Tumor-infiltrating normal B cells revealed by immunoglobulin repertoire clonotype analysis are highly prognostic and crucial for antitumor immune responses in DLBCL

Zijun Y. Xu-Monette[1*], Yong Li[2], Thomas Snyder[3], Tiantian Yu[1], Tingxun Lu[1], Alexandar Tzankov[4], Carlo Visco[5], Govind Bhagat[6], Wenbin Qian[7], Karen Dybkaer[8], April Chiu[9], Wayne Tam[10], Youli Zu[11], Eric D. Hsi[12], Fredrick B. Hagemeister[13], Yingjun Wang[14], Heounjeong Go[15], Maurilio Ponzoni[16], Andrés J.M. Ferreri[16], Michael B. Møller[17], Benjamin M. Parsons[18], Xiangshan Fan[19], J. Han van Krieken[20], Miguel A. Piris[21], Jane N. Winter[22], Qingyan Au[23], Ilan Kirsch[3], Mingzhi Zhang[14], John Shaughnessy[24], Bing Xu[25], and Ken H. Young[1,26*]

[1]Hematopathology Division and Department of Pathology, Duke University Medical Center, Durham, NC, USA;
[2]Department of Medicine, Baylor College of Medicine, Houston, TX, USA;
[3]Adaptive Biotechnologies, Seattle, WA, USA;
[4]Institute of Pathology, University Hospital Basel, Switzerland;
[5]Department of Hematology, University of Verona, Verona, Italy;
[6]Columbia University Irving Medical Center and New York Presbyterian Hospital, New York, NY, USA;
[7]Department of Hematology, The Second Affiliated Hospital, College of Medicine, Zhejiang University, Hangzhou, Zhejiang, China;
[8]Aalborg University Hospital, Aalborg, Denmark;
[9]Mayo Clinic, Rochester, MN, USA;
[10]Weill Medical College of Cornell University, New York, NY, USA;
[11]The Methodist Hospital, Houston, TX, USA;
[12]Wake Forest University, Winston-Salem, NC, USA;
[13]Department of Lymphoma/Myeloma, The University of Texas MD Anderson Cancer Center, Houston, TX, USA;
[14]The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China;
[15]Asan Medical Center, Ulsan University College of Medicine, Seoul, Korea;
[16]San Raffaele H. Scientific Institute, Milan, Italy;
[17]Odense University Hospital, Odense, Denmark;

[18]Gundersen Lutheran Health System, La Crosse, WI, USA;
[19]Pathology Center, Anhui Medical University and the first Affiliated Hospital, Hefei, China;
[20] Radboud University Nijmegen Medical Centre, Nijmegen, Netherlands;
[21]Hospital Universitario Marqués de Valdecilla, Santander, Spain;
[22]Feinberg School of Medicine, Northwestern University, Chicago, IL, USA;
[23]NeoGenomics Laboratories, Aliso Viejo, California, USA;
[24]Myeloma Center, Winthrop P. Rockefeller Institute, University of Arkansas for Medical Sciences, Little Rock, Arkansas, USA;
[25]The First Affiliated Hospital of Xiamen University, Xiamen, Fujian, China;
[26]Duke Cancer Institute, Durham, NC, USA.

***Correspondence:**

Zijun Y. Xu-Monette, PhD, zijun.xumonette@duke.edu; **Ken H. Young, MD, PhD**, Division of Hematopathology and Department of Pathology, Duke University Medical Center, Duke Cancer Institute, Duke School of Medicine, Durham, NC 27710, USA. Email: ken.young@duke.edu; Phone: 1-919-668-7568; Fax: 1-919-684-1856.

Running Title: TIL-Bs play a critical role in determining DLBCL prognosis

Keywords: TIL-B, BCR clonotype, TME, GEP, multiplex IHC, EcoTyper, immunoglobulin, DLBCL, ABC, GCB, B-cell activation, helper T cell, EZB, germinal center, IGH, CDR3, immune response, cytokine, chemokine

Abstract Word Count: 250

Main Text Word Count: 5107

Number of Figures: 6

Number of References: 50

Number of Supplementary Tables: 11

Number of Supplementary Figures: 16

Supplementary Excel File: 1

Abstract

Purpose: Tumor-infiltrating B lymphocytes (TIL-Bs) have demonstrated prognostic and predictive significance in solid cancers. In this study, we aimed to distinguish TIL-Bs from malignant B-cells in diffuse large B-cell lymphoma (DLBCL) and determine the clinical and biological significance.

Patients and Methods: A total of 269 patients with *de novo* DLBCL from the International DLBCL R-CHOP Consortium Program were studied. Ultra-deep sequencing of the immunoglobulin genes was performed to determine B-cell clonotypes. The frequencies and numbers of TIL-B clonotypes in individual repertoires were correlated with patient survival, gene-expression profiling (GEP) data, and frequencies of DLBCL-infiltrating immune cells quantified by fluorescent multiplex immunohistochemistry at single-cell resolution.

Results: TIL-B abundance, evaluated by frequencies of normal B-cell clonotypes in the immunoglobulin repertoires, showed remarkably positive associations with significantly better survival of patients in our sequenced cohorts. DLBCLs with high versus low TIL-B abundance displayed distinct GEP signatures, increased pre-memory B cell state and naïve CD4 T cell state fractions, and higher $CD4^+$ T-cell infiltration. TIL-B frequency, as a new biomarker in DLBCL, outperformed the germinal center B-cell-like/activated B-cell-like classification and TIL-T frequency. The identified TIL-B-high GEP signature, including genes highly expressed in normal germinal center B cells and T cells and those upregulated during T-dependent B-cell activation, showed significant favorable prognostic effects in several external validation cohorts.

Conclusions: TIL-B frequency is a significant prognostic factor in DLBCL and plays a crucial role in antitumor immune responses. This study provides novel insights into the prognostic determinants in DLBCL and TIL-B functions with important therapeutic implications.

Statement of translational relevance:

Tumor-infiltrating B lymphocytes (TIL-Bs) have shown significantly favorable prognostic effects in solid cancers where TIL-Bs are easily quantifiable. This study provides the first evidence that the total frequency of TIL-Bs in the immunoglobulin repertoires and clonotypic diversity have a significant prognostic impact in DLBCL, the most common B-cell lymphoma. The stratification by TIL-Bs was also biologically significant, demonstrating prominent GEP signatures, increased tumor-infiltrating CD4 T cells (by fluorescent mIHC), specific cell state enrichment (by EcoTyper bulk GEP analysis), and alteration of PD-1's prognostic effects. The remarkable ability of TIL-B frequency to stratify DLBCL patients will inform studies on predictive biomarkers and mechanisms underlying the prognostic significance of GCB/ABC classification. The important role of normal B cells in antitumor immunity may underlie their clinical significance in DLBCL, which has significant implications for current and future therapies (such as BCR-targeted therapies, CAR T-cell therapy, and bispecific antibodies).

Introduction

Tumor-infiltrating B lymphocytes (TIL-Bs) have received renewed attention. Unlike T cells, which can only recognize processed neoantigens presented on major histocompatibility complexes (MHC; or CD1 for NKT) to T-cell receptors (TCRs) and require co-stimulatory signals from antigen-presenting cells (APCs) for activation, B cells use their surface immunoglobulins to directly bind native antigens, proteins, polysaccharides, lipids, and nucleic acids, and can become activated either T-independently or T-dependently (1, 2). Unlike TCRs, B-cell receptors (BCRs) can undergo somatic hypermutation (SHM) and isotype switching in germinal centers (GCs) after T-dependent B-cell activation. Finally, plasmablasts and plasma cells (effector B cells) differentiated from activated B cells secrete antibodies.

High density of TIL-Bs and mature tertiary lymphoid structures, which are ectopic lymphoid aggregates in non-lymphoid tissues, have consistently shown favorable prognostic effects in solid cancers (3, 4). The diversity of BCR repertoires is also correlated with better prognosis in some solid tumors in a previous study using TCGA RNA-seq data of 9522 tumors (5). Tumor-reactive antibodies have been discovered in a few studies, and are associated with unfavorable (6) or favorable (7) clinical outcomes. Intriguingly, the antibodies discovered to date are predominantly against self-antigens (8), suggesting a break in peripheral B-cell tolerance that can be caused by increased self-antigen concentrations and lead to “epitope spreading” (9) involving T cells and linked recognition. It is proposed that activated self-reactive B cells can combat cancer immunoediting, in that they can prime polyclonal T cells with broad linked antigen reactivity through epitope spreading (8).

The clinical significance TIL-Bs in B-cell lymphomas is unknown, as TIL-Bs are not easily identifiable. Diffuse large B-cell lymphoma (DLBCL) arising from mature B cells is the most

common lymphoma. DLBCL can be classified into two major cell-of-origin subtypes by gene expression profiling (GEP), GC B-cell-like (GCB) and activated B-cell-like (ABC) (10), and various genetic (11, 12) and lymphoma ecosystem (LE) (13) subtypes. With the standard treatment, rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisone (R-CHOP), ~40% of DLBCL patients are not cured. Autologous stem-cell transplantation and chimeric antigen receptor (CAR) T cells improve DLBCL prognosis; however, resistance mechanisms and common relapses after CAR-T therapies (14, 15) remain to be better understood.

A recent single-cell RNA-sequencing (scRNA-seq) study identified TIL-Bs in 17 DLBCL cases based on gene-expression data (16). The average frequency was 7.5% (range in 13 positive cases: 1-60%). Naïve and memory subclusters of TIL-Bs, but not most of malignant B-cell subclusters (16), highly expressed 3 gene signatures of normal GC B-cells: precursor-memory (preM), INT (intermediate dark-zone/light-zone stage)-d, and INT-e (17). Another study detected TIL-Bs in one of 4 scRNA-seq DLBCL cases with a frequency of 10.7%, estimated by doubling the proportion of kappa/lambda light chain–non-restricted B-cell clusters, and in 4 of 9 flow cytometry DLBCL cases with an average frequency of 2.43% (range in 4 positive cases: 0.1-11.7%) (18).

Mature B cells have clonotypic immunoglobulins with unique gene rearrangements of immunoglobulin heavy (IGH) and light chain gene segments and differences in SHMs (19, 20). In this study, we performed immunoglobulin gene (IG) ultra-deep sequencing, distinguished normal B cells from DLBCL cells through clonotype analysis, and analyzed the frequencies and clonotypic diversity of bulk TIL-Bs for their prognostic and immunological effects in DLBCL.

Patients and Methods

145 **Patients**

146 The study cohort from the International Lymphoma Consortium Program included 269 patients
147 with *de novo* DLBCL treated with R-CHOP. Based on bulk GEP data (GSE31312), 107 and 108
148 cases of the study cohort were classified as GCB and ABC (21), respectively, and 171 cases
149 were assigned to a LE subtype (15, 1, 9, 26, 9, 58, 30, 14, 9 cases as LE1 to LE9, respectively)
150 using Lymphoma EcoTyper software (13). Based on genetic alterations identified by targeted
151 next-generation sequencing (NGS) (22), 71 cases were genetically subtyped using the
152 LymphGen algorithm (23). This study was conducted in accordance with the principles of the
153 Declaration of Helsinki. Material transfer agreements and data collection protocols were
154 approved as minimal to no risk or exempt by the institutional review board of each participating
155 institution.

156

157 **Ultra-deep IG sequencing**

158 DNAs extracted from formalin-fixed, paraffin-embedded (FFPE) diagnostic tissue sections
159 underwent IG ultra-deep sequencing using the immunoSEQ^{TM} platform (Adaptive
160 Biotechnologies, Seattle, WA) as described previously (24). The IGH and IG kappa or lambda
161 light chain (IGK/L) loci were independently analyzed by hsIGH and hsIGKL assays, respectively,
162 with an average of 260ng of genomic DNA and 162.08x sequencing depth of coverage.

163

164 After NGS, the raw sequencing data were processed by a complexity filter and nearest neighbor
165 algorithm to remove technical failures and correct sequencing errors. Specific V, D, and J genes
166 were identified according to criteria established by the International ImMunoGeneTics (IMGT)
167 collaboration with a standard algorithm to identify V, D, and J gene segments. When a specific
168 gene could not be identified unambiguously, the V, D, or J family gene was listed, or as
169 “unresolved”. Point mutations occurring in the IG variable regions (IGV) were identified by
170 comparisons with IMGT germline sequences.

Sequence results were clustered into distinct clonotypes with perfectly matched nucleotide sequences in the complementarity determining region 3 (CDR3) region using altorithms from Adaptive Biotechnologies (which have been replicated by others in BioConductor [RRID:SCR_006442]). The translated amino acid sequence was identified as productive, out-of-frame, or containing a premature stop codon. For a clone to be included in the results, its unique nucleotide sequence must be identified at least twice by the sequencer. The frequency of each clone (the percentage of normalized reads divided by the total number of normalized reads) was reported by ImmunoSEQ analyzer to determine the relative expansion in individual sequence repertoires. Clones with >5% frequency in a sequence repertoire were identified as index-trackable sequences. Shannon's entropy was calculated from the frequencies of all productive sequences divided by the logarithm of the total number of unique productive sequences.

Multiplex immunofluorescence

Fluorescent multiplex immunohistochemistry (mIHC) was performed using a MultiOmyx^{TM} immunofluorescence platform (NeoGenomics Laboratories, Inc., Aliso Viejo, CA) for a large DLBCL cohort including 218 cases in this study. FFPE tissue microarrays were stained with 13 antibodies conjugated to either cyanine 3 (cy3) or cyanine 5 (cy5) and 4′,6-diamidino-2-phenylindole (DAPI) in eight staining rounds. Stained images were collected on an INCell analyzer 2200 microscope (GE Healthcare Life Sciences) equipped with high-efficiency fluorochrome-specific filter sets for DAPI, cy3, and cy5 and analyzed in multiple stages by applying a rigid registration algorithm, nuclear segmentation algorithms, morphologic and shape algorithms, and tissue-quality algorithms as described previously (25). PD-1 (antibody RRID:AB_2894867), PD-L1 (RRID:AB_2827816), PD-L2 (RRID:AB_2799999), and CTLA-4 (RRID:AB_10988256) expressions in B cells (CD20^{+}; RRID:AB_2144403), T cells (CD3^{+}; RRID:AB_2631163), macrophages (CD68^{+}**;** RRID:AB_627158), and NK cells (CD56^{+};

RRID:AB_2864402) were evaluated by co-expression analysis. T-cell infiltration was evaluated by T-cell count divided by the total counts of $PAX5^+$ cells, $CD3^+$ cells, macrophages, and NK cells in each sample.

GEP analysis

GEP was performed using the Affymetrix GeneChip Human Genome HG-U133 Plus 2.0 (21) ~~A~~ Bayesian model was used to classified GCB/ABC subtype (21). CEL files were pre-processed and normalized using the Robust Multi-chip Average (RMA) implemented in R package affy (release 3.16). Two-class unpaired Significance Analysis of Microarrays (SAM; RRID:SCR_010951) was performed to identify significantly differentially expressed genes (DEGs) using R package samr (version 3.0). CLUSTER 3.0 software (RRID:SCR_013505) and JAVA TREEVIEW (RRID:SCR_016916) were used to cluster and visualize differential expression. Expression Analysis Systematic Explorer (EASE, a desktop version of DAVID; RRID:SCR_013361) software was used to categorize over-represented biological pathways using Gene Ontology terms.

Statistical analysis

The overall survival (OS) and progression-free survival (PFS) durations were calculated from the time of diagnosis to death from any cause and to progression or death from any cause, respectively, or censored at the last follow-up. Univariate survival analysis was performed using the Kaplan-Meier method. Multivariate analysis was performed using the Cox model. Fisher's exact test was used to compare the frequencies of the categorical variables. The unpaired/paired 2-tailed *t* test, unpaired two-sided Mann-Whitney U test, and Wilcoxon matched-pairs signed rank test were used to compare the distribution and mean/median values between two groups. Statistical analyses were performed using GraphPad Prism version 9 and IBM SPSS version 26.0. *P*-values of ≤0.05 were considered statistically significant.

Data availability

Major IGH sequencing data are provided in a Supplementary Excel file. GEP data have been deposited in Gene Expression Omnibus (GEO) GSE31312. mIHC and other generated data are available from the corresponding authors on reasonable request.

Results

Frequencies of normal B-cell clonotypes in the IGH repertoires are highly prognostic in DLBCL

High throughput ultra-deep sequencing of IGH genes was performed using multiplexed forward and reverse primers to amplify all rearranged VDJ and DJ sequences in the DNA samples. Bioinformatics pipelines were used to determine clonotypes, normalize the sequence read counts, and calculate the frequency of each clonotype in individual sequence repertoires. A predominant productive IGH-VDJ sequence in a repertoire was identified as the diagnostic clonotype (successfully in 138 cases and analyzed for SHM and prognostic significance in our previous study (24)), representing expanded malignant B cells. Its derivative sequences were also likely from neoplasms, while other clonally unrelated productive IGH-VDJ clonotypes were identified as normal B-cell clonotypes likely derived from non-transformed TIL-Bs.

For example, in **Figure 1A**, a diagnostic IGH sequence (SHM^+) and 2 derivative sequences with additional SHMs were identified as neoplastic B-cell clonotypes, and 42 unrelated productive VDJ sequences were identified as normal B-cell clonotypes. Since mature B cells in secondary lymphoid organs should have only one successful VDJ rearrangement, other types of IGH sequences (DJ-only; VDJ with nonsense or out-of-frame mutations) could be derived from non-

functional alleles originally rearranged during B cell development or failure of VDJ sequencing using FFPE tissues. **Figure 1B** shows the frequencies of the five sequence types in the 138individual repertoires. The mean (**Figure 1C**) and median frequencies of bulk normal B-cell clonotypes were 8.48% and 3.94%, respectively.

In these 138 DLBCL cases, frequencies of bulk normal B-cell clonotypes showed significant effects on OS and PFS across a wide range of cutoffs (≥0.019%-16.7% for high frequencies *vs*. 0% to ≤16.35% for low frequencies). **Figure 2A** shows the OS/PFS using two cutoffs: 1% and 2.4%. The effects remained highly significant after excluding four patients with DLBCL/high grade B-cell lymphoma with *MYC* and *BCL2* rearrangements (26) from the study cohort (Supplementary Figure 1A) and in multivariate analysis adjusted for clinicopathological parameters (Supplementary Tables 1-2).

High normal B-cell clonotype frequencies display prominent immune gene signatures

Bulk GEP data of DLBCL cases with and without high frequencies (≥2.4%) of normal B-cell clonotypes were compared, resulting in 427 significantly DEGs (348 DEGs using a 1% cutoff; Supplementary Figure 1B, Supplementary Tables 3-4), suggesting the biological meaningfulness of the stratification.

Immune response was the most enriched Gene Ontology biological process (Supplementary Table 5). Among the top DEGs visualized in **Figure 2B**, *FDCSP* (a regulator of B-cell responses (27); 19.9 fold), *CXCL13* (B-cell chemoattractant; 11.7 fold), and *CD1C* (presenting lipid and glycolipid antigens; 8.2 fold) showed the highest upregulation. T-cell signature was especially enriched, including *CCL19* (T/B cell migration (1, 2); 5.3 fold upregulation), *ENPP2* (T cell/ dendritic cell (DC) migration (28, 29)), *IL7R* (important for B-cell development and T-cell

especially memory T-cell survival (1, 2)), *CCL21* (T-cell chemoattractant), many TCR-signaling genes, T-helper (Th) 2 response-related genes including *IL4R*, *MAF* (highest in follicular Th [T_{FH}] and Th17 cells), and *DEF6*, Th1 response-related genes including *STAT4, IL2RG*, and *IL12RB1*, *CD84* (SLAM family member), *CD40LG* (CD40L), *CD27* (co-stimulatory for $CD8^+$ T cell activation (1); memory B-cell marker), *FOXP3* (regulatory T cell [Treg] marker), and *BCL11B* (30). In addition, *CTLA4*, *TIGIT*, *NLRC3, DGKA* (31), and *RASA3* (32) which suggest immunosuppression and *EVI2B* potentially relevant for B-cell activation (33) were also upregulated.

As **Figure 2C** illustrates, these genes are those upregulated during T-dependent B-cell activation (Supplementary Table 4) (1, 2). In this process, B/T cells are positioned by chemokines: CXCL13 in B-follicles, EBI2 ligand in outer follicle and interfollicular regions (34), CCL19 and CCL21 in T-cell zones. B cells present processed antigens to antigen-specific CD4 T cells that are activated independently by dendritic cells (DCs). Upon antigen recognition, the SLAM family and SAP binding are induced to prolong B cell:T cell interactions. Upon TCR stimulation, activated T cells express CD40L and cytokines to help with B-cell activation. Consequently, activated B cells proliferate and differentiate, first in extrafollicular foci and subsequently GCs, where B cells interact with follicular dendritic cells (FDCs) and T_{FH} cells expressing cytokines and co-stimulators, and some activated Th cells develop into T_{FH} cells (1, 2).

To further understand the cellular differences manifested in bulk GEP, we implemented Lymphoma EcoTyper to assign the cell states in each DLBCL (Supplementary Figure 1C). DLBCLs with high frequencies (≥2.4%) of normal B-cell clonotypes had significantly increased frequencies of the LE7 ecotype, B cell state S2 (preM (13)), CD4 T cell state S3 (naïve), T_{FH} cell

297 state S3 (*HLA-A*, *RPS7*, and *RPL13A* upregulation might contribute to the assignment), and
298 endothelial cell state S3 hich were associated with favorable clinical outcome significantly in
299 three to four DLBCL cohorts by a previous study (13), and significantly less frequent assignment
300 (according to the most abundant cell state) of B cell state S5 (plasmablast (13); this cell state
301 also expresses gene sets of normal dark-zone GC B cells (17)), T_{FH} cell state S1, and
302 endothelial cell state S1. In addition, DLBCLs with versus without >1% normal B-cell clonotypes
303 had a higher frequency of mast cell state S1 (**Figure 2D**, Supplementary Figure 1D). However,
304 the software does not differentiate malignant and normal B cells. The gene signature
305 upregulated in TIL-B-high patients (Supplementary Figure 1B) included both PreM-upregulated
306 and PreM-downregulated (*LCK*, *CD27*, *IL4R*, *TOX*, *SH3KBP1*, *SNX29P2*) gene signatures (17).
307
308 **The prognostic effects of TIL-B frequencies outweigh that of GCB/ABC**
309 In the prognostic analysis within the GCB/ABC subtypes, high frequencies of normal B-cell
310 clonotypes showed significantly favorable effects in both subtypes (**Figure 3A**). In contrast, the
311 GCB/ABC subtypes had no significant prognostic effects in either the TIL-B-low or TIL-B-high
312 patient groups, although they had starkly distinct GEP (Supplementary Figure 2).
313
314 The GCB subtype had a significantly higher proportion of TIL-B-high patients than the ABC
315 subtype. We further differentiated SHM^{-} (likely naïve B cells; however, we acknowledged that
316 not full IGV genes were sequenced) and SHM^{+} IGHV clonotypes (possibly memory/GC B-cells
317 or plasmablasts), and found that the ABC compared with GCB subtype had a significantly lower
318 median frequency of SHM^{-} clonotypes, and that ABC-DLBCLs had more SHM^{+} than SHM^{-}
319 normal B-cell clonotypes (**Figure 3A**).
320
321 A prominent GEP signature was identified for high TIL-B frequencies in GCB-DLBCL

(Supplementary Figure 3A, Supplementary Table 6). In contrast, in ABC-DLBCL, only a small number of DEGs showed statistical significance. TIL-B-low ABC cases had heterogeneously high expression of immune genes (**Figure 3B**, Supplementary Figure 3A-B), suggesting expression by other mechanisms compounding the analysis. We further compared patients with high *vs*. low frequencies of both SHM^{+} and SHM^{-} normal B-cell clonotypes. DEGs significant in both ABC- and GCB-DLBCL (**Figure 3C-D**, Supplementary Table 7) included *CXCL13*, *IL7R*, *CCL21*, *BCL11B*, *LPAR1*, *FOXP3*, *CTLA4*, *CD226* (which can attenuate Treg suppressive capacity (35)) and *MAF* (Th2 responses/T_{FH}/Th17). Only the signature identified in GCB-DLBCL included many genes involved in antigen degradation and presentation (which could suggest epitope spreading in TIL-B-high patients).

A53, MCD, and ST2 genetic subtypes showed significantly lower frequencies of normal B-cell clonotypes; however, the numbers of cases were too small. Only in the EZB subtype, high frequencies of normal B-cell clonotypes showed significant prognostic effects (**Figure 3E**).

Clonotypic diversity of TIL-Bs show significant favorable prognostic effects and immune gene signatures

High frequencies of normal B-cell clonotypes were significantly associated with high Shannon's entropy and normal B-cell clonotype numbers ($P < 0.0001$). Shannon's entropy correlated with significant prognostic effects (Supplementary Figure 4A), but not with GEP signatures. The number of normal B-cell clonotypes showed significant prognostic effects (significant with a wide range of cutoffs, ≥2 to ≥62) and distinct GEP signatures (**Figure 4A**, Supplementary Figure 4A-B**;** cutoff, ≥10 clonotypes) resembling those of normal B-cell clonotype frequencies in overall DLBCL and the GCB/ABC subsets. Additional DEGs identified included *SH2D1A* (SAP (1)), *KLRK1* (NKG2D), *RORA* (Th17; paradoxically, *IL17F* was downregulated), *PECAM1* (CD31),

347 etc. (Supplementary Table 8).

348

349 Furthermore, ABC had lower normal B-cell SHM^{-} clonotypic diversity than the GCB subtype.
350 The diversity status of normal B-cell SHM^{-} and SHM^{+} clonotypes (cutoff, ≥5) was consistent in
351 most GCB-DLBCL cases. In contrast, 12 ABC-DLBCL cases were high in SHM^{+}-clonotypic-
352 diversity yet low in SHM^{-}-clonotypic-diversity and had poor OS (**Figure 4B**; in 9 cases, the total
353 frequencies of normal B-cell SHM^{+} clonotypes were low). SHM^{-}-clonotypic-diversity-high cases
354 almost always had ≥5 SHM^{+} clonotypes of normal B-cells; SHM^{-}-clonotypic-diversity-low GCB-
355 DLBCL cases almost always had <5 SHM^{+} clonotypes of normal B-cells. Normal B-cell SHM^{-}
356 clonotypic diversity correlated with significantly higher frequencies of bulk normal B-cell
357 clonotypes and better patient survival in both GCB- and ABC-DLBCL (Supplementary Figure
358 5A). In contrast, in ABC-DLBCL, normal B-cell SHM^{+} clonotype diversity was not associated
359 with significant prognostic effects, nor with significantly higher frequencies of total or SHM^{-}
360 normal B-cell clonotypes as in GCB-DLBCL (Supplementary Figure 5B).

361

362 To further characterize the bulk TIL-Bs, we evaluated the clonal dominance and affinity potential
363 of the top three expanded normal B-cell clonotypes (**Figure 4C**). The high dominance of these
364 clonotypes in the normal B-cell compartment was associated with significantly poorer survival
365 and lower frequencies/diversity of the bulk normal B-cell clonotypes (**Figure 4C**, Supplementary
366 Figure 6A-C). The SHM status of the second/third largest normal B-cell clonotype was
367 associated with less clonal expansion and lower frequencies of bulk normal B-cell clonotypes,
368 and SHM percentage ≥1.28% of the second largest clonotype was associated with significantly
369 poorer survival (Supplementary Figure 6B-C). However, only the dominance of the largest
370 clonotype showed a significant GEP signature (Supplementary Figure 6A), whose up/down-
371 regulation was opposite to the signatures identified by high frequencies/diversity of bulk normal
372 B-cell clonotypes (Supplementary Table 8).

Comparisons among genetic subtypes showed some but not very significant differences in clonotype numbers, clonal dominance and SHM% of the second largest clonotypes, yet the case numbers were very small (Supplementary Figure 6D). We did not find significant associations between numbers of mutated genes by targeted NGS and Shannon's entropy or normal B-cell clonotype numbers/frequencies.

Increased TIL-B frequencies are associated with increased $CD4^+$ T cell infiltration in DLBCL

To visualize and quantify DLBCL and immune cells, fluorescent mIHC was performed on FFPE tissues (25). **Figure 5A** shows the H&E staining of lymph node tissues and three representative multispectral images of fluorescent mIHC for the example case in **Figure 1A** (a GCB-DLBCL; EZB subtype with *BCL2* and *BCL6* but not *MYC* gene rearrangements). In the immune landscape, we noticed some $CD20^+PAX5^-$ cells (enlarged circles in **Figure 5A**) clustered next to some $CD4^+CD3^-CD68^-$ cells (36). We speculate that these $CD20^+PAX5^-$ cells were normal pre-plasma cells in the GC light zone or at an even more differentiated stage, based on previous studies using normal reactive lymphoid tissues: PAX5 immunohistochemistry found that plasma cells were PAX5-negative (37), as well as the bulk of centrocytes within the GC light zone, likely pre-plasma cells as they were $IRF4^+$ and $Blimp\text{-}1^+$ indicating a commitment to the plasma cell lineage (38). In previous CD20 staining studies, most plasmablasts/plasma cells were CD20-negative (39); however most tonsil plasma cells were $CD20^+$ (40), and TLR9 stimulation of $CD27^+$ memory B cells isolated from tonsils or the peripheral blood *in vitro* generated $CD20^+$ pre-plasmablasts and subsequent $CD20^+$ plasmablasts/plasma cells favoring IgM secretion (41).

We analyzed the digitally quantified immune-cell data (25) and found that high frequencies of normal B-cell clonotypes were associated with significantly higher counts and frequencies of $CD3^+CD4^+$ T cells (also significant for Treg, memory, $FOXP3^-$, and $CD45RO^-$ CD4 TIL-T subsets), as well as a higher percentage of $FOXP3^+$ and CTLA-4^+ within the $CD3^+CD4^+$ TIL-T subset (Supplementary Figure 7A), but not with $CD3^+CD8^+$, macrophages, or NK cells. The clonotypic diversity of TIL-Bs showed similar, but less significant, associations.

Analysis in GCB/ABC-DLBCL found that in GCB-DLBCL, high frequencies and diversity of normal B-cell clonotypes were associated with significantly increased $CD3^+CD4^+$ and $CD3^+CD8^+$ TIL-Ts. In ABC-DLBCL, only high frequencies (but not the diversity) were associated with significantly increased $CD3^+CD4^+$ TIL-Ts (**Figure 5B**, **Figure 4C**, Supplementary Figure 7A). The high dominance of the largest normal B-cell clonotype was associated with significantly lower frequencies and counts of both $CD3^+CD4^+$ and $CD3^+CD8^+$ cells in overall DLBCL and the GCB/ABC subtypes (**Figure 4C**, Supplementary Figure 7B).

Owing to these associations, we performed two-factor prognostic analyses as shown in **Figure 5C** and Supplementary Figure 7C. TIL-B-low patients had poorer survival regardless of sufficient or deficient CD4/CD8 T-cell infiltration (using the cutoff from our previous study (25)). However, patients with sufficient TIL-Bs but not CD4 TIL-Ts (only 4 cases) also had significantly poorer OS, suggesting that both TIL-Bs and TIL-Ts were important.

We also examined the prognostic effects of PD-1 (owing to the clinical significance) in the context of TIL-B abundance. PD-1 expression (in T cells or B cells) exhibited significant adverse effects in TIL-B-low patients (**Figure 5C**, Supplementary Figure 8A). In contrast, in TIL-B-high ABC-DLBCL, PD-1^+ expression in $CD20^+$ B cells (cutoff, ≥5%; not differentiating malignant and normal B cells) and high PD-$1^+CD4^+CD3^+$ cell densities (absolute cell counts divided by the area

of imaged regions) were associated with significantly better OS and higher frequencies of normal B-cell clonotypes (**Figure 5C**, Supplementary Figure 8B). Such significantly favorable prognostic effects of PD-1 expression were not observed in TIL-B-high GCB-DLBCL.

Normal B-cell clonotype analysis in other sequenced cases also reveals significant favorable prognostic effects and gene signatures

IGH ultra-deep sequencing also resulted in sequence repertoires for 79 DLBCL cases, but with a predominant (median, 87%) DJ-only sequence without a productive VDJ sequence meeting the diagnostic clonotype criteria (24). As frequencies of all productive IGH-VDJ sequences in these repertoires were low, their total frequencies were used to estimate TIL-B frequencies for these 'DJ-only' cases, which turned out to show similar prognostic (favorable, Supplementary Figure 9A) and biological effects (including immune-response-enriched GEP signatures and increased $CD4^+$ TIL-Ts, Supplementary Table 5, **Figure 5D,** Supplementary Figure 9B-C) as in the 'VDJ' clonotype cases. Uniquely in these DJ-only cases but not their GCB subset (**Figure 3D**), estimated TIL-B-high status was associated with downregulation of *BLNK*, *IRF4* (possibly attributable to the association with GCB subtype; **Figure 5D,** Supplementary Table 9), and *CD79B* (Supplementary Table 10).

Moreover, we estimated TIL-B abundance by IG light-chain clonotypes (median/mean: 3.1%/8.2%) in 163 cases with a diagnostic IGK/L sequence identified (24). Similar prognostic and biological effects of normal B-cell clonotypes with the median as the cutof are shown in Supplementary Figure 10 and Supplementary Table 11.

In cases with both IG heavy and light chain clonotypes analyzed (**Figure 6A**), normal B-cell groups were discordant in 29% of IGH-VDJ cases and 38% of DJ-only cases, which could be

caused by different cutoffs for TIL-B-high, differences between IGK/L and IGH in gene rearrangements and the number of gene segments (light-chain clonotypes are more shared (20)), receptor editing (1, 2, 42), and other analytical factors.

Importantly, using IGK/L cases that were not included in the IGH cohort as a validation cohort, frequencies of normal B-cell clonotypes showed significant prognostic effects (**Figure 6A**).

Gene signatures identified by TIL-B frequencies show significant prognostic effects in external DLBCL cohorts

As no TIL-B data are available from previous DLBCL studies, to validate the role of TIL-Bs in external DLBCL cohorts, we used an indirect approach using gene signatures associated with TIL-B-high. We performed unsupervised clustering in NCI's RNA-seq cohort (11) and two GEO datasets (GSE10846 and GSE98588) using gene signatures identified by TIL-B-high status in our 138 VDJ cases, 79 DJ-only cases, and 163 IGK/L cases (referred as Signatures 1, 2, and 3, respectively), and correlate visualized clusters to the survival of patients treated with R-CHOP-like regimens. Each Signature showed significantly favorable effects in 2-3 validation cohorts (**Figure 6B-C**; Supplementary Figures 11-14); Signature 1 was probably the best. In GSE98588, Signature 1 was association with "host response (HR)" subtype (43) and a lower frequency of "BCR/proliferation" subtype; in GSE10846, the ABC subtype was associated with low Signature 1 expression (**Figure 6C**, Supplementary Figure 12).

To better understand the biology of these signatures, we performed similar unsupervised clustering in another GEO dataset, GSE12195, which had purified GC centroblasts and centrocytes as normal controls. Compared with Signatures 2-3, Signature 1 (**Figure 6D**) had a much higher proportion of genes highly expressed in normal GC B-cells (62% *vs*. 27.2% in

Signature 2 and 22% in Signature 3. Supplementary Figures 13A, 14A). Only Signature 2 had GC B-cell signature genes that were downregulated in TLB-high DLBCL cases (including *BLNK*, *CD79* and *IRF4*; Supplementary Figure 13A-C).

Furthermore, we performed clustering in GSE65135 with purified B cells and T cells from tonsils as normal controls for disaggregated follicular lymph node biopsies and peripheral blood mononuclear cells (PMBCs) from DLBCL and extranodal marginal zone lymphoma (EMZL) patients, which resulted in further dissected signatures. Most of the dissected B-cell signature subsets in Signature 1 (**Figure 6E**) and Signature 3 (Supplementary Figures 14B) but not Signature 2 (Supplementary Figure 13B) overlapped with their T-cell signature subsets. Signature 1 also included a B-cell signature subset not overlapping with the T-cell signature (yellow box in **Figure 6E**), but differed from Signature 2 in upregulation in TIL-B-high patients. Signatures 2 and 3 had a higher proportion of T-cell signature subsets (56.9% and 51%, respectively) than Signature 1 (32.8%). For genes not expressed in B and T cells, PMBC signatures (cyan boxes in **Figure 6E** and Supplementary Figures 13B, 14B) were identified, which accounted for a higher proportion of Signature 1 (11.6%) than Signatures 2 and 3 (6.5% and 6% respectively) and were mainly involved in antigen presentation (ubiquitin, ubiquitin ligases, proteasome/lysosomal degradation, endocytosis, trafficking, sorting) and adhesion.

Discussion

In this study, we assessed TIL-B-high status in DLBCLs based on the frequencies of normal B-cell clonotypes in IGH sequence repertoires, and found that high frequencies were significantly associated with not only favorable prognostic effects, but also distinct GEP signatures, B/CD4 T/endothelial cell states, and increased $CD4^+$ TIL-Ts (our immunophenotyping panel did not

evaluate DCs). The prognostic and biological effects of TIL-B-high status were consistent across different patient cohorts, supporting that TIL-B frequency is a new prognostic biomarker in DLBCL and an important player in anti-lymphoma immune responses. Both clonotypic diversity and clonal expansion after TIL-B activation might contribute to the high frequency of TIL-Bs. The cutoff used in this study, 2.4% of individual IGH repertoires, could indicate 2.4%-4.8% of total B-cells (as the repertoires may include different second alleles; **Figure 2C**). To calculate TIL-B frequencies, we once used the counts of productive IGH-VDJ sequences as the denominators, which also resulted in significant prognostic effects with an optimal cutoff of ≥4.5%, yet the GEP analysis identified less significant DEGs (Supplementary Figure 15, Supplementary Table 4, which also had very significant prognostic effects in two validation GEO cohorts; data not shown), therefore we adopted the current evaluation method to stratify patients. It is possible that in some cases, normal B-cell clonotype frequencies in the entire sequence repertoire may correlate better with the TIL-B abundance, probably because our DNA samples were extracted from FFPE tissues with increased cross-linking and fragmented DNAs.

Despite the association of TIL-B-high status with increased TIL-Ts, our results suggest that TIL-Bs are not simply bystanders of antitumor T-cell responses. In our cohort, a large number of patients with sufficient TIL-Ts but not TIL-Bs had as poor survival as TIL-T-deficient patients; TIL-T abundance itself showed much less prognostic significance in our cohorts with limited cutoff options and significance only in ABC- but not GCB-DLBCL (Supplementary Figure 16); ABC compared with GCB subset of TIL-B-low cases had increased $CD8^+$ T cells (Supplementary Figure 10C) but not improvement in OS/PFS. Likewise, in GSE10846 and the NCI validation cohort (11), T-cell gene signature alone did not predict good prognosis (white boxes in **Figure 7B** and Supplementary Figures 11, 14).

B-cell functions include antigen presentation, T-cell priming (44) (controversial) and

differentiation, generation of CD4 T-cell memory (45), antibody-mediated immune responses, transcytosis (46), and suppression of myeloid-derived suppressor cells (47), which vary among different B-cell subsets. Non-protein antigens (polysaccharides, lipids, and nucleic acids) may activate extrafollicular B-cell responses without T-cell help, whereas in T-dependent B-cell activation, B cells internalize protein antigens and present processed protein fragments, which are different from the epitopes recognized by BCRs, on their MHC-class II to Th cells. Our results suggest that T-dependent B-cell activation was the main contributor to the favorable prognostic effects of TIL-Bs, as cases with TIL-B^{+}TIL-T^{-} or high dominance of the largest normal B-cell clonotype likely expanded T cell-independently had poor prognosis. TIL-Bs may be important for T-cell responses, as CD4 TIL-Ts were significantly increased in TIL-B-high cases, many T-cell activation and Th response genes were upregulated, and PD-1 expression had adverse prognostic effects in TIL-B-low patients. However, we did not observe significant associations of TIL-B clonotype frequencies/diversity with genomic mutation numbers (neoantigens numbers), suggesting that complex linked recognition may underlie the association between TIL-B and TIL-T frequencies. The T-dependent antigens for B-cell activation in DLBCL warrant further investigation.

The identification of TIL-B biomarker in DLBCL has significant implications. It could inform biomarker studies to consider TIL-Bs as a confounding factor, albeit low abundance, owing to their critical role. For example, the prognostic difference between GCB and ABC in our IGH sequencing cohort appeared to result from the higher proportion of TIL-B-high patients in GCB-DLBCL; however, validation studies are needed. Moreover, it could inform treatment strategies. For example, many genes related to antigen degradation and processing were upregulated in TIL-B-high patients, suggesting that proteasome inhibitors may dampen TIL-B functions.

A previous study showed that a single dose of rituximab did not significantly affect the $CD19^{+}$ B-

cell percentage in secondary lymphoid organs as in the peripheral blood, and the remaining B cells ($CD19^{+}CD20^{-}$) predominantly had a switched memory phenotype ($IgD^{-}CD27^{+}$) with altered Ig subclass production after stimulation (48). A recent study (49) found that CD20 is an organizer of the IgD-class nanocluster (including CD19, CD81, CD40, etc.) and a gatekeeper of the resting B-cell state (so is IgD (50)), and rituximab treatment of Ramos and naïve human B cells resulted in transient B-cell activation (49). In GSE65135, BCR-related genes and *IL4R* were downregulated immediately after rituximab treatment in one EMZL patient's PMBCs; however, some genes in the B-cell signature were upregulated (*C16orf54*, *RASSF5*, *TES*, *TXNIP*, *GVINP1*, **Figure 6D**), so were *GIMAP4*, *GIMAP6, GIMAP7*, *GIMAP8*, and *SAMHD1* in the T-cell signature (Supplementary Figures 13-14).

In summary, we showed for the first time that high frequencies of DLBCL-infiltrating normal B cells (associated with clonotypic diversity) have significant favorable prognostic and immunological effects in a large cohort of DLBCL patients. We conclude that normal B-cell activation and function through interactions with T cells are critical for optimal antitumor immune responses and clinical outcome of patients with DLBCL, and have significant clinical and therapeutic implications.

Authors' Disclosure

T.S. and I.K. were employees of Adaptive Biotechnologies during their contributions to this research**.** QA is an employee of NeoGenomics Laboratories, Inc..

Other authors declare no potential conflicts of interest.

Authors' contributions

Z.Y.X-M and K.H.Y. designed the study; Z.Y.X-M., T.S., Q.A., I.K. and K.H.Y. performed the experiments and analyzed the data, Y.L., T.Y., and T.L. participated in light-chain clonotype

analysis and/or validation cohort analysis; A.T., C.V., G.B., W.Q., K.D., A.C., W.T., Y.Z., E.D.H., F.B.H., Y.W., H.G., M.P., A.JM.F., M.B.M, B.M.P., X.F., J.H.vK, M.A.P., J.N.W., Q.A., I.K., M.Z., J.S., B.X., and K.H.Y. contributed to the patient samples, clinical data, and analytical tools; and Z.Y.X-M. and K.H.Y. wrote the manuscript with contributions and approval from all the authors.

Acknowledgements

This study was supported by grants R01CA233490, R01CA187415, R01CA138688, and RC1CA146299 from the National Cancer Institute/National Institutes of Health. KHY was also supported by The Duke University Institutional Research Grant Award and the Hagemeister Lymphoma Foundation.

The authors thank the patients, their families, supporting staff, technicians, and the availability of the LymphGen and Ecotyper software. Xiaoya Xu-Monette created Figure 2C using Adobe Photoshop software.

References

1. Murphy K., Weaver C. *Janeway's Immunobiology*, Tenth Edition. New York, *W. W. Norton & Company*; 2022.
2. Abbas AK, Lichtman AH, Pillai S. *Cellular and Molecular Immunology*, Tenth Edition. Philadelphia, PA: Elsevier; 2022.
3. Fridman WH, Meylan M, Petitprez F, Sun CM, Italiano A, Sautès-Fridman C. B cells and tertiary lymphoid structures as determinants of tumour immune contexture and clinical outcome. Nat Rev Clin Oncol 2022;19:441-457.
4. Schumacher TN, Thommen DS. Tertiary lymphoid structures in cancer. Science 2022;375:eabf9419.
5. Yu K, Ravoor A, Malats N, Pineda S, Sirota M. A Pan-Cancer Analysis of Tumor-Infiltrating B Cell Repertoires. Front Immunol 2021;12:790119.
6. Garaud S, Zayakin P, Buisseret L, Rulle U, Silina K, de Wind A, et al. Antigen Specificity and Clinical Significance of IgG and IgA Autoantibodies Produced in situ by Tumor-Infiltrating B Cells in Breast Cancer. Front Immunol 2018;9:2660.
7. Meylan M, Petitprez F, Becht E, Bougoüin A, Pupier G, Calvez A, et al. Tertiary lymphoid structures generate and propagate anti-tumor antibody-producing plasma cells in renal cell cancer. Immunity 2022;55:527-541 e525.
8. Laumont CM, Banville AC, Gilardi M, Hollern DP, Nelson BH. Tumour-infiltrating B cells: immunological mechanisms, clinical impact and therapeutic opportunities. Nat Rev Cancer 2022;22:414-430.

9. Vanderlugt CJ, Miller SD. Epitope spreading. Curr Opin Immunol 1996;8:831-836.
10. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature 2000;403:503-511.
11. Schmitz R, Wright GW, Huang DW, Johnson CA, Phelan JD, Wang JQ, et al. Genetics and Pathogenesis of Diffuse Large B-Cell Lymphoma. N Engl J Med 2018;378:1396-1407.
12. Chapuy B, Stewart C, Dunford AJ, Kim J, Kamburov A, Redd RA, et al. Molecular subtypes of diffuse large B cell lymphoma are associated with distinct pathogenic mechanisms and outcomes. Nat Med 2018;24:679-690.
13. Steen CB, Luca BA, Esfahani MS, Azizi A, Sworder BJ, Nabet BY, et al. The landscape of tumor cell states and ecosystems in diffuse large B cell lymphoma. Cancer Cell 2021;39:1422-1437 e1410.
14. Chong EA, Ruella M, Schuster SJ. Five-Year Outcomes for Refractory B-Cell Lymphomas with CAR T-Cell Therapy. N Engl J Med 2021;384:673-674.
15. Spiegel JY, Dahiya S, Jain MD, Tamaresis J, Nastoupil LJ, Jacobs MT, et al. Outcomes of patients with large B-cell lymphoma progressing after axicabtagene ciloleucel therapy. Blood 2021;137:1832-1835.
16. Ye X, Wang L, Nie M, Wang Y, Dong S, Ren W, et al. A single-cell atlas of diffuse large B cell lymphoma. Cell Rep 2022;39:110713.
17. Holmes AB, Corinaldesi C, Shen Q, Kumar R, Compagno N, Wang Z, et al. Single-cell analysis of germinal-center B cells informs on lymphoma cell of origin and outcome. J Exp Med 2020;217.
18. Roider T, Seufert J, Uvarovskii A, Frauhammer F, Bordas M, Abedpour N, et al. Dissecting intratumour heterogeneity of nodal B-cell lymphomas at the transcriptional, genetic and drug-response levels. Nat Cell Biol 2020;22:896-906.
19. Riley RL, Wylie DE, Klinman NR. B cell repertoire diversification precedes immunoglobulin receptor expression. J Exp Med 1983;158:1733-1738.
20. Soto C, Bombardi RG, Branchizio A, Kose N, Matta P, Sevy AM, et al. High frequency of shared clonotypes in human B cell receptor repertoires. Nature 2019;566:398-402.
21. Visco C, Li Y, Xu-Monette ZY, Miranda RN, Green TM, Li Y, et al. Comprehensive gene expression profiling and immunohistochemical studies support application of immunophenotypic algorithm for molecular subtype classification in diffuse large B-cell lymphoma: a report from the International DLBCL Rituximab-CHOP Consortium Program Study. Leukemia 2012;26:2103-2113.
22. Xu-Monette ZY, Wei L, Fang X, Au Q, Nunns H, Nagy M, et al. Genetic Subtyping and Phenotypic Characterization of the Immune Microenvironment and MYC/BCL2 Double Expression Reveal Heterogeneity in Diffuse Large B-cell Lymphoma. Clin Cancer Res 2022;28:972-983.
23. Wright GW, Huang DW, Phelan JD, Coulibaly ZA, Roulland S, Young RM, et al. A Probabilistic Classification Tool for Genetic Subtypes of Diffuse Large B Cell Lymphoma with Therapeutic Implications. Cancer Cell 2020;37:551-568 e514.
24. Xu-Monette ZY, Li J, Xia Y, Crossley B, Bremel RD, Miao Y, et al. Immunoglobulin somatic hypermutation has clinical impact in DLBCL and potential implications for immune checkpoint blockade and neoantigen-based immunotherapies. J Immunother Cancer 2019;7:272.
25. Xu-Monette ZY, Xiao M, Au Q, Padmanabhan R, Xu B, Hoe N, et al. Immune Profiling and Quantitative Analysis Decipher the Clinical Role of Immune-Checkpoint Expression in the Tumor Immune Microenvironment of DLBCL. Cancer Immunol Res 2019;7:644-657.

26. Alaggio R, Amador C, Anagnostopoulos I, Attygalle AD, Araujo IBO, Berti E, et al. The 5th edition of the World Health Organization Classification of Haematolymphoid Tumours: Lymphoid Neoplasms. Leukemia 2022;36:1720-1748.
27. Al-Alwan M, Du Q, Hou S, Nashed B, Fan Y, Yang X, et al. Follicular dendritic cell secreted protein (FDC-SP) regulates germinal center and antibody responses. J Immunol 2007;178:7859-7867.
28. Lee JH, Choi SY, Park SY, Jung NC, Noh KE, Nam JH, et al. Enpp2 Expression by Dendritic Cells Is a Key Regulator in Migration. Biomedicines 2021;9.
29. Lin S, Haque A, Raeman R, Guo L, He P, Denning TL, et al. Autotaxin determines colitis severity in mice and is secreted by B cells in the colon. FASEB J 2019;33:3623-3635.
30. Cismasiu VB, Ghanta S, Duque J, Albu DI, Chen HM, Kasturi R, et al. BCL11B participates in the activation of IL2 gene expression in CD4+ T lymphocytes. Blood 2006;108:2695-2702.
31. Baldanzi G, Pighini A, Bettio V, Rainero E, Traini S, Chianale F, et al. SAP-mediated inhibition of diacylglycerol kinase α regulates TCR-induced diacylglycerol signaling. J Immunol 2011;187:5941-5951.
32. Johansen KH, Golec DP, Huang B, Park C, Thomsen JH, Preite S, et al. A CRISPR screen targeting PI3K effectors identifies RASA3 as a negative regulator of LFA-1-mediated adhesion in T cells. Sci Signal 2022;15:eabl9169.
33. Cerutti A, Puga I, Cols M. New helping friends for B cells. Eur J Immunol 2012;42:1956-1968.
34. Cyster JG. B cell follicles and antigen encounters of the third kind. Nat Immunol 2010;11:989-996.
35. Wang N, Yi H, Fang L, Jin J, Ma Q, Shen Y, et al. CD226 Attenuates Treg Proliferation via Akt and Erk Signaling in an EAE Model. Front Immunol 2020;11:1883.
36. Mebius RE, Rennert P, Weissman IL. Developing lymph nodes collect CD4+CD3-LTbeta+ cells that can differentiate to APC, NK cells, and follicular cells but not T or B cells. Immunity 1997;7:493-504.
37. Krenacs L, Himmelmann AW, Quintanilla-Martinez L, Fest T, Riva A, Wellmann A, et al. Transcription factor B-cell-specific activator protein (BSAP) is differentially expressed in B cells and in subsets of B-cell lymphomas. Blood 1998;92:1308-1316.
38. Cattoretti G, Shaknovich R, Smith PM, Jäck HM, Murty VV, Alobeid B. Stages of germinal center transit are defined by B cell transcription factor coexpression and relative abundance. J Immunol 2006;177:6930-6939.
39. Leandro MJ. B-cell subpopulations in humans and their differential susceptibility to depletion with anti-CD20 monoclonal antibodies. Arthritis Res Ther 2013;15 Suppl 1:S3.
40. Forsthuber TG, Cimbora DM, Ratchford JN, Katz E, Stüve O. B cell-based therapies in CNS autoimmunity: differentiating CD19 and CD20 as therapeutic targets. Ther Adv Neurol Disord 2018;11:1756286418761697.
41. Geffroy-Luseau A, Chiron D, Descamps G, Jégo G, Amiot M, Pellat-Deceunynck C. TLR9 ligand induces the generation of CD20+ plasmablasts and plasma cells from CD27+ memory B-cells. Front Immunol 2011;2:83.
42. Vettermann C, Schlissel MS. Allelic exclusion of immunoglobulin genes: models and mechanisms. Immunol Rev 2010;237:22-42.
43. Monti S, Savage KJ, Kutok JL, Feuerhake F, Kurtin P, Mihm M, et al. Molecular profiling of diffuse large B-cell lymphoma identifies robust subtypes including one characterized by host inflammatory response. Blood 2005;105:1851-1861.
44. Castiglioni P, Gerloni M, Cortez-Gonzalez X, Zanetti M. CD8 T cell priming by B lymphocytes is CD4 help dependent. Eur J Immunol 2005;35:1360-1370.

45. Whitmire JK, Asano MS, Kaech SM, Sarkar S, Hannum LG, Shlomchik MJ, et al. Requirement of B cells for generating CD4+ T cell memory. J Immunol 2009;182:1868-1876.
46. Biswas S, Mandal G, Payne KK, Anadon CM, Gatenbee CD, Chaurio RA, et al. IgA transcytosis and antigen recognition govern ovarian cancer immunity. Nature 2021;591:464-470.
47. Vito A, Salem O, El-Sayes N, MacFawn IP, Portillo AL, Milne K, et al. Immune checkpoint blockade in triple negative breast cancer influenced by B cells through myeloid-derived suppressor cells. Commun Biol 2021;4:859.
48. Kamburova EG, Koenen HJ, Borgman KJ, ten Berge IJ, Joosten I, Hilbrands LB. A single dose of rituximab does not deplete B cells in secondary lymphoid organs but alters phenotype and function. Am J Transplant 2013;13:1503-1511.
49. Kläsener K, Jellusova J, Andrieux G, Salzer U, Böhler C, Steiner SN, et al. CD20 as a gatekeeper of the resting state of human B cells. Proc Natl Acad Sci U S A 2021;118.
50. Noviski M, Mueller JL, Satterthwaite A, Garrett-Sinha LA, Brombacher F, Zikherman J. IgM and IgD B cell receptors differentially respond to endogenous antigens and control B cell fate. Elife 2018;7.

Figure legends

Figure 1. Composition of immunoglobulin heavy chain gene (IGH) sequence repertoires amplified from DNA samples extracted from DLBCL diagnostic tissues. (**A**) Left: A scatterplot illustrating the diverse clonotypes in an example IGH repertoire, which included a predominant productive IGH sequence and its two derivative sequences likely derived from lymphoma cells, other productive IGH sequences likely derived from normal B cells, unproductive VDJ sequences due to either out-of-frame insertion/deletions or mutations resulting in premature Stop codons, and DJ-only sequences (V genes were inferred from IMGT) that could be derived from non-functional alleles or fragmented V genes in FFPE tissues. Note: The size of dots reflects, but are not scaled to, the normalized read count of each sequence. Middle: A stacked column chart showing the fraction of five categories of IGH sequences in this repertoire calculated by the normalized sequence read counts. Right: a column chart showing the normalized read counts of normal B-cell-derived clonotypes (range, 2-532 cell counts [2 was the cutoff to be included in the clonotype analysis]. For comparison, counts for the diagnostic clonotype and two derivative clonotypes were 11,709, 37 and 8, respectively). The somatic hypermutation (SHM) status of each clonotype is indicated below the chart. (**B**) A heatmap to visualize the Log2 transformed frequencies of five categories of sequences calculated from normalized read counts in individual IGH repertoires of 138 patients with DLBCL. Each column represents one patient's repertoire. (**C**) A pie chart showing the mean frequencies of five categories of IGH sequences in the 138 repertoires.

Figure 2. Prognostic and gene expression profiling (GEP) analyses for the frequencies of normal B-cell IGH clonotypes in 138 patients with DLBCL. (**A**) Overall survival (OS) and progression-free survival (PFS) curves for high frequencies of normal B-cell clonotypes using a 2.4% cutoff and OS using a 1% cutoff. (**B**) Heatmap showing the significantly differentially expressed genes between patients with and without high frequencies (≥2.4%) of normal B-cell

clonotypes with q < 0.01 and local false discovery rate <0.025 by SAM. (**C**) Illustration of T-dependent B-cell activation suggested by the GEP signature associated with high frequencies of normal B-cell clonotypes. For simple illustration, molecules associated with T:B cell interaction and activation are depicted on a same T cell and a same B cell, even though the T_{FH} cell is not extrafollicular. Most molecules illustrated are included the heatmap in panel **B**, except for IL6ST with a higher q (0.0157, local false discovery rate: 0.0257), EBI2 and SAP which are significant in other TIL-B-associated signatures identified in this study (Supplementary Table 4 and Supplementary Table 8). (**D**) Comparisons between DLBCLs with and without high frequencies (≥2.4%) of normal B-cell clonotypes regarding B cell state (S) abundance and the T cell state assignment for DLBCL cases. Significance: **: $P < 0.01$; ***: $P < 0.005$. In the scatter plot, each dot represent one patient. The stacked column chart shows the proportion of patients with a T cell state assignment according to the most abundant T cell state in that case. Abbreviation: U.C. unclassified.

Figure 3. Prognostic and GEP analyses for normal B-cell IGH clonotypes in DLBCL cell-of-origin and genetic subtypes. (**A**) Left: High frequencies of normal B-cell clonotypes showed significant favorable prognostic effects in both GCB and ABC subtypes, whereas GCB/ABC cell-of-origin did not show significant effects in two groups stratified by normal B-cell clonotype frequencies. Right: Scatter plots showing the distribution (dots) and mean (bars) frequencies of SHM^- IGH normal B-cell clonotypes in GCB and ABC subtypes of DLBCL; in the ABC subtype, SHM^+ IGH normal B-cell clonotypes had higher frequencies of SHM^+ than SHM^- IGH normal B-cell clonotypes by Wilcoxon matched paired signed rank test. (**B**) Scatter plots showing the distribution and mean transcriptional expression levels of immune genes in GCB/ABC patients with high/low frequencies of normal B-cell (B_n) clonotypes. ABC compared with GCB patients with B_n Low status had heterogeneously higher expression of immune genes. *P* values by Mann-Whitney test (2-tailed) are also provided. The expression values on the y axis are median

centered log2(Robust Multi-chip Average [RMA] expression measure) processed and normalized by R package affy. (**C**) Heatmap visualizing differentially expressed genes (DEGs) between ABC-DLBCL cases with both high frequencies of SHM^{-} and SHM^{+} IGH normal B-cell clonotypes and those with both low. These genes were not statistically significant in SAM analysis. (**D**) Left: Heatmap of significant DEGs of GCB-DLBCL cases with frequencies of SHM^{-} and SHM^{+} IGH normal B-cell clonotypes both being high *vs*. both low ($q < 0.01$, local false discovery rate ≤ 0.03). Right: Heatmap showing significant DEGs of high *vs*. low frequencies of productive VDJ clonotypes in the sequence repertoires of 'DJ-only' GCB-DLBCL cases. (**E**) A scatter plot showing the distribution (dots) and mean (bars) frequencies of normal B-cell clonotypes in genetic subtypes. Due to the small case numbers, A53, MCD and ST2 cases were combined for statistical significance in comparisons. *P* values by Mann-Whitney test (2-tailed) are also provided. High frequencies of normal B-cell clonotypes correlated with significantly better OS in the EZB genetic subtype.

Figure 4. Prognostic and GEP analyses for IGH clonotypic diversity of TIL-Bs in 138 patients with DLBCL. (**A**) High IGH clonotypic diversity of TIL-Bs (≥10 clonotypes) was associated with significantly better survival. (**B**) Survival analysis for IGH clonotypic diversity of TIL-Bs dissecting SHM^{+} and SHM^{-} clonotypes in GCB- and ABC-DLBCL. With a cutoff of ≥5 clonotypes for high diversity, the diversity status of SHM^{+} and SHM^{-} normal B-cell clonotypes was consistent in most GCB-DLBCL cases; in contrast, 28.6% of ABC-DLBCL cases with ≥5 SHM^{+} clonotypes (12 patients) had low-diversity (<5) of SHM^{-} clonotypes of normal B cells and significantly poorer survival. The scatter plot on the right side of survival curves in ABC-DLBCL shows that the ABC subtype had significantly lower numbers of SHM^{-} IGH normal B-cell clonotypes compared to the GCB subtype (unpaired *t* test; *P* value by Mann-Whitney test is also provided). (**C**) Top: A plot showing mutation status of the top 3 most expanded normal B-cell clonotypes in 131 patients with normal B-cell clonotypes (each column represents one patient).

Middle: High dominance of the largest normal B-cell clonotype in bulk TIL-Bs was associated with significantly poorer OS, lower clonotype numbers, and lower total frequencies of normal B-cell clonotypes. Bottom: High clonotypic diversity (≥10) of normal B-cells were associated with significantly higher frequencies of $CD3^+CD4^+$ T cells in GCB-DLBCL. High dominance of the largest normal B-cell clonotype was associated with significantly lower frequencies of $CD3^+CD4^+$ T cells in the GCB/ABC subtypes and with significantly lower frequencies of $CD3^+CD8^+$ T cells in the GCB subtype only. Note: the bars indicate mean levels; *P* values by Mann-Whitney test (2-tailed) are also provided.

Figure 5. Fluorescent multiplex immunohistochemistry (mIHC) and correlative and prognostic analysis for the quantified immunofluoresence. (**A**) Representative mIHC images and corresponding H&E staining of a tissue core from the example case shown in Figure 1A. Original magnification 20x; each field of view: 2560x2160 pixels (~0.832x0.702mm). Two representative areas enriched with $PAX5^-CD20^+$ cells are marked by white circles in three images and enlarged for the third image, which can also differentiate the CD68 status of $CD4^+CD3^-$ cells. (**B**) High frequencies of normal B-cell clonotypes (cutoff: ≥2.4%) were associated with significantly higher frequencies of $CD3^+CD4^+$ TIL-Ts in the GCB/ABC subtypes of DLBCL (and overall DLBCL) and with higher frequencies of $CD3^+CD8^+$ T cells in the GCB subtype only (but not in the ABC subtype and overall DLBCL). Note: the bars indicate mean levels; *P* values by Mann-Whitney test (2-tailed) are also provided. (**C**) Left: Prognostic analysis with two factors in overall cohort. One factor is high/low frequency status of normal B-cell clonotypes (cutoff: ≥2.4%), and the other factor is deficiency/proficiency status of $CD4^+CD3^+$ cells (cutoff: >1.6% of total cell counts of $PAX5^+$ cells, $CD3^+$ cells, $CD68^+$ cells, and $CD56^+$ cells), PD-1^+ percentage expression in $CD3^+$ cells (cutoff: >50% of $CD3^+$ cells), or PD-1^+ percentage expression in $CD20^+$ cells (cutoff: ≥5% of $CD20^+$ cells). Right: In ABC-DLBCL, PD-1^+ expression in $CD20^+$ cells showed opposite prognostic effects in TIL-B-low and TIL-B-high

cases, as well as associations of opposite trends of frequencies of normal B-cell clonotypes. (**D**) In 'DJ-only' cases (cases with a predominant IGH DJ-only sequence without a diagnostic VDJ sequence identified), GCB compared with ABC subset had a significantly higher mean frequency of productive VDJ clonotypes in the IGH sequence repertoires. High frequencies of productive VDJ clonotypes in these repertoires were associated with significantly higher frequencies of $CD3^{+}CD4^{+}$ TIL-Ts quantified by mIHC.

Figure 6. Validation of the prognostic effects of TIL-Bs in other DLBCL cohorts. (**A**) Cases with TIL-Bs already evaluated by IG heavy-chain clonotypes were viewed as the training cohort, and new cases with only IG light-chain clonotypes analyzed were used as a validation cohort. High frequencies of normal B-cell light-chain clonotypes were associated with significantly better OS in both the training and validation cohorts. The cutoff for high frequency was same as the median value in the training cohort and the overall light-chain cases (Supplementary Figure 10). The plot below illustrates the relationship between the analyzed IGH and IG light-chain cases. Each column represents one patient. (**B**) Validating the prognostic effect of a gene signature identified in this study (Signature 1, Supplementary Table 3) in 233 DLBCL cases treated with R-CHOP-like regimen from the GSE10846 dataset (LLMPP validation cohort). Patients with high expression identified by unsupervised clustering had significantly better overall survival. However, a subcluster shown in a white box had no significant prognostic effect albeit upregulation of T-cell signature genes. Notes: Not all genes were labeled due to space limitation. For labeled genes, blue color indicates high expression in normal B-cell signatures in figure panels **D** and/or **E**; green color indicates high expression in the T-cell but not B-cell signature in panel **E**. Some genes have both color, due to the different clustering results in panels **D** and **E**. (**C**) Validating Signature 1 in the GSE98588 dataset (Harvard validation cohort, 137 cases. Not all cases had OS and PFS data). Visualized cluster of high expression by unsupervised clustering was associated with significantly better OS and PFS and higher

proportion of HR ("host response") subtype in this cohort. BCR/proliferation subtype of DLBCL was associated with low expression of the gene signature identified in this study. (**D-E**) Unsupervised clustering in the GSE12195 and GSE65135 datasets using the same gene signature identified in this study. Blue color was used for labeling genes showing high expression in B cells (purified germinal center B cells in the panel **D**, and tonsil B cells in the panel **E**). For the rest of genes, genes showing high expression in tonsil T cells are in green color, genes shown in the PMBC cluster are in cyan color, and the remaining follicular lymphoma cluster genes are in red color. Abbreviations: CB, centroblasts; CC, centrocytes; PBMC, peripheral blood mononuclear cell; FL, follicular lymphoma; LN, lymph node.

Figure 1

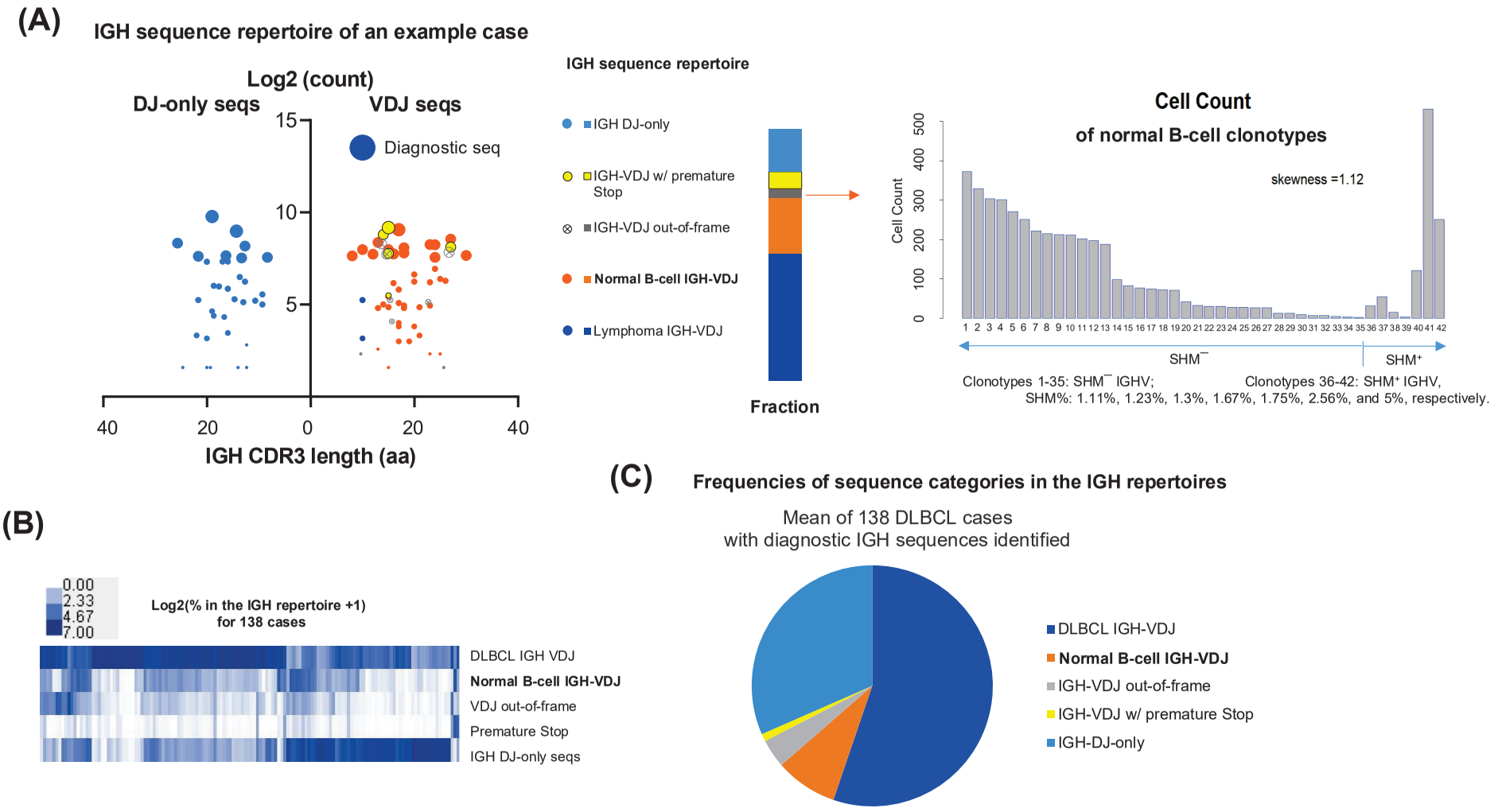

Figure 2

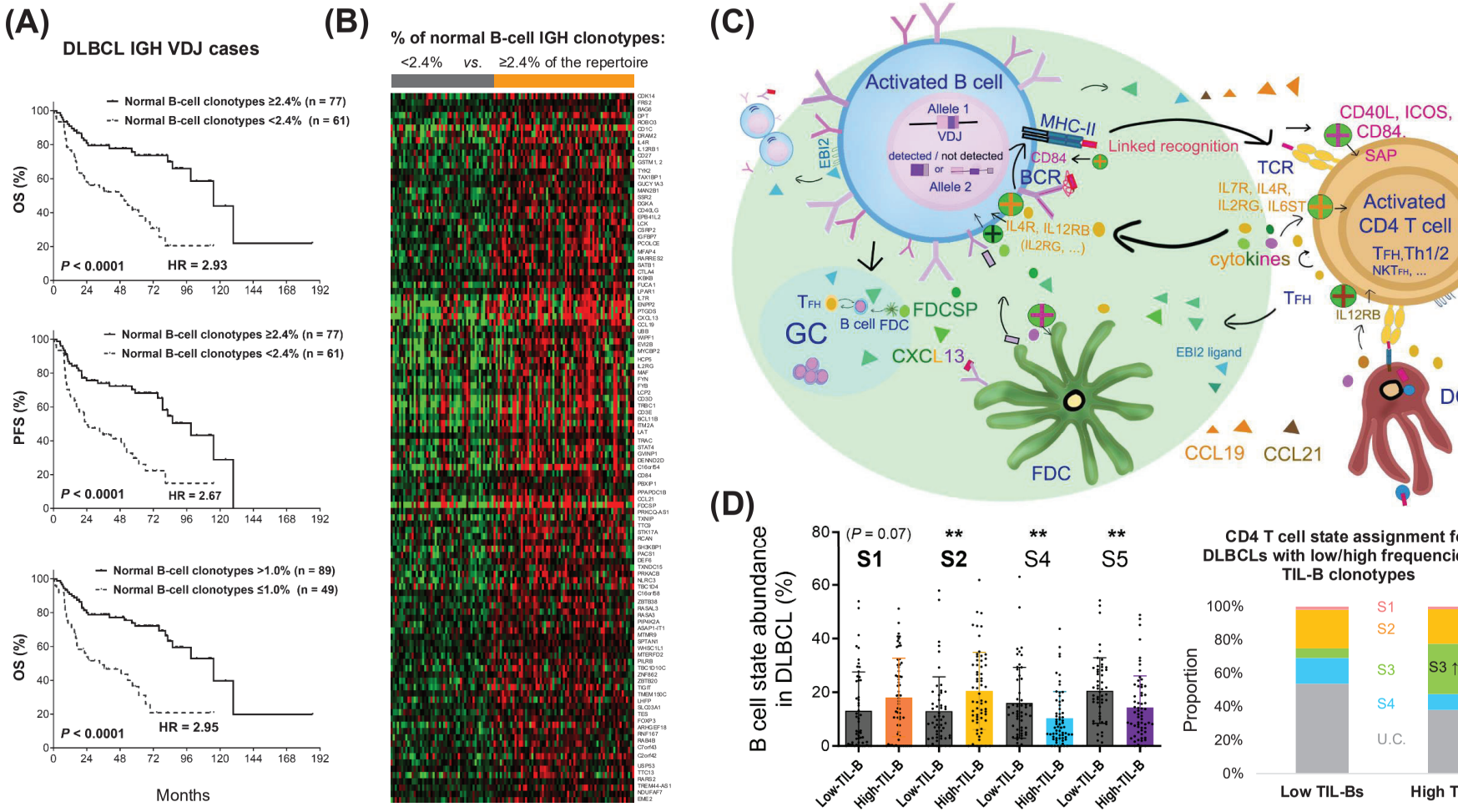

Figure 3

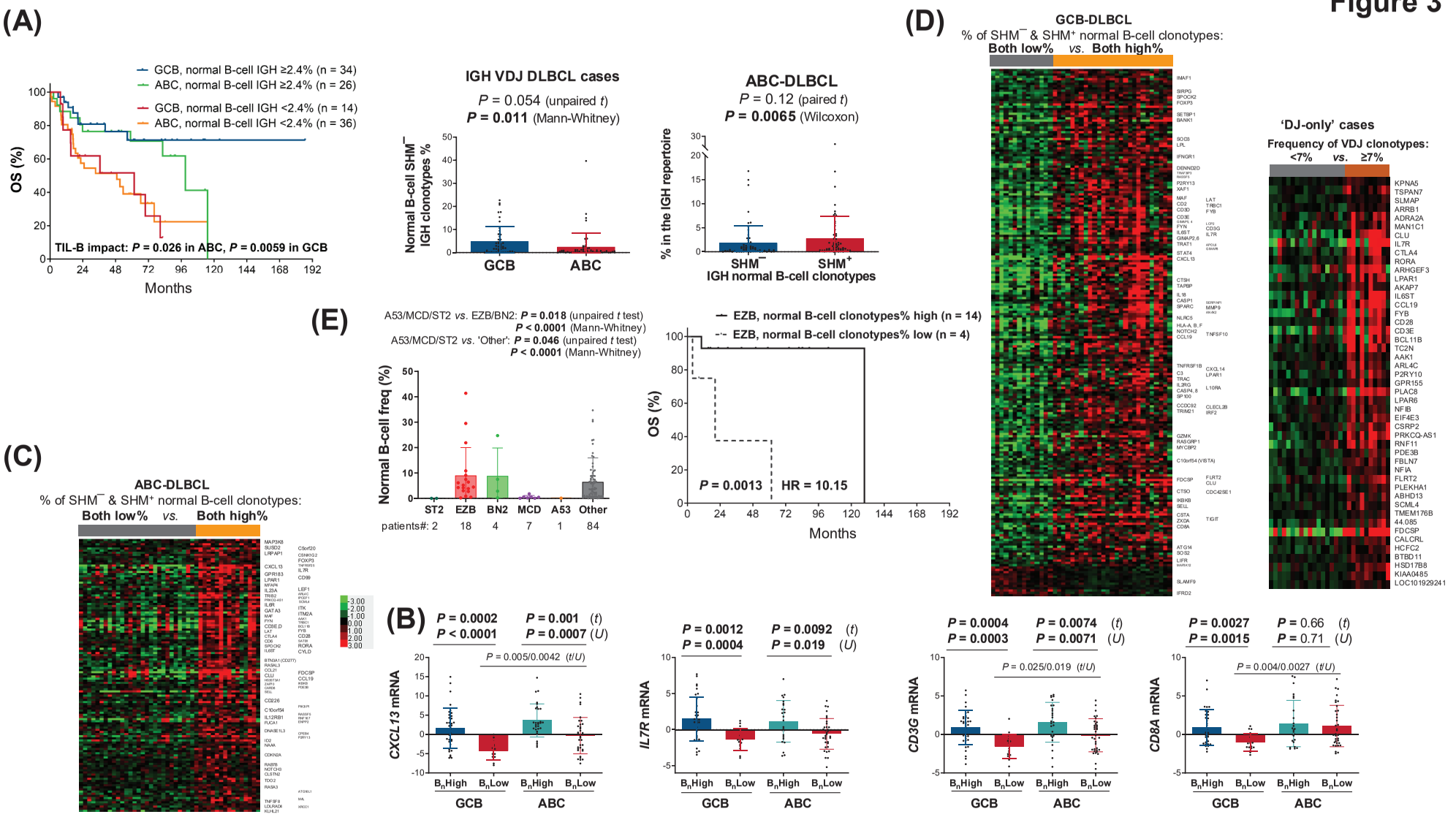

Figure 4

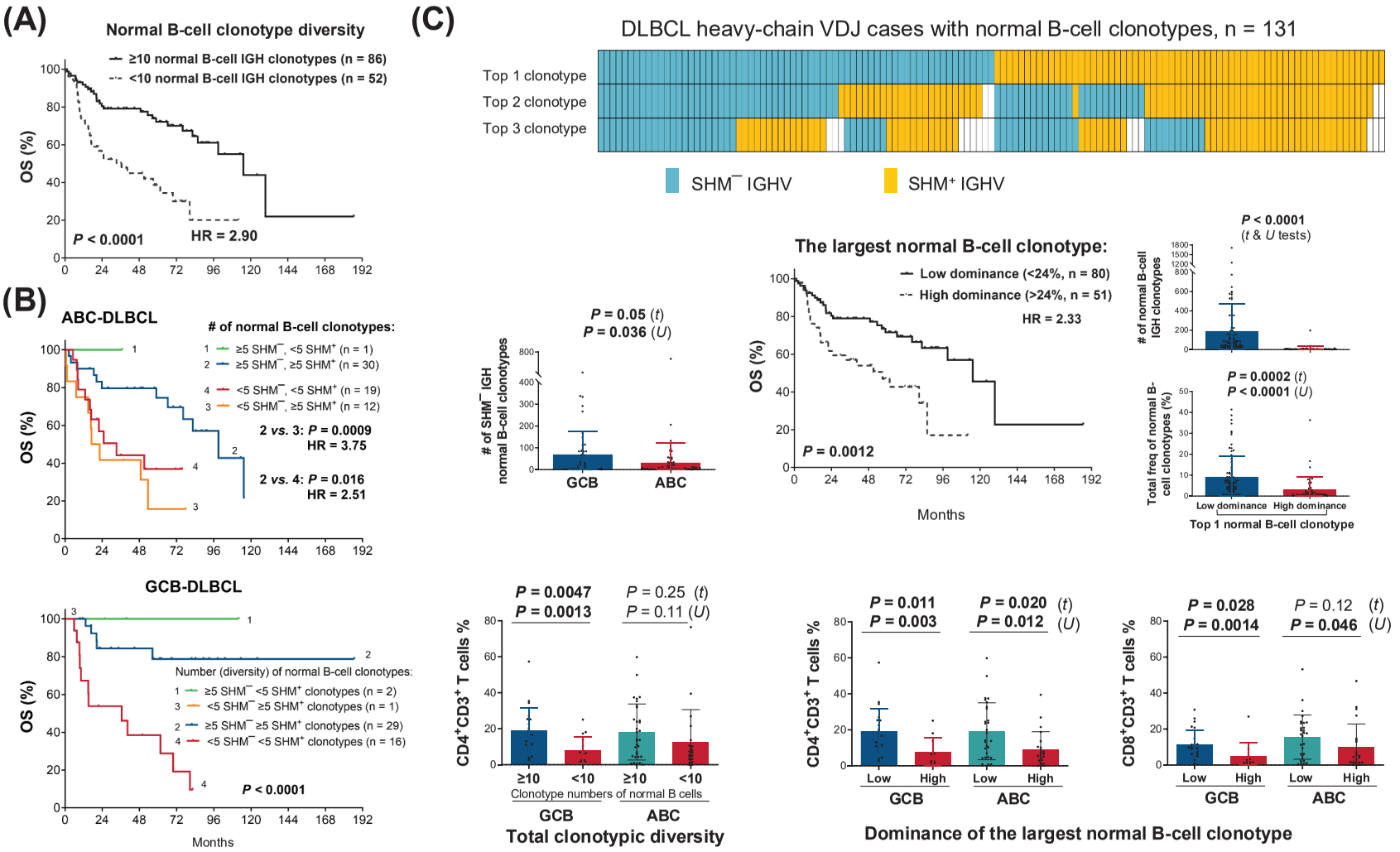

Figure 5

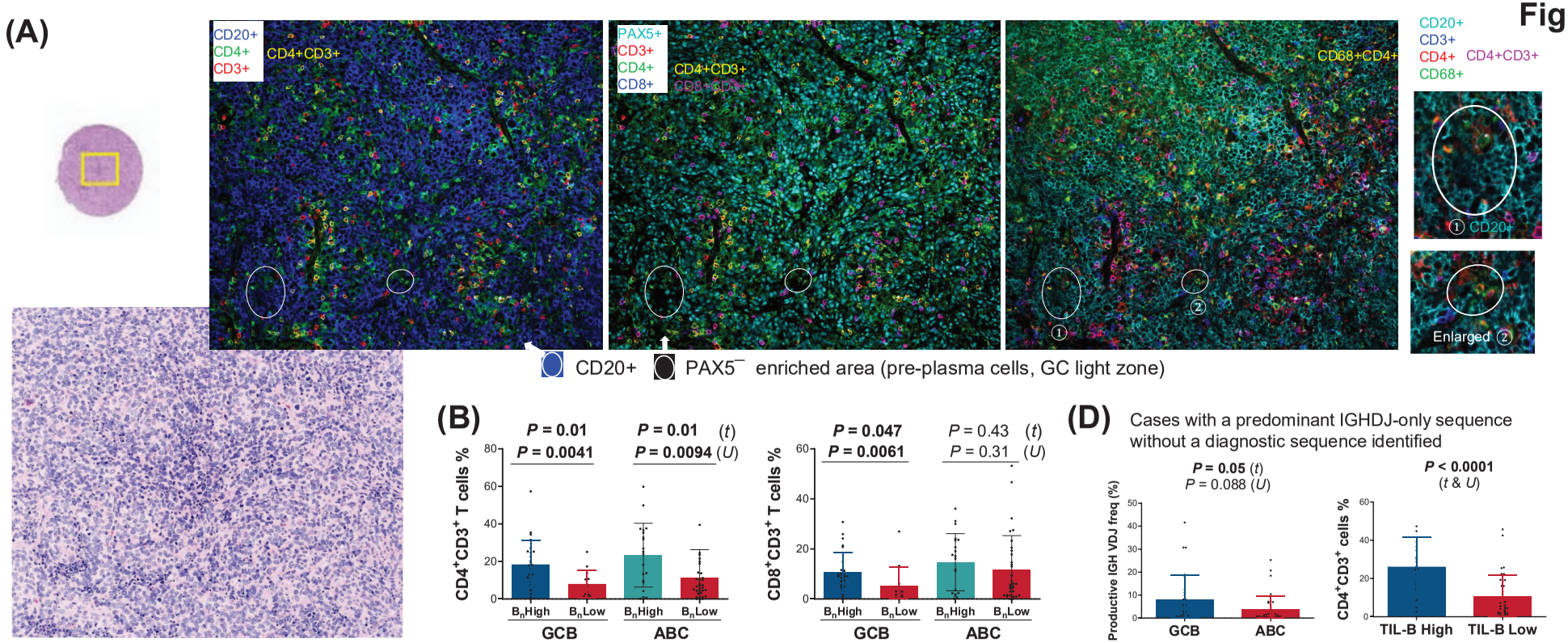

Figure 6