1200/jco.2014.57.3329.

Crump M, Neelapu SS, Farooq U, Van Den Neste E, Kuruvilla J, Westin J, Link BK, Hay A, Cerhan JR, Zhu L, Boussetta S, Feng L, Maurer MJ, Navale L, Wiezorek J, Go WY, & Gisselbrecht C. (2017). Outcomes in refractory diffuse large B-cell lymphoma: results from the international SCHOLAR-1 study. *Blood* **130(16):** 1800-1808 DOI: 10.1182/blood-2017-03-769620.

Czuczman MS, Grillo-López AJ, White CA, Saleh M, Gordon L, LoBuglio AF, Jonas C, Klippenstein D, Dallaire B, & Varns C. (1999). Treatment of patients with low-grade B-cell lymphoma with the combination of chimeric anti-CD20 monoclonal antibody and CHOP chemotherapy. *Journal of Clinical Oncology* **17(1):** 268-276 DOI: 10.1200/jco.1999.17.1.268.

Efremova M, Finotello F, Rieder D, & Trajanoski Z. (2017). Neoantigens Generated by Individual Mutations and Their Role in Cancer Immunity and Immunotherapy. *Frontiers in Immunology* **8:** 1679 DOI: 10.3389/fimmu.2017.01679.

Fairhurst RM, Wang CX, Sieling PA, Modlin RL, & Braun J. (1998). CD1-restricted T cells and resistance to polysaccharide-encapsulated bacteria. *Immunology Today* **19(6):** 257-259 DOI: 10.1016/s0167-5699(97)01235-8.

Farkona S, Diamandis EP, & Blasutig IM. (2016). Cancer immunotherapy: the beginning of the end of cancer? *BMC Medicine* **14:** 73 DOI: 10.1186/s12916-016-0623-5.

Feng C, Song C, Liu Y, Qian F, Gao Y, Ning Z, Wang Q, Jiang Y, Li Y, Li M, Chen J, Zhang J, & Li C. (2020). KnockTF: a comprehensive human gene expression profile database with knockdown/knockout of transcription factors. *Nucleic Acids Research* **48(D1):** D93-d100 DOI: 10.1093/nar/gkz881.

Finn OJ. (2012). Immuno-oncology: understanding the function and dysfunction of the immune system in cancer. *Ann Oncol* **23 Suppl 8(Suppl 8):** viii6-9 DOI: 10.1093/annonc/mds256.

He D, Wang D, Lu P, Yang N, Xue Z, Zhu X, Zhang P, & Fan G. (2021). Single-cell RNA sequencing reveals heterogeneous tumor and immune cell populations in early-stage lung adenocarcinomas harboring EGFR mutations. *Oncogene* **40(2):** 355-368 DOI: 10.1038/s41388-020-01528-0.

Hung JH, Yang TH, Hu Z, Weng Z, & DeLisi C. (2012). Gene set enrichment analysis: performance evaluation and usage guidelines. *Brief Bioinform* **13(3):** 281-291 DOI: 10.1093/bib/bbr049.

Keane C, Gill D, Vari F, Cross D, Griffiths L, & Gandhi M. (2013). CD4+ Tumor infiltrating lymphocytes are prognostic and independent of R-IPI in patients with DLBCL receiving R-CHOP chemo-immunotherapy. *Am J Hematol* **88(4):** 273-276 DOI: https://doi.org/10.1002/ajh.23398.

Khalifa KA, Badawy HM, Radwan WM, Shehata MA, & Bassuoni MA. (2014). CD14+ HLA-DR low/− monocytes as indicator of disease aggressiveness in B-cell non-Hodgkin lymphoma. *International Journal of Laboratory Hematology* **36(6):** 650-655 DOI: https://doi.org/10.1111/ijlh.12203.

Khalil DN, Smith EL, Brentjens RJ, & Wolchok JD. (2016). The future of cancer treatment: immunomodulation, CARs and combination immunotherapy. *Nature Reviews: Clinical Oncology* **13(6):** 394 DOI: 10.1038/nrclinonc.2016.65.

Kotlov N, Bagaev A, Revuelta MV, Phillip JM, Cacciapuoti MT, Antysheva Z, Svekolkin V, Tikhonova E, Miheecheva N, Kuzkina N, Nos G, Tabbo F, Frenkel F, Ghione P, Tsiper M, Almog N, Fowler N, Melnick AM, Leonard JP, Inghirami G, & Cerchietti L. (2021). Clinical and Biological Subtypes of B-cell Lymphoma Revealed by Microenvironmental Signatures. *Cancer Discov* **11(6):** 1468-1489 DOI: 10.1158/2159-8290.Cd-20-0839.

Leivonen SK, Pollari M, Brück O, Pellinen T, Autio M, Karjalainen-Lindsberg ML, Mannisto S, Kellokumpu-Lehtinen PL, Kallioniemi O, Mustjoki S, & Leppä S. (2019). T-cell inflamed tumor microenvironment predicts favorable prognosis in primary testicular lymphoma. *Haematologica* **104(2):** 338-346 DOI: 10.3324/haematol.2018.200105.

Lenz G, & Staudt LM. (2010). Aggressive lymphomas. *N Engl J Med* **362(15):** 1417-1429 DOI: 10.1056/NEJMra0807082.

Lepore M, de Lalla C, Mori L, Dellabona P, De Libero G, & Casorati G. (2015). Targeting leukemia by CD1c-restricted T cells specific for a novel lipid antigen. *Oncoimmunology* **4(3):** e970463 DOI: 10.4161/21624011.2014.970463.

Li S, Young KH, & Medeiros LJ. (2018). Diffuse large B-cell lymphoma. *Pathology* **50(1):** 74-87 DOI: 10.1016/j.pathol.2017.09.006.

Li T, Fu J, Zeng Z, Cohen D, Li J, Chen Q, Li B, & Liu XS. (2020). TIMER2.0 for analysis of tumor-infiltrating immune cells. *Nucleic Acids Res* **48(W1):** W509-w514 DOI: 10.1093/nar/gkaa407.

Lossos IS, & Morgensztern D. (2006). Prognostic biomarkers in diffuse large B-cell lymphoma. *J Clin Oncol* **24(6):** 995-1007 DOI: 10.1200/jco.2005.02.4786.

Marchesi F, Cirillo M, Bianchi A, Gately M, Olimpieri OM, Cerchiara E, Renzi D, Micera A, Balzamino BO, Bonini S, Onetti Muda A, & Avvisati G. (2015). High density of CD68+/CD163+ tumour-associated macrophages (M2-TAM) at diagnosis is significantly correlated to unfavorable prognostic factors and to poor clinical outcomes in patients with diffuse large B-cell lymphoma. *Hematol Oncol* **33(2):** 110-112 DOI: https://doi.org/10.1002/hon.2142.

Martin LH, Calabi F, & Milstein C. (1986). Isolation of CD1 genes: a family of major histocompatibility complex-related differentiation antigens. *Proc Natl Acad Sci U S A* **83(23):** 9154-9158 DOI: 10.1073/pnas.83.23.9154.

Matias A, Suvi-Katri L, Oscar B, Satu M, Judit Mészáros J, Marja-Liisa K-L, Klaus B, Harald H, Teijo P, & Sirpa L. (2020). Immune cell constitution in the tumor microenvironment predicts the outcome in diffuse large B-cell lymphoma. *Haematologica* **106(3):** 718-729 DOI: 10.3324/haematol.2019.243626.

Morin RD, Mendez-Lago M, Mungall AJ, Goya R, Mungall KL, Corbett RD, Johnson NA, Severson TM, Chiu R, Field M, Jackman S, Krzywinski M, Scott DW, Trinh DL, Tamura-Wells J, Li S, Firme MR, Rogic S, Griffith M, Chan S, Yakovenko O, Meyer IM, Zhao EY, Smailus D, Moksa M, Chittaranjan S, Rimsza L, Brooks-Wilson A, Spinelli JJ, Ben-Neriah S, Meissner B, Woolcock B, Boyle M, McDonald H, Tam A, Zhao Y, Delaney A, Zeng T, Tse K, Butterfield Y, Birol I, Holt R, Schein J, Horsman DE, Moore R, Jones SJ, Connors JM, Hirst M, Gascoyne RD, & Marra MA. (2011). Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. *Nature* **476(7360):** 298-303 DOI: 10.1038/nature10351.

Morton LM, Wang SS, Devesa SS, Hartge P, Weisenburger DD, & Linet MS. (2006). Lymphoma incidence patterns by WHO subtype in the United States, 1992-2001. *Blood* **107(1):** 265-276 DOI: 10.1182/blood-2005-06-2508.

Nam SJ, Go H, Paik JH, Kim TM, Heo DS, Kim CW, & Jeon YK. (2014). An increase of M2 macrophages predicts poor prognosis in patients with diffuse large B-cell lymphoma treated with rituximab, cyclophosphamide, doxorubicin, vincristine and prednisone. *Leuk Lymphoma* **55(11):** 2466-2476 DOI: 10.3109/10428194.2013.879713.

Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, & Alizadeh AA. (2015). Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods* **12(5):** 453-457 DOI: 10.1038/nmeth.3337.

Nicholas NS, Apollonio B, & Ramsay AG. (2016). Tumor microenvironment (TME)-driven immune suppression in B cell malignancy. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1863(3):** 471-482 DOI: https://doi.org/10.1016/j.bbamcr.2015.11.003.

Reinhold WC, Sunshine M, Liu H, Varma S, Kohn KW, Morris J, Doroshow J, & Pommier Y. (2012). CellMiner: a web-based suite of genomic and pharmacologic tools to explore transcript and drug patterns in the NCI-60 cell line set. *Cancer Res* **72(14):** 3499-3511 DOI: 10.1158/0008-5472.Can-12-1370.

Rizvi H, Sanchez-Vega F, La K, Chatila W, Jonsson P, Halpenny D, Plodkowski A, Long N, Sauter JL, Rekhtman N, Hollmann T, Schalper KA, Gainor JF, Shen R, Ni A, Arbour KC, Merghoub T, Wolchok J, Snyder A, Chaft JE, Kris MG, Rudin CM, Socci ND, Berger MF, Taylor BS, Zehir A, Solit DB, Arcila ME, Ladanyi M, Riely GJ, Schultz N, & Hellmann MD. (2018). Molecular Determinants of Response to Anti-Programmed Cell Death (PD)-1 and Anti-Programmed Death-Ligand 1 (PD-L1) Blockade in Patients With Non-Small-Cell Lung Cancer Profiled With Targeted Next-Generation Sequencing. *Journal of Clinical Oncology* **36(7):** 633-641 DOI: 10.1200/jco.2017.75.3384.

Roschewski M, Staudt LM, & Wilson WH. (2014). Diffuse large B-cell lymphoma-treatment approaches in the molecular era. *Nature Reviews: Clinical Oncology* **11(1):** 12-23 DOI: 10.1038/nrclinonc.2013.197.

Samstein RM, Lee CH, Shoushtari AN, Hellmann MD, Shen R, Janjigian YY, Barron DA, Zehir A, Jordan EJ, Omuro A, Kaley TJ, Kendall SM, Motzer RJ, Hakimi AA, Voss MH, Russo P, Rosenberg J, Iyer G, Bochner BH, Bajorin DF, Al-Ahmadie HA, Chaft JE, Rudin CM, Riely GJ, Baxi S, Ho AL, Wong RJ, Pfister DG, Wolchok JD, Barker CA, Gutin PH, Brennan CW, Tabar V, Mellinghoff

IK, DeAngelis LM, Ariyan CE, Lee N, Tap WD, Gounder MM, D'Angelo SP, Saltz L, Stadler ZK, Scher HI, Baselga J, Razavi P, Klebanoff CA, Yaeger R, Segal NH, Ku GY, DeMatteo RP, Ladanyi M, Rizvi NA, Berger MF, Riaz N, Solit DB, Chan TA, & Morris LGT. (2019). Tumor mutational load predicts survival after immunotherapy across multiple cancer types. *Nat Genet* **51(2):** 202-206 DOI: 10.1038/s41588-018-0312-8.

Sari R, Idun F, Minna T, Heli V, Maria T, Olav Y, Marja-Liisa K-L, Jan D, Erlend S, Harald H, & Sirpa L. (2015). Prognostic influence of macrophages in patients with diffuse large B-cell lymphoma: a correlative study from a Nordic phase II trial. *Haematologica* **100(2):** 238-245 DOI: 10.3324/haematol.2014.113472.

Schmitz R, Wright GW, Huang DW, Johnson CA, Phelan JD, Wang JQ, Roulland S, Kasbekar M, Young RM, Shaffer AL, Hodson DJ, Xiao W, Yu X, Yang Y, Zhao H, Xu W, Liu X, Zhou B, Du W, Chan WC, Jaffe ES, Gascoyne RD, Connors JM, Campo E, Lopez-Guillermo A, Rosenwald A, Ott G, Delabie J, Rimsza LM, Tay Kuang Wei K, Zelenetz AD, Leonard JP, Bartlett NL, Tran B, Shetty J, Zhao Y, Soppet DR, Pittaluga S, Wilson WH, & Staudt LM. (2018). Genetics and Pathogenesis of Diffuse Large B-Cell Lymphoma. *N Engl J Med* **378(15):** 1396-1407 DOI: 10.1056/NEJMoa1801445.

Schumacher TN, & Schreiber RD. (2015). Neoantigens in cancer immunotherapy. **348(6230):** 69-74 DOI: doi:10.1126/science.aaa4971.

Sehn LH, & Gascoyne RD. (2015). Diffuse large B-cell lymphoma: optimizing outcome in the context of clinical and biologic heterogeneity. *Blood* **125(1):** 22-32 DOI: 10.1182/blood-2014-05-577189.

Shain KH, Dalton WS, & Tao J. (2015). The tumor microenvironment shapes hallmarks of mature B-cell malignancies. *Oncogene* **34(36):** 4673-4682 DOI: 10.1038/onc.2014.403.

Smith A, Crouch S, Lax S, Li J, Painter D, Howell D, Patmore R, Jack A, & Roman E. (2015). Lymphoma incidence, survival and prevalence 2004-2014: sub-type analyses from the UK's Haematological Malignancy Research Network. *British Journal of Cancer* **112(9):** 1575-1584 DOI: 10.1038/bjc.2015.94.

Solimando AG, Annese T, Tamma R, Ingravallo G, Maiorano E, Vacca A, Specchia G, & Ribatti D. (2020). New Insights into Diffuse Large B-Cell Lymphoma Pathobiology. *Cancers* **12(7):** 1869 DOI:

Steen CB, Luca BA, Esfahani MS, Azizi A, Sworder BJ, Nabet BY, Kurtz DM, Liu CL, Khameneh F, Advani RH, Natkunam Y, Myklebust JH, Diehn M, Gentles AJ, Newman AM, & Alizadeh AA. (2021). The landscape of tumor cell states and ecosystems in diffuse large B cell lymphoma. *Cancer Cell* **39(10):** 1422-1437.e1410 DOI: https://doi.org/10.1016/j.ccell.2021.08.011.

Tabarkiewicz J, Rybojad P, Jablonka A, & Rolinski J. (2008). CD1c+ and CD303+ dendritic cells in peripheral blood, lymph nodes and tumor tissue of patients with non-small cell lung cancer. *Oncol Rep* **19(1):** 237-243 DOI:

Teras LR, DeSantis CE, Cerhan JR, Morton LM, Jemal A, & Flowers CR. (2016). 2016 US lymphoid malignancy statistics by World Health Organization subtypes. *CA: A Cancer Journal for Clinicians* **66(6):** 443-459 DOI: 10.3322/caac.21357.

Thanarajasingam G, Thanarajasingam U, & Ansell SM. (2016). Immune checkpoint blockade in lymphoid malignancies. **283(12):** 2233-2244 DOI: https://doi.org/10.1111/febs.13668.

Topalian SL, Wolchok JD, Chan TA, Mellman I, Palucka K, Banchereau J, Rosenberg SA, & Dane Wittrup K. (2015). Immunotherapy: The path to win the war on cancer? *Cell* **161(2):** 185-186 DOI: 10.1016/j.cell.2015.03.045.

Tran E, Robbins PF, & Rosenberg SA. (2017). 'Final common pathway' of human cancer immunotherapy: targeting random somatic mutations. *Nat Immunol* **18(3):** 255-262 DOI: 10.1038/ni.3682.

van Cruijsen H, van der Veldt AA, Vroling L, Oosterhoff D, Broxterman HJ, Scheper RJ, Giaccone G, Haanen JB, van den Eertwegh AJ, Boven E, Hoekman K, & de Gruijl TD. (2008). Sunitinib-induced myeloid lineage redistribution in renal cell cancer patients: CD1c+ dendritic cell frequency predicts progression-free survival. *Clinical Cancer Research* **14(18):** 5884-5892 DOI: 10.1158/1078-0432.Ccr-08-0656.

Vose JM, Link BK, Grossbard ML, Czuczman M, Grillo-Lopez A, Gilman P, Lowe A, Kunkel LA, & Fisher RI. (2001). Phase II study of rituximab in combination with chop chemotherapy in patients with previously untreated, aggressive non-Hodgkin's lymphoma. *Journal of Clinical Oncology* **19(2):** 389-397 DOI: 10.1200/jco.2001.19.2.389.

Wada N, Zaki MAA, Hori Y, Hashimoto K, Tsukaguchi M, Tatsumi Y, Ishikawa J, Tominaga N, Sakoda H, Take H, Tsudo M, Kuwayama M, Morii E, & Aozasa K. (2012). Tumour-associated macrophages in diffuse large B-cell lymphoma: a study of the Osaka Lymphoma Study Group. *Histopathology* **60(2):** 313-319 DOI: https://doi.org/10.1111/j.1365-2559.2011.04096.x.

Wang L, Fan J, Ye W, Han J, Zhang Y, Zhao L, Duan J, Yin D, & Yi Y. (2019). The Expression of ILT4 in Myeloid Dendritic Cells in Patients with Hepatocellular Carcinoma. *Immunological Investigations* **48(7):** 704-718 DOI: 10.1080/08820139.2019.1571507.

Wang S, He Z, Wang X, Li H, & Liu XS. (2019). Antigen presentation and tumor immunogenicity in cancer immunotherapy response prediction. *Elife* **8**DOI: 10.7554/eLife.49020.

Wang X, & Li M. (2019). Correlate tumor mutation burden with immune signatures in human cancers. *BMC Immunology* **20(1):** 4 DOI: 10.1186/s12865-018-0285-5.

Zheng Z, Venkatapathy S, Rao G, & Harrington CA. (2002). Expression profiling of B cell chronic lymphocytic leukemia suggests deficient CD1-mediated immunity, polarized cytokine response, altered adhesion and increased intracellular protein transport and processing of leukemic cells. *Leukemia* **16(12):** 2429-2437 DOI: 10.1038/sj.leu.2402711.

Figure 1

The workflow of the study

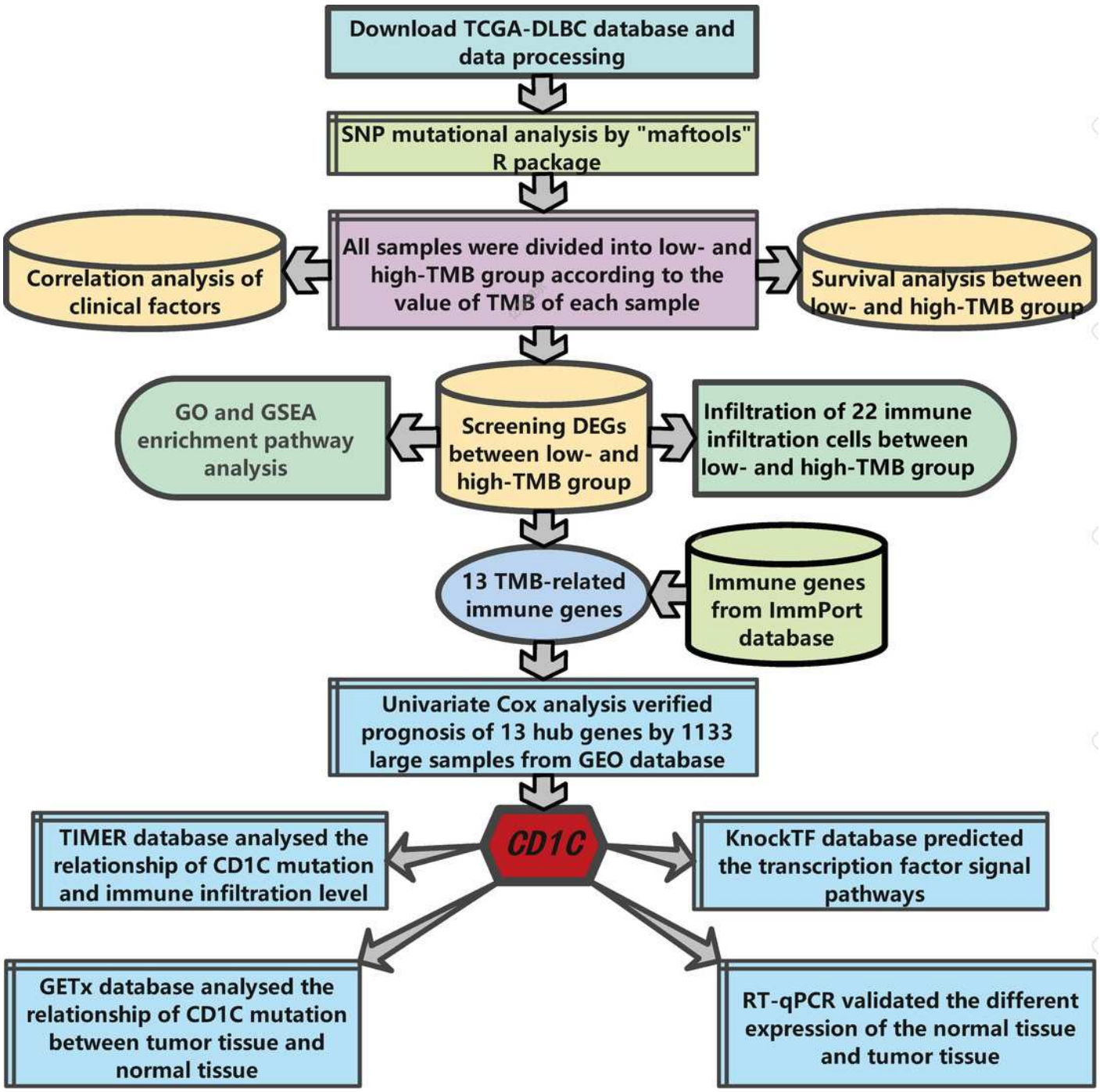

Figure 2

Summary of the mutation information with statistical.

(A-C) Classification of mutation types according to different categories, in which missense mutation accounts for the most fraction, SNP showed more frequency than insertion or deletion, and C>T was the most common of SNV. (D, E) Tumor mutation burden in specific samples. (F) The top 10 mutated genes in DLBCL. (G) The landscape of mutation profiles in DLBCL samples. Mutation information of each gene in each sample was shown in the waterfall plot, in which various colors with annotations at the bottom represented the different mutation types. The barplot above the legend exhibited the mutation burden. (H) The coincident and exclusive associations across mutated genes. (SNP, single nucleotide polymorphism; SNV, single nucleotide variants; DLBCL, diffuse large B cell lymphoma).

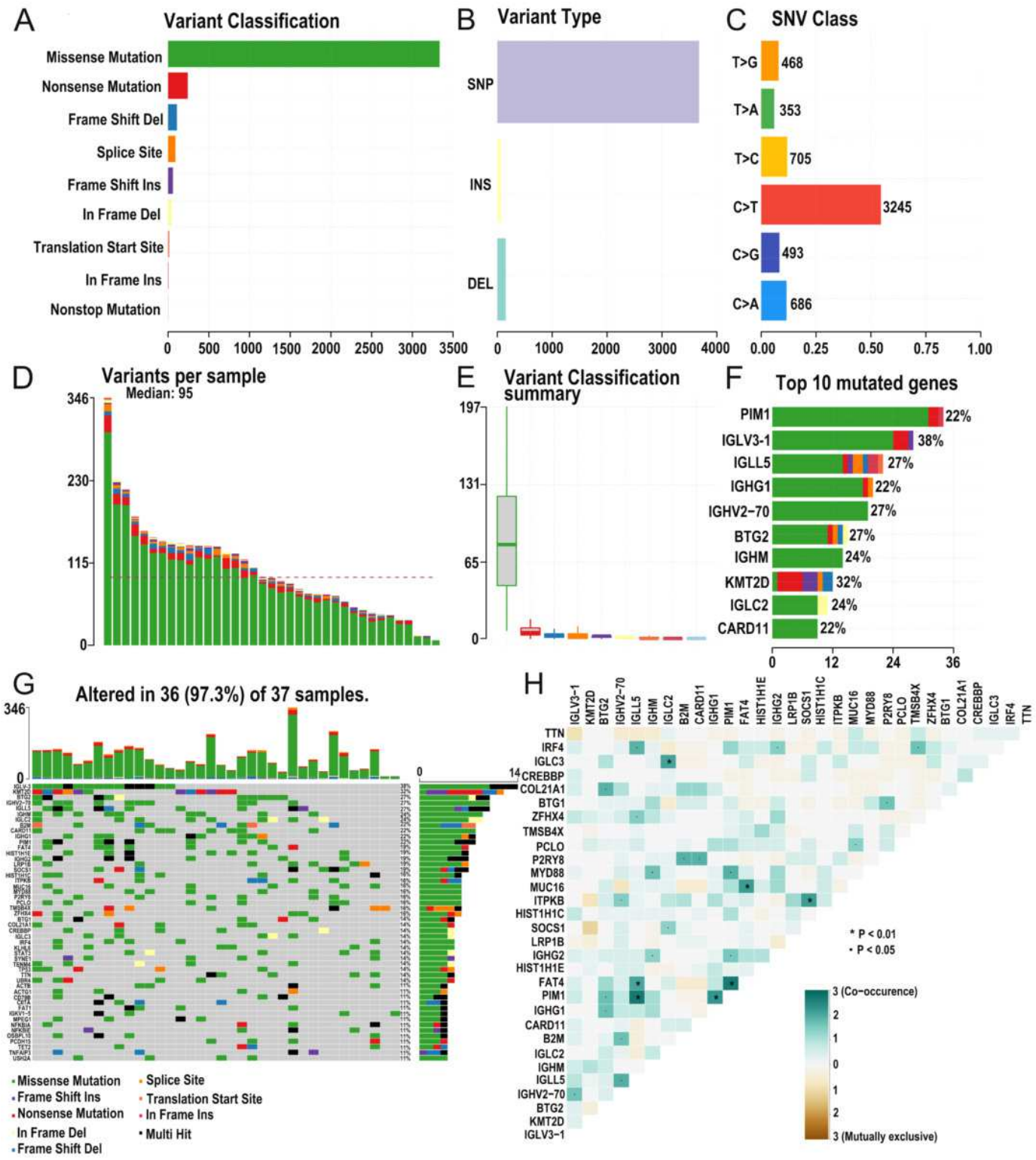
A
Variant Classification
Missense Mutation
Nonsense Mutation
Frame Shift Del
Splice Site
Frame Shift Ins
In Frame Del
Translation Start Site
In Frame Ins
Nonstop Mutation
0 500 1000 1500 2000 2500 3000 3500
B
Variant Type
SNP
INS
DEL
0 1000 2000 3000 4000
C
SNV Class
T>G 468
T>A 353
T>C 705
C>T 3245
C>G 493
C>A 686
0.00 0.25 0.50 0.75 1.00
D
Variants per sample
Median: 95
346
230
115
0
E
Variant Classification summary
197
131
65
0
F
Top 10 mutated genes
PIM1 22%
IGLV3-1 38%
IGLL5 27%
IGHG1 22%
IGHV2-70 27%
BTG2 27%
IGHM 24%
KMT2D 32%
IGLC2 24%
CARD11 22%
0 12 24 36
G
Altered in 36 (97.3%) of 37 samples.
346
0
14
Missense Mutation
Frame Shift Ins
Nonsense Mutation
In Frame Del
Frame Shift Del
Splice Site
Translation Start Site
In Frame Ins
Multi Hit
H
IGLV3-1
KMT2D
BTG2
IGHV2-70
IGLL5
IGHM
IGLC2
B2M
CARD11
IGHG1
PIM1
FAT4
HIST1H1E
IGHG2
LRP1B
SOCS1
HIST1H1C
ITPKB
MUC16
MYD88
P2RY8
PCLO
TMSB4X
ZFHX4
BTG1
COL21A1
CREBBP
IGLC3
IRF4
TTN
* P < 0.01
• P < 0.05
3 (Co-occurence)
2
1
0
1
2
3 (Mutually exclusive)

Figure 3

Distribution of TMB samples and prognosis of TMB.

(A) Distribution of TMB samples, those above the median value represented the samples with high mutation, and those below the median value represented the samples with low mutation. (B) Higher TMB levels correlated with poor survival outcomes with a p-value = 0.076.

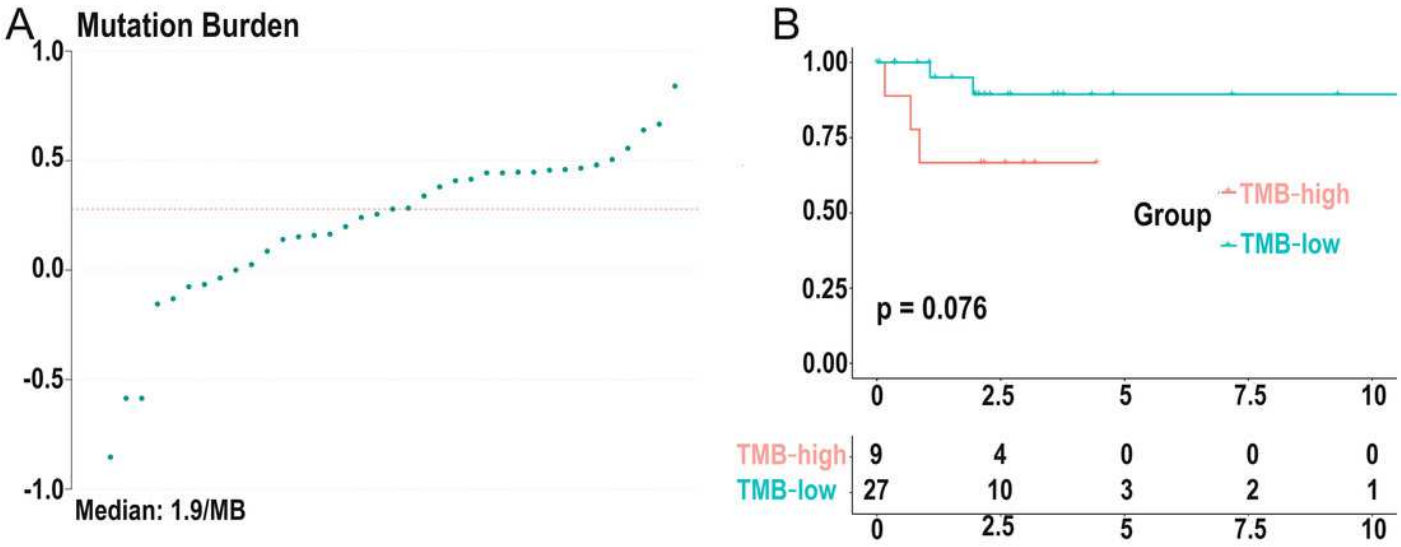

Figure 4

Comparisons of gene expression profiles in low- and high-TMB groups and enrichment pathway analysis.

(A) A total of 62 DEGs were shown in the heatmap plot. Vertical and horizontal axes represented genes and DLBCL samples respectively. Gene expression levels with higher and lower were displayed in red and blue, respectively. Color bars at the top of the heatmap represent sample types, with red and green indicating low- and high-TMB samples, respectively. (B) Volcano plot of all DEGs were drawn with | log2 (FC) | > 1 and p-value <0.05. Each symbol represented a gene, and red, grey and blue indicate upregulated, normal and downregulated genes, respectively. (C) GOplot revealed that these differentially expressed genes were involved in immune-related pathways. Different colors represented different GO terms, and the depth of gene color represented log2 (FC). (D) The DEGs enrichment analysis information (red colour represents the pathway for CD1C gene enrichment). (E, F) GSEA analysis shown that high-TMB related crosstalk, including one carbon pool by folate and rig-I like receptor signaling pathway-creative diagnostics. (G) GSEA analysis shown that low-TMB related crosstalk, including tight junction. (NES represented a normal enrichment score. ES represented enrichment score. DEGs, differentially expressed genes; TMB, tumor mutation burden; GO, gene ontology; GSEA, gene set enrichment analysis).

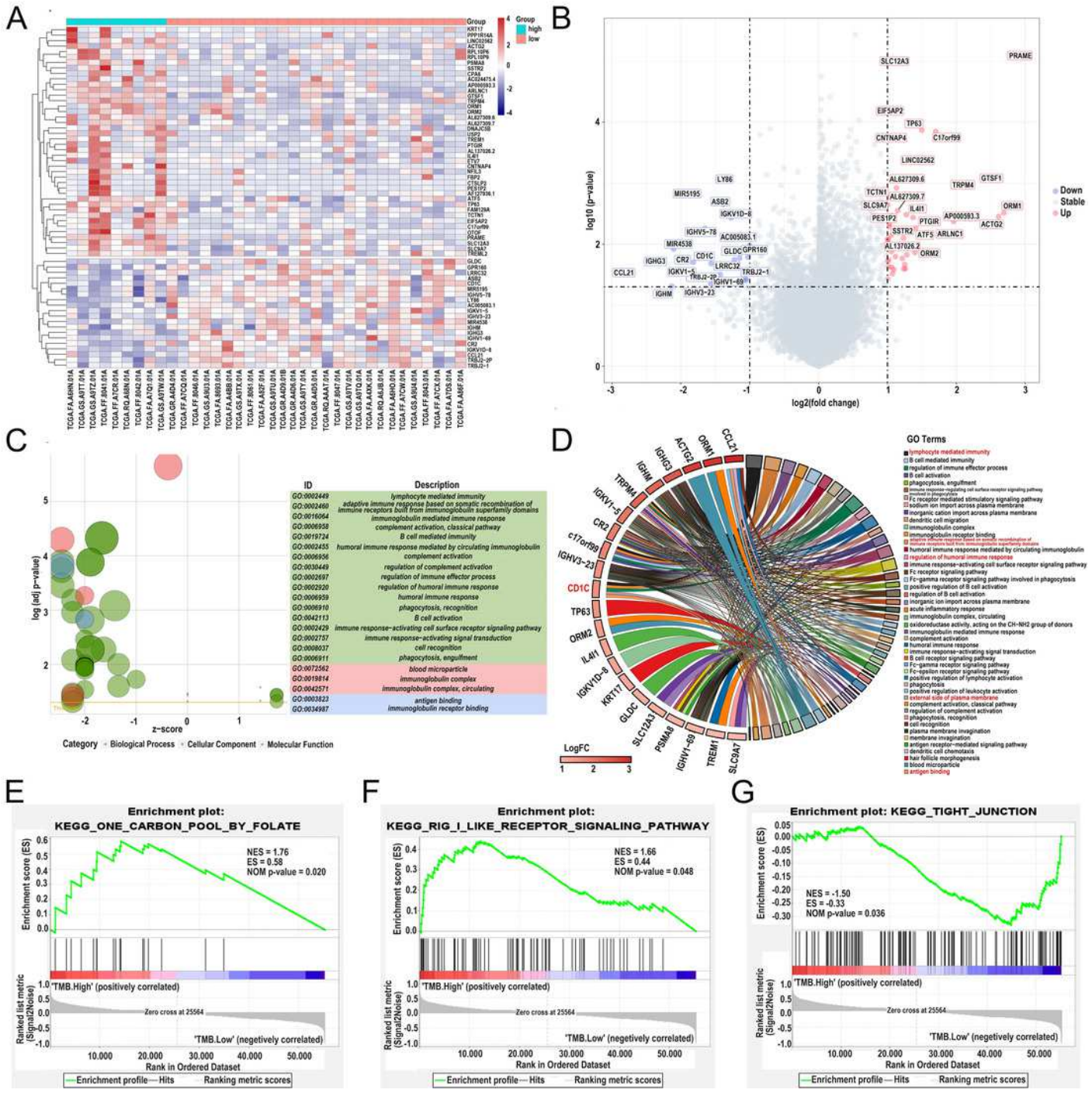
A
Group
high
low
B
log10 (p-value)
log2(fold change)
Down
Stable
Up
C
log (adj p-value)
z-score
ID
Description
GO:0002449 lymphocyte mediated immunity
GO:0002460 adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains
GO:0016064 immunoglobulin mediated immune response
GO:0006958 complement activation, classical pathway
GO:0019724 B cell mediated immunity
GO:0002455 humoral immune response mediated by circulating immunoglobulin
GO:0006956 complement activation
GO:0030449 regulation of complement activation
GO:0002697 regulation of immune effector process
GO:0002920 regulation of humoral immune response
GO:0006959 humoral immune response
GO:0006910 phagocytosis, recognition
GO:0042113 B cell activation
GO:0002429 immune response-activating cell surface receptor signaling pathway
GO:0002757 immune response-activating signal transduction
GO:0008037 cell recognition
GO:0006911 phagocytosis, engulfment
GO:0072562 blood microparticle
GO:0019814 immunoglobulin complex
GO:0042571 immunoglobulin complex, circulating
GO:0003823 antigen binding
GO:0034987 immunoglobulin receptor binding
Category Biological Process Cellular Component Molecular Function
D
GO Terms
LogFC
E
Enrichment plot:
KEGG_ONE_CARBON_POOL_BY_FOLATE
NES = 1.76
ES = 0.58
NOM p-value = 0.020
F
Enrichment plot:
KEGG_RIG_I_LIKE_RECEPTOR_SIGNALING_PATHWAY
NES = 1.66
ES = 0.44
NOM p-value = 0.048
G
Enrichment plot: KEGG_TIGHT_JUNCTION
NES = -1.50
ES = -0.33
NOM p-value = 0.036
Enrichment score (ES)
Ranked list metric (Signal2Noise)
'TMB.High' (positively correlated)
Zero cross at 25564
'TMB.Low' (negetively correlated)
Rank in Ordered Dataset
Enrichment profile
Hits
Ranking metric scores

Figure 5

Relationship between TMB and immune infiltration.

(A) The stacked bar chart showed the distribution of 22 types of immune cells in each sample, the horizontal axis represented the sample name, and the vertical axis represented the proportion of 22 types of immune cells. (B) The boxplot was arranged according to the content of immune cells in all DLBCL samples, among which B cells naïve accounted for the largest proportion. (C) The difference analysis of the heatmap showed the distribution of immune cells in the low- and high-TMB samples. (D) The boxplot showed differentially infiltrated immune cells between low- and high-TMB groups, with green, represented the high-TMB group and red represented the low-TMB group.

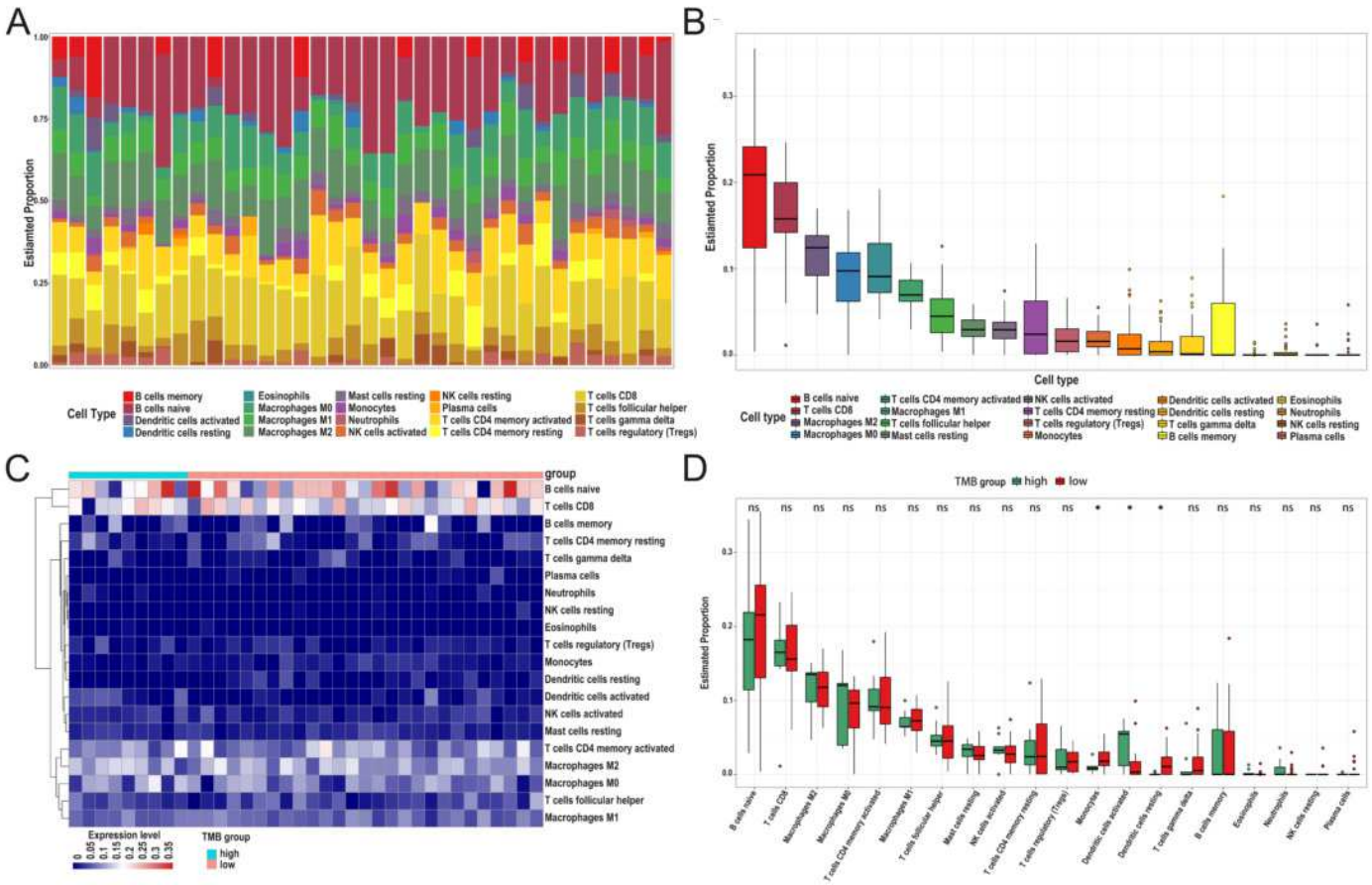

Figure 6

Identification of important TMB-related immune genes for DLBCL prognosis.

(A) Venn diagram showed that a total of 13 differential immune genes were associated with tumor mutation burden and immune infiltration. Kaplan-Meier survival analysis showed a relationship between the expression of CD1C, CCL21, ORM1, CR2, SSTR2 and the prognosis, suggesting that downregulation of CD1C was associated with better survival outcomes. (B) CD1C (p-value = 0.0012) (C) CCL21 (p-value = 0.0862) (D) ORM1 (p-value = 0.1745) (E) CR2 (p-value = 0.3151) (F) SSTR2 (p-value = 0.0715).

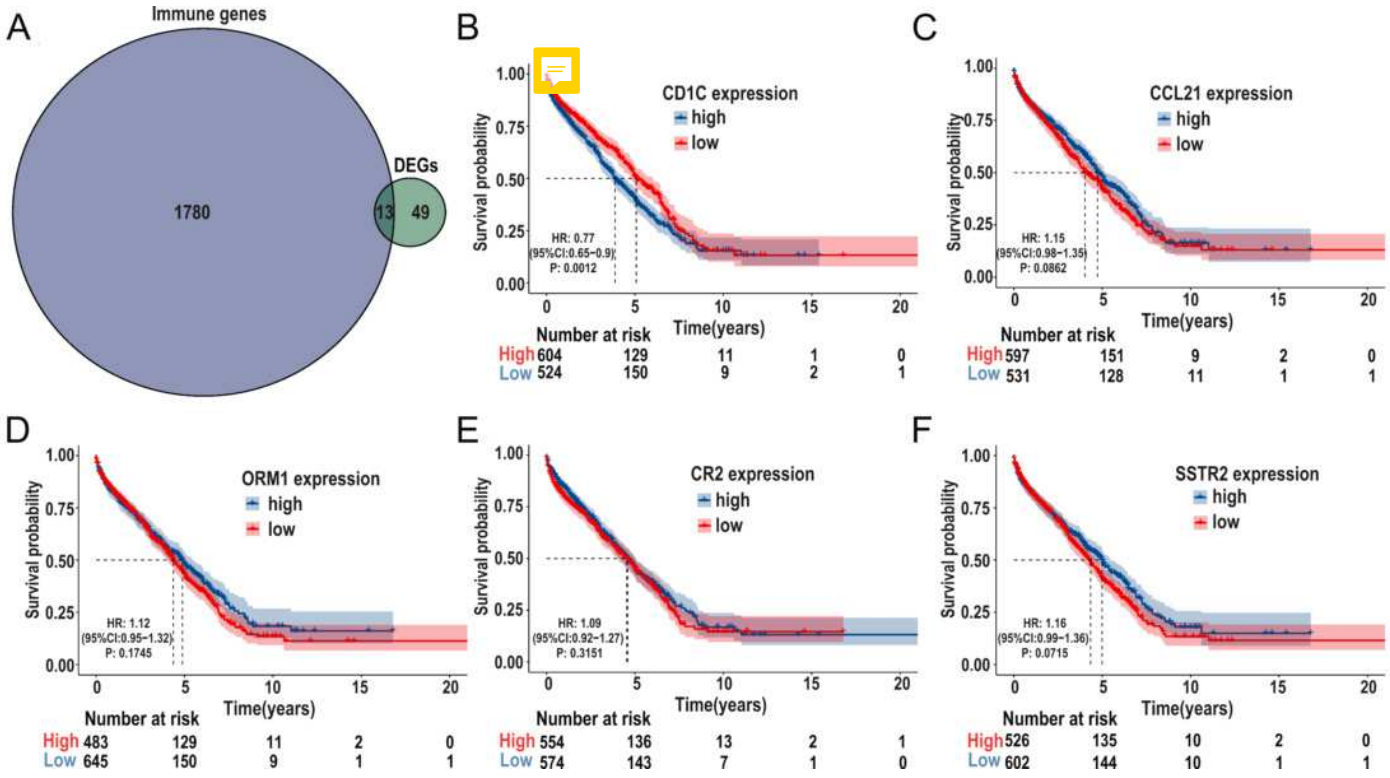

Figure 7

Correlations between the CNV of CCD[illegible] immune cells infiltration, and prognosis using TIMER database.

(A) CD1C expression between normal and tumor tissues in various cancers. (B) The expression of CD1C was correlated with 6 types of immune infiltrating cells, of which the correlation with B cells was the highest (cor = 0.693, p = 1.44e−03). (C) High amplification of CD1C in B cells and dendritic cells ($p<0.01$). (D) High levels of CD8+ T cells and dendritic cells indicated a good prognosis.

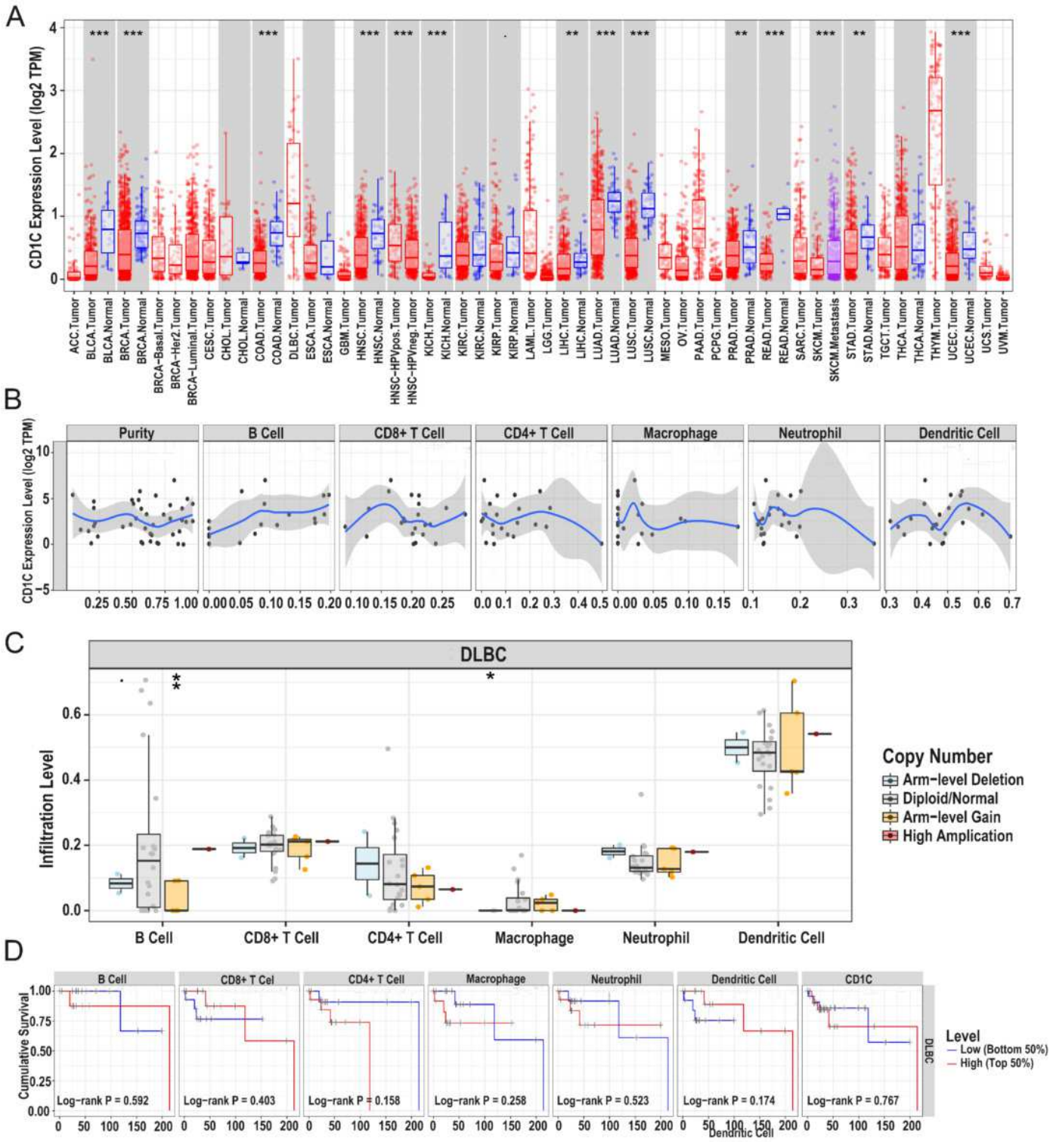
A
CD1C Expression Level (log2 TPM)
ACC.Tumor
BLCA.Tumor
BLCA.Normal
BRCA.Tumor
BRCA.Normal
BRCA−Basal.Tumor
BRCA−Her2.Tumor
BRCA−Luminal.Tumor
CESC.Tumor
CHOL.Tumor
CHOL.Normal
COAD.Tumor
COAD.Normal
DLBC.Tumor
ESCA.Tumor
ESCA.Normal
GBM.Tumor
HNSC.Tumor
HNSC.Normal
HNSC−HPVpos.Tumor
HNSC−HPVneg.Tumor
KICH.Tumor
KICH.Normal
KIRC.Tumor
KIRC.Normal
KIRP.Tumor
KIRP.Normal
LAML.Tumor
LGG.Tumor
LIHC.Tumor
LIHC.Normal
LUAD.Tumor
LUAD.Normal
LUSC.Tumor
LUSC.Normal
MESO.Tumor
OV.Tumor
PAAD.Tumor
PCPG.Tumor
PRAD.Tumor
PRAD.Normal
READ.Tumor
READ.Normal
SARC.Tumor
SKCM.Tumor
SKCM.Metastasis
STAD.Tumor
STAD.Normal
TGCT.Tumor
THCA.Tumor
THCA.Normal
THYM.Tumor
UCEC.Tumor
UCEC.Normal
UCS.Tumor
UVM.Tumor
B
CD1C Expression Level (log2 TPM)
Purity
B Cell
CD8+ T Cell
CD4+ T Cell
Macrophage
Neutrophil
Dendritic Cell
C
DLBC
Infiltration Level
B Cell
CD8+ T Cell
CD4+ T Cell
Macrophage
Neutrophil
Dendritic Cell
Copy Number
Arm−level Deletion
Diploid/Normal
Arm−level Gain
High Amplication
D
Cumulative Survival
B Cell
CD8+ T Cel
CD4+ T Cell
Macrophage
Neutrophil
Dendritic Cell
CD1C
Log−rank P = 0.592
Log−rank P = 0.403
Log−rank P = 0.158
Log−rank P = 0.258
Log−rank P = 0.523
Log−rank P = 0.174
Log−rank P = 0.767
Level
Low (Bottom 50%)
High (Top 50%)
DLBC

Figure 8

Analysis of the relationship between TMB target gene and drug sensitivity.

(A) Scatter plot of the top 20 clinical drug sensitivity and CD1C expression based on | cor | value. (B) The lollipop plot also showed the relationship between CD1C expression and drug sensitivity, with the p-value indicating signiûcance and cor indicating correlation.

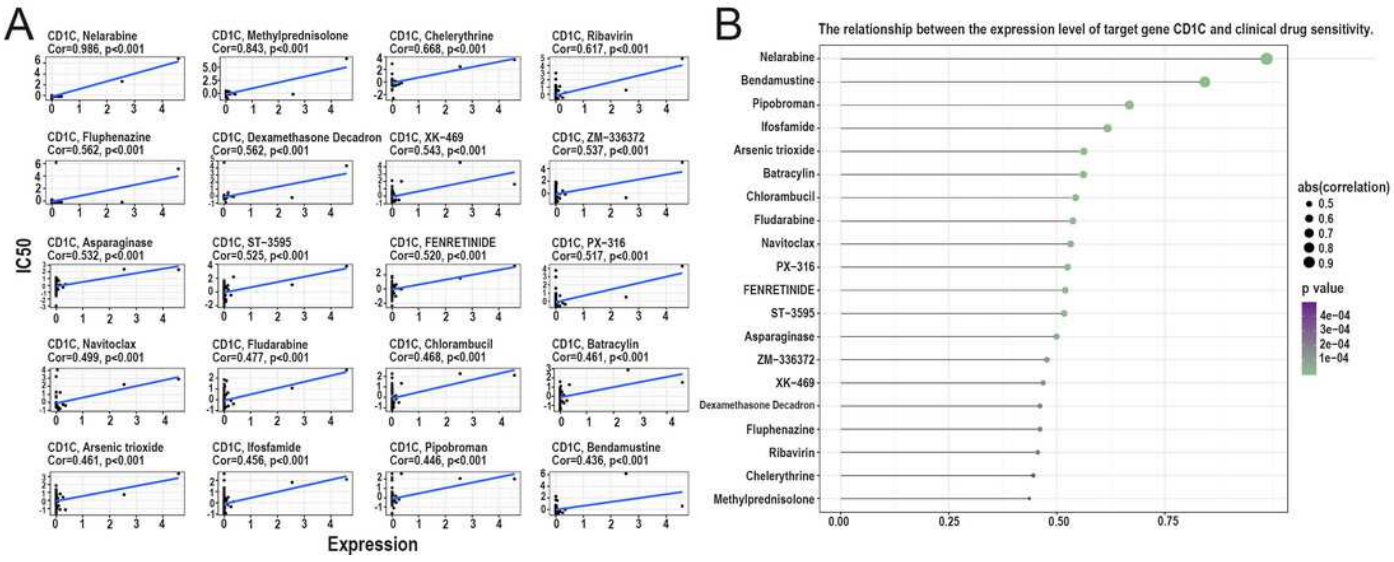

Figure 9

Validation of mRNA expression of the final key gene CD1C related to the degree of in FFPE samples (p-value = 0.019458).

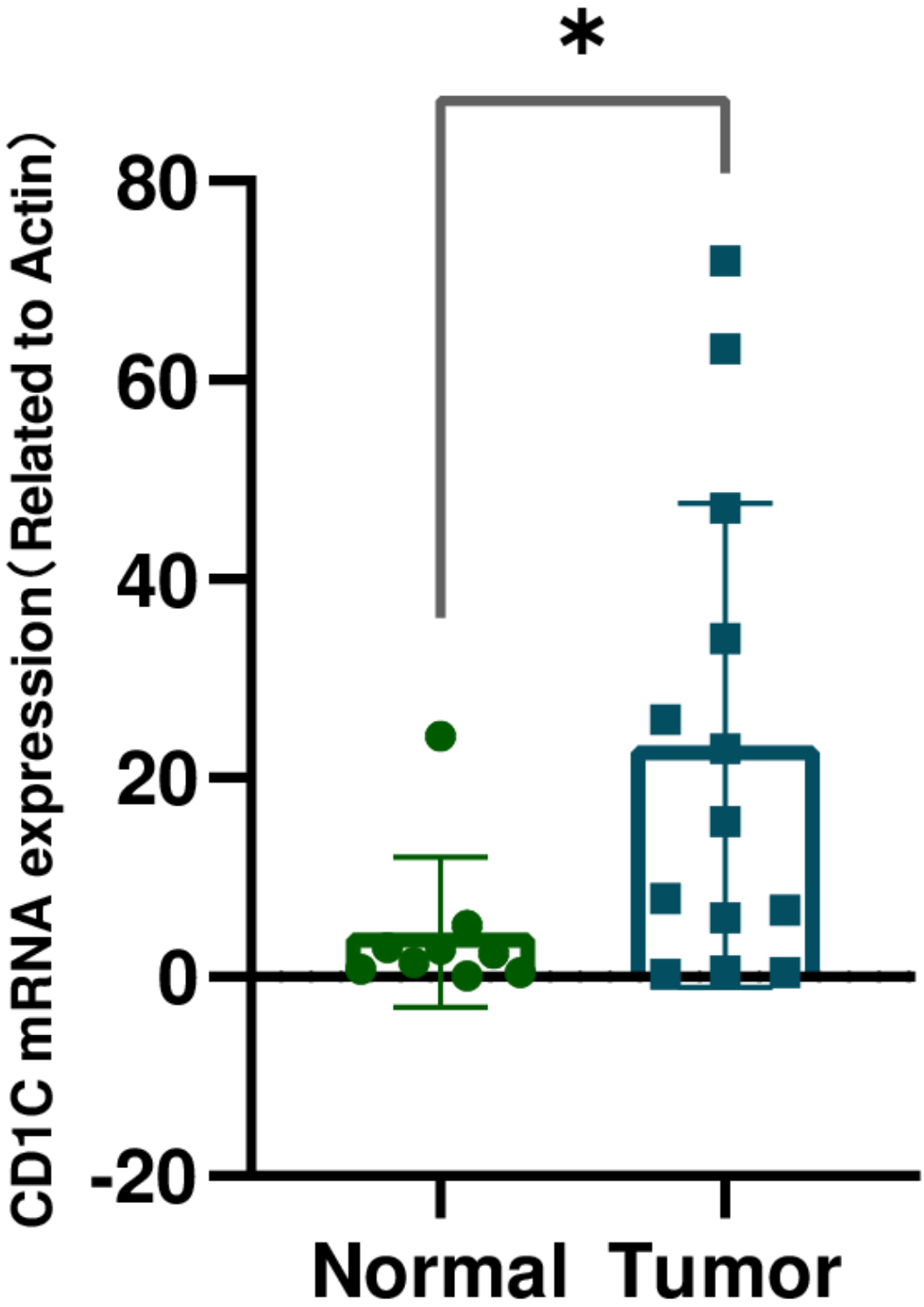
*
CD1C mRNA expression（Related to Actin）
80
60
40
20
0
-20
Normal
Tumor

Table 1(on next page)

Differentially expressed genes between low-TMB and high-TMB groups.

Table 1. Differentially expressed genes between low-TMB and high-TMB group.

Genesymbol	logFC	AveExpr	t	P.Value	adj.P.Val	B
PRAME	2.922405	1.672344	5.237783	8.14E-06	0.184374	2.785421
SLC12A3	1.100708	0.871668	5.168989	1.00E-05	0.184374	2.580099
EIF5AP2	1.037704	0.7309014	4.540696	6.55E-05	0.446506	0.725776
TP63	1.498609	1.72559	4.29477	0.000135	0.446506	0.015043
C17orf99	1.701547	1.3456096	4.276447	0.000143	0.446506	-0.03746
CNTNAP4	1.047761	0.3234867	4.189251	0.000184	0.446506	-0.28634
LINC02562	1.440104	0.682375	3.902082	0.000421	0.597317	-1.09357
LY86	-1.36579	4.6202479	-3.65037	0.000859	0.628234	-1.78272
GTSF1	2.491803	3.3402724	3.647969	0.000865	0.628234	-1.78919
TRPM4	2.112655	2.5636372	3.589926	0.001018	0.685802	-1.94525
AL627309.6	1.130337	1.5450149	3.531998	0.001195	0.691416	-2.09986
MIR5195	-1.90974	4.7542273	-3.45277	0.001487	0.77525	-2.30939
TCTN1	1.056882	1.3681228	3.418234	0.001635	0.799445	-2.40003
ASB2	-1.38053	3.2815049	-3.30631	0.002216	0.84459	-2.69058
SLC9A7	1.001124	3.0139012	3.296243	0.002277	0.84459	-2.71648
AL627309.7	1.147779	1.3959171	3.215333	0.002829	0.902441	-2.92307
ORM1	2.689855	2.1889532	3.185368	0.003065	0.934341	-2.99888
IL4I1	1.275177	5.8378104	3.159182	0.003286	0.945784	-3.06482
PES1P2	1.059463	0.3808462	3.144121	0.003419	0.955589	-3.10261
ACTG2	2.614232	2.8996668	3.131267	0.003538	0.96048	-3.13478
PTGIR	1.373384	2.2671073	3.119858	0.003646	0.96048	-3.16327
IGKV1D-8	-1.26761	1.4899645	-3.11549	0.003688	0.96048	-3.17417
AP000593.3	1.96134	1.6328582	3.061174	0.004254	0.979536	-3.30891
SSTR2	1.027005	1.0412898	3.029081	0.004626	0.999998	-3.38788
AF127936.1	1.028164	0.7326819	2.986073	0.005172	0.999998	-3.49296
ATF5	1.412084	6.6579457	2.967457	0.005427	0.999998	-3.53817
IGHV5-78	-1.64874	4.4896813	-2.9616	0.00551	0.999998	-3.55236
ARLNC1	1.842147	2.1616004	2.883652	0.006729	0.999998	-3.7396
OTOF	1.054667	0.5352778	2.858965	0.007165	0.999998	-3.79827
AL137026.2	1.234373	0.6788708	2.818765	0.007932	0.999998	-3.89314
CTSLP2	1.010085	0.3901514	2.797078	0.008378	0.999998	-3.94397
NFIL3	1.002554	3.2248748	2.796697	0.008386	0.999998	-3.94486
AC005083.1	-1.00218	1.4038589	-2.70415	0.010562	0.999998	-4.15895
MIR4538	-2.1038	2.6705054	-2.65241	0.011997	0.999998	-4.2766
PSMA8	1.058776	0.8946082	2.611077	0.01327	0.999998	-4.3695
ORM2	1.39914	1.4257549	2.597839	0.013704	0.999998	-4.39905
DNAJC5B	1.282866	2.7580284	2.563015	0.014907	0.999998	-4.47631

AC024475.4	1.115043	0.6922118	2.532595	0.016037	0.999998	-4.54321
GPR160	-1.02655	2.3532824	-2.52493	0.016334	0.999998	-4.55998
GLDC	-1.14561	1.4819458	-2.50791	0.017011	0.999998	-4.5971
KRT17	1.215934	0.8071803	2.505569	0.017106	0.999998	-4.60218
FAM129A	1.01797	2.6609447	2.484965	0.017965	0.999998	-4.64685
LRRC32	-1.22325	3.030123	-2.48445	0.017987	0.999998	-4.64797
IGHG3	-2.28464	5.678394	-2.48286	0.018055	0.999998	-4.65141
ETV7	1.003919	2.1068744	2.45744	0.019174	0.999998	-4.70613
CR2	-1.8187	3.1337495	-2.44458	0.019763	0.999998	-4.73365
CD1C	-1.56179	2.9112972	-2.42331	0.020775	0.999998	-4.77898
TREML2	1.251118	2.1785378	2.409244	0.02147	0.999998	-4.80879
TREM1	1.013126	1.0170469	2.369286	0.02356	0.999998	-4.89281
USP2	1.059806	0.9148026	2.362492	0.023933	0.999998	-4.907
RPL10P6	1.261382	2.1649398	2.339933	0.025211	0.999998	-4.95389
RPL10P9	1.236246	3.0773338	2.337468	0.025354	0.999998	-4.95899
PPP1R14A	1.094098	1.1728365	2.324857	0.026099	0.999998	-4.98504
CCL21	-2.82779	5.2870774	-2.27481	0.029253	0.999998	-5.08742
FBP2	1.072932	0.6428013	2.256376	0.030498	0.999998	-5.1247
TRBJ2-2P	-1.42691	3.7776966	-2.24323	0.031415	0.999998	-5.15115
TRBJ2-1	-1.05611	2.6341454	-2.16798	0.037154	0.999998	-5.30037
IGKV1-5	-1.88826	3.856146	-2.16686	0.037245	0.999998	-5.30255
CPA6	1.006875	0.7805608	2.165899	0.037325	0.999998	-5.30444
IGHV1-69	-1.05659	1.4366253	-2.14157	0.039378	0.999998	-5.35181
IGHV3-23	-1.56581	3.7508495	-2.08555	0.044491	0.999998	-5.45934
IGHM	-2.13098	9.540321	-2.04243	0.048817	0.999998	-5.54061

2

Table 2(on next page)

Transcription factors regulating CD1C in the Haematopoietic and lymphoid tissue by knockTF database.

Table 2. Transcription factors regulating CD1C in the Haematopoietic and lymphoid tissue by knockTF database.

Target Gene	TF	Knock-Method	Tissue Type	Biosample Name	Fold Change	Log2FC
CD1C	CREB1	shRNA	Haematopoietic and lymphoid tissue	K562	0.43886	-1.18818
CD1C	AHR	siRNA	Haematopoietic and lymphoid tissue	THP-1	0.36694	-1.4464
CD1C	TOX	shRNA	Haematopoietic and lymphoid tissue	CCRF-CEM	0.32445	-1.62392