Molecular characterization of the GZL spectrum by GEP: supplementary file

Table of contents :

I. Supplementary Methods

A. *Cohort construction*

Grey zone lymphoma (GZL) and EBV-positive LBCL with polymorphic morphology (polymorphic-EBV-L) cases came from the LYSA and the Centre for Lymphoid Cancer (CLC; BC Cancer, Canada) and have been all reviewed by an international panel of experts in hematopathology. Latent EBV-infection of the malignant cells was evaluated in all cases by *in situ* hybridization for EBER.

The review process of the LYSA cases has been previously described[1]. Briefly, between 2014-2016, after the review of 233 cases, 107 cases (19 Group 0, 79 bona-fide-GZL [group 1 and 2], and 9 Group 3) were included as part of the GZL spectrum and 32 cases were classified as polymorphic-EBV-L as reported in Sarkozy et al[1]. Among these 139 reported cases, FFPE samples were available for 84 cases that were finally included in the gene expression profiling study (62 GZL-spectrum and 22 EBV polymorphic). In 2017, 79 new cases were reviewed, of which 17 were included, leading to a total of 101 LYSA cases: 71 GZL (13 Group 0, 54 bona-fide-GZL, 4 Group 3) and 30 polymorphic-EBV-L (Supplementary Figure 1).

GZL from the CLC (N=38) were extracted from the local database and reviewed during an international meeting in August 2018 with 3 hematopathologists from the CLC (GS, KT, TT), 1 pathologist (ATG) and 1 hematologist (CS) from the LYSA. Out of the 38 cases, 16 were included as GZL, of which 12 had a specimen that could be retrieved from the CLC tissue archives or referring hospitals (see flow chart for exclusion criteria). Among these 12 cases (9 bona-fide-GZL and 3 group 0), 3 were EBV positive and then labelled as polymorphic-EBV-L. Additionally, CD30 staining was performed on in-house CLC tissue microarrays (TMAs) of PMBCL and DLBCL (N=782 cases). Among the 782 PMBCL/DLBCL cases analyzed, 17 (2%) had CD30 expression on 100% of tumor cells and were included as Group 3 GZL, as previously described[1].

Among the 130 GZL, 6 samples failed RNA extraction (RNA amount < 50 ng), 3 failed library construction and 9 did not pass QC thresholds for RNA-seq analysis (see additional methods bellow), leading to a final cohort of 112 cases (see additional methods below).

Fourteen PMBCL cases, all reviewed by expert-hematopathologists and previously reported[2], were included in the study and submitted to RNA sequencing. These cases were selected based on the availability of RNA extracted from FFPE samples, previous R-CHOP treatment and the presence of consensual clinical, morphological and molecular PMBCL features, as previously published[3]. As well, 16 cases of classic Hodgkin lymphoma (cHL) were selected based on the availability of RNA extracted from FFPE samples, and a central expert pathological review. Two cHL cases failed library construction and/or RNA sequencing.

This study was conducted with the approval from institutional boards according to the declaration of Helsinki (LYSA: 2017-006B, BCCRC: H18-01460).

B. TMA (tissue micro-array) construction and antibodies information

The same cohort of GZL and polymorphic-EBV-L was used for both TMA construction and RNA-seq analysis. TMAs were constructed using standard techniques, containing three (1mm) cores per case. In total, 71 and 18 of the GZL and polymorphic-EBV-L cases could be placed on a TMA. Whole tissue sections were available for 15 GZ and 6 polymorphic-EBV-L samples. The cohort of 14 PMBCL cases was used for both TMA and RNA-seq analyses (see Supplementary table 1 for clinical and biological characteristics).
For cHL, a distinct cohort of 22 cases was used for TMA construction (see supplementary table 1 for clinical and biological characteristics).
Antibodies used to assess the tumor microenvironment (TME) composition are presented in the table below.

Antigen	Antibody clone	Manufacturer
CD3	2GV6	Ventana
CD4	SP35	Ventana
CD8	C8/144B	Sigma
LAG3	D2G40	Cell Signaling Technology
FOXP3	236A/E7	Abcam
PD1	MRQ-22	Cell Marque
CD68	KP1	Dako
CD163	10D6	Novocastra
HLA-ABC	EMR8-5	Abcam
HLA-DP/DQ/DR	CR3/43	Dako
PD-L1	SP142	Abcam
PD-L2	D7U8C	Cell Signaling Technology

For MHC-I and -II scoring, moderate to strong membranous staining of 90% or more of positive cells was considered as positive, and less than 90% of positive cells, cytoplasmic staining or weakly positive cases were considered as negative. For PD-L1 and PD-L2 expression, a histoscore (H; range 0-300) was calculated by multiplying the percentage of positive tumor cells with staining intensity (score 1-3).

C. *Library construction*

The quality of the input RNA was assessed by running the RNA samples on an Agilent Bioanalyzer RNA 6000 Nano Chip to determine the RNA Integrity Number (RIN). All samples were rRNA depleted with the Human NEBNEXT rRNA Depletion kit and libraries were generated by using the NEBNext Ultra II Directional RNA Library Prep Kit for Illumina. All procedures were performed according to the manufacturer's protocol with the exception that all samples were fragmented for 7 minutes regardless of RIN. Indexed libraries from 16 cases were pooled and sequenced on the Illumina NextSeq 550 using 75-bp paired-end reads.

D. *RNA sequencing*

The proportion of usable bases was used to identify poor quality samples, defined as those with a value less than 2 median absolute deviations below the median. The samples excluded using this threshold included 7 GZL, 2 polymorphic-EBV-L and 1 cHL (GZ-045, GZ-092, GZ-178, GZ-181, EBV-194, GZ-230, EBV-254, GZ-274, GZ-299, GE0013A).

E. *Nanostring DLBCL90 assay*

The DLBCL90 assay was used to assign PMCBL vs DLBCL status as previously reported[3,4].

F. *Outcome analysis*

Time to progression (TTP) was calculated from the date of diagnosis to the date of progression, a change of therapy that was not initially scheduled and due to lymphoma progression (radiotherapy, high-dose therapy with autologous stem cell transplantation and other unplanned treatments), or death from any cause. Disease specific survival (DSS) was calculated from date of diagnosis to date of death related to lymphoma. Statistical analysis was performed with SAS software. Categorical variables were summarized using frequencies and percentages and compared using the chi-square test.

Given the heterogeneity of the treatment, outcome correlates were analyzed for the population of patients homogeneously treated with either R-CHOP or ABVD regimens (60/105, 57%).

G. *References*

1. Sarkozy C, Copie-Bergman C, Damotte D, et al. Gray-zone Lymphoma between cHL and Large B-Cell Lymphoma. *Am. J. Surg. Pathol.* 2019;43(3):.
2. Mottok A, Hung SS, Chavez EA, et al. Integrative genomic analysis identifies key pathogenic mechanisms in primary mediastinal large B-cell lymphoma. *Blood*. 2019;134(10):802–813.
3. Mottok A, Wright G, Rosenwald A, et al. Molecular classi fi cation of primary

mediastinal large B-cell lymphoma using routinely available tissue specimens. 2019;132(22):2401–2406.

4. Ennishi D, Jiang A, Boyle M, et al. Double-hit gene expression signature defines a distinct subgroup of germinal center B-cell-like diffuse large B-cell lymphoma. *J. Clin. Oncol.* 2019;37(3):190–201.
5. Sarkozy C, Copie-Bergman C, Damotte D, et al. Gray-zone Lymphoma between cHL and Large B-Cell Lymphoma. *Am. J. Surg. Pathol.* 2019;43(3):341–351.

II. Supplementary Tables

Supplementary Table 1. Clinical characteristics of cHL and PMBCL cohort.

Characteristics	PMBCL, N=14	cHL RNA-seq, N=16*	cHL TMA, N=22
Centre	BCCRC	BCCRC	BCCRC
Median Age (range)	37 (25-45)	28.5 (17-67)	37 (20-75)
>=60y	0	2	4
Sex: F/M	9 (64%)	8 (50%)	6 (27%)
Histology classification			
Nodular sclerosis	NA	14 (88%)	12
Mixed cellularity	NA	2	9
Lymphocyte-rich	NA	0	1
EBV (yes/no)	0	3 (19%)	5 (23%)
Mediastinal involvement			
Thymic	14 (100%)	12 (75%)	8
Mediastinal, non-thymic	0	0	0
Non-mediastinal	0	4	13
Ann Arbor Stage			
1-2	11 (79%)	9 (56%)	12 (57%)
3-4	3 (21%)	7 (44%)	9 (43%)
B symptoms	9 (64%)	5 (31%)	7 (33%)
Bulky (>10cm)	11 (79%)	0 (0%)	

cHL: classic Hodgkin lymphoma; PMBCL: primary mediastinal B cell lymphoma

*2 cHL cases failed RNAseq QC metrics (2 EBV^{pos} cHL) and are not presented here.

Supplementary Table 2. Outcome for GZL, polymorphic-EBV-L, cHL and PMBCL patients treated with R-CHOP or ABVD. To provide a comparison with cHL and PMBCL treated with ABVD and RCHOP respectively, we included in this table only bona-fide-GZL, group 0, group 3 and polymorphic-EBV-L treated with either ABVD or R-CHOP.

Pathology	N	2-year TTP	2-year DSS
cHL	30	82% (68-98%)	100% (100-100%)
Group 0	2*	NA	NA
Bona-fide-GZL (Group 1-2)	34	60% (45-80%)	70% (58-90%)
Group 3	18	94% (84-100%)	94% (84-100%)
PMBCL	14	86% (69-100%)	93% (80-100%)
Polymorphic-EBV-L	20	67% (48-93%)	80% (60-100%)

cHL: classic Hodgkin lymphoma ; PMBCL: primary mediastinal B cell lymphoma ; GZL: grey zone lymphoma ; R-CHOP: rituximab, cyclophosphamide, oncovin, adriamycin, prednisone ; ABVD: adriamycin, bleomycin, vinblastin, dacarbazin; TTP: time to progression ; DSS: disease specific survival.

*Among group 0 cases, 9 were mainly treated intensively with R-ACVBP/R-COPADEM or escBEACOPP regimens and are not included here.

Supplementary Table 3. Gene set enrichment analysis using principal component score.

See Excel file “sum_GSEA_ALL.xls”.

Supplementary Table 4. Clinical characteristics of thymic and non-thymic bona-fide-GZL cases.

Characteristics	Thymic, N=33	Non-Thymic, N=19	Chi2, p-value
Median age (range)	39y (19-60)	68y (44-87)	<0.001
Sex: F/M (/130)	19/14	12/7	0.8
GZL classification 1 2	 21 (64%) 12 (36%)	 9 (47%) 10 (53%)	 0.25
EBV (yes/no)	0	0	-
Mediastinal involvement Thymic Mediastinal, non-thymic (MNT) Non-mediastinal	 33 0 0	 0 6 13	 -
Bulky (10 cm)	12 (43%)	4 (21%)	0.1
Ann Arbor Stage 1-2 3-4	 20 (65%) 11 (35%)	 3 (16%) 16 (84%)	 <0.001
Extra-nodal site No Yes : - Spleen - Liver - Lung isolated*	 23 10* (30%) 1 (10%) 2 (20%) 5 (50%)	 8 11 (58%) 5 (45%) 1 (10%) 2 (18%)	0.05
LDH >UNL	14 (52%)	13 (68%)	0.15
aaIPI 0-1 2-3	 20 (71%) 8 (29%)	 7 (37%) 11 (63%)	 0.028

*Extra-nodal site was lung only in 50% of thymic cases *versus* 18% of non-thymic cases.

Supplementary Table 5. Clinical characteristics of polymorphic-EBV-L versus GZL.

Characteristics	Polymorphic-EBV-L (N=27)	GZL (N=67)	p.val (Chi2)
Centre			0.9
CLC	3 (11%)	8 (12%)	
LYSA	24 (89%)	59 (88%)	
Median Age (range)	53 (16-79)	46 (14-90)	0.1
>=60y	11 (41%)	16 (24%)	
Sex (/112)			0.03
F	8 (30%)	36 (54%)	
M	19 (70%)	31 (46%)	
Path_Group (/112)			<0.001
0*	18 (67%)	11 (16%)	
1	4 (15%)	31 (46%)	
2	5 (18%)	25 (38%)	
Mediastinal involvement	(/26)	(/62)	0.09
Mediastinal	15 (58%)	47 (76%)	
Anterior (Thymic)	7 (27%)	41 (66%)	0.0007
Non-Anterior	8 (31%)	6 (10%)	
Non-mediastinum	11 (42%)	15 (24%)	
Ann Arbor_Stage 3-4	(/21)	(/62)	0.49
1-2	8 (38%)	29 (47%)	
3-4	13 (62%)	33 (53%)	
Hemoglobin	(/22)	(/59)	0.57
<12g/dL	12 (55%)	28 (47%)	
LDH	(/21)	(/56)	0.74
>UNL	10 (45%)	29 (52%)	
aaIPI	(/22)	(/58)	0.18
0-1	10 (45%)	36 (62%)	
2-3	12 (55%)	22 (38%)	
Bulky	(/16)	(/58)	0.13
>10cm	2 (12%)	18 (31%)	

*Bona-fide-GZL and group 0 were included in this comparison.

Supplementary Table 6. Gene set enrichment analysis (Excel file: "gsea_report_Hallmark_EBV.xls") for differential expression between grey-zone lymphoma (group 0-1-2) and polymorphic-EBV-LBCL.

Supplementary Table 7. Genomic aberrations of the 9p24 JAK2/PDL12 locus

Cohort	Bona-fide GZ*, N=56	Bona-fide - thymic, N=33	Bona-fide -non thymic, N=19	Group 0, N=11
JAK2-PDL12 CN Gain Amp all	/47 14 (30%) 15 (32%) 29 (62%)	/27 9 (33%) 8 (30%) 17 (63%)	/19 4 (21%) 7 (37%) 11 (58%)	/10 3 (30%) 3 (30%) 6 (60%)
JAK2-PDL12 BA	/47 9 (19%)	/27 6 (22%)	/19 3 (16%)	/10 1 (10%)
JAK2-PDL12 all	/47 32 (68%)	/27 18 (67%)	/19 13 (68%)	/10 6 (60%)

*4 cases do not have thymic status
See ref[5] for methods

Additional Supplementary Files

1. merged_DESeq2_results: merged differential expression (DE) results between (1) GZL vs cHL, (2) GZL vs PMBCL, and (3) cHL vs PMBCL
2. thymic_vs_non_DESeq2_results: DE between thymic vs. non-thymic Bona-Fide-GZL cases
3. EBV_pos_vs_neg_Group0_1_2_DESeq2_results: DE between EBV-positive polymorphic DLBCL vs GZL (Group 0 and Bona-fide)
4. Group0_vs_cHL: DE between EBV-positive polymorphic DLBCL vs GZL (Group 0 and Bona-fide)

III. Supplementary Figure Legends

Supplementary Figure 1. Schematic flow chart of cases selected for the study cohort. CLC: Centre for Lymphoid Cancer; QC: quality control; FFPE: formalin-fixed and paraffin-embedded; PMBCL: primary mediastinal B cell lymphoma; DLBCL: diffuse large B cell lymphoma.

Supplementary Figure 2. Scree plot describing the percentage of variance explained by the first 20 principal components.

Supplementary Figure 3. Low small B cell content in the tumor micro-environment of bona-fide grey zone lymphoma. (A) GZL cases present a down regulation of specific B cell markers, suggesting smaller B cell content in their TME. Box plots show expression levels (normalized and log-transformed) of CD79A, CD79B, VPREB3 and IGKV3OR2-268 in the RNA-seq data.
(B) IHC staining of CD20 for a representative bona-fide-GZL case showing low small B cell content in the TME.

Supplementary Figure 4. TME composition in thymic and non-thymic bona-fide-GZL. (A) Thymic GZL and PMBCL have a complete loss of MHC-I expression (IHC). cHL: classical Hodgkin lymphoma; Bona-F_GZ_thymic: grey zone group 1-2 (bona-fide-GZL) with a thymic involvement; Bona-F_GZ_non_thymic: grey zone group 1-2 (bona-fide-GZL) without thymic involvement; PMBCL: primary mediastinal B-cell lymphoma.
(B) TME description in thymic bona-fide-GZL, non-thymic bona-fide-GZL, cHL and PMBCL. Non-thymic GZL have intermediate features between cHL and large B cell lymphoma, with an intermediate infiltration by T cells as shown by CD3, CD4 and CD8 staining. They also present a GZL-specific feature of strong macrophage (M2) involvement (CD163).

Supplementary Figure 5. Comparison of Group 0 to cHL and bona-fide-GZL (group 1-2)
(A) Small B cell content in group 0 *versus* cHL cases (EBV negative). Group 0 present less small B cell content (scored here as CD20 within the TME) compared to cHL (p=0.0062, top right

panel). Conversely, group 0 cases have a stronger enrichment in CD163 positive cells, or macrophages (top left and middle panels show CD163 positivity, and the ratio of CD163 positivity to CD68 positivity, respectively). This inverse correlation was confirmed at the patient level using mRNA expression (lower panel, spearman).

(B) TME in group 0, bona-fide-GZL (group 1-2) and cHL assessed by IHC. This analysis highlights the intermediate nature of the group 0 TME, with the same enrichment in T cells as compared to cHL, but a stronger enrichment in macrophages, further suggesting that group 0 might more closely resemble bona-fide-GZL.

(C) Pre-ranked GSEA based on differential expression score between group 0 and cHL (EBV negative) cases showed that group 0 had a stronger enrichment for the GZL signature genes compared to cHL, suggesting that these cases should be part of the GZL spectrum.

Supplementary Figure 6. Group 3 and PMBCL. (A) Principal component analysis within group 3 and PMBCL. PC2, reflecting 13% of the variance, could discriminate the PMBCL and group 3: group 3 cases had a high PC2 score, and PMBCL a low PC2 score. Interestingly, the group 3 cases with a PMBCL morphology (N=3) were located at the border between class 3 and PMBCL. (B) Unsupervised clustering of the group 3 and PMBCL samples, using expression of the Lymph3Cx gene list (Mottok et al, Blood 2018). The DLBCL and PMBCL genes each clustered together, validating the approach. The cases clustered based on the presence of thymic involvement, with a stronger DLBCL signature for extra-thymic cases and a stronger PMBCL signature for thymic cases. (C) We used the correlation score with PC2 to perform a pre-ranked GSEA with the Lymph3Cx and GZL signature gene lists. The GZL signature was not significantly associated with PC2 score (*i.e.* PMBCL vs group 3). On the contrary, the Lymph3Cx signature genes were significantly correlated with PC2 score: cases with high PC2 score (group 3) had higher expression of the DLBCL signature, as opposed to cases with a low PC2 score which had a stronger PMBCL signature.

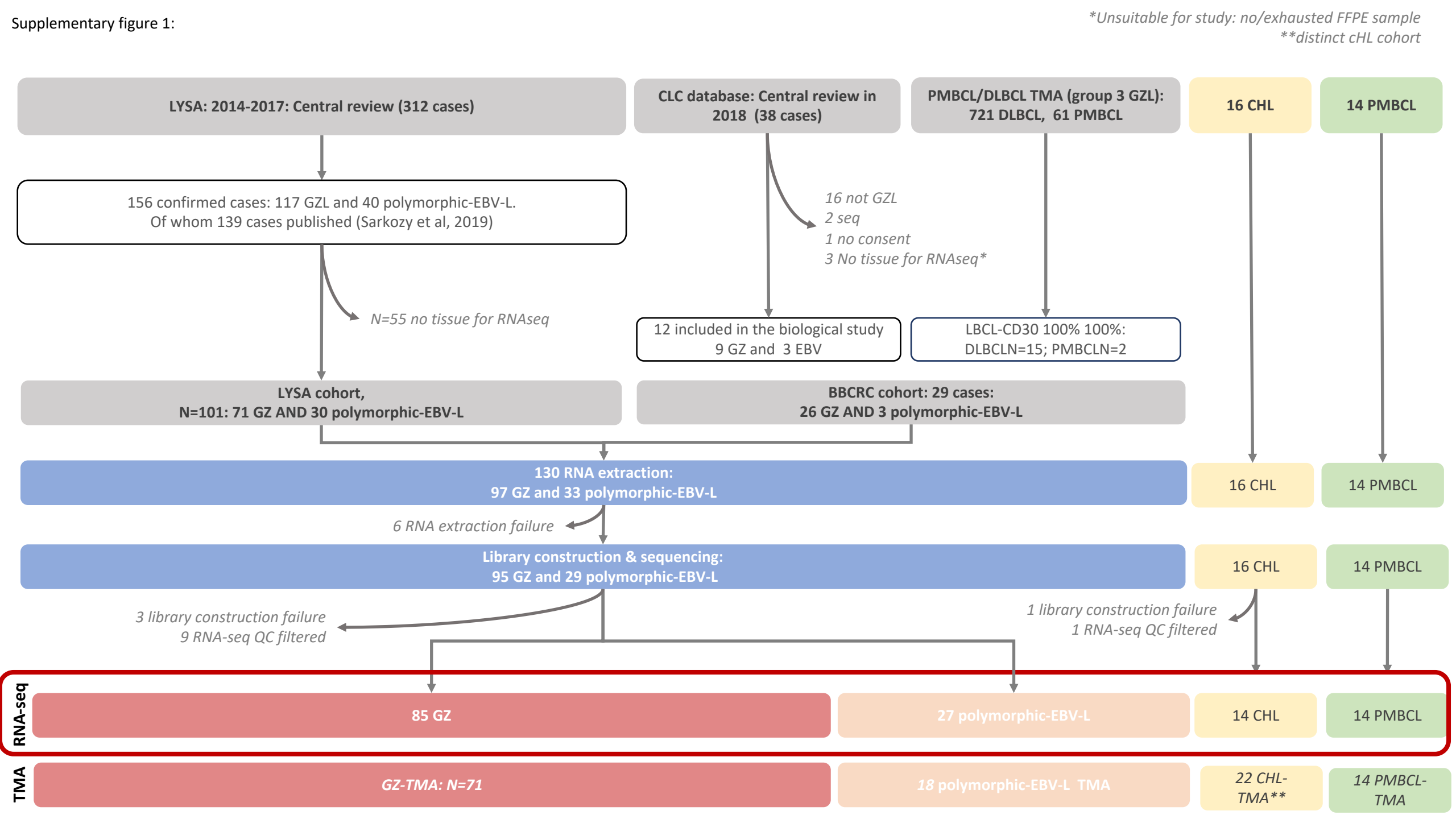
Supplementary Figure 7. EBV related molecular and immuno-phenotypic characterization.

(A) Volcano plot summarizing differential expression between grey zone lymphoma (group 0 and bona-fide-GZL) and polymorphic-EBV-L cases. The most significantly up-regulated genes

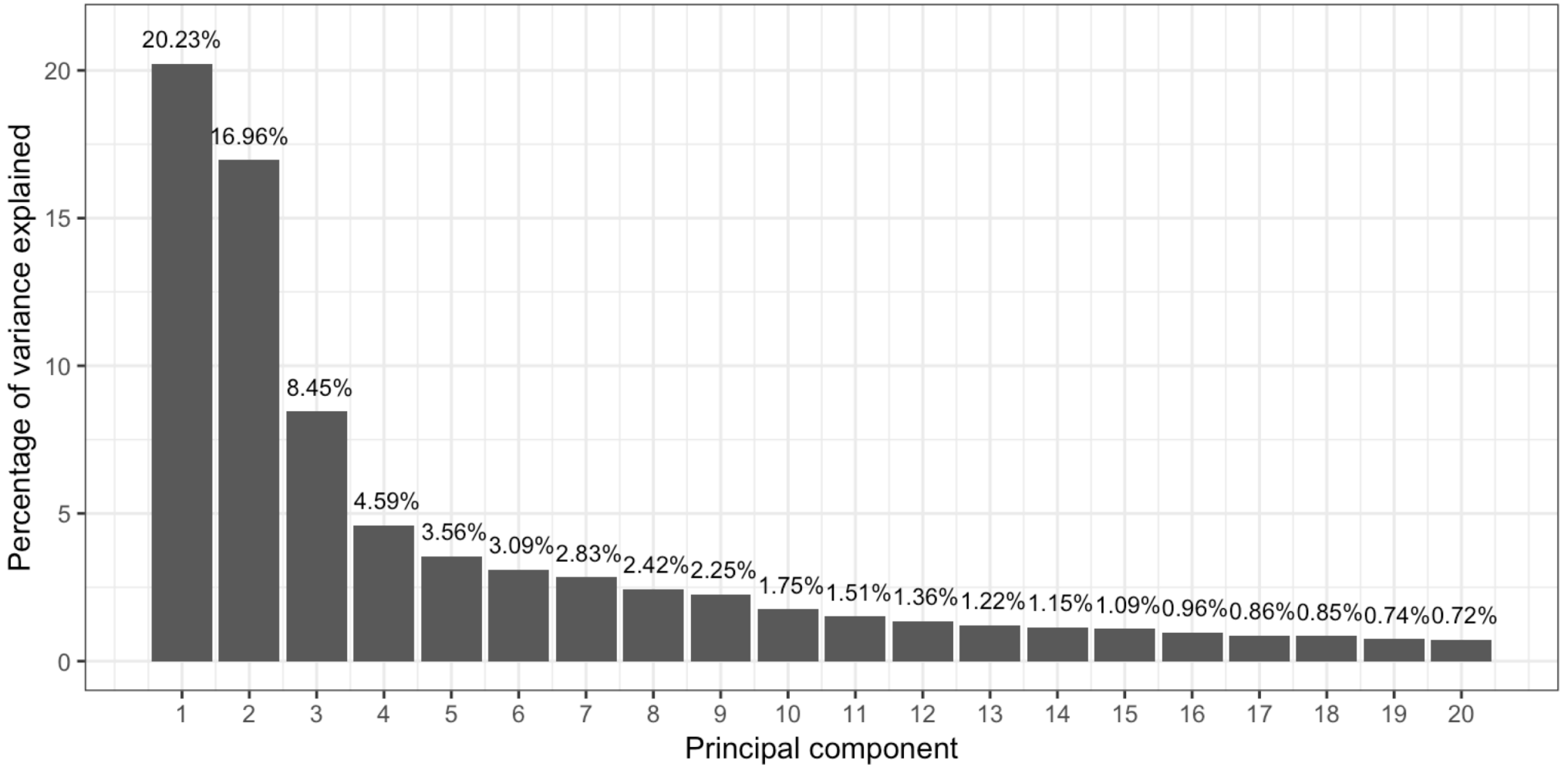
in GZL cases are highlighted in blue, and most up-regulated in polymorphic-EBV-L are highlighted in red. In the lower panel, two representative GSEA signatures are shown: EBV negative cases were enriched in the GZL-specific signature genes, whereas EBV positive cases were enriched in microenvironment-related signatures.

(B) TME composition according to EBV status. EBV positive cases are represented in red, and GZL negative in black.

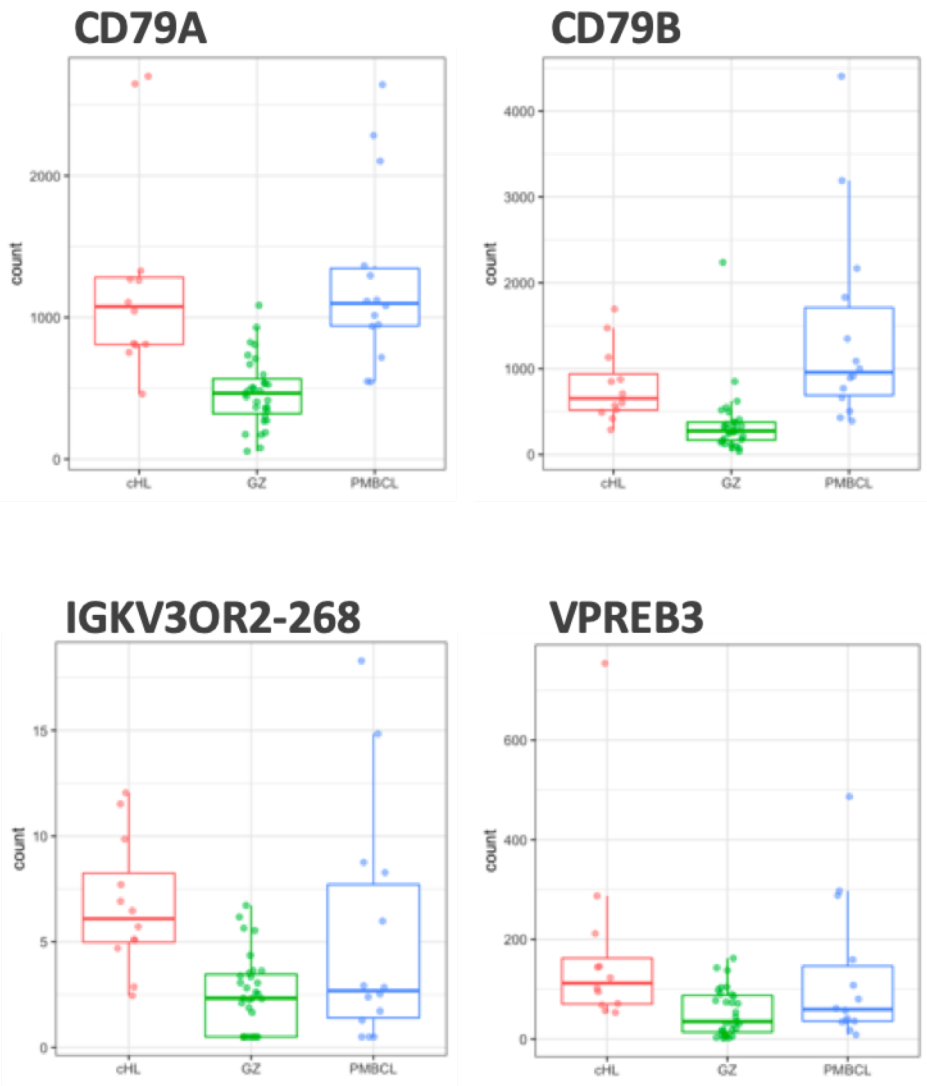
Supplementary figure 1:



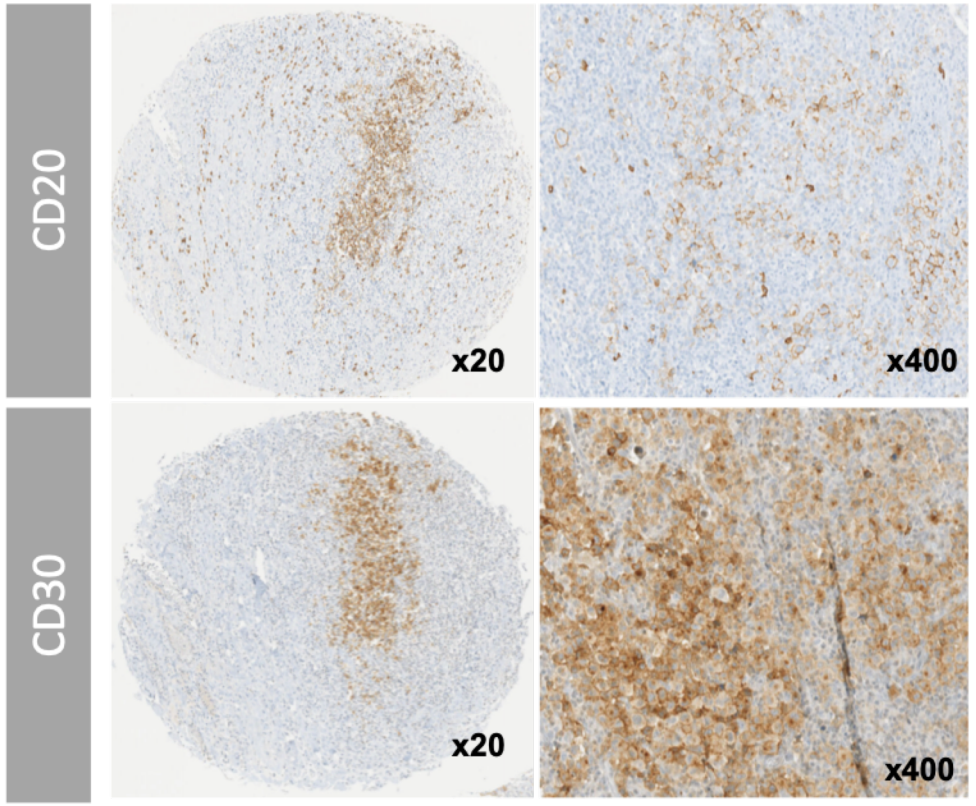
Supplementary figure 2:



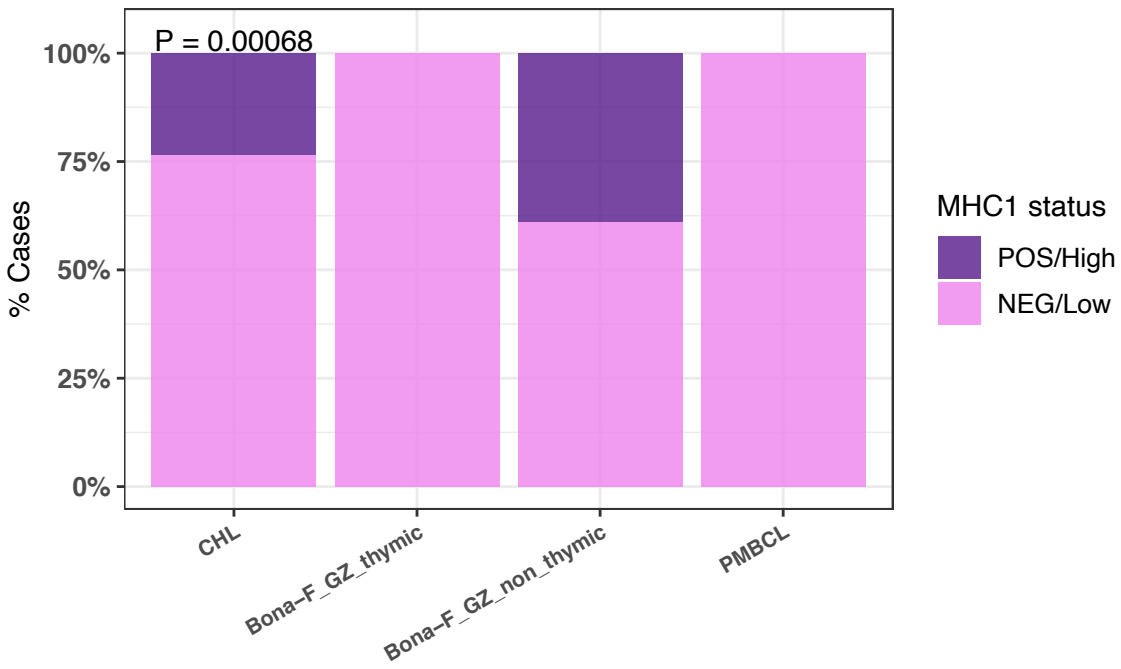
Supplementary figure 3A



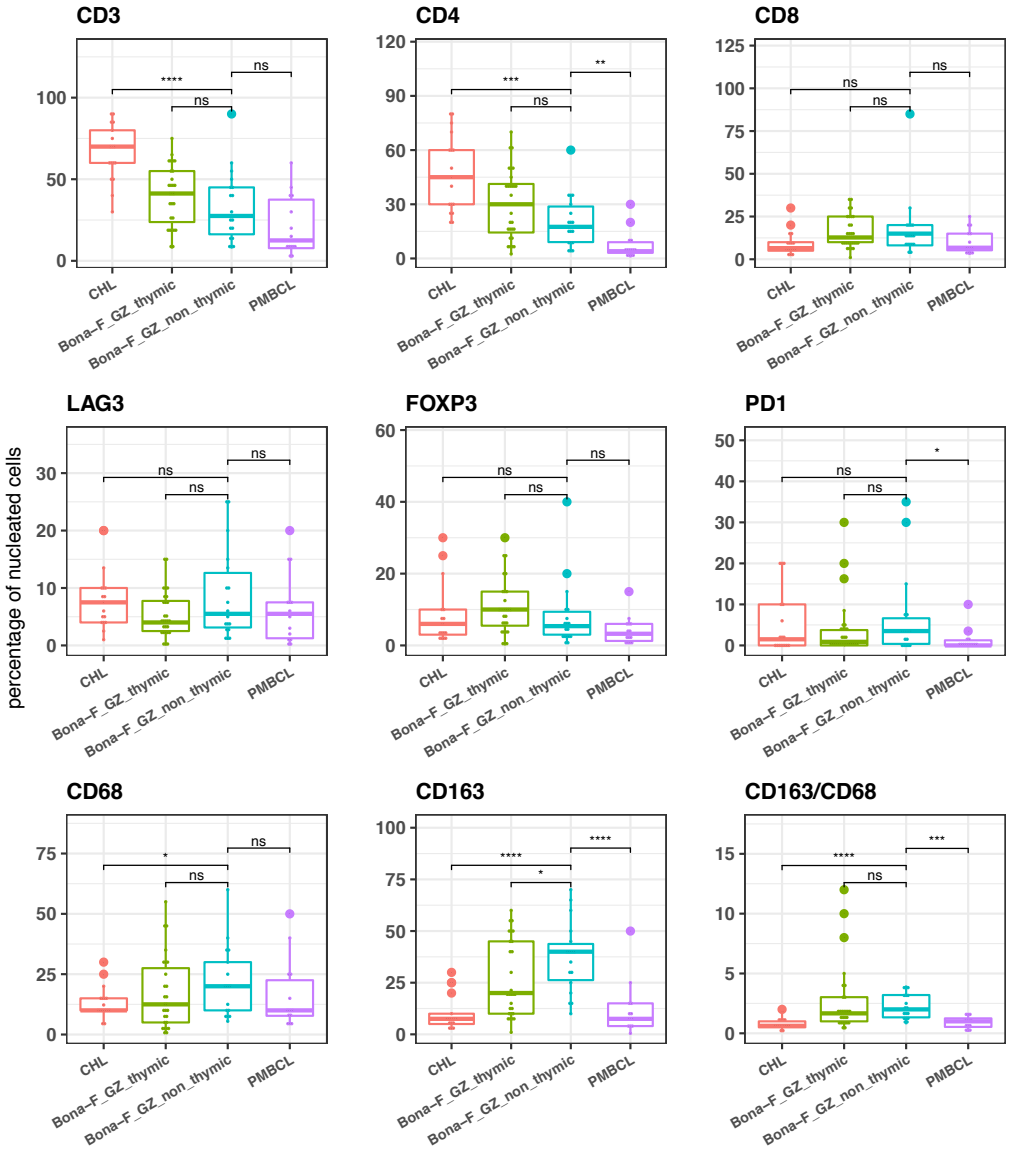
Supplementary figure 3B



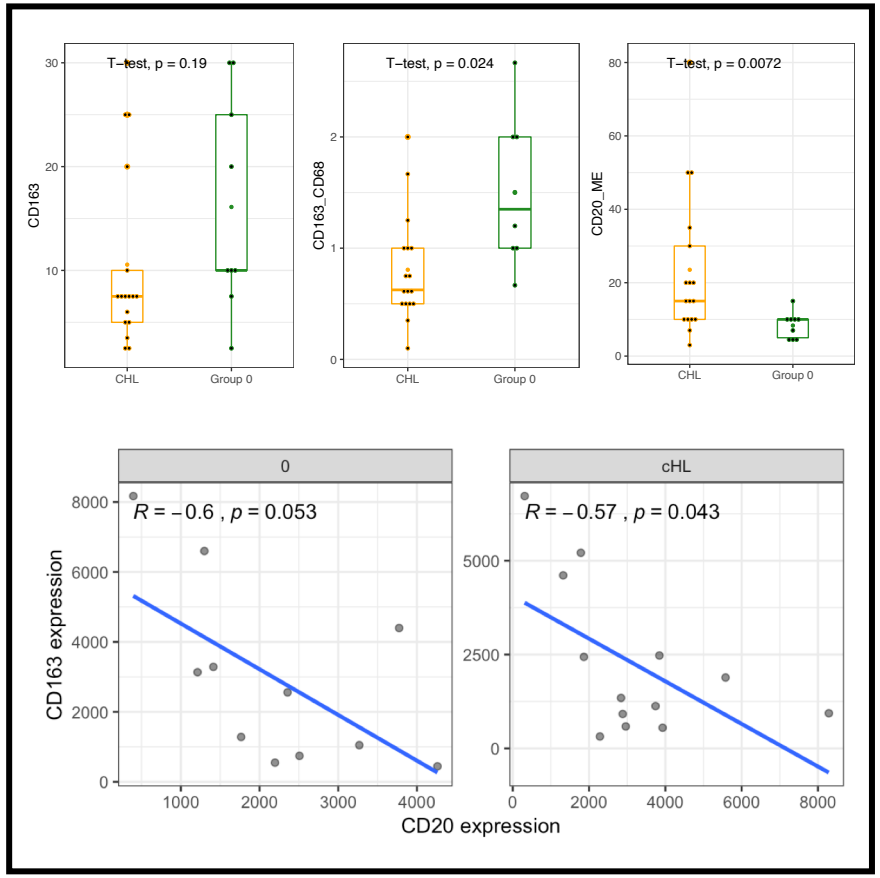
Supplementary Figure 4A



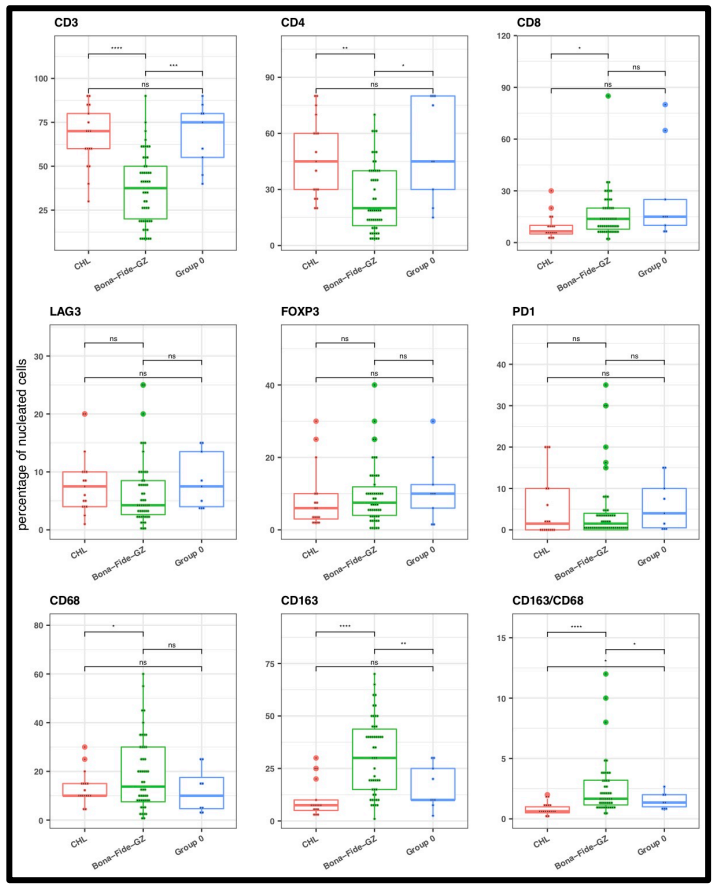
Supplementary Figure 4B



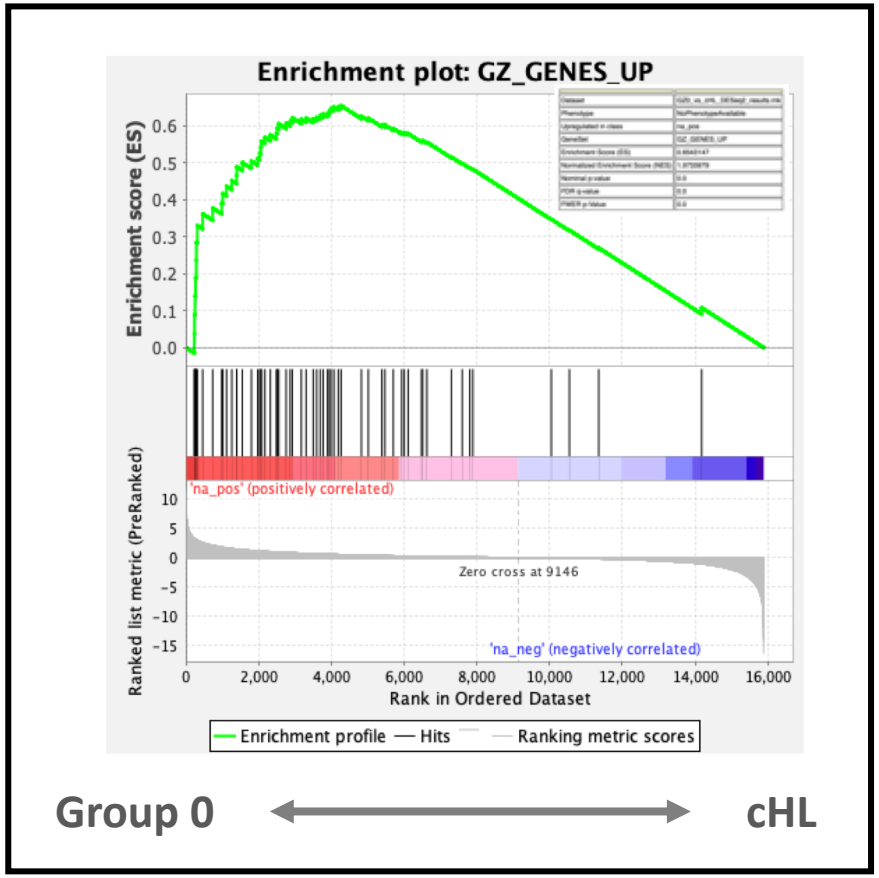
Supplementary Figure 5A:



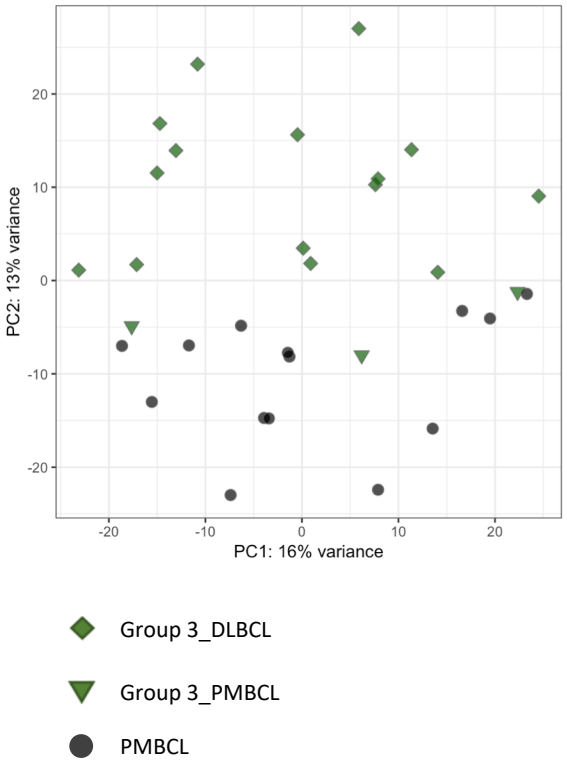
Supplementary Figure 5B:



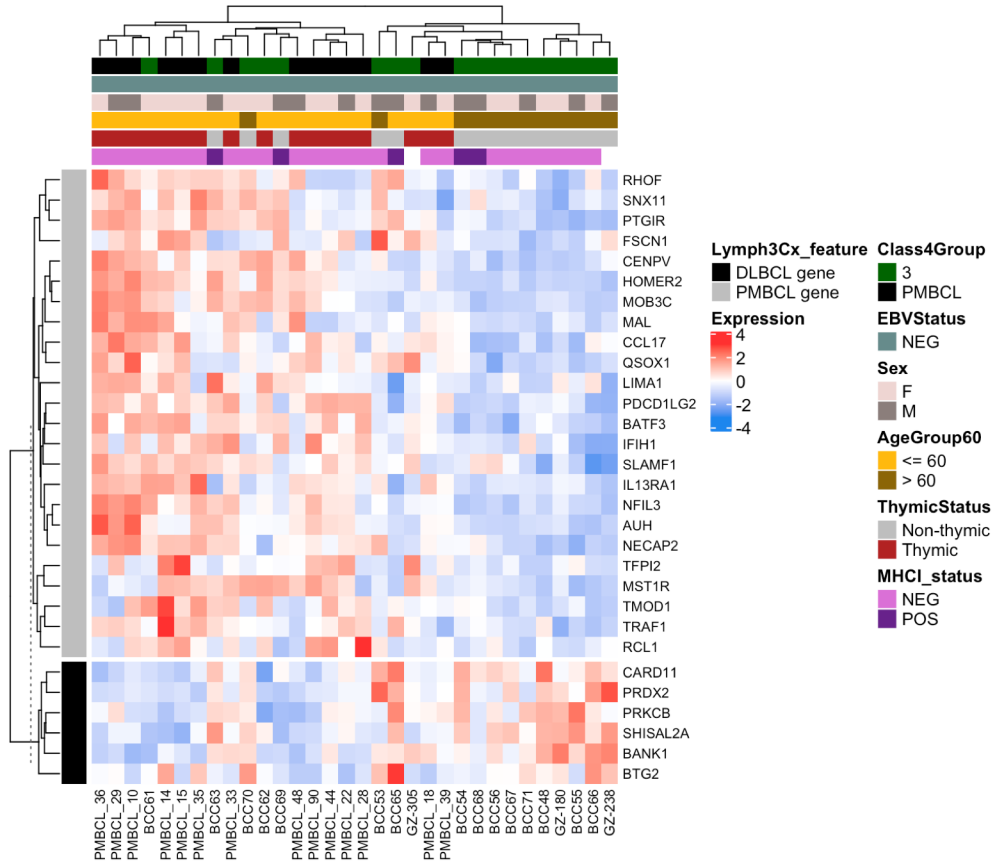
Supplementary Figure 5C:



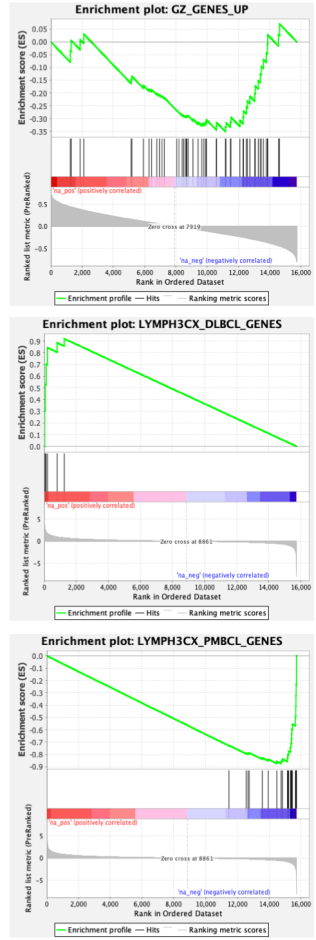
Supplementary Figure 6A:



Supplementary Figure 6B:

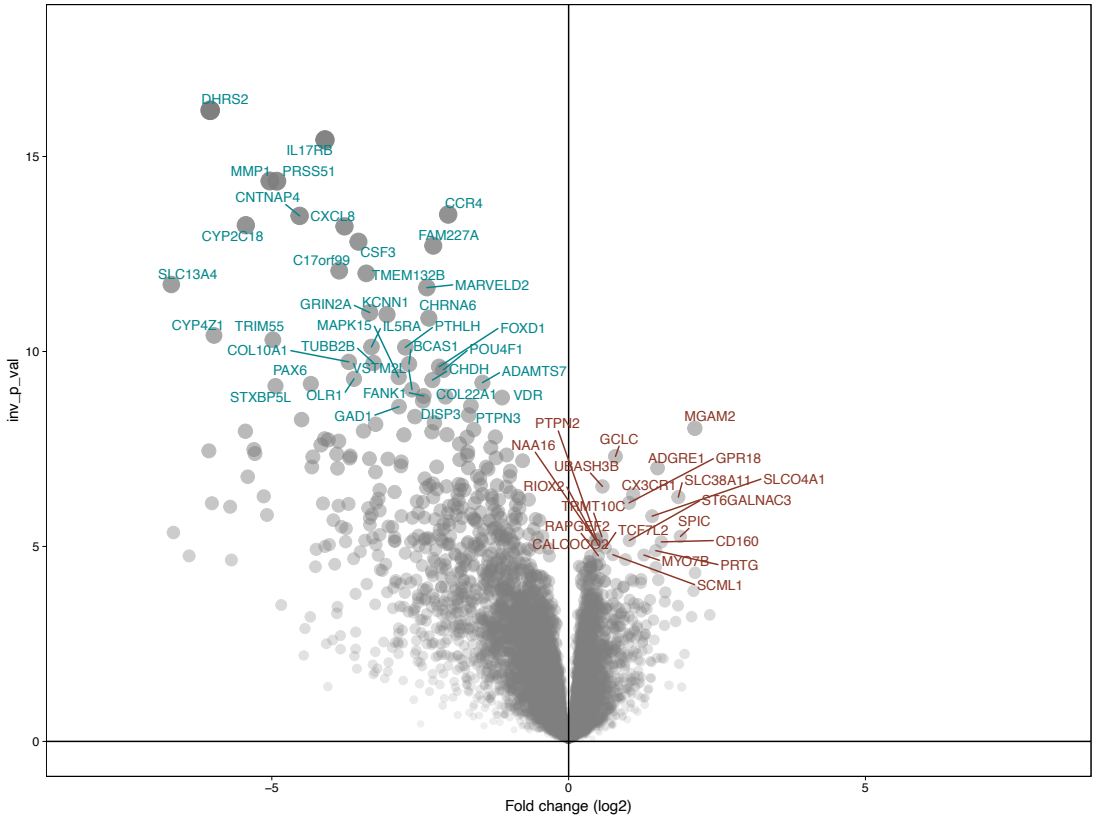


Supplementary Figure 6C:

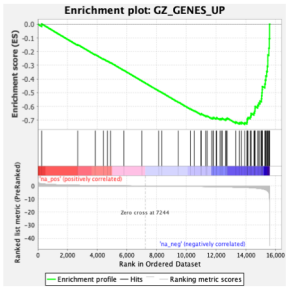


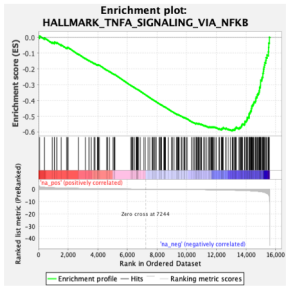
PC2_high ⟷ PC2_low

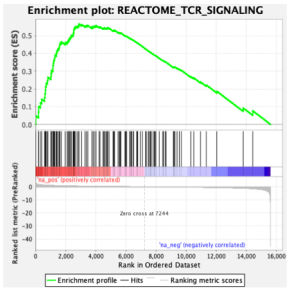
Supplementary Figure 7A

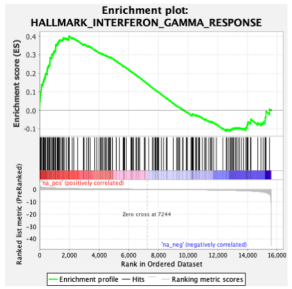


GZ ⟷ Polymorphic-EBV









Supplementary Figure 7B