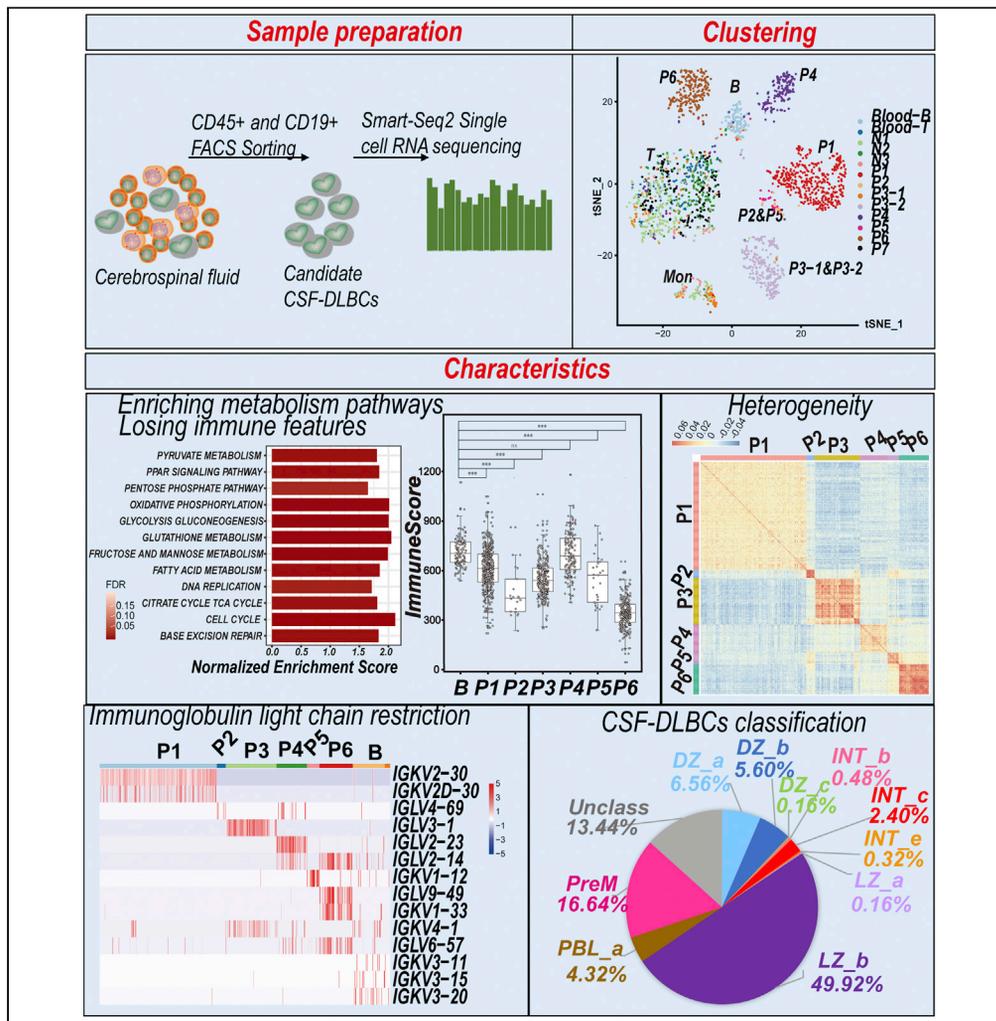


Article

Single-cell transcriptome analysis of diffuse large B cells in cerebrospinal fluid of central nervous system lymphoma



Haoyu Ruan, Zhe Wang, Yue Zhai, ..., Xuan Deng, Chao Zhang, Ming Guan

zhangchao@shsmu.edu.cn (C.Z.)
guanming88@yahoo.com (M.G.)

Highlights

We first investigate single-cell transcriptome characteristics of CSF-DLBCs

CSF-DLBCs exhibited active cell proliferation and energy metabolism properties

CSF-DLBCs had cell cycle state and cancer-testis antigen expression heterogeneity

CSF-DLBCs heterogeneity was also shown on classification by sc-GC B-cell signature



Article

Single-cell transcriptome analysis of diffuse large B cells in cerebrospinal fluid of central nervous system lymphoma

Haoyu Ruan,^{1,8,9} Zhe Wang,^{2,9} Yue Zhai,⁷ Ying Xu,⁷ Linyu Pi,⁷ Jihong Zheng,⁷ Yihang Zhou,⁷ Cong Zhang,² Ruofan Huang,³ Kun Chen,⁴ Xiangyu Li,⁴ Weizhe Ma,⁵ Zhiyuan Wu,⁴ Jie Shen,⁶ Xuan Deng,¹ Chao Zhang,^{2,*} and Ming Guan^{1,10,*}

SUMMARY

Diffuse large B cells in the cerebrospinal fluid (CSF-DLBCs) have offered great promise for the diagnostics and therapeutics of central nervous system lymphoma (CNSL) leptomeningeal involvement. To explore the phenotypic states of CSF-DLBCs, we analyzed the transcriptomes of more than one thousand CSF-DLBCs from six patients with CNSL diffuse large B-cell lymphoma (DLBCL) using Smart-seq2 single-cell RNA sequencing. CSF-DLBCs were defined based on abundant expression of B-cell markers, the active cell proliferation and energy metabolism properties, and immunoglobulin light chain restriction. We identified inherent heterogeneity of CSF-DLBCs, which were mainly manifested in cell cycle state, cancer-testis antigen expression, and classification based on single-cell germinal center B-cell signature. In addition, the 16 upregulated genes in CSF-DLBCs compared to various normal B cells showed great possibility in the homing effect of the CNS-DLBCL for the leptomeninges. Our results will provide insight into the mechanism research and diagnostic direction of CNSL-DLBCL leptomeningeal involvement.

INTRODUCTION

Primary central nervous system lymphoma (PCNSL) is a rare and aggressive extranodal non-Hodgkin lymphoma (NHL), accounting for up to 1% of NHL and about 3% of all primary brain tumors (Villano et al., 2011). PCNSL is confined to the brain, eyes, spinal cord, or leptomeninges without systemic involvement and regarded as an “immune-privileged” lymphoma (Han and Batchelor, 2017). The majority of PCNSL cases (>95%) are diffuse large B-cell lymphoma (DLBCL) with expression of B-cell markers (Giannini et al., 2014). In contrast to PCNSL-DLBCL, systemic DLBCL at diagnosis or relapse involved both within and outside the central nervous system (CNS) is defined as secondary CNS lymphoma (Baraniskin et al., 2018). Due to the poor central nervous system penetration of drugs and the prolonged overall survival of patients, the incidence of CNS-DLBCL has been increasing in recent decades (Baraniskin et al., 2018).

For patients with suspected CNS lymphoma, a histopathologic diagnosis by stereotactic brain biopsy is the gold standard (Khatib et al., 2014). However, brain biopsy is an invasive method with a risk of complications, and decreased sensitivity of biopsies as a result of the administration of corticosteroids can delay the initiation of systemic therapy (Onder et al., 2015). Considerable recent improvements in neuroimaging techniques have provided the requisite sensitivity for diagnosis of central nervous system lymphoma (CNSL) and are able to define the site and extent of the disease, but neuroimaging findings are not specific (Nabavizadeh et al., 2016). In recent years, multimodal investigations of cerebrospinal fluid (CSF) have greatly facilitated the CNSL diagnosis. Positive cytopathological examination of the CSF is still regarded as the “gold standard” for a definitive diagnosis of CNSL leptomeningeal involvement (Baraniskin and Schroers, 2014). In addition, the advent of immunophenotypic, molecular genetic mutations and interleukins further expand the diagnostic value of CSF (Hiemcke-Jiwa et al., 2018; Sasagawa et al., 2015). CSF obtained through lumbar puncture has the advantage of minimal risk and can be sampled multiple times, which is beneficial to monitor the progress of CNSL. The treatment of CNSL has evolved from the use of whole-brain radiotherapy alone to multimodality regimens that include chemotherapy, monoclonal

¹Department of Clinical Laboratory, Huashan Hospital, Fudan University, Shanghai, China

²Department of Plastic and Reconstructive Surgery, Shanghai Institute of Precision Medicine, Shanghai Ninth People's Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China

³Department of Oncology, Huashan Hospital, Fudan University, Shanghai, China

⁴Department of Clinical Laboratory, Huashan Hospital North, Fudan University, Shanghai, China

⁵Central Laboratory, Huashan Hospital, Fudan University, Shanghai, China

⁶10K Genomics Technology Co., Ltd., 333 Guiping Road, Shanghai 200233, China

⁷Translational Medical Center for Stem Cell Therapy and Institute for Regenerative Medicine, Shanghai East Hospital, Shanghai Key Laboratory of Signaling and Disease Research, Frontier Science Center for Stem Cell Research, School of Life Sciences and Technology, Tongji University, Shanghai, China

⁸Department of Laboratory Medicine, the First Affiliated Hospital of Nanjing Medical University, Nanjing, China

⁹These authors contributed equally

¹⁰Lead contact

*Correspondence: zhangchao@shsmu.edu.cn (C.Z.), guanming88@yahoo.com (M.G.)

<https://doi.org/10.1016/j.isci.2021.102972>



Table 1. Summary of cell-type identity in scRNA-seq results of six patient CSF samples, three normal CSF samples, and blood-T/B samples.

Patient ID	Diagnostics	Number of sequenced cells	Number of QC ^d filtered cells	T Cells	B Cells	Monocytes	Diffuse large B cells
P1	SCNSL ^a -DLBCL	519	425	1	1	1	422
P2	PCNSL ^b -DLBCL ^c	36	32	0	1	9	22
P3-1	PCNSL-DLBCL	348	118	61	6	48	3
P3-2		384	252	13	0	4	235
P4	PCNSL-DLBCL	480	197	36	6	0	155
P5	PCNSL-DLBCL	48	41	4	3	2	32
P6	SCNSL-DLBCL	336	228	1	0	3	224
P7	PCNSL-DLBCL	480	188	188	0	0	0
N1	Control	288	213	176	1	36	0
N2	Control	240	145	138	3	4	0
N3	Control	83	41	27	0	14	0
Blood T	Normal	168	113	113	0	0	0
Blood B	Normal	168	121	2	118	0	0
Total		3,591	2,114	760	139	121	1,093

^aSCNSL: secondary central nervous system lymphoma.

^bPCNSL: primary central nervous system lymphoma.

^cDLBCL: diffuse large B-cell lymphoma.

^dQC: quality control.

antibodies, and autologous stem cell transplantation (Chukwueke and Nayak, 2019; Nayak and Batchelor, 2013). The development of new technologies and approaches has improved the diagnosis and therapy of CNSL, but the overall prognosis for PCNSL remains relatively poor with a mean survival period of less than 5 years (Camilleri-Broet et al., 2006).

Until now, a comprehensive understanding of CNSL-DLBCL mechanisms is still lacking. To make breakthroughs in tackling the clinical challenge of CNSL-DLBCL, we analyzed the transcriptome characteristics of diffuse large B cells (DLBCs) in CSF (CSF-DLBCs) from six patients with CNSL-DLBCL. In order to avoid the disadvantages of RNA sequencing in bulk, we obtained the transcriptional profiles of CSF-DLBCs by Smart-seq2 single-cell RNA sequencing (scRNA-seq), which could capture the heterogeneity of DLBCs. Our study is the first one to investigate systematic and comprehensive characterization of more than one thousand CSF-DLBCs at the single-cell transcriptome level.

RESULTS

Profiles of individual cells in CSF

In this study, we enrolled patient CSF samples from seven patients with CNSL-DLBCL (P1-P7) and performed scRNA-seq on 2,631 target cells (Table 1). In addition, some normal CSF cells and blood T/B cells were also included for data analysis (BioprojectID PRJNA602172). According to the selection criteria, 2,114 cells with high-quality transcriptome data were retained for subsequent analysis, including 1,481 cells from CSF samples of patients with CNSL (Table 1). Cell-cycle-based correction of CSF cell data was clustered by t-distributed stochastic neighbor embedding (t-SNE) (van der Maaten and Hinton, 2008). On the basis of their preferential or distinctive marker gene expression (Figures 1A–1C and S2), three clusters of non-malignant cells were annotated as T cells (760 cells), B cells (139 cells), or monocytes (121 cells; Table 1). The majority of patient CSF cells strongly clustered according to the patient of origin, in addition to some normal leukocytes (Figure 1A). At the molecular level, we defined 1,093 CSF-DLBCs (422 cells from P1, 22 from P2, 238 from P3, 155 from P4, 32 from P5, 224 from P6; Table 1) with transcriptome signatures for B-cell markers and proliferation genes (Holmes et al., 2020) associated with the S-G2-M stages of the cell cycle (Figures 1B, 1C, and S2), especially *MKI67*, a classical proliferation marker commonly used in immunohistochemistry of clinical examinations. Patient (P) P3-1 and P3-2 samples were collected from the same patient within a two-month time interval. The proportion of DLBCs in the P3-1 CSF sample was 2% by cytopathological analysis, whereas in the P3-2 CSF sample, it was 60%, indicating tumor progression

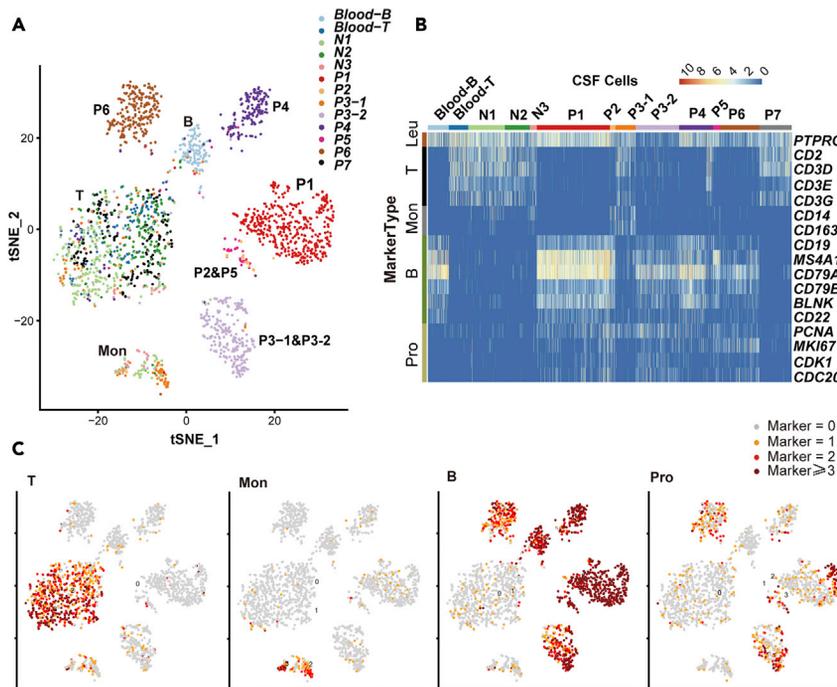


Figure 1. Clustering and analysis of single-cell expression data of CSF (cerebrospinal fluid) samples

(A) 2D representation of sample correlations by t-SNE dimensionality reduction including eight patient CSF samples (P), three normal CSF samples (N), blood T and B cells.

(B) Heatmap showing expression of selected gene panels in different samples.

(C) Feature plots demonstrating the expression of selected gene panels on the t-SNE plot (Figure 1A). Scaled expression levels are depicted by the number of expressing markers. No marker expression, gray; one marker expression, orange; two markers expression, red; more than two markers expression, dark red. Leukocyte (Leu) marker gene: *PTPRC*; T-cell (T) marker genes: *CD2*, *CD3D/E/G*; monocyte (Mon) marker genes: *CD14*, *CD68*, *CD163*; B-cell (B) marker genes: *CD19*, *MS4A1*, *CD79A*, *CD79B*, *BLNK*, *CD22*; proliferation (Pro)-associated genes: *PCNA*, *MKI67*, *CDK1*, and *CDC20*. See also Figures S1 and S2.

during the two months. The P3-1 sample had three DLBCs (the rest are leukocytes) in the P3-2 DLBC cluster though they had undergone independent cell sorting, library construction, and sequencing (Figure 1A). There was no significant heterogeneity in mapping quality and gene coverage across patient samples. The clustering pattern was not driven by technical variability and batch effect.

Transcriptome signatures of diffuse large B cells in CSF

We performed gene set enrichment analysis (GSEA) to further determine the functional enrichment in CSF-DLBCs compared to normal B cells in B-cell cluster (Figure 1A). We discovered that cell proliferation category was significantly enriched in CSF-DLBCs from six patients consisting of cell cycle and DNA replication and repair pathway (p value < 0.05; Figures 2A and S3). The metabolism category significantly enriched in CSF-DLBCs from P1, P2, and P3 (p value < 0.05; Figures 2A and S3) contained pentose phosphate, glycolysis gluconeogenesis, pyruvate metabolism, tricarboxylic acid (TCA) cycle, glutathione metabolism, fatty acid metabolism, fructose and mannose metabolism, and PPAR (proliferator-activated receptor) signaling pathway. These energy metabolism pathways are critical for tumor growth and the energy demand in the brain. The enriched energy metabolism pathways of patient P4, P5, and P6 CSF-DLBCs were not as obvious and significant as other patients (p value < 0.05; Figures 2A and S3). The observed downregulation of the antigen processing and presentation pathway and the B-cell receptor (BCR) signaling pathway in CSF-DLBCs (p value < 0.05; Figures 2A and S3) suggested their decreased capacity as immune cells. To further delineate immune characteristics of CSF-DLBCs, the ImmuneScore was computed based on ESTIMATE R package (Yoshihara et al., 2013). The result showed normal B cells had higher ImmuneScore than CSF-DLBCs, except for those from P4, which also indicated the loss of features related to their immune identity (Figure 2B; p value < 0.001, Kruskal-Wallis test).

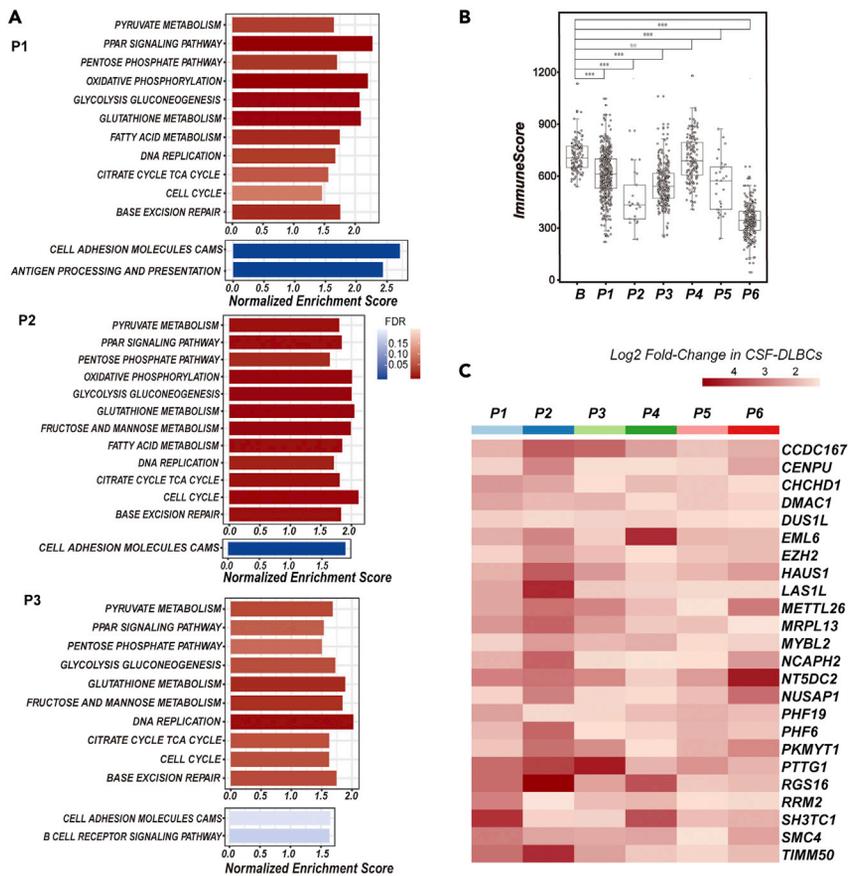


Figure 2. Characteristics of DLBCs (diffuse large B cells) in CSF (cerebrospinal fluid) using single-cell transcriptome analysis

(A) GSEA showing significantly upregulated (red gradient) or downregulated (blue gradient) KEGG pathways in CSF-DLBCs of patient P1, P2, and P3 compared to normal B cells in B-cell cluster (Figure 1A), p value < 0.05 .

(B) The immune signature of cells quantified by the ImmuneScore computed from the ESTIMATE algorithm, showing the significant difference between the B-cell group and the individual patient CSF-DLBC group (** p value < 0.001 , Kruskal-Wallis test). Data are represented as mean \pm SEM.

(C) Heatmap showing the 24 selected genes (Table S3) upregulated in CSF-DLBCs of all six patients compared to normal B cells.

See also Figure S3.

To address the fundamental differences in the expression program of the pure CSF-DLBC population, we employed the DESeq2 method to assess whether a gene is differentially expressed between normal B cells and CSF-DLBCs within each patient. It has been reported that B cells show no transcriptional changes between blood and CSF samples (Schafflick et al., 2020); therefore, we compared CSF-DLBCs to normal B cells of B-cell cluster (Figure 1A). Overall, 3516 genes were identified as differentially expressed genes (DEGs) in at least 1 sample, and 167 DEGs were identified in all six patients (adjusted p value < 0.05 ; $|\log_2(\text{-fold-change})| \geq 1$). In order to determine the genes specifically expressed in CSF-DLBCs, we selected out 45 genes upregulated in all six patients and the percentage of cells where the gene is detected in normal B cells was fewer than 5% (the percentage of cells where the gene is detected [PCT] $< 5\%$; Table S2). We also performed the DEG analysis between CSF-DLBCs and 30 normal CSF B cells from GEO databases (GSE138266) (Schafflick et al., 2020). Among the 45 genes, the 24 genes (adjusted p value < 0.05 ; $|\log_2(\text{-fold-change})| \geq 1$, PCT $< 5\%$; Table S3), namely, *CCDC167*, *CENPU*, *CHCHD1*, *DMAC1*, *DUS1L*, *EML6*, *EZH2*, *HAUS1*, *LAS1L*, *METTL26*, *MRPL13*, *MYBL2*, *NCAPH2*, *NT5DC2*, *NUSAP1*, *PHF19*, *PHF6*, *PKMYT1*, *PTTG1*, *RGS16*, *RRM2*, *SH3TC1*, *SMC4*, and *TIMM50* were also defined as DEGs upregulated in DLBCs compared to normal CSF B cells (Figure 2C). *EZH2* (enhancer of zeste homolog 2) functions as a transcriptional repressor by methylating *H3K27* (histone H3 at lysine 27), whose mutations and over-expression have

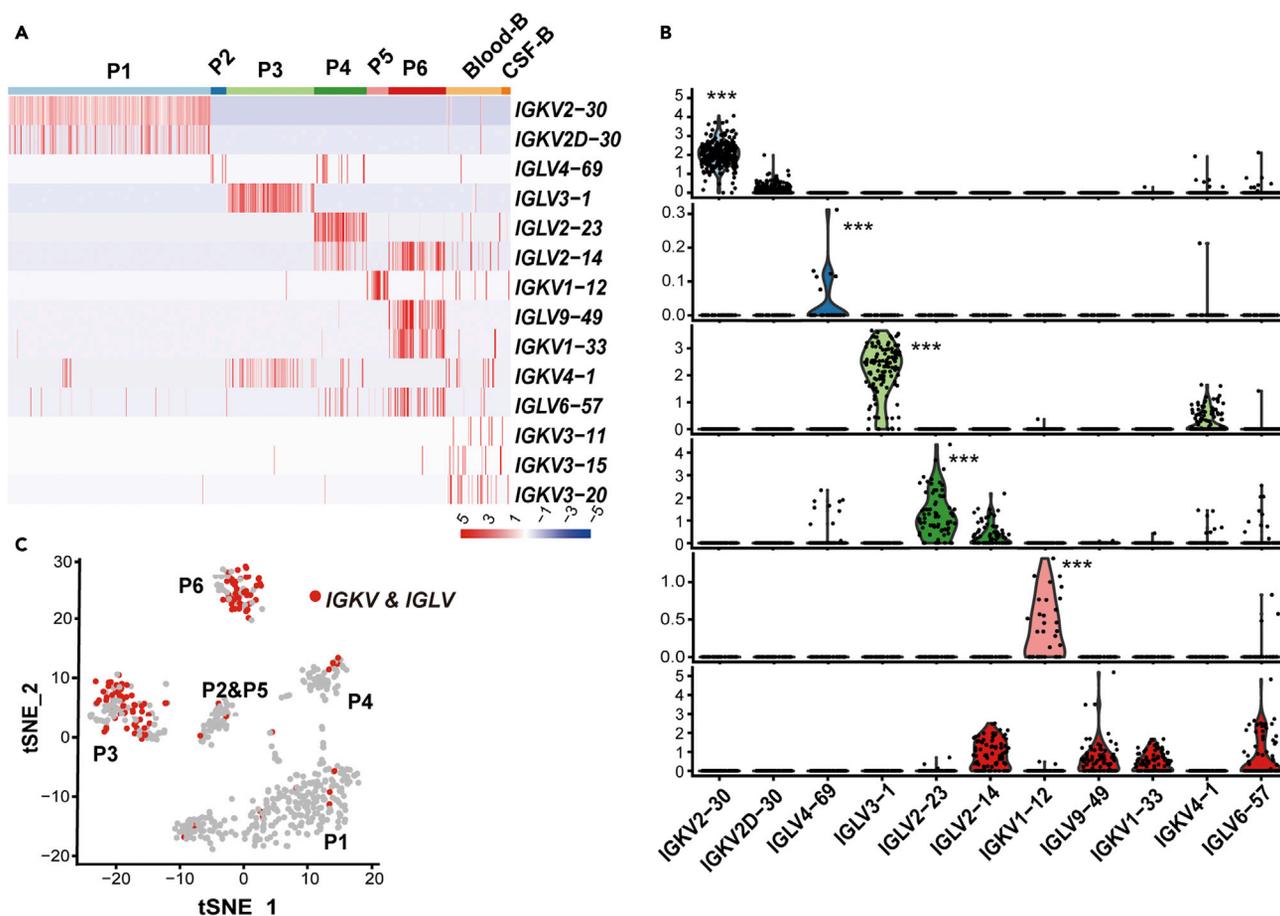


Figure 3. Characterization of immunoglobulin light chain variable molecules in CSF-DLBCs (cerebrospinal fluid diffuse large B cells)

(A) Heatmap displaying the distribution of expression of immunoglobulin light chain variable molecules in CSF-DLBCs (IGKV, immunoglobulin kappa variable; IGLV, immunoglobulin lambda variable).

(B) Violin plots displaying the distribution of immunoglobulin light chain variable molecule expression in CSF-DLBCs of six patients (**p value < 0.001, Wilcoxon rank-sum test).

(C) Cells (red) expressing both IGKV and IGLV shown on the t-SNE plot of 624 CSF-DLBCs (Figure S4A).

See also Figure S4.

been shown to be oncogenic and promote tumor progression in many cancers, especially associated with DLBCL deriving from germinal center (GC) B cells (Chase and Cross, 2011; Velichutina et al., 2010), although previous study showed up to 96% PCNSLs have similarity with DLBCLs of the activated B-cell (ABC) type (Hiemcke-Jiwa et al., 2018). In addition, the genes *PHF19* (Ning et al., 2018), *CENPU* (Zhang et al., 2018), *PKMYT1* (Schmidt et al., 2017), *NCAPH2* (Wallace et al., 2019), *PTTG1* (Huang et al., 2018), *SMC4* (Steffensen et al., 2001), *LAS1L* (Castle et al., 2010), *MRPL* (Cai et al., 2021), *MYBL2* (Iness et al., 2019), *NUSAP1* (Simonetti et al., 2019), *PHF6* (Warmerdam et al., 2020), and *RRM2* (Shu et al., 2020) play important roles in cell cycle and proliferation. The 24 genes deserve further study due to their expression specificity in CSF-DLBCs.

The analysis of variable region of light chain (VL) in CSF-DLBCs

When analyzing the differences of CSF-DLBCs among six patients with CNSL-DLBCL, we further removed cells that had fewer than 1000 covered genes, and 624 CSF-DLBCs were retained (289 cells from P1, 22 from P2, 125 from P3, 75 from P4, 31 from P5, 82 from P6, Figure S4). The immunoglobulin light chain restriction (LCR) indicates monoclonality of the proliferating mature B cells. CSF-DLBCs of a patient had monoclonality of the variable region of the light chain (Figure 3A). The VL molecule mainly expressed in P1 is *IGKV2-30* (IGKV, immunoglobulin kappa variable), in P2 is *IGLV4-69* (IGLV, immunoglobulin lambda variable), in P3 is

IGLV3-1, in P4 is *IGLV2-23*, and in P5 is *IGKV1-12* (Figures 3A and 3B). Additionally, other VL molecules were detected in CSF-DLBCs with lower expression (Figures 3A and 3B). P6 was atypical; the expression levels of four VL molecules were similar in P6 CSF-DLBCs (*IGLV2-14*, *IGLV9-49*, *IGLV6-57*, and *IGKV1-33*, Figures 3A and 3B).

As we know, DLBCL exhibits allelic exclusion in which only a single class of light chain is expressed (either lambda λ or kappa κ). However, there have been increasing numbers of reports that a double class of light-chain gene rearrangements can occur in B-cell malignant neoplasms (Xu, 2006). P6 is such a sample with 60.98% (50/82) of CSF-DLBCs expressing κ and λ light chains (Figure 3C). P3 also had 58/125 CSF-DLBCs expressing dual κ/λ light chains (Figure 3C), though the transcription level of *IGKV4-1* was lower than *IGLV3-1*. The remaining patients had few CSF-DLBCs with dual κ/λ light chain expression (9/289 cells of P1, 1/22 of P2, 7/75 of P4, 2/31 of P5, Figure 3C).

For VL expression of normal B cells, normal B cells were polyclonal and only 6 of 92 (6.5%) cells had dual κ/λ light-chain transcriptions (Figure 3A), which is consistent with previous reports that 0.2%–3.4% of normal maturing B cells have dual κ/λ light-chain expression (Diaw et al., 2000; Giachino et al., 1995).

Gene expression heterogeneity of CSF-DLBCs

The analysis of cell-to-cell correlation showed significant heterogeneity between CSF-DLBCs within a given patient in spite of the monoclonality of VL (correlation coefficients ranging from -0.043 to 0.693 ; Figure 4A). The correlations between CSF-DLBCs within individual patients (intra-patient) were much higher than those among different patients (inter-patient; mean correlation coefficient -0.012 vs. 0.025 , p value $< 2.2 \times 10^{-16}$, Wilcoxon rank-sum test; Figures 4A and 4B). CSF-DLBCs from patients who had undergone chemotherapy (P1, P3, P6) showed considerably greater intercellular heterogeneity than those from patients who had not received chemotherapy (P2, P4, P5), which is consistent with the view that chemotherapy promotes the progression of tumor (mean correlation coefficient 0.0244 vs. 0.0345 , p value $< 2.2 \times 10^{-16}$, Wilcoxon rank-sum test; Figure 4B).

Cell cycle heterogeneity of CSF-DLBCs

A total of 811 differentially expressed genes exclusively or preferentially expressed in one individual patient with CNSL-DLBCL were identified (p value < 0.05 , fold-change > 1.5 ; Figure 4C and Table S4). From the differential gene expression list, we could see many cell cycle-related genes upregulated in P3 and P6 CSF-DLBCs (Figure 4C). To characterize this different proliferation state of CSF-DLBCs, we used gene signatures to denote G1/S or G2/M phases (Tirosh et al., 2016). Cell cycle phase-specific signatures were highly expressed in a subset of CSF-DLBCs, distinguishing cycling cells from noncycling cells (Figure S5A). These signatures revealed variability in the fraction of cycling cells across six patients. The proportion of CSF-DLBCs in the cycling state was higher in P3 (44%) and P6 (50%) than that in the other patients (Figure 4D). In addition, compared to non-cycling cells, the cell cycle genes *TOP2A*, *CCNB2*, *CDC20*, and *SMC4* were greatly upregulated in cycling cells regardless of the tumor proliferation state of the patient (Figures S5B and S5C). These genes are candidates for proliferation markers as *MKI67* in the diagnosis of CSF-DLBCs and greatly important for CSF-DLBC proliferation.

Cancer-testis antigen heterogeneity of CSF-DLBCs

The differentially expressed genes list among patients also included cancer-testis antigens (CTAs, Figure 4C). CTAs have particular characteristics of high immunogenicity with restricted expression in normal male germ cells and offer extraordinary opportunities for cancer diagnosis and immunotherapy (Salmani-nejad et al., 2016). Several studies have evaluated the expression of CTAs in NHL to date Hudolin et al. (2013) (Inaoka et al., 2012), but little is known about the expression of CTAs in CSF-DLBCs. We examined the expression of 276 selected CTAs (<http://www.cta.lncc.br/modelo.php>) in CSF-DLBCs and discovered substantial inter-tumor heterogeneity and intra-tumor heterogeneity of CTAs (Figure S6). The number of expressed CTAs in CSF-DLBCs is different among patients, and patient P1 exhibited expression of the most CTAs (Figure 4E). In addition, the expression of *PAGE5*, *TDRD1*, *CTAG2*, *MAEL*, *CT45A1*, *PAGE2B*, and *MAGEA9B* was greatly restricted to P1 CSF-DLBCs (Figure S6). *ATAD2* (29%, 181/624) and *MPHOSPH10* (29.17%, 182/624) were ubiquitously and highly expressed in CSF-DLBCs of all patients and had the potential to serve as immunotherapy targets (Figure S6).

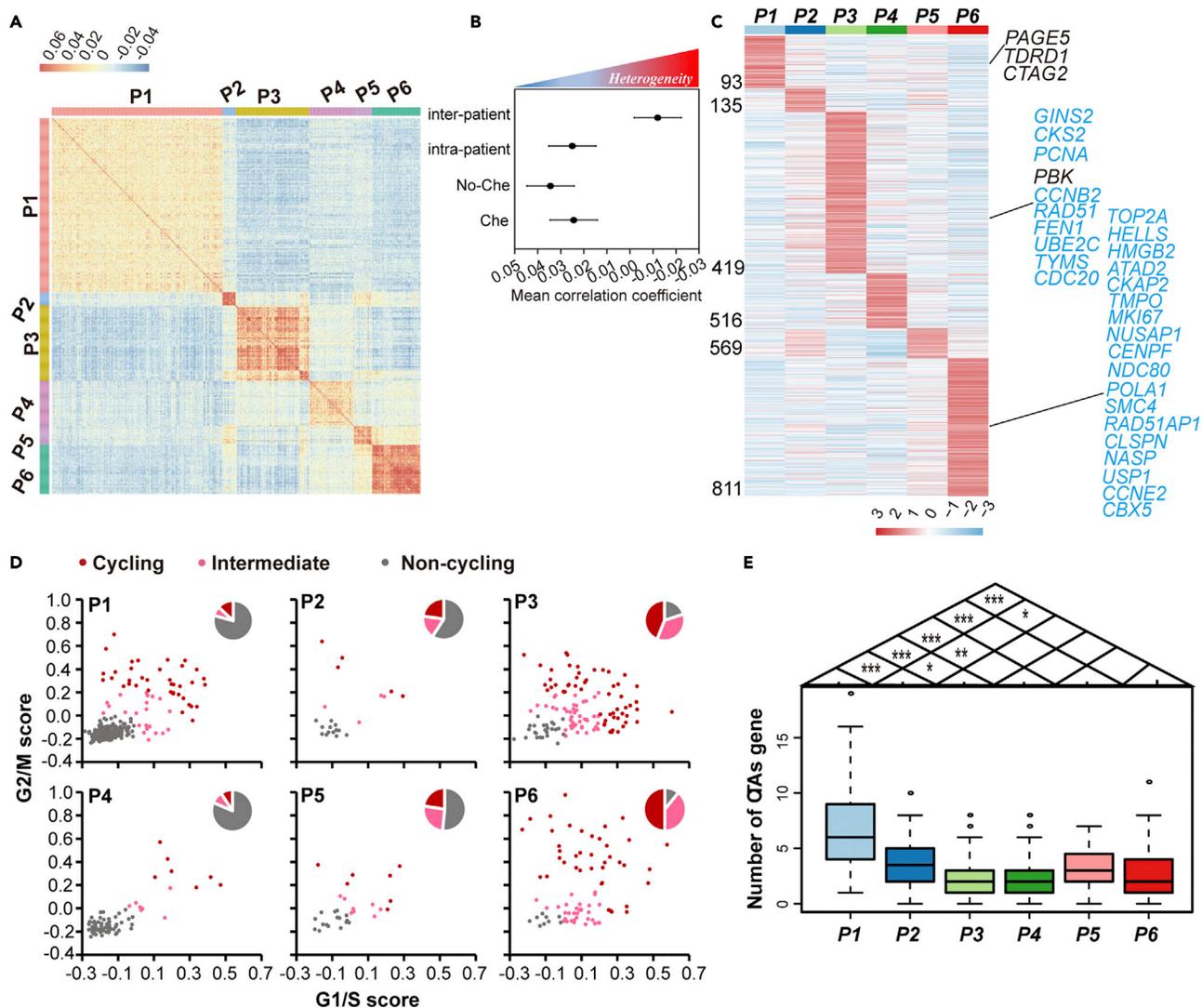


Figure 4. The heterogeneity of diffuse large B cells in cerebrospinal fluid (CSF-DLBCLs) among different patients and within individual patients
(A) The pairwise correlations between the expression profiles of CSF-DLBCLs from six patients with CNSL-DLBCL (central nervous system lymphoma-diffuse large B-cell lymphoma).

(B) Heterogeneity analysis showing the mean correlation coefficient for CSF-DLBCLs within patients with individual CNSL-DLBCL (intra-patient), among patients with CNSL-DLBCL (inter-patient), and within patients with individual CNSL-DLBCL with chemotherapy (P1, P3, P6; Che) or without chemotherapy (P2, P4, P5; No-Che). Data are represented as mean \pm SEM.

(C) Heatmap of differentially expressed genes (adjusted p value < 0.05, fold-change > 1.5) that are exclusively or preferentially expressed in one individual patient with CNSL-DLBCL. The names of selected genes are labeled. Gene names marked in blue (black) are cell cycle-related genes (cancer-testis antigens; CTAs).

(D) Estimation of the cell cycle state of every CSF-DLBCL (circles) on the basis of relative expression of G1/S (x axis) and G2/M (y axis) gene sets in different CNSL-DLBCL patients. Cells are colored by presumed cell cycle states. Red, cycling cells (score > 0.2); pink, intermediate (0 < score \leq 0.2); gray, noncycling cells (score \leq 0).

(E) Boxplot showing the number of CTAs (y axis) expressed in CSF-DLBCLs from six patients with CNSL-DLBCL (x axis) and the difference between any two patients (***p value < 0.001, **p value < 0.01, *p value < 0.05, Wilcoxon rank-sum test). Data are represented as mean \pm SEM.

See also Figures S5 and S6.

Germinal center B cell signatures define CSF-DLBC subgroups

The DLBCL has been classified into two subtypes by the cell-of-origin (COO) classification: the GC B cell-like (GCB) DLBCL which apparently originates from LZ B cells and the activated B-cell-like (ABC) DLBCL which commits to plasmablast (PBL) differentiation, respectively (Alizadeh et al., 2000). Up to 96% of PCNSL cases are classified as ABC type (Hiemcke-Jiwa et al., 2018). In order to define the origin of CSF-DLBCLs, we

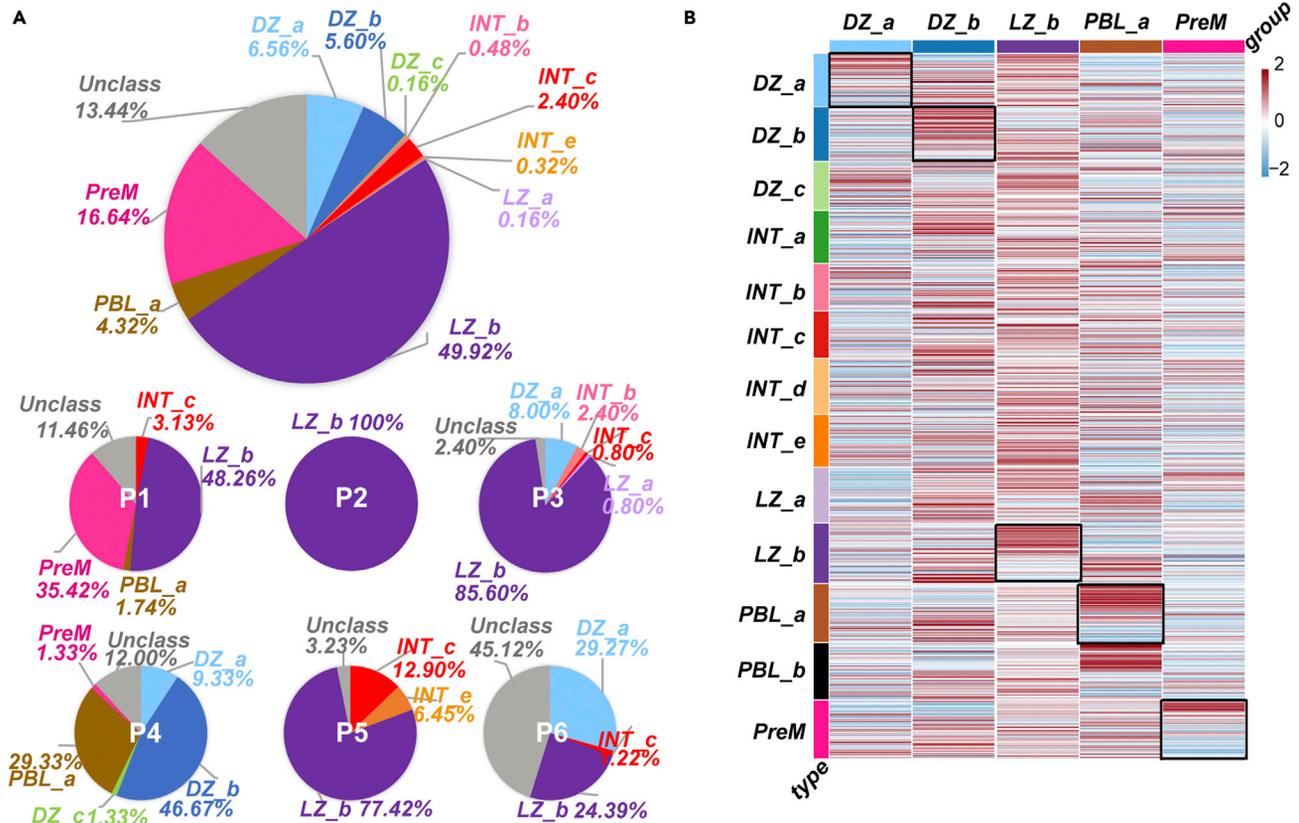


Figure 5. The classification of CSF-DLBCs based on the signatures of single-cell germinal center (sc-GC) cluster

(A) Distribution of the CSF-DLBCs by Seurat v3 Integration according to sc-GC cluster assignments (Holmes et al., 2020, Figure 2). Each cluster was labeled as DZ (dark zone), INT (intermediate), LZ (light zone), PBL (plasmablast), or PreM (precursor memory B cells). Gray depicts the fraction of CSF-DLBCs that remain unclassified (Unclass) by Seurat v3 Integration (the score of sc-GC classification less than 0.3).

(B) Heatmap summarizing the sc-GC-based classification of CSF-DLBCs. Each column represents a sc-GC cluster of CSF-DLBCs, and each row displays the relative mean expression of a sc-GC cluster gene. The sc-RNAseq signatures of a specific sc-GC cluster on the left of the heatmap include the top 50 upregulated and the top 50 downregulated genes in each sc-GC cluster (Holmes et al., 2020, Table S2).

See also Figure S7.

apply the Seurat v3 Integration method (<https://satijalab.org/seurat/v3.2/integration.html>) to classify CSF-DLBCs based on the scRNA-seq signatures of thirteen single-cell GC (sc-GC) clusters (GSM4148370 and GSM4148371) of DZ (dark zone), INT (intermediate), LZ (light zone), PBL, or PreM (precursor memory B cells) (Holmes et al., 2020). The results showed that in both, ~85% of the 624 CSF-DLBCs could be assigned to one sc-GC cluster with high confidence (prediction score ≥ 0.3 , Table S5). The majority of CSF-DLBCs were shown to be related with LZ_b cluster (49.92%, 312/624), especially in P1 (48.26%, 139/288), P2 (100%, 22/22), P3 (85.60%, 107/125), P5 (77.42%, 24/31), and P6 (24.39%, 20/82) (Figures 5A and S7A). P4 is a special patient with PCNSL; 57.33% of 75 CSF-DLBCs were classified into the DZ cluster, whereas 29.33% of cells were labeled as PBL_a (Figures 5A and S7A). Of note, 35.42% of P1 CSF-DLBCs and 1.33% of P4 CSF-DLBCs were assigned to PreM cluster (Figures 5A and S7A), which were consistent with previous observations associating as many as one-quarter of ABC-DLBCL cases to PreM B cells (Holmes et al., 2020). In total, clusters DZ_a, DZ_b, LZ_b, PBL_a, and PreM displayed significant enrichment for CSF-DLBCs. In addition, the top 50 upregulated and downregulated genes from each sc-GC cluster signature (Holmes et al., 2020) were used to test the five clusters of CSF-DLBCs, and the results showed that every CSF-DLBC was properly classified using Seurat v3 Integration based on the sc-RNAseq signatures (Figure 5B).

We also performed the DEG analysis between our CSF-DLBCs assigned to DZ_a (41 cells), DZ_b (35 cells), LZ_b (312 cells), PBL_a (27 cells), or PreM (104 cells) cluster and normal GC B cells in DZ_a, DZ_b, LZ_b, PBL_a, or PreM cluster, respectively, by DESeq2 method. We paid attention to these 24 genes upregulated

specifically in CSF-DLBCs compared to normal blood B cells and normal CSF B cells (Tables S2 and S3) and found some genes still served as DEGs in the five paired clusters, especially in the LZ_b cluster (Table S6). As the major sc-GC classification of CSF-DLBCs, the 16 upregulated genes, namely, *CCDC167*, *EML6*, *EZH2*, *HAUS1*, *LAS1L*, *METTL26*, *NCAPH2*, *NT5DC2*, *NUSAP1*, *PHF19*, *PKMYT1*, *PTTG1*, *RRM2*, *SH3TC1*, *SMC4*, and *TIMM50* in CSF-DLBCs of LZ_b cluster (Table S6) have great importance in the CNS-DLBCL. In addition, we performed the comparison of transcriptional programs by the GSEA method between the CSF-DLBCs and GC B cells assigned to the LZ_b cluster. We discovered that cell cycle pathway, DNA replication and repair pathway, and metabolism category containing pyruvate metabolism, citrate cycle TCA cycle, glutathione metabolism, fatty acid metabolism, and PPAR signaling pathway were enriched in CSF-DLBCs, whereas the antigen processing and presentation pathway and the BCR signaling pathway were downregulated in CSF-DLBCs (FDR q -value < 0.05 ; Figure S7B), which were same as functional enrichments when DLBCs compared to B cells of B-cell cluster (Figures 2A and S3).

DISCUSSION

In this study, we analyzed the transcriptomes of CSF-DLBCs at the single-cell level from six patients with CNSL-DLBCL. CSF-DLBCs mainly clustered according to patient of origin, whereas P2 and P5 CSF-DLBCs were in the same cluster but they had different variable regions of light chain molecules (Figures 3A and 3B). Immunoglobulin LCR is a typical feature of DLBCs. Individual patient DLBCs showed monoclonality of VL, and three CSF-DLBCs in the P3-1 sample were same as CSF-DLBCs in the P3-2 sample expressing *IGLV3-1*. Therefore, we analyzed the transcriptomes of CSF-DLBCs based on individual patients.

scRNA-Seq enabled us to directly compare the transcriptomes of entirely pure normal B cells and CSF-DLBCs. Compared to normal B cells, gene expression in CSF-DLBCs was enriched in the cell proliferation and energy metabolism pathways, which are critical for tumor growth and energy demand in the CSF-DLBCs. As reported, consensus cluster classification has grouped DLBCLs into the BCR/proliferation cluster (BCR-DLBCL), the OxPhos cluster (OxPhos-DLBCL), and the host response (HR) cluster with a brisk host inflammatory response (Monti et al., 2005). Compared to BCR-DLBCL, OxPhos-DLBCL has enhanced oxidative phosphorylation, TCA cycle, fatty acid oxidation program, PPAR signaling pathway, pyruvate metabolism, glucose-derived metabolites, and glutathione synthesis but does not display active/functional BCR signaling (Caro et al., 2012). Based on the active metabolism characteristics of CSF-DLBCs, CNS-DLBCL might be grouped into the OxPhos-DLBCL cluster with downregulated BCR signaling pathway, especially in patients P1, P2, and P3 (Figures 2A and S3). As is well known, adhesive cell-cell and cell-matrix interactions generally play important roles in tumor metastases and drug resistance (Wu et al., 2015). Whether the general downregulated cell adhesion molecule pathway (Figures 2A and S3) in CSF-DLBCs affects CNS-DLBCL metastases and drug resistance deserves further study.

Heterogeneities of CSF-DLBCs

We identified the transcriptome heterogeneity of CSF-DLBCs in the cell cycle state, the expression of CTAs, and the classification by sc-GC B-cell signatures. In addition, the heterogeneities were also observed in other aspects. First, the energy metabolism activity in CSF-DLBCs was different among patients to some extent (Figures 2A and S3). Second, HLA-II (human leukocyte antigen class II) molecules (*HLA-DRB5*, *HLA-DQA1*, *HLA-DQB1*, *HLA-DMB*, *HLA-DRA*) were highly expressed in P3 and P4 but not in other patients (Figure S8). The frequent loss of HLA-II in CNS-DLBCL is mainly due to homozygous deletions in the HLA region, which affects the onset and modulation of immune response for lack of activated CD4⁺ T lymphocytes (Jordanova et al., 2003). Third, the patient clinical characteristics (Table S1) including age, sex, intracranial tumor sites, therapies received or not before sample collection, and the time interval from disease diagnosis to sample collection were different, which also contributed to the heterogeneity of CSF-DLBCs among six patients with CNS-DLBCL.

Primary and secondary CNS lymphomas are completely different entities despite the share of CNS involvement: this fact should be taken into account in the different analyses. In the analyses of GSEA, ImmuneScore, the cycling state, CTA expression and CSF-DLBC classification by sc-GC B-cell signature, there were no obvious difference between PCNSL and SCNSL. The great heterogeneities were mainly manifested between different patients. The analysis of cell-to-cell correlation showed CSF-DLBCs from patients with SCNSL (P1, P6) and showed considerably greater intercellular heterogeneity than that from patients with PCNSL (P2, P3, P4, P5) (mean correlation coefficient: 0.0213 vs. 0.0397, p value $< 2.2 \times 10^{-16}$, Wilcoxon rank-sum test). The DEG analysis showed 155 genes expressed specially in P1 and P6 (SCNSL) CSF-DLBCs,

whereas 41 genes were upregulated in DLBCs of all 4 patients with PCNSL compared to normal B cells (the screening criteria were same as those in [Figure 2C](#), [Tables S2](#), and [S3](#)). The difference between PCNSL and SCNSL CSF-DLBCs needs to include more patients to analyze.

Taken together, our findings firstly investigated systematic and comprehensive transcriptome characteristics of more than one thousand CSF-DLBCs at the single-cell level. We discovered that most CSF-DLBCs displayed immunoglobulin LCR, active cell proliferation and energy metabolism properties, and inherent heterogeneity which were shown on the cell cycle state, CTA expression, and classification by sc-GC B-cell signature. Despite a few patients, our study revealed biological insights of the intricate machinery responsible for CNS-DLBCL progression, and a larger validation cohort is required in the future.

Limitations of the study

Our study totally performed scRNA-seq on 2,631 target cells from 7 patients and retained 2,114 cells (1,093 CSF-DLBCs) for subsequent analysis; the number of cells analyzed is relatively small for lack of various immune cells in the CSF tumor microenvironment. Our study mainly focused on CSF-DLBCs; the machinery of CSF-DLBC immune evasion has been still unknown. The analysis of the immune cell characteristics in pathological CSF samples will be informative to understand a systems-level view of the CNS-DLBCL tumor microenvironment. In addition, although CSF-DLBCs serving as one type of circulating tumor cells could reflect the characteristics of CNSL from one aspect, they were in absence of information on tumor cells from the primary location. The gene expression of DLBCs at CSF might be substantially different from that of primary and leptomeningeal site which is worthy of further study.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- [KEY RESOURCES TABLE](#)
- [RESOURCE AVAILABILITY](#)
 - Lead contact
 - Materials availability
 - Data and code availability
- [EXPERIMENTAL MODEL AND SUBJECT DETAILS](#)
- [METHOD DETAILS](#)
 - Single-cell preparation and SMART-seq2 library construction
 - Generation of gene expression matrix
 - Population identification
 - Analysis of differential expression and gene enrichment
 - Cell cycle analysis
- [QUANTIFICATION AND STATISTICAL ANALYSIS](#)

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2021.102972>.

ACKNOWLEDGMENTS

This work was funded by the National Key Research and Development Program of China, grant number 2017YFA0103902 & 2019YFA0111400; the National Natural Science Foundation of China, grant number 31771283 and 82072367; the Innovative Research Team of High-level Local Universities in Shanghai, grant number SSMU-ZDCX20180700; the Key Laboratory Program of the Education Commission of Shanghai Municipality, grant number ZDSYS14005; the Innovation Group Project of Shanghai Municipal Health Commission, grant number 2019CXJQ03; and Shanghai Municipal Key Clinical Specialty (Laboratory Medicine), grant number shslczdk03303. M.G. was funded by the Program for Shanghai Municipal Leading Talent (2015).

AUTHOR CONTRIBUTIONS

H.R. and M.G. designed the work. H.R. and Z.W. analyzed the data. H.R., R.H., K.C., and X.L. interpreted the patient data. H.R., Y.Z., Y.X., L.P., J.Z., Y.Z., C.Z., W.M., and J.S. performed the acquisition of data. H.R. was a

major contributor in writing the manuscript. Z.W., X.D., C.Z., and M.G. reviewed and edited the manuscript. C.Z. and M.G. acquired funding. All authors read and approved the final manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: December 15, 2020

Revised: August 3, 2021

Accepted: August 6, 2021

Published: September 24, 2021

REFERENCES

- Alizadeh, A.A., Eisen, M.B., Davis, R.E., Ma, C., Lossos, I.S., Rosenwald, A., Boldrick, J.C., Sabet, H., Tran, T., Yu, X., et al. (2000). Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 403, 503–511.
- Baraniskin, A., Chomiak, M., Ahle, G., Gress, T., Buchholz, M., Turewicz, M., Eisenacher, M., Margold, M., Schlegel, U., Schmiegel, W., et al. (2018). MicroRNA-30c as a novel diagnostic biomarker for primary and secondary B-cell lymphoma of the CNS. *J. Neurooncol.* 137, 463–468.
- Baraniskin, A., and Schroers, R. (2014). Modern cerebrospinal fluid analyses for the diagnosis of diffuse large B-cell lymphoma of the CNS. *CNS Oncol.* 3, 77–85.
- Cai, M., Li, H., Chen, R., and Zhou, X. (2021). MRPL13 promotes tumor cell proliferation, migration and EMT process in breast cancer through the PI3K-AKT-mTOR pathway. *Cancer Manag. Res.* 13, 2009–2024.
- Camilleri-Broet, S., Criniere, E., Broet, P., Delwail, V., Mokhtari, K., Moreau, A., Kujas, M., Raphael, M., Iraqi, W., Sautes-Fridman, C., et al. (2006). A uniform activated B-cell-like immunophenotype might explain the poor prognosis of primary central nervous system lymphomas: analysis of 83 cases. *Blood* 107, 190–196.
- Caro, P., Kishan, A.U., Norberg, E., Stanley, I.A., Chapuy, B., Ficarro, S.B., Polak, K., Tondera, D., Gounarides, J., Yin, H., et al. (2012). Metabolic signatures uncover distinct targets in molecular subsets of diffuse large B cell lymphoma. *Cancer Cell* 22, 547–560.
- Castle, C.D., Cassimere, E.K., Lee, J., and Denicourt, C. (2010). Las1L is a nucleolar protein required for cell proliferation and ribosome biogenesis. *Mol. Cell. Biol.* 30, 4404–4414.
- Chase, A., and Cross, N.C. (2011). Aberrations of EZH2 in cancer. *Clin. Cancer Res.* 17, 2613–2618.
- Chukwueke, U.N., and Nayak, L. (2019). Central nervous system lymphoma. *Hematol. Oncol. Clin. North Am.* 33, 597–611.
- Diaw, L., Siwarski, D., DuBois, W., Jones, G., and Huppi, K. (2000). Double producers of kappa and lambda define a subset of B cells in mouse plasmacytomas. *Mol. Immunol.* 37, 775–781.
- Giachino, C., Padovan, E., and Lanzavecchia, A. (1995). kappa+lambda+ dual receptor B cells are present in the human peripheral repertoire. *J. Exp. Med.* 181, 1245–1250.
- Giannini, C., Dogan, A., and Salomao, D.R. (2014). CNS lymphoma: a practical diagnostic approach. *J. Neuropathol. Exp. Neurol.* 73, 478–494.
- Han, C.H., and Batchelor, T.T. (2017). Diagnosis and management of primary central nervous system lymphoma. *Cancer* 123, 4314–4324.
- Hiemcke-Jiwa, L.S., Leguit, R.J., Snijders, T.J., Jiwa, N.M., Kuiper, J.J.W., de Weger, R.A., Minnema, M.C., and Huibers, M.M.H. (2018). Molecular analysis in liquid biopsies for diagnostics of primary central nervous system lymphoma: Review of literature and future opportunities. *Crit. Rev. Oncol. Hematol.* 127, 56–65.
- Holmes, A.B., Corinaldesi, C., Shen, Q., Kumar, R., Compagno, N., Wang, Z., Nitzan, M., Grunstein, E., Pasqualucci, L., Dalla-Favera, R., et al. (2020). Single-cell analysis of germinal-center B cells informs on lymphoma cell of origin and outcome. *J. Exp. Med.* 217, e20200483.
- Huang, J.L., Cao, S.W., Ou, Q.S., Yang, B., Zheng, S.H., Tang, J., Chen, J., Hu, Y.W., Zheng, L., and Wang, Q. (2018). The long non-coding RNA PTTG3P promotes cell growth and metastasis via up-regulating PTTG1 and activating PI3K/AKT signaling in hepatocellular carcinoma. *Mol. Cancer* 17, 93.
- Hudolin, T., Kastelan, Z., Ilic, I., Levarda-Hudolin, K., Basic-Jukic, N., Rieken, M., Spagnoli, G.C., Juretic, A., and Mengus, C. (2013). Immunohistochemical analysis of the expression of MAGE-A and NY-ESO-1 cancer/testis antigens in diffuse large B-cell testicular lymphoma. *J. Transl. Med.* 11, 123.
- Inaoka, R.J., Jungbluth, A.A., Gnjatic, S., Ritter, E., Hanson, N.C., Frosina, D., Tassello, J., Etto, L.Y., Bortoluzzo, A.B., Alves, A.C., et al. (2012). Cancer/testis antigens expression and autologous serological response in a set of Brazilian non-Hodgkin's lymphoma patients. *Cancer Immunol. Immunother.* 61, 2207–2214.
- Iness, A.N., Felthousen, J., Ananthapadmanabhan, V., Sesay, F., Saini, S., Guiley, K.Z., Rubin, S.M., Dozmorov, M., and Litovchick, L. (2019). The cell cycle regulatory DREAM complex is disrupted by high expression of oncogenic B-Myb. *Oncogene* 38, 1080–1092.
- Jordanova, E.S., Philippo, K., Giphart, M.J., Schuurings, E., and Kluijn, P.M. (2003). Mutations in the HLA class II genes leading to loss of expression of HLA-DR and HLA-DQ in diffuse large B-cell lymphoma. *Immunogenetics* 55, 203–209.
- Khatab, S., Spliet, W., and Woerdeman, P.A. (2014). Frameless image-guided stereotactic brain biopsies: emphasis on diagnostic yield. *Acta Neurochir. (Wien)* 156, 1441–1450.
- Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550.
- Monti, S., Savage, K.J., Kutok, J.L., Feuerhake, F., Kurtin, P., Mihm, M., Wu, B., Pasqualucci, L., Neuberg, D., Aguiar, R.C., et al. (2005). Molecular profiling of diffuse large B-cell lymphoma identifies robust subtypes including one characterized by host inflammatory response. *Blood* 105, 1851–1861.
- Nabavizadeh, S.A., Vossough, A., Hajmomenian, M., Assadsangabi, R., and Mohan, S. (2016). Neuroimaging in central nervous system lymphoma. *Hematol. Oncol. Clin. North Am.* 30, 799–821.
- Nayak, L., and Batchelor, T.T. (2013). Recent advances in treatment of primary central nervous system lymphoma. *Curr. Treat. Options Oncol.* 14, 539–552.
- Ning, F., Wang, C., Niu, S., Xu, H., Xia, K., and Wang, N. (2018). Transcription factor Phf19 positively regulates germinal center reactions that underlies its role in rheumatoid arthritis. *Am. J. Transl. Res.* 10, 200–211.
- Onder, E., Arikok, A.T., Onder, S., Han, U., Sorar, M., Kertmen, H., Yilmaz, E.D., Fesli, R., and Alper, M. (2015). Corticosteroid pre-treated primary CNS lymphoma: a detailed analysis of stereotactic biopsy findings and consideration of interobserver variability. *Int. J. Clin. Exp. Pathol.* 8, 7798–7808.
- Ruan, H., Zhou, Y., Shen, J., Zhai, Y., Xu, Y., Pi, L., Huang, R., Chen, K., Li, X., Ma, W., et al. (2020). Circulating tumor cell characterization of lung cancer brain metastases in the cerebrospinal fluid through single-cell transcriptome analysis. *Clin. Transl. Med.* 10, e246.
- Salmaninejad, A., Zamani, M.R., Pourvahedi, M., Golchehre, Z., Hosseini Bereshneh, A., and Rezaei, N. (2016). Cancer/testis antigens: expression, regulation, tumor invasion, and use in

immunotherapy of cancers. *Immunol. Invest.* **45**, 619–640.

Sasagawa, Y., Akai, T., Tachibana, O., and Iizuka, H. (2015). Diagnostic value of interleukin-10 in cerebrospinal fluid for diffuse large B-cell lymphoma of the central nervous system. *J. Neurooncol.* **121**, 177–183.

Schafflick, D., Xu, C.A., Hartlehnert, M., Cole, M., Schulte-Mecklenbeck, A., Lautwein, T., Wolbert, J., Heming, M., Meuth, S.G., Kuhlmann, T., et al. (2020). Integrated single cell analysis of blood and cerebrospinal fluid leukocytes in multiple sclerosis. *Nat. Commun.* **11**, 247.

Schmidt, M., Rohe, A., Platzer, C., Najjar, A., Erdmann, F., and Sippl, W. (2017). Regulation of G2/M transition by inhibition of WEE1 and PKMYT1 kinases. *Molecules* **22**, 2045.

Shu, Z., Li, Z., Huang, H., Chen, Y., Fan, J., Yu, L., Wu, Z., Tian, L., Qi, Q., Peng, S., et al. (2020). Cell-cycle-dependent phosphorylation of RRM1 ensures efficient DNA replication and regulates cancer vulnerability to ATR inhibition. *Oncogene* **39**, 5721–5733.

Simonetti, G., Padella, A., do Valle, I.F., Fontana, M.C., Fonzi, E., Bruno, S., Baldazzi, C., Guadagnuolo, V., Manfrini, M., Ferrari, A., et al. (2019). Aneuploid acute myeloid leukemia exhibits a signature of genomic alterations in the cell cycle and protein degradation machinery. *Cancer* **125**, 712–725.

Steffensen, S., Coelho, P.A., Cobbe, N., Vass, S., Costa, M., Hassan, B., Prokopenko, S.N., Bellen, H., Heck, M.M., and Sunkel, C.E. (2001). A role for *Drosophila* SMC4 in the resolution of sister chromatids in mitosis. *Curr. Biol.* **11**, 295–307.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U S A* **102**, 15545–15550.

Tirosh, I., Izar, B., Prakadan, S.M., Wadsworth, M.H., 2nd, Treacy, D., Trombetta, J.J., Rotem, A., Rodman, C., Lian, C., Murphy, G., et al. (2016). Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196.

van der Maaten, L., and Hinton, G. (2008). Visualizing Data using t-SNE. *J. Mach Learn Res.* **9**, 2579–2605.

Velichutina, I., Shaknovich, R., Geng, H., Johnson, N.A., Gascoyne, R.D., Melnick, A.M., and Elemento, O. (2010). EZH2-mediated epigenetic silencing in germinal center B cells contributes to proliferation and lymphomagenesis. *Blood* **116**, 5247–5255.

Villano, J.L., Koshy, M., Shaikh, H., Dolecek, T.A., and McCarthy, B.J. (2011). Age, gender, and racial differences in incidence and survival in primary CNS lymphoma. *Br. J. Cancer* **105**, 1414–1418.

Wallace, H.A., Rana, V., Nguyen, H.Q., and Bosco, G. (2019). Condensin II subunit NCAPH2 associates with shelterin protein TRF1 and is required for telomere stability. *J. Cell. Physiol.* **234**, 20755–20768.

Warmerdam, D.O., Alonso-de Vega, I., Wiegant, W.W., van den Broek, B., Rother, M.B., Wolthuis, R.M., Freire, R., van Attikum, H., Medema, R.H., and Smits, V.A. (2020). PHF6 promotes non-homologous end joining and G2 checkpoint recovery. *EMBO Rep.* **21**, e48460.

Wu, Y., Xu, X., Miao, X., Zhu, X., Yin, H., He, Y., Li, C., Liu, Y., Chen, Y., Lu, X., et al. (2015). Sam68 regulates cell proliferation and cell adhesion-mediated drug resistance (CAM-DR) via the AKT pathway in non-Hodgkin's lymphoma. *Cell Prolif.* **48**, 682–690.

Xu, D. (2006). Dual surface immunoglobulin light-chain expression in B-cell lymphoproliferative disorders. *Arch. Pathol. Lab Med.* **130**, 853–856.

Yoshihara, K., Shahmoradgoli, M., Martinez, E., Vegesna, R., Kim, H., Torres-Garcia, W., Trevino, V., Shen, H., Laird, P.W., Levine, D.A., et al. (2013). Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* **4**, 2612.

Zhang, Q., Li, Y.D., Zhang, S.X., and Shi, Y.Y. (2018). Centromere protein U promotes cell proliferation, migration and invasion involving Wnt/beta-catenin signaling pathway in non-small cell lung cancer. *Eur. Rev. Med. Pharmacol. Sci.* **22**, 7768–7777.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
CD45	BD Biosciences	Cat#560973; RRID: AB_10565969
CD19	BD Biosciences	Cat#564456; RRID: AB_2744309
CellTrace Calcein Blue AM	Life Technologies	Cat#C34853
Deposited Data		
Single-cell RNA-seq data of DLBCs	Smart-seq2	GSE175510
Software and Algorithms		
STAR version 2.7	R package	https://github.com/alexdobin/STAR
Seurat package version 3.1.1	R package	https://cran.r-project.org/web/packages/Seurat/index.html
Seurat v3 Integration method	R package	https://satijalab.org/seurat/v3.2/integration.html
DESeq2 v3.9	R package	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
ClusterProfile	R package	https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Ming Guan (guanming88@yahoo.com).

Materials availability

This study did not generate new unique materials

Data and code availability

1. Data

The accession number for original data of single-cell RNA sequencing reported in this study is Gene Expression Omnibus (GEO): GSE175510.

2. Code

This study does not report original code. All codes were used in this study in alignment with recommendations made by authors of R packages in their respective user's guide, which can be accessed at [key resources table](#).

3. Additional information requests

Any additional information required to reanalyze the data used in this study is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

All human CSF sample (P: P1-P7) materials used in this study were collected with written informed consent. The proposed studies were approved by Institutional Review Board of Huashan Hospital (HIRB). Clinical information of patients is listed in [Table S1](#), including diagnosis, age, gender, tumor sites, developmental stage, chemotherapy received or not, etc. The seven CNSL patients included six males and one female, and their ages ranged from 32-64. Sex and gender have no influence on the study.

METHOD DETAILS

Single-cell preparation and SMART-seq2 library construction

Patient CSF samples (P: P1-P7) were diagnosed through cytopathology (Figure S1A) and 3 mL samples were remained for fluorescence-activated cell sorting (FACS). P1-P6 CSF cells sorted were CD45 (Catalog number Cat#560973, BD Biosciences) and CD19 (Cat#564456, BD Biosciences) positive, and had a larger cell diameter, showing greater forward scatter height (FSC-H) than the normal CSF leukocytes (Figure S1B). Whereas P7 CSF cells was only sorted based on CD45 positive. Live candidate CSF cells (CellTrace Calcein Blue AM+, Cat#C34853, Life Technologies, CA) were sorted into pre-prepared 96-well plates by FACS for SMART-seq2 scRNA-seq (Table S1 and Figure S1B). All antibody and labeling dye were used according to manufacturer recommendations. The construction of SMART-seq2 library was performed as the following modifications (Ruan et al., 2020): (1) RNA was reverse transcribed and amplified using Maxima H Minus Reverse Transcriptase (Cat#00724792, Thermo Fisher Scientific, MA) and KAPA HiFi Hot Start Ready Mix (Cat#KE2502, KAPA Biosystems, MA), (2) cDNA library was purified and quantified using Agencourt XP DNA beads (Cat#A63852, Beckman Coulter, CA) and a high sensitivity dsDNA Quant Kit (Cat#Q32854, Life Technologies, CA), (3) full length cDNA libraries were tagmented and only 3' end sequence (500-1000 bp) was enriched for sequencing on an Illumina HiSeqX machine.

Generation of gene expression matrix

Sequenced reads were mapped to hg38 using the STAR (version 2.7; <https://github.com/alexdobin/STAR>) with the default parameters. These uniquely mapped reads in the genome were used, and reads aligned to more than one locus were discarded. The expression level of gene was quantified by the number of counts.

In addition to 8 patient CSF samples, 3 normal CSF samples, one blood T sample and one blood B sample from the NCBI BioProject database under accession number PRJNA602172 were enrolled in the data analysis (Ruan et al., 2020). Then, in the gene expression matrix from 12 samples, genes expressed (counts > 0) in less than 10 cells were filtered out. Cells were removed according to the following criteria: (1) cells had fewer than 400 genes; (2) cells had over 20% mitochondrial-gene counts. A filtered gene expression matrix including 2,114 cells were used in the following analysis (Seurat package version 3.1.1; <https://cran.r-project.org/web/packages/Seurat/index.html>). In addition, when analyzing the differences of CSF-DLBCs between six CNSL-DLBCL patients, we further removed cells with fewer than 1000 genes and 624 CSF-DLBCs were retained.

Population identification

After filtration, a merged expression matrix of 12 samples was used for cell clustering by the Seurat 3.1.1 (Seurat package version 3.1.1), adapting the typical pipeline. In brief, gene expression was normalized by the "NormalizeData" function. Highly variable genes were calculated by the "FindVariableGenes" method with the default parameters. Significant principal components were used for downstream graph-based and semi-supervised clustering into distinct populations (FindClusters function in R), and t-SNE dimensionality reduction was used to project these populations in two dimensions. To identify marker genes, the clusters were compared pairwise for differential gene expression using the Wilcoxon rank-sum test for single-cell gene expression (FindAllMarkers function, min.pct = 0.1, logFC.threshold = 0.25). Subsequently, cell clusters were annotated manually based on known markers.

Analysis of differential expression and gene enrichment

DESeq2 (R package DESeq2 v3.9; <https://bioconductor.org/packages/release/bioc/html/DESeq2.html>) was used to detect DEGs (differentially expressed genes) between target samples (Love et al., 2014). Functional enrichment analysis of KEGG (kyoto encyclopedia of genes and genomes) pathways were performed by GSEA (ClusterProfiler; <https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html>) (Subramanian et al., 2005).

Cell cycle analysis

CellCycleScoring function included in Seurat 3.1.1 (<https://cran.r-project.org/web/packages/Seurat/index.html>) was used to perform cell cycle assignment. We identified cells that had either S.Score or G2M.Score > 0.2 as cycling cells; cells that had either 0 < S.Score or G2M.Score ≤ 0.2 as intermediate cells; and the other cells as non-cycling cells.



QUANTIFICATION AND STATISTICAL ANALYSIS

Data are presented as the mean \pm standard error of the mean (SEM). Samples were analyzed using Wilcoxon Rank-Sum test or Kruskal walls test for two groups and P-value < 0.05 was considered significant.