Molecular Attributes Underlying Central Nervous System and Systemic Relapse in Diffuse Large B-cell Lymphoma


Haematologica 2020 [Epub ahead of print]

doi:10.3324/haematol.2020.255950

Publisher's Disclaimer.
E-publishing ahead of print is increasingly important for the rapid dissemination of science. Haematologica is, therefore, E-publishing PDF files of an early version of manuscripts that have completed a regular peer review and have been accepted for publication. E-publishing of this PDF file has been approved by the authors. After having E-published Ahead of Print, manuscripts will then undergo technical and English editing, typesetting, proof correction and be presented for the authors' final approval; the final version of the manuscript will then appear in print on a regular issue of the journal. All legal disclaimers that apply to the journal also pertain to this production process.
Molecular attributes underlying central nervous system and systemic relapse in diffuse large B-cell lymphoma

Short title: CNS and systemic relapse in DLBCL

Keren Isaev,1 Daisuke Ennishi,2 Laura Hilton,3 Brian Skinnider,2 Karen L. Mungall,4 Andrew J. Mungall,4 Mehran Bakthiari,1 Rosemarie Tremblay-LeMay,5 Anjali Silva,1 Susana Ben-Neriah,2 Merrill Boyle,2 Diego Villa,2 Marco A. Marra,4 Christian Steidl,2 Randy D. Gascoyne,2 Ryan D. Morin,3 Kerry J. Savage,2 David W. Scott2*, Robert Kridel1*

1Princess Margaret Cancer Centre - University Health Network, Toronto, ON, Canada; 2Centre for Lymphoid Cancer, BC Cancer, Vancouver, BC, Canada; 3Simon Fraser University, Burnaby, BC, Canada; 4Canada's Michael Smith Genome Sciences Centre, BC Cancer, Vancouver, BC, Canada; 5Laboratory Medicine Program - University Health Network, Toronto, ON, Canada.

Corresponding author:

Dr. Robert Kridel
Princess Margaret Cancer Centre – University Health Network
700 University Avenue, DMOH OPG Rm 6-714, Toronto, ON M5G 2M9, Canada
E-mail: robert.kridel@uhn.ca; Phone: 416-946-2243; Fax: 416-946-4520

* These authors contributed equally to this work.

Text word count: 1,503
Number of figures and tables: 3
Number of references: 15
Refractory or relapsed diffuse large B-cell lymphoma (DLBCL) is challenging to treat in general and progression in the central nervous system (CNS) is associated with particularly dismal outcomes. Indeed, the median survival of patients diagnosed with CNS relapse is typically 2-4 months.\(^1\) The incidence of CNS relapse is 2-5% in unselected patient cohorts treated with modern immunochemotherapy,\(^2,3\) presenting a challenge for risk prediction. Individual risk factors for CNS relapse are well established, and include clinical parameters such as age, stage, elevated lactate dehydrogenase and performance status.\(^1,4\) Moreover, the involvement of certain extranodal sites such as uterus/testis or kidney/adrenal gland confers higher risk of CNS relapse.\(^5\) These individual clinical risk factors have been integrated into a risk score (CNS International Prognostic Index, CNS-IPI).\(^3\) Molecular attributes such as double-hit translocation status, BCL2/MYC dual protein expression as well as the activated B-cell-like (ABC) subtype (or non-GCB), have also been associated with a higher risk of CNS relapse.\(^6,7\) However, the major shortcoming of available risk prediction models is their limited discriminative power. While the CNS-IPI classifies 12% of patients into the high-risk group, only 10% of patients within this high-risk group experience CNS relapse.\(^3\)

Herein, to comprehensively assess molecular determinants of CNS relapse, we chose three distinct approaches. Firstly, we performed exome sequencing in paired tissue biopsies (peripheral and CNS parenchymal tumors) and reconstructed clonal phylogenies that underlie CNS invasion. Secondly, we assembled a dataset to assess the association of specific gene mutations with CNS relapse. Lastly, we performed gene expression and pathway analysis on an enriched cohort of diagnostic tissue samples from 222 patients with diverging clinical outcomes. An overview of all samples used in this study is shown in Supplemental Figure 1.
To describe changes in clonal structure associated with CNS relapse, we used the PyClone computational tool to infer clonal structures in five paired CNS samples (Supplemental Table 1). Recurrent gene mutations included $PIM1$ ($n = 4$), $CD44$ ($n = 3$), $ETV6$ ($n = 3$), $MALAT1$ ($n = 3$), $ARHGAP5$ ($n = 2$), $CD79B$ ($n = 2$), $CXCR4$ ($n = 2$) and $RUNX1T1$ ($n = 2$). We observed striking clonal divergence in all five cases, with biopsies from the time of CNS invasion characterized by discrete sets of mutations (Figure 1). In general, the expanding subclones were not sampled at the time point of diagnosis, with the main exception being case LY_CNSrel_006. The number of mutations found in CNS-relapsing clones tended to be higher than the number of shared mutations, and higher than the number of mutations exclusively found at diagnosis (Supplemental Figure 2). The most frequently mutated gene was $PIM1$, with 22 exonic, intronic and untranslated region mutations identified in samples from four patients, in keeping with $PIM1$ being recognized as a known target of aberrant somatic hypermutation in DLBCL. These mutations were mostly subclonal (95%), identified in either expanding (14/22 mutations, 64%) or extinguished clones (7/22 mutations, 32%). The ETS family member transcription factor $ETV6$ was found to be mutated in expanding clones in two cases, and in an ancestral clone in another case, with all mutations mapping to the exon 1 donor splice site. Our findings document the existence of intratumoral heterogeneity that is associated with the emergence of subclones that are observed in the CNS. They also suggest that the disease leading to relapse has features that are distinct from the disease found at a primary diagnosis.

Next, we asked whether a larger cohort of diagnostic samples could delineate gene mutations associated with either CNS or systemic relapse. We compiled mutation data from 223 diagnostic DLBCL samples, derived from several datasets of both population-based and clinical
trial series. Information on 45 genes was available, for 72 cases with CNS relapse, 62 cases with systemic relapse and 89 cases without relapse (Supplemental Table 2). The strongest signal for enrichment within cases with subsequent CNS relapse was found for \textit{MYD88}, although this gene was almost as frequently mutated in cases with subsequent systemic relapse (Figure 2). On the other hand, several gene mutations appeared most robustly associated with systemic relapse, and were comparatively uncommon in cases with subsequent CNS relapse. This pattern was most striking for \textit{TP53}, with mutations found in 45\% of patients experiencing systemic relapse, in 20\% of patients with no relapse, and in only 11\% of patients with subsequent CNS relapse. Furthermore, we found that certain gene mutations were relatively depleted in CNS relapse cases, compared to cases with either systemic or no relapse. The most striking example was \textit{KMT2D}, mutated in 22\% of patients with CNS relapse, compared to 37 and 39\% of patients with no relapse and systemic relapse, respectively.

Lastly, we asked whether specific pathways were enriched in diagnostic samples from patients who subsequently experienced either CNS or systemic relapse. We included a total of 222 patients with samples available for gene expression profiling: 50 patients with CNS relapse (n = 39) or CNS involvement documented at diagnosis (n = 11), 64 patients with systemic relapse and 108 patients without relapse (Supplemental Table 3). We performed gene set enrichment analysis to identify biological processes deregulated in the three risk groups (Supplemental Table 4). Interestingly, we observed a significant downregulation of pathways in ABC-DLBCL when comparing systemic vs. no relapse, and an upregulation of pathways in GCB-DLBCL (Figure 3A). Downregulation of pathways was also observed in CNS vs. no relapse in ABC-DLBCL but not GCB-DLBCL. We observed that systemic relapse in GCB-DLBCL appeared to be driven by MYC,
metabolic and protein translation signatures, to a higher degree than CNS relapse (Figure 3B).

We assessed whether the double-hit signature allowed the delineation of varying risks of CNS and/or systemic relapse in our cohort. Overall, 39 out of 90 classifiable GCB cases (43%) were positive for DHITsig. The percentage of DHITsig-positive cases was similar between systemic (64%) and CNS relapse (58%), and lower in cases without relapse (29%, \( \chi^2 \) test \( P = 0.008 \), Supplemental Figure 3). These results link expression of the DHIT signature with CNS relapse, mirroring the increased risk of CNS relapse conferred by dual expression of MYC and BCL2. ABC-DLBCL cases with subsequent relapse were characterized by depletion of stromal signatures, an effect that was most pronounced in those ABC-DLBCL cases with systemic relapse. In summary, while cases with subsequent CNS relapse displayed similar pathway enrichment patterns to cases with subsequent systemic relapse, there are likely additional CNS-specific processes not captured through transcriptome profiling.

Herein, we present a comprehensive description of clonal evolution, gene mutations and gene expression changes associated with CNS and systemic relapse in DLBCL. Our analysis of clonal evolution patterns underlying CNS relapse revealed clear clonal divergence. In general, mutations specifically associated with the relapsing clone were not found at measurable levels in preceding diagnostic specimens. Our cohort for exome sequencing of paired samples was limited due to inherent challenges associated with collecting biospecimens of this nature, as a limited number of patients undergo sampling of brain parenchyma at the time of relapse. Further, a divergent mode of clonal evolution is not unique to CNS relapse, but rather a general property of cancers that harbor underlying intra-tumoral heterogeneity. Divergent evolution has also been demonstrated in DLBCL, although – to the best of our knowledge – no prior study
Isaev et al.

has applied state-of-the-art phylogenetic reconstruction tools to trace clonal trajectories in the specific context of CNS relapse. The phylogenetic portraits that we could draw exert a cautionary tale, illustrating that efforts to improve patient outcomes will need to take into consideration the molecular landscape at the time of lymphoma relapse. Moreover, well-known challenges to predict CNS relapse may be explained by both inter-patient and intra-tumor heterogeneity that is universally recognized in DLBCL.

Alterations of ABC subtype-associated genes are frequently seen in primary CNS lymphoma,\textsuperscript{12} which is often of an ABC phenotype,\textsuperscript{13} and are also characteristic of the recently identified C5/MCD subtype of DLBCL that is associated with extranodal involvement and poor outcome.\textsuperscript{14,15} In the gene mutation analysis, the strongest enrichment was seen for $\textit{MYD88}$, although this gene was almost as often mutated in cases with subsequent systemic relapse. Conversely, we found that several gene mutations that were more commonly seen in patients experiencing systemic, as opposed to CNS relapse (e.g. $\textit{TP53}$). Other mutations such as $\textit{KTM2D}$, characteristic of the C3/EZB subtype of DLBCL, appeared to be comparatively infrequent in CNS relapse cases, perhaps suggesting that the $\textit{KMT2D}$-mutant phenotype is not conducive to establishing tumors in the CNS. Lastly, we found that biological pathways were differentially enriched between clinical risk groups in the ABC vs. GCB subtype of DLBCL. Signals underlying CNS relapse were overall weaker compared with signals underlying systemic relapse, potentially reflective of clonal divergence and resulting phenotypic shifts that may accompany CNS relapse. It seems unlikely that a gene expression-based biomarker can be developed to positively identify patients at highest risk of CNS relapse, beyond the information that is already contained within the transcriptional footprints that define known DLBCL subtypes. In contrast, large-scale,
integrative analyses and in-depth characterization of clonal trajectories hold the promise to increase our ability to understand dissemination of DLBCL into the CNS.

Acknowledgments

This work was supported by an Innovation Grant (award # 703505) from the Canadian Cancer Society Research Institute (D.W.S.), by a Genome Canada Large-Scale Applied Research Project (Genome Canada #13124, Genome BC #271LYM, Canadian Institutes of Health Research #GP1-155873 and the BC Cancer Foundation) (C.S., D.W.S, M.A.M.), the Ontario Research Fund (R.K.), and by the Princess Margaret Cancer Foundation (R.K.). We thank the expert staff of the Michael Smith’s Genome Sciences Centre at BC Cancer in Vancouver and The Centre for Applied Genomics at SickKids in Toronto for generating exome sequencing data and gene expression. We thank Dr. Osvaldo Espin-Garcia for expert statistical advice.

Competing Interests

The authors declare no competing conflicts of interest relating to the work presented herein.
References


**Figure legends**

**Figure 1**: Clonal evolution patterns associated with central nervous system (CNS) relapse. PyClone analysis to retrace clonal phylogenies in five cases with documented CNS dissemination. The leftmost column shows variant allele frequencies (VAF) at diagnosis vs. VAF in CNS sample. The middle column depicts clusters of mutation cellular prevalence. The rightmost column shows time sweep plots that graphically illustrated clonal tides over the course of relapse. Representative DLBCL genes are labelled.

**Figure 2**: Gene mutations associated with CNS and systemic relapse. The barplot shows the proportion of samples harboring given gene mutations. Shown are 45 genes. The total sample size is 223 samples (n = 72 with CNS relapse, n = 62 with systemic relapse and n = 89 without relapse). The line plot shows differences between proportions of mutated samples, when comparing either CNS vs. no relapse, or CNS vs. systemic relapse. The lines indicate the ranges of the credible intervals for group differences, and the small circles represent the point estimates of the group differences.

**Figure 3**: Pathways associated with CNS and systemic relapse. A. Number of gene sets enriched by contrast and by COO, based on differential gene expression analysis. Only gene sets with adjusted \( P \) value of < 0.05 are shown. B. Heatmap showing the enrichment scores (ES) of the most down- and upregulated pathways by contrast and by COO. Pathways are grouped by biological theme.
**Supplemental Methods**

*Exome sequencing, detection of single nucleotide variants and clonal analysis*

Ten formalin-fixed and paraffin-embedded (FFPE) tissue DNA samples from five patients were submitted to the BC Genome Sciences Centre for library construction, Agilent SureSelect V6 exome capture and Illumina sequencing (Supplemental Table 1). Single nucleotide variants were called by Mutect2\(^1\) (version 2.1) using tumor only mode. Hg19 coordinates in VCF files were left-normalized using bcftools (samtools v1.9).\(^2\) Single nucleotide variants (SNVs) and indels were annotated by Annovar using the following datasets: ensGene, gnomad211, cosmic68 and avsnp142 (Version date 16 Apr 2018).\(^3\) Variants marked as population variants (dbSNP) were used further to infer copy number status around these regions using CNVkit.\(^4\) Additional filters were applied to obtain somatic variants. This included a minimum depth of 100 for all but one sample with low coverage (LY_CNSrel_003_T1, minimum depth of 30 was required for this sample), variant allele frequency greater than 0.1 as well as not being annotated in dbSNP or having a population allele frequency > 0.001. Variants mapping to sex chromosomes were excluded. CNVkit was used in tumor only mode with a flat reference to infer copy number changes in these ten samples using BAM files. Discrete copy number segments were first inferred using the haar algorithm.\(^5\) VCF files with variants marked as potential germline SNPs were used in the “call” function to infer allele-specific copy number.

Copy number segments were merged with annotated somatic variants using Bedtools (v2.27.1)\(^6\) to generate the input data for PyClone (v0.13.1)\(^7\) that includes read counts, allele-specific copy number status and tumor content. Paired analysis of clonal evolution was performed using Pyclone’s multisample mode. If a variant was identified as a somatic mutation in only one of the two samples, bam-readcount (https://github.com/genome/bam-readcount) was used to retrieve read counts for that position. Variants with no read coverage in one of the two paired samples were discarded for this analysis. Variants in the *IGH* locus were also removed and the final list included only those mutations in copy neutral regions (total copy number status = 2). PyClone was run on each pair using default parameters, in addition to tumor content, as determined by examination of corresponding H&E sections. Mutations were then clustered based on inferred cellular prevalence. To increase the confidence of clustering, we removed mutations that formed a cluster on their own, as well as mutations with a standard deviation of cellular prevalence > 0.15. PyClone was re-run using this new set of mutations. Citup\(^8\) was used to infer clone phylogenies based on mutation cellular frequencies obtained from PyClone. The maximum number of nodes was set to the one obtained by the final number of clusters from PyClone. The R package timescape (https://github.com/maiasmith/timescape) was used to visualize clonal dynamics and clonal phylogenies.

*Identification of gene mutations associated with CNS or systemic relapse*

In order to identify associations between specific gene mutations and the three above-mentioned clinical risk groups, we compiled a partially overlapping dataset with mutation data of 45 genes in 223 diagnostic samples (n = 72 with CNS relapse, n = 62 with systemic relapse and n = 89 without relapse). This dataset was derived by combining our own mutation data\(^9\) with two publicly available datasets (Reddy et al.\(^10\) and Klanova et al.\(^11\)). Enrichment across clinical groups was evaluated with a Bayesian implementation of the proportion test, using the BayesianFirstAid R package (v0.1).\(^12\) To that effect, the bayes.prop.test function was run with default parameters.
Bayesian data analysis was used as it has several advantages over traditional null hypothesis significance testing, given that it allows to describe the uncertainty of parameter estimates and that it does not rely on arbitrary $P$ value cut-offs.\textsuperscript{13}

**Case selection for gene expression profiling**

Given the relatively infrequent occurrence of CNS relapse, we put emphasis on accruing a cohort of diagnostic samples that was enriched for outcome events. Samples from 222 patients were retrospectively identified from the population-based BC Cancer lymphoma biorepository. Events were considered when they occurred within 1 year following diagnosis, given that CNS relapse is typically reported to occur early.\textsuperscript{14–16} Moreover, in this discovery study, we assumed that gene expression profiles were more likely to reveal predictive features of relapse if the time between diagnosis and relapse was short. Hence, patients were selected to fall into 3 different clinical groups: 1) Cases with documented CNS relapse ($n=50$) within 1 year of diagnosis of de novo DLBCL who were treated with R-CHOP (or R-CEOP) ($n=39$), or those who had concurrent CNS involvement at diagnosis ($n=11$); 2) Cases with refractory disease or systemic relapse ($n=64$) within the first year following diagnosis and prior treatment with R-CHOP, but who did not have evidence of CNS involvement at any point in time; and 3) Cases with neither CNS nor systemic relapse for at least 5 years following diagnosis and who had received R-CHOP ($n=108$). We balanced the 3 clinical groups with regards to cell-of-origin (COO), given that the ABC subtype has been reported as a risk factor for CNS and systemic relapse,\textsuperscript{11,17} and is defined by a distinct transcriptional footprint.\textsuperscript{18} Regarding CNS prophylaxis, prior to September 2002, intrathecal chemotherapy was recommended for all patients with involvement of specific extranodal sites (bone marrow if involvement by large cell lymphoma, epidural, testes or sinus). After September 2002, this guideline was restricted to only patients with sinus involvement. Beginning in 2013, high dose methotrexate ($3.5\text{g/m}^2$) was recommended for patients with testicular and kidney involvement and select other high-risk patients.

All pathological specimens were centrally reviewed by expert hematopathologists at BC Cancer. All cases were required to yield sufficient RNA from a diagnostic formalin-fixed paraffin-embedded (FFPE) tissue block. COO was determined using the Lymph2Cx assay, as previously described.\textsuperscript{19,20} Up to five $10\mu m$ tissue sections were cut from FFPE tissue blocks and extracted using the Qiagen AllPrep DNA/RNA FFPE Kit. Gene expression profiling using Affymetrix Human Gene 2.0 ST arrays was performed by The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, Canada.

**Gene expression analysis**

One sample (CNR7010T1) was an outlier based on average probe intensity and principal component analysis, but was retained in the study as its omission did not alter results in a significant way. We used the \textit{oligo} R package (v1.46.0) to perform background correction and quantile normalization using the Robust Multi-array Average (RMA) method.\textsuperscript{21} Multiple probes for a given gene were averaged to provide a single measurement per gene and per sample. The dataset was then filtered (genes retained if at least 5\% of samples had an intensity of greater than 1.5, and if the coefficient of variation was between 0.15 and 5). Differential gene expression analysis was performed for each COO subtype (ABC and GCB) using the \textit{limma} R package (v3.42.0).\textsuperscript{22} In order to identify and remove latent variation, we applied Surrogate Variable
Analysis using the svy R package (v3.34.0), with the number of surrogate variables identified as 1 and 0 for ABC and GCB-DLBCL, respectively (parameters: method = leek, vfilter = 1,000).\textsuperscript{23}

**Double-hit signature**

Calls for the recently published double hit gene expression signature were available for 74 of the 96 germinal centre B-cell-like (GCB) cases, based on the DLBCL90 NanoString assay presented in Ennishi et al.\textsuperscript{9} For all 96 GCB cases, we also computed double hit signature calls using the Affymetrix dataset and the PRPS package (available from https://github.com/ajiangsfu/PRPS). When comparing to the NanoString-based assay, the accuracy for our calls was 86\% and 93\%, when including or excluding indeterminate cases, respectively.

**Pathway enrichment analysis**

For pathway enrichment analysis, we followed the recommendations outlined by Reimand et al.\textsuperscript{24} For each of the three contrasts (CNS relapse vs. no relapse, systemic relapse vs. no relapse, CNS relapse vs. systemic relapse), we generated a gene list that was ranked by a descending moderated t-statistic and used as input for Gene Set Enrichment Analysis (GSEA, v4.0.3).\textsuperscript{25} GSEA was run with default parameters (number of permutations 1000, set size 15-500, enrichment statistic weighted), using the following gene sets: hallmark (h.all.v7.1), curated canonical pathways (c2.cp.v7.1), GO biological processes (c5.bp.v7.1), oncogenic signatures (c6.all.v7.1) and immunologic signatures (c7.all.v7.1). In addition, we supplemented these gene sets with a publicly available list of signatures that are relevant in the lymphoid context.\textsuperscript{26}

**Data Sharing Statement**

Mutation data can be found in Supplemental Table 2. Qualified researchers may obtain access to the original data used in this study. Microarray data have been deposited into the European Genome-phenome Archive (study accession EGAD00010001909). For exome sequencing data, please contact the senior authors of this study.

**References**


**Supplemental Figure 1:** Flow diagram illustrating all patients described in this study and their relationship with profiling techniques used, clinical risk groups and the source of samples and/or data.
**Supplemental Figure 2:** Number of mutations found to be shared, exclusively seen in T1 (i.e. extra-CNS biopsy) or T2 (CNS biopsy).
**Supplemental Figure 3:** Proportion of cases positive (in salmon color) for the double-hit signature, by clinical risk group.
**Supplemental Table 1**: Cases used for exome sequencing.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Source</th>
<th>Site timepoint 1</th>
<th>Site timepoint 2</th>
<th>Treatment between biopsies</th>
<th>Time to CNS relapse (years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LY_CNSrel_001</td>
<td>UHN</td>
<td>Breast</td>
<td>Brain</td>
<td>R-CHOP</td>
<td>2.78</td>
</tr>
<tr>
<td>LY_CNSrel_003</td>
<td>UHN</td>
<td>Lymph node</td>
<td>Brain</td>
<td>CHOP or CHOP-like</td>
<td>5.37</td>
</tr>
<tr>
<td>LY_CNSrel_005</td>
<td>BC Cancer</td>
<td>Stomach</td>
<td>Brain</td>
<td>R-CHOP</td>
<td>0.38</td>
</tr>
<tr>
<td>LY_CNSrel_006</td>
<td>BC Cancer</td>
<td>Testis</td>
<td>Brain</td>
<td>CHOP or CHOP-like</td>
<td>0.72</td>
</tr>
<tr>
<td>LY_CNSrel_007</td>
<td>BC Cancer</td>
<td>Testis</td>
<td>Brain</td>
<td>CNS involvement at diagnosis</td>
<td>0.08</td>
</tr>
</tbody>
</table>
**Supplemental Table 2:** Mutation data from 223 diagnostic DLBCL samples, by clinical group and source.

(see separate Excel file)
**Supplemental Table 3:** Patient included in gene expression study, characteristics by clinical group.

<table>
<thead>
<tr>
<th>Clinical Group</th>
<th>CNS relapse (n=50)</th>
<th>Systemic relapse (n=64)</th>
<th>No relapse (n=108)</th>
<th>CNS vs. systemic relapse</th>
<th>CNS vs. no relapse</th>
<th>Systemic vs. no relapse</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age ≤ 60</strong></td>
<td>19 (38%)</td>
<td>21 (33%)</td>
<td>40 (37%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Age &gt; 60</strong></td>
<td>31 (62%)</td>
<td>43 (67%)</td>
<td>68 (63%)</td>
<td>0.693</td>
<td>1.000</td>
<td>0.623</td>
<td></td>
</tr>
<tr>
<td><strong>ECOG ≤ 1</strong></td>
<td>12 (24%)</td>
<td>25 (39%)</td>
<td>65 (61%)</td>
<td>0.108</td>
<td></td>
<td></td>
<td>&lt;0.001</td>
</tr>
<tr>
<td><strong>ECOG &gt; 1</strong></td>
<td>38 (76%)</td>
<td>39 (61%)</td>
<td>42 (39%)</td>
<td></td>
<td></td>
<td></td>
<td>0.007</td>
</tr>
<tr>
<td><strong>Stage I-II</strong></td>
<td>8 (16%)</td>
<td>12 (19%)</td>
<td>47 (43.5%)</td>
<td>0.806</td>
<td>1.000</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td><strong>Stage III-IV</strong></td>
<td>42 (84%)</td>
<td>52 (81.2%)</td>
<td>61 (56.5%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>LDH normal</strong></td>
<td>10 (21%)</td>
<td>14 (23%)</td>
<td>55 (56%)</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>LDH elevated</strong></td>
<td>37 (79%)</td>
<td>47 (77%)</td>
<td>44 (44%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Nb extranodal sites 0-1</strong></td>
<td>22 (44%)</td>
<td>48 (75%)</td>
<td>89 (82%)</td>
<td>0.001</td>
<td></td>
<td></td>
<td>&lt;0.001</td>
</tr>
<tr>
<td><strong>Nb extranodal sites &gt; 1</strong></td>
<td>28 (56%)</td>
<td>16 (25%)</td>
<td>19 (18%)</td>
<td></td>
<td></td>
<td></td>
<td>0.248</td>
</tr>
<tr>
<td><strong>IPI low</strong></td>
<td>2 (4%)</td>
<td>7 (11%)</td>
<td>33 (31%)</td>
<td>0.288</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>IPI intermediate</strong></td>
<td>19 (38%)</td>
<td>27 (42%)</td>
<td>56 (52%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>IPI high</strong></td>
<td>29 (58%)</td>
<td>30 (47%)</td>
<td>19 (18%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>CNS-IPI low</strong></td>
<td>2 (4%)</td>
<td>6 (10%)</td>
<td>22 (22%)</td>
<td>0.352</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>CNS-IPI intermediate</strong></td>
<td>16 (34%)</td>
<td>25 (41%)</td>
<td>57 (58%)</td>
<td></td>
<td></td>
<td></td>
<td>&lt;0.001</td>
</tr>
<tr>
<td><strong>CNS-IPI high</strong></td>
<td>29 (62%)</td>
<td>30 (49%)</td>
<td>20 (20%)</td>
<td></td>
<td></td>
<td></td>
<td>0.001</td>
</tr>
<tr>
<td><strong>Non-double hit</strong></td>
<td>35 (88%)</td>
<td>48 (80%)</td>
<td>95 (91%)</td>
<td>0.420</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Double hit</strong></td>
<td>5 (12%)</td>
<td>12 (20%)</td>
<td>9 (9%)</td>
<td></td>
<td></td>
<td></td>
<td>0.051</td>
</tr>
<tr>
<td><strong>Cell-of-origin ABC</strong></td>
<td>19 (40%)</td>
<td>31 (48%)</td>
<td>36 (33%)</td>
<td>0.458</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Cell-of-origin GCB</strong></td>
<td>19 (40%)</td>
<td>25 (39%)</td>
<td>52 (48%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Cell-of-origin Unclassified</strong></td>
<td>10 (21%)</td>
<td>8 (12%)</td>
<td>20 (19%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>R-CHOP</strong></td>
<td>34 (68%)</td>
<td>64 (100%)</td>
<td>108 (100%)</td>
<td>&lt;0.001</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>R-CEOP</strong></td>
<td>5 (10%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td></td>
<td></td>
<td></td>
<td>&lt;0.001</td>
</tr>
<tr>
<td><strong>CNS invasion at diagnosis</strong></td>
<td>11 (22%)</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td></td>
<td></td>
<td></td>
<td>1.000</td>
</tr>
</tbody>
</table>

ABC, activated B-cell subtype; CNS-IPI, Central Nervous System - International Prognostic Index; ECOG, Eastern Cooperative Oncology Group; GCB, germinal centre B-cell subtype; LDH, lactate dehydrogenase; IPI, International Prognostic Index.
**Supplemental Table 4**: Results from gene set enrichment analysis.
(see separate Excel file)