

A three-gene signature based on *MYC*, *BCL-2* and *NFKBIA* improves risk stratification in diffuse large B-cell lymphoma

Enrico Derenzini,^{1,2} Saveria Mazzara,³ Federica Melle,³ Giovanna Motta,³ Marco Fabbri,³ Riccardo Bruna,¹ Claudio Agostinelli,⁴ Alessandra Cesano,⁵ Chiara Antonia Corsini,⁶ Ning Chen,⁵ Simona Righi,⁴ Elena Sabattini,⁴ Annalisa Chiappella,⁷ Angelica Calleri,³ Stefano Fiori,³ Valentina Tabanelli,³ Antonello Cabras,⁸ Giancarlo Pruneri,⁸ Umberto Vitolo,⁹ Alessandro Massimo Gianni,¹ Alessandro Rambaldi,¹⁰ Paolo Corradini,⁷ Pier Luigi Zinzani,¹¹ Corrado Tarella^{1,2} and Stefano Pileri³

¹Onco-Hematology Division, IEO European Institute of Oncology IRCCS, Milan and European Institute of Oncology IRCCS, Milan, Italy; ²Department of Health Sciences, University of Milan, Milan, Italy; ³Division of Diagnostic Hematopathology, IEO European Institute of Oncology IRCCS, Milan and European Institute of Oncology IRCCS, Milan, Italy; ⁴Hematopathology Unit, Department of Experimental, Diagnostic, and Specialty Medicine (DIMES), Bologna University School of Medicine, Bologna, Italy; ⁵NanoString Technologies Inc, Seattle, WA, USA; ⁶Laboratory of Hematology-Oncology, IEO European Institute of Oncology IRCCS, Milan and European Institute of Oncology IRCCS, Milan, Italy; ⁷Division of Hematology, Fondazione IRCCS Istituto Nazionale dei Tumori, University of Milan, Milan, Italy; ⁸Department of Pathology, Fondazione IRCCS Istituto Nazionale dei Tumori di Milano, Milan, Italy; ⁹Multidisciplinary Oncology Outpatient Clinic, Candiolo Cancer Institute, FPO-IRCCS, Candiolo, Italy; ¹⁰Hematology and Bone Marrow Transplant Unit, ASST-Papa Giovanni XXIII, Bergamo, Italy and ¹¹Institute of Hematology and Medical Oncology “L. e A. Seragnoli”, Department of Experimental, Diagnostic, and Specialty Medicine (DIMES), Bologna University School of Medicine, Bologna, Italy

©2021 Ferrata Storti Foundation. This is an open-access paper. doi:10.3324/haematol.2019.236455

Received: August 25, 2019.

Accepted: August 3, 2020.

Pre-published: August 13, 2020.

Correspondence: *ENRICO DERENZINI* - enrico.derenzini@ieo.it

STEFANO PILERI - stefano.pileri@ieo.it

Supplemental information

Supplemental methods

Original R-HDS0305 and DLCL04 trials, and validation cohorts

Two hundred and forty-six and 399 patients with high risk were enrolled respectively in the R-HDS trial¹ and in the DLCL04 trial². Median follow-up was 5 years in the R-HDS0305 study and 72 months in the DLCL04 trial. The results of both studies did not support the role of first line intensification in DLBCL. Five-year overall survival (OS) rates were similar in the 2 studies with 74% and 77% 5-year OS in the no transplantation groups vs 77% and 78% 5-year OS in the transplantation groups of the R-HDS0305 and DLCL04 trials respectively. The overall outcome of the patients analyzed in the present study was superimposable to the outcome of the 2 original studies^{1,2} (5-year OS 78%).

We validated our results in 3 independent cohorts, including a real-life cohort and 2 *in silico* validation datasets: a dataset from the recent study from Sha and coworkers including 928 patients (469 treated with R-CHOP and 459 with R-CHOP plus Bortezomib)²⁷; a public gene expression dataset [Affymetrix Human Genome U133 Plus 2.0 Array, GSE10846³⁶, available in the Gene Expression Omnibus (GEO) Database (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE10846>), including 233 patients treated with R-CHOP; an additional validation cohort including 102 consecutive DLBCL NOS cases with available FFPE tissue, treated with R-CHOP/R-CHOP-like regimens in “real-life” at the S. Orsola-Malpighi Hospital, Bologna (Italy) and European Institute of Oncology (Milan, Italy) from 2007 to 2018.

Characteristics of the 3 validation cohorts used in this study are summarized in table S1.

Immunohistochemistry

Immunohistochemistry (IHC) was centralized in Bologna for the DLCL04 trial and in Milan for the RHDS0305 and real-life control group. The antibodies source and dilutions are shown in Table S2. At both sites, antigen retrieval was carried on PT-links by the high pH solution (Dako Agilent). All IHC tests were performed on AutoStainer Plus platforms, using the LSAB+ REAL Detection System (Dako Agilent). The IHC preparations were counterstained with Gill's haematoxylin and mounted in Kaiser's glycerin. The IHC results were independently evaluated by 4 expert haematopathologists (CA, SAP, ES, VT). The Hans' algorithm was used for the COO assessment, while the cut-off values of 50% and 40% positive neoplastic cells were applied for BCL2 and MYC, respectively (according to the Revised 4th Edition of the WHO Classification of Tumours of Haematopoietic and Lymphoid Tissue)³. In case of discrepant results among the observers, the IHC preparations were jointly reviewed at a multi-head microscope until consensus was reached.

FISH analysis

FISH studies were conducted on paraffin sections using the following probes: Vysis LSI MYC dual color break-apart, Vysis LSI BCL2 dual color break-apart, Vysis LSI BCL6 dual color break-apart and Vysis LSI IGH/MYC/CEP8 Tri-color FISH probe kit. In brief, the slides were deparaffinized, co-denatured with probe, hybridized overnight, washed and then mounted with DAPI/Antifade. For each probe, 200 interphase nuclei were analyzed to detect rearrangement and numerical abnormalities. Cut-off values were established for each probe by assessing 10 normal controls (reactive lymph nodes) and choosing values 3SD above the mean. Gains were considered when a

pattern of three or four copies of the gene were identified, whereas more than four copies were considered as amplifications.

NanoString methods

Total RNA was extracted from three sections 15- μ m-thick of each FFPE sample using RecoverAll Total Nucleic Acid Isolation Kit for FFPE (Thermo Fisher). Yield and quality of RNA extracted was assessed. Quantitative RNA analysis was performed using NanoDrop ND-1000 Spectrophotometer (NanoDrop Technologies, Rockland, DE, USA). RNA quality was scored according to DV200 values (percentage of RNA fragments \geq 200 nucleotides), utilizing the Agilent 2100 BioAnalyzer.

Gene expression was measured on the NanoString nCounter Analysis System (NanoString Technologies, Seattle, WA, USA).

Gene expression data were analyzed by the NanoString Company using a modified RUO version of the NanoString Lymphoma Subtyping Test (LST) algorithm to determine the Cell-of-Origin molecular subtype of each sample⁴. The system computes the relative abundance of each mRNA transcript of interest, through a multiplexed hybridization assay and digital readouts of fluorescent barcoded probes, which are hybridized to each transcript. An nCounter CodeSet (NanoString Technologies) containing capture and reporter probes (the latter attached to a color barcode) was hybridized to 200 ng of total RNA for 20 hours at 65 °C, according to the manufacturer's instructions. Hybridized samples were loaded into the nCounter Prep Station for post-hybridization processing. Target mRNA was assessed with nCounter Digital Analyzer.

The quality control and normalization of NanoString nCounter data were performed using R package NanoStringNorm. The Raw NanoString counts for each gene were subjected to a technical normalization considering positive and negative probes. A normalization factor was calculated by obtaining the geometric mean of the positive controls used for each sample and applied to the raw counts of the nCounter output data to eliminate variability that was unrelated to the samples. The

resulting data were normalized again with the geometric mean of the housekeeping genes (*ISY1*, *R3HDM1*, *TRIM56*, *UBXN4* and *WDR55*). Normalized data were log₂-transformed for further analyses. Statistical analyses were calculated with the R statistical programming environment (v3.5.0).

Statistical analyses

Recursive Partitioning Analysis

This algorithm models the association between response and covariates building a tree that resembles a division that is most prognostic for survival. The tree-structure model was performed using the R package party (<https://cran.r-project.org/web/packages/party/index.html>), and applied to OS data for survival prediction analysis. For T-GEP analyses and classification of patients into high and low MBN expression groups, high and low expressors were defined based on the median values of mRNA expression.

Correlations and differences in patient characteristics

Differences between groups were analyzed with the χ^2 and Fisher's exact test. A p value ≤ 0.05 was considered as statistically significant.

Random Forest Classifier

A Random Forest (RF) classifier was constructed using with the R randomForest package (<https://cran.r-project.org/web/packages/randomForest/index.html>) including the expression values of the three informative genes. The discovery cohort (n=186) was used as a training dataset, the model was built using gene expressions as prediction variables and the MBN groups as the categorical outcome. To give an estimate of the model performance in an unbiased manner a splitting procedure was introduced; we randomly generate the training (80%) and test (20%) partitions from the discovery cohort. By applying the RF model, samples were classified into the two MBN subgroups and

discriminated according to expression values of the three genes. An independent validation set, based on real-life cases (n=102), was used to confirm the robustness and transferability of the classifier. The performance of the classifier was assessed by accuracy (ACC), sensitivity (SE) and specificity (SP). Moreover, a receiver operating characteristic (ROC) was constructed to evaluate the classification efficiency of the RF classifier using the ROCR library (<https://cran.r-project.org/web/packages/ROCR/index.html>) considering the area under ROC (AUC) as classification performance metric.

List of genes and target sequences

GeneName	ProbeID	Comments	TargetSeq
1	UBXN4 NM_014607.3	HOUSEKEEPING	CATCGCGACGGCCAAAAGGAGCGGCGGGTCTTCGTGGTGTTCGTGGCAGGTGATGATGAACAGTCTACACAGATGGCTGCAAGTTGGGAAGATGATAAA
2	ISY1 NM_020701.2	HOUSEKEEPING	GGCAAACATCAGTGTCTGTGGGTAGTTGGAATCTTCAGTTCCTGTGAGCGTCGCGTCTTCTGGGCTGTGGAGTTTCTTGGACAGGGGCCGCGGGGCT
3	R3HDM1 NM_015361.2	HOUSEKEEPING	CCTGTGTTCCCAAGAGAATTACATTATTGACAAAAGACTCCAAGACGAGGATGCAGTAGTACCCAGCAGAGGCGCCAGATATTTAGAGTTAATAAAGAT
4	WDR55 NM_017706.4	HOUSEKEEPING	CTACCTCTTCAATTGGAATGGCTTTGGGGCCACAAGTGACCGCTTTGCCCTGAGAGCTGAATCTATCGACTGCATGGTTCCAGTCACCGAGAGTCTGCTG
5	TRIM56 NM_030961.1	HOUSEKEEPING	GTGGAGGCCGAGGACATTTTCTGAAGGGCAGGGGTGGCAACTTTTCAACATGAGTGCCAAACTGCTAACCCGTCTTCTAGTGTGTGAGAATAGGGAC
6	MYC NM_002467.3	ENDOGENOUS	TCGGACACCGAGGAGAATGTCAAGAGGCGAACACACAACGTCTTGGAGCGCCAAGGAGGAACGAGCTAAAACGGAGCTTTTTTGGCCCTGCGTGACCAGA
7	PKI3CA NM_006218.2	ENDOGENOUS	CCTCAGGCTTGAAGAGTGTCTGAATTATGTCCTCTGCAAAAAGGCCACTGTGGTTGAATTGGGAGAACCAGACATCATGTCTCAGAGTTACTGTTTCAGAAC

8 PIM2 NM_006875.2 ENDOGENOUS
GCCATCCAGCACTGCCATTCCCGTGGAGTTGTCCATCGTGACATCAAGGATGAG
AACATCCTGATAGACCTACGCCGTGGCTGTGCCAAACTCATTGATT

9 IRF4 NM_002460.1 ENDOGENOUS
GGGCACTGTTTAAAGGAAAGTTCCGAGAAGGCATCGACAAGCCGGACCCTCCC
ACCTGGAAGACGCGCCTGCGGTGCGCTTTGAACAAGAGCAATGACTT

10 NFKBIA NM_020529.1 ENDOGENOUS
GGATGAGGAGAGCTATGACACAGAGTCAGAGTTCACGGAGTTCACAGAGGACG
AGCTGCCCTATGATGACTGTGTGTTTGGAGGCCAGCGTCTGACGTTA

11 STAT3 NM_139276.2 ENDOGENOUS
AGACTTGGGCTTACCATTGGGTTTAAATCATAGGGACCTAGGGCGAGGGTTCAG
GGCTTCTCTGGAGCAGATATTGTCAAGTTCATGGCCTTAGGTAGCA

12 TNFRSF13B NM_012452.2 ENDOGENOUS
TGCAAAACCATTTGCAACCATCAGAGCCAGCGCACCTGTGCAGCCTTCTGCAGG
TCACTCAGCTGCCGCAAGGAGCAAGGCAAGTTCTATGACCATCTCC

13 S1PR2 NM_004230.2 ENDOGENOUS
TCCCGCCAGGTGGCCTCGGCCTTCATCGTCATCCTCTGTTGCGCCATTGTGGTGG
AAAACCTTCTGGTGCTCATTGCGGTGGCCCGAAACAGCAAGTTCC

14 MME NM_000902.2 ENDOGENOUS
GGATTGTAGGTGCAAGCTGTCCAGAGAAAAGAGTCCTTGTTCCAGCCCTATTCT
GCCACTCCTGACAGGGTGACCTTGGGTATTTGCAATATTCCTTTGG

15 ASB13 NM_024701.3 ENDOGENOUS
GGACACGTAGGCGGTACCACTAAGGTTTTGGTAATGAGCCATTCAAACCGACAG
CAGTGTGAAGGTGTGTCAAGGTGTATATTCTCGTGGCTCGGCATTCC

16 BCL2 NM_000657.2 ENDOGENOUS
GTGAAGCAGAAGTCTGGGAATCGATCTGGAAATCCTCCTAATTTTTACTCCCTCT
CCCCGCGACTCCTGATTCATTGGGAAGTTTCAAATCAGCTATAAC

17 CYB5R NM_016229.3 ENDOGENOUS
CCATGTCTTAGGGCTTCCTGTAGGTA ACTATGTCCAGCTCTTGGCAAAAATCGAT
AATGAATTGGTGGTCAGGGCTTACACCCCTGTCTCCAGTGATGAT

18 MAML3 NM_018717.4 ENDOGENOUS
TGGAAGCCATCAACAATTTGCCAGTAACATGCCACTGCCTTCAGCTTCTCCTCT
TCACCAACTTGACCTGAAACCTTCTTTGCCCTTGCAGAACAGTGG

19 SERPINA9 NM_001042518.1 ENDOGENOUS
CCACTAAATCCTAGGTGGGAAATGGCCTGTAACTGATGGCACATTGCTAATGC
ACAAGAAATAACAAACCACATCCCTCTTTCTGTTCTGAGGGTGCAT

20 MYBL1 XM_034274.14 ENDOGENOUS
CTCCTTTTAAGAATGCGCTTGCTGCTCAGGAGAAAAAATATGGACCTCTTAAAA
TTGTGTCCCAGCCACTTGCTTTCTTGGGAAGAAGATATTCGGGAAGT

21 RAB7L1 NM_001135664.1 ENDOGENOUS
CATTGGAATTGTCTCCTGACTACTGTCCAGTAAGGAGGCCCATTTGTCACCTTAGAA
AAGACACCTGGAACCCATGTGCATTTCTGCATCTCCTGGATTAGC

22 LIMD1 NM_014240.2 ENDOGENOUS
AAGGCAAGTCTCAGGAACCCATGCAGGTACATCGCTTGCACCTGTTTTTAGCTT
ATTTAATGACGGGCTTTTGGGAAGAGCTGCCCGCATACTGAGAGAC

23 ITPKB NM_002221.3 ENDOGENOUS
GGTTTGCGCCTCTGGGCATGTAGTCTACACAGGACCTGAGAATCTGAGAAACTG
CAGCCGCACGGTTGTTTATGGAGCTTTGGGCGGGGGCTGAGCCCGC

24 PTEN NM_000314.4 ENDOGENOUS
TCTTGACCAATGGCTAAGTGAAGATGACAATCATGTTGCAGCAATTCACCTGTAA
AGCTGGAAAGGGACGAACTGGTGTAAATGATATGTGCATATTTATTA

25 CREB3L2 NM_001253775.1 ENDOGENOUS
CGCACTTCTCAGAACTTCTGGATGAGTTTTCCAGAACGTCTTGGGTCAGCTCCT
GAATGATCCTTTCCTCTCAGAGAAGAGTGTGTCAATGGAGGTGGA

26 CCDC50 NM_174908.3 ENDOGENOUS
AGGACATAGCTCGCCTTTTGGCAAGAAAAGGAGTTACAGGAAGAGAAAAAGAGA
AAGAAACACTTTCAGAGTTCCTGCAACCCGTGCTTATGCAGATAG

Supplementary Figures Legends

Figure S1. Overall survival curves according to COO defined by immunohistochemistry (Hans algorithm) or by NanoString-based T-GEP (Lymph2Cx signature). P values were calculated with the log rank test.

- A) OS of the discovery cohort (R-HDS0305+DLCL04; n=186 patients) according to the COO defined by IHC (Hans algorithm), showing no significant differences in OS between GCB and non-GCB DLBCL subtypes.
- B) OS of the discovery cohort (R-HDS0305+DLCL04; n=186 patients) according to the COO defined by nanostring-based T-GEP, showing a significantly worse outcome for ABC-derived DLBCL as compared to GCB and unclassified subgroups.

- C) OS of patients treated with chemoimmunotherapy in the absence of ASCT consolidation in the discovery cohort (R-HDS0305 + DLCL04; n=105), according to the COO as determined by T-GEP.
- D) OS of patients treated with chemoimmunotherapy followed by ASCT consolidation in the discovery cohort (R-HDS0305 + DLCL04; n=81), according to the COO as determined by T-GEP.

Figure S2. Prognostic impact of MYC/BCL-2 status, correlation between T-GEP and immunohistochemistry, and correlation with the COO.

- A) Box plot graphs showing significant correlation and concordance between NanoString and immunohistochemistry in the determination of BCL-2 and MYC levels. mRNA levels detected by NanoString in the BCL-2 and MYC negative and positive subgroups as classified by immunohistochemistry (applying a standard 50% and 40% cut-off for BCL-2 and MYC respectively) are represented here. All cases but one (n=98) from the DLCL04 trial were evaluable for both T-GEP and IHC. In the RHDS0305 trial, although NanoString GEP was performed in all cases (n=87), evaluable tissue for additional IHC stainings besides the Hans classification was obtained only in 43 instances for MYC and 82 instances for BCL-2. For this reason the total number of cases evaluable for MYC and BCL-2 IHC in the discovery cohort was 141 and 180 respectively. P values were calculated with the Student's T test.
- B) OS according to the *MYC* and *BCL-2* status in the discovery cohort (R-HDS+DLCL04; n=186 patients). *MYC* and *BCL-2* low and high expressors were defined according to the median values of mRNA expression. DEXPmRNA: double expressors. P value was calculated with the log rank test.

- C) OS according to the *MYC* and *BCL-2* status assessed by T-GEP in the GCB/U patient subgroup. P value was calculated with the log rank test.
- D) OS according to the *MYC* and *BCL-2* status assessed by T-GEP in the ABC patient subgroup. P value was calculated with the log rank test.
- E) OS according to *BCL-2* levels assessed by T-GEP in the GCB/U patient subgroup. P value was calculated with the log rank test. Low and high expressors were defined according to the median values of mRNA expression.
- F) OS according to *MYC* levels assessed by T-GEP in the GCB/U patient subgroup. P value was calculated with the log rank test. Low and high expressors were defined according to the median values of mRNA expression.

Figure S3. Prognostic impact of the additional targets included in the NanoString panel. P values were calculated with the log rank test.

- A) OS according to *NFKBIA* levels as determined by NanoString in the discovery cohort (DLCL04 + R-HDS, n=186). *NFKBIA* low and high expressors were defined according to the median values of mRNA expression.
- B) OS according to *STAT3* levels as determined by NanoString in the discovery cohort (DLCL04 + R-HDS, n=186). *STAT3* low and high expressors were defined according to the median values of mRNA expression.

Figure S4. Univariate Z-score analyses.

Bar graph depicting all genes ranked according to their predictive power in univariate Z-score analysis.

Figure S5. Progression-free survival in the discovery cohort and in the Sha's validation cohort according to the MBN signature

- A) PFS of the discovery cohort according to the MBN-signature, showing significant differences between MBN-Sig low vs MBN-Sig high patients subsets. P values were calculated with the log rank test.
- B) PFS of the Sha's cohort according to the MBN-signature, showing significant differences between MBN-Sig low vs MBN-Sig high patients subsets. P values were calculated with the log rank test.

Figure S6. Survival curves according to MBN-signature in two additional validation cohorts.

- A) OS of the Lenz's validation cohort (n=233 patients) according to the MBN-signature showing significant differences in outcome between MBN-Sig low vs high patients subsets. P values were calculated with the log rank test.
- B) OS of the real-life validation cohort (n=102 patients) according to the MBN-signature showing significant differences in outcome between MBN-Sig low vs high patients subsets. P values were calculated with the log rank test.
- C) Frequencies of MBN-Sig high vs low cases in ABC and GCB/U subsets in the Lenz's validation cohort (n=233 patients)
- D) Frequencies of MBN-Sig high vs low cases in ABC and GCB/U subsets in the real life validation cohort (n=102 patients)

Figure S7. The MBN signature identifies a significant proportion of poor prognosis DLBCL subsets enriched in DH, MHG and ABC DLBCL cases.

- A) Graphs depicting proportions of MBN-Sig high vs. MBN-Sig low subgroups in ABC-derived DLBCL vs non-ABC, in HG-BCL w/DH vs non-DH, and in cases with intermediate-high aaIPI (2) vs high aaIPI (3) in the discovery cohort. P values were calculated with the χ^2 test.
- B) Graphs depicting proportions of MBN-Sig high vs. MBN-Sig low subgroups in ABC-derived DLBCL vs non-ABC, in HG-BCL w/DH vs non-DH, in MHG vs non-MHG, and in cases with low IPI (0-2) vs high IPI (3-5) in the Sha's cohort. P values were calculated with the χ^2 test.

Figure S8. Prognostic value of consolidation ASCT in the MBN-Sig high subgroup of the discovery cohort.

- A) PFS of patients treated with or without ASCT consolidation in the MBN-Sig high subgroup (discovery cohort). P value was calculated with the log rank test.
- B) OS of patients treated with or without ASCT consolidation in the MBN-Sig high subgroup (discovery cohort). P value was calculated with the log rank test.

Supplementary Tables

Table S1. Patients characteristics in the validation cohorts

Variable			
Cohort	<i>Sha cohort</i>	<i>Lenz cohort</i>	<i>Real-life</i>
N° of patients	469	233	102
Immuno-CHT alone	n/a	n/a	102 (100%)
Immuno-CHT + ASCT	n/a	n/a	-
Median age, y (range)	66 (24-86)	61 (17-92)	61 (17-88)
COO			
ABC	129 (28%)	93 (40%)	36 (35%)
GCB	277 (59%)	107 (46%)	49 (48%)
Unclassified	63 (13%)	33 (14%)	17 (17%)
COO Hans IHC			
Non-GCB	n/a	n/a	47 (46%)
GCB-like	n/a	n/a	23 (23%)
			32 (31%) n/a
Stage (Ann Arbor)	I-IV	I-IV	I-IV
IPI score			
0-2	239 (51%)	101 (43%)	58 (57%)
3-5	230 (49%)	63 (27%)	43 (42%)
		69 (30%) n/a	1 (1%) n/a

n/a: not available

Table S2. Antibodies source and IHC conditions

Antibody	Source	Clone	Dilution
CD10	Leica	56C6	1:40
CD5	Dako	4C7	1:40
CD20	Dako	L26	1:150
KI-67	Dako	MIB-1	1:100
BCL-2	Dako	124	1:100
c-MYC	Abcam	Y69	1:100
MUM1/IRF4	Kindly provided by Prof. Falini	2C10-2D6	1:4
BCL6	Kindly provided by Prof. Falini	1G1	Undiluted

Table S3. Distribution of prognostic biomarkers and therapeutic targets according to the COO in the discovery cohort. P value was calculated with chi-square test. High and low subgroups were defined based on the median mRNA expression values.

Factor	ABC	GCB/Unclassified	p-value
<i>MYC</i>			
Low	15	78	0.1
High	25	68	
<i>BCL2</i>			
Low	6	87	<0.001
High	34	59	
<i>MYC-BCL-2</i> DEXPmRNA			
No	17	119	<0.001
Yes	23	27	
<i>STAT3</i>			
Low	21	73	0.9
High	19	73	
<i>NFKBIA</i>			
Low	30	66	0.001
High	10	80	
<i>PIK3CA</i>			
Low	22	71	0.59
High	18	75	
<i>PTEN</i>			
Low	24	69	0.21
High	16	77	

Table S4. Multivariate Cox regression analysis (discovery cohort)

Variable	Hazard Ratio (95% CI)
COO Nano GCB/U ABC	ref 2.81 (1.35-5.83)
DEXP mRNA no yes	ref 0.92 (0.45-1.86)
aaIPI 2 3	ref 2.92 (1.52-5.60)
STAT3 Low High	ref 0.52 (0.26-1.04)
MBN Signature Low High	ref 3.01 (1.27-7.15)

Table S5. Multivariate Cox regression analysis (Sha cohort)

Variable	Hazard Ratio (95% CI)
COO GCB/U ABC	ref 1.63 (0.97-2.75)
DEXP mRNA no yes	ref 0.99 (0.59-1.67)
IPI 0-2 3-5	ref 2.01 (1.27-3.19)
STAT3 Low High	ref 0.42 (0.26-0.68)
MBN Signature Low High	ref 2.46 (1.46-4.14)

References

- 1- Cortelazzo S, Tarella C, Gianni AM, Ladetto M, Barbui AM, Rossi A et al. Randomized Trial Comparing R-CHOP Versus High-Dose Sequential Chemotherapy in High-Risk Patients With Diffuse Large B-Cell Lymphomas. *J Clin Oncol*. 2016; 34(33): 4015-4022.
- 2- Chiappella A, Martelli M, Angelucci E, Brusamolino E, Evangelista A, Carella AM et al. Rituximab-dose-dense chemotherapy with or without high-dose chemotherapy plus autologous stem-cell transplantation in high-risk diffuse large B-cell lymphoma (DLCL04): final results of a multicentre, open-label, randomised, controlled, phase 3 study. *Lancet Oncol*. 2017; 18(8): 1076-1088.
- 3- Swerdlow SH, Campo E, Harris NL, Jaffe ES, Pileri SA, Stein H, Thiele J: WHO Classification of Tumour of Haematopoietic and Lymphoid Tissues, Revised 4th edition. 2017. IARC Press, Lyon.
- 4- Scott DW, Wright GW, Williams PM, Lih CJ, Walsh W, Jaffe ES et al. Determining cell-of-origin subtypes of diffuse large B-cell lymphoma using gene expression in formalin-fixed paraffin-embedded tissue. *Blood*. 2014; 123(8): 1214-7.

Figure S1

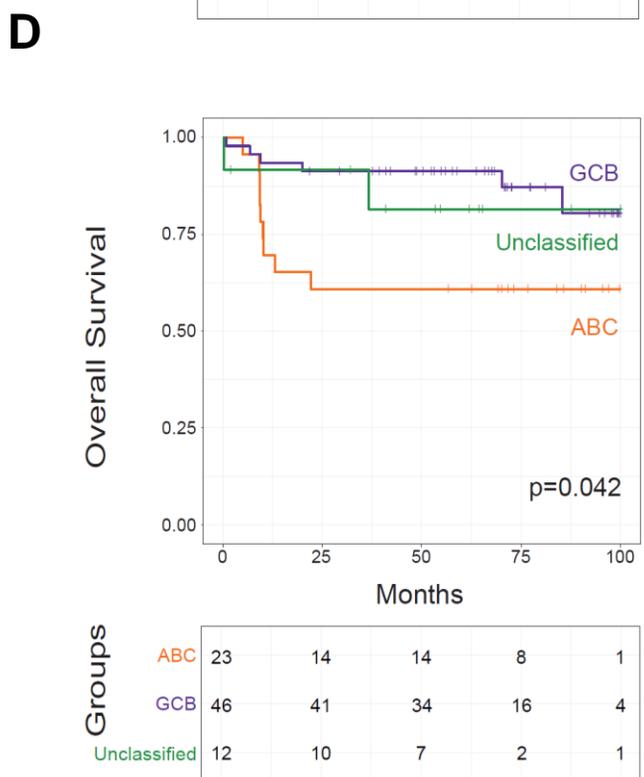
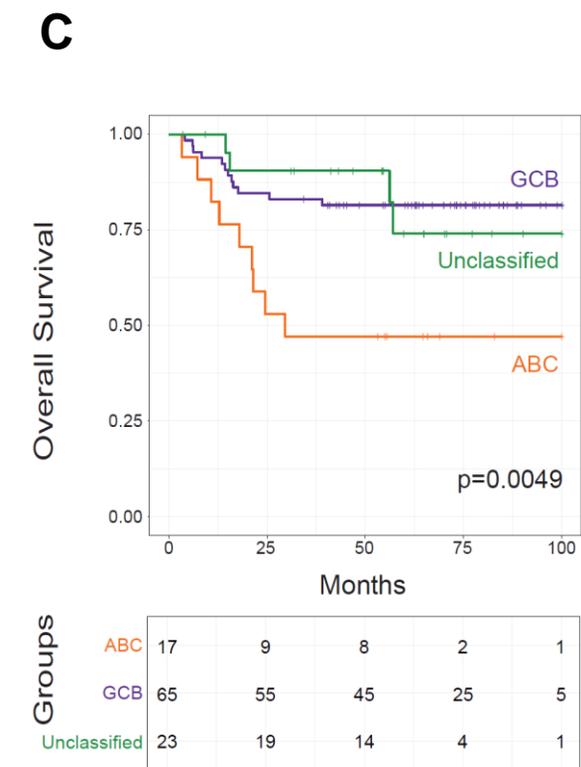
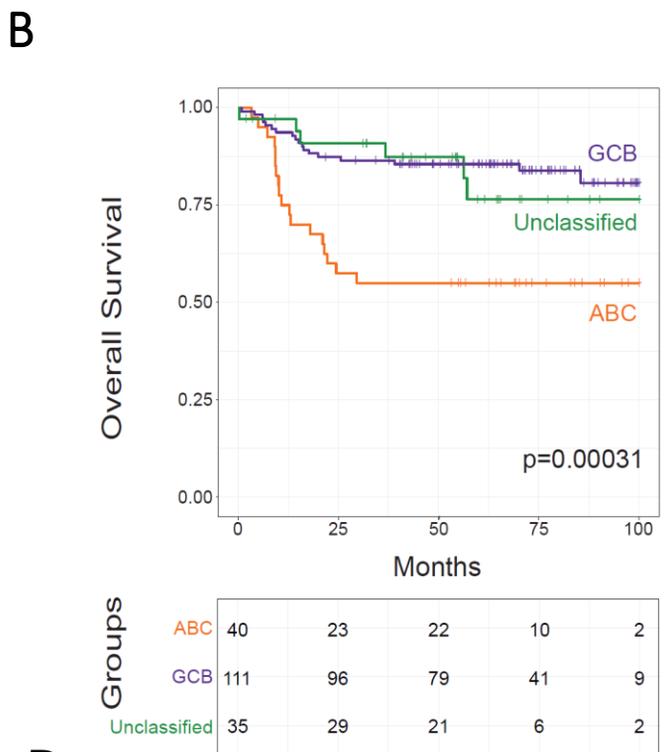
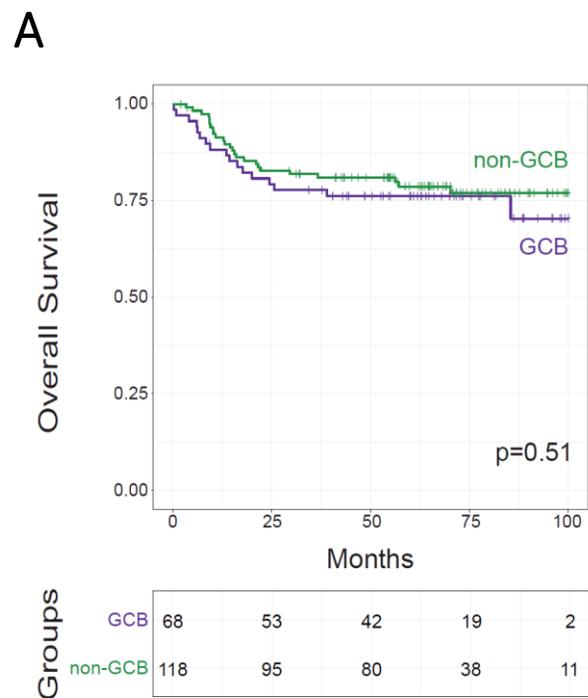


Figure S2

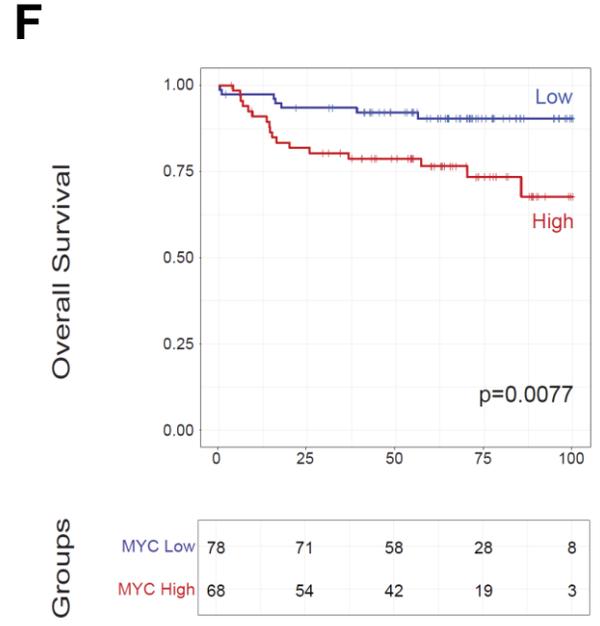
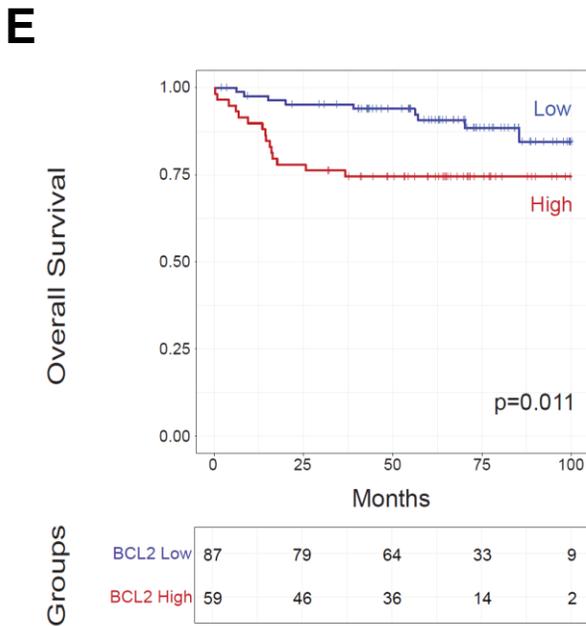
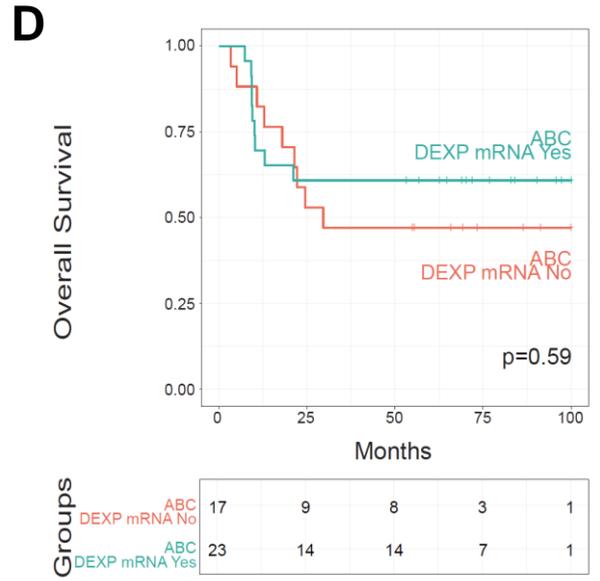
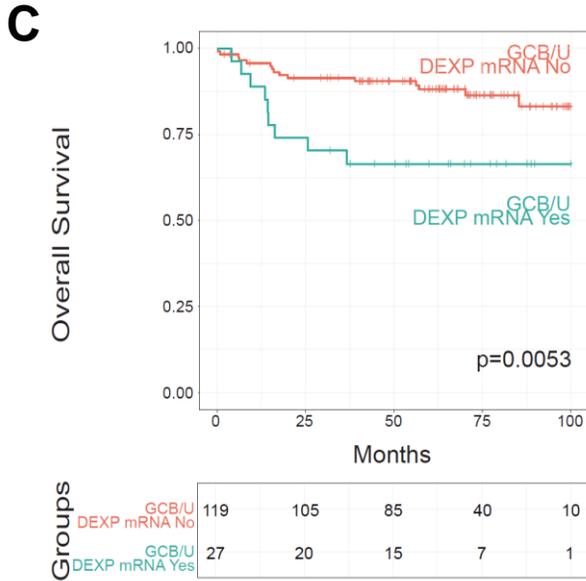
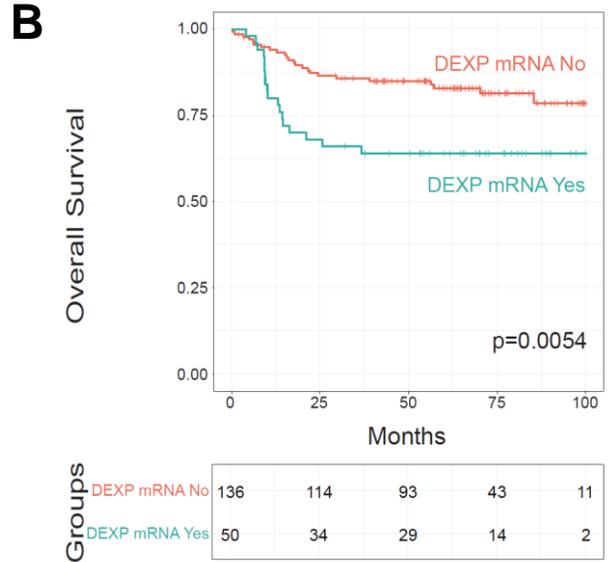
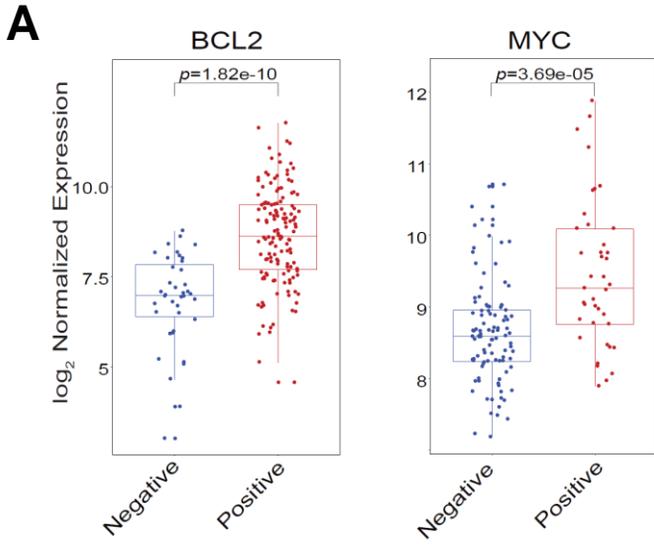
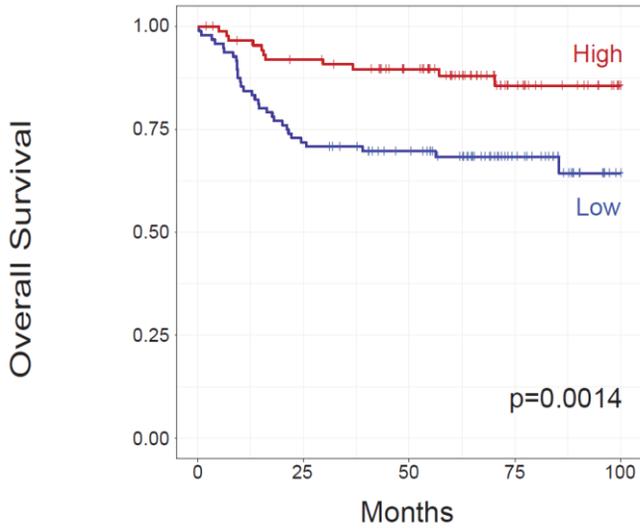


Figure S3

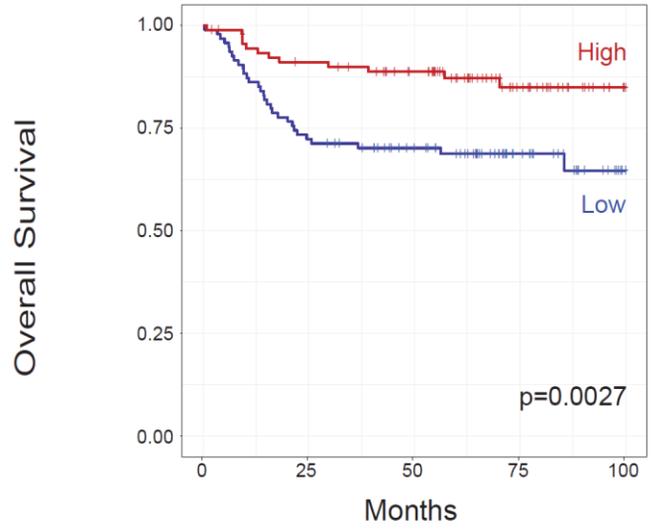
A



Groups

NFKBIA Low	96	69	57	29	4
NFKBIA High	90	79	65	28	9

B



Groups

STAT3 Low	94	68	54	25	4
STAT3 High	92	80	68	32	9

Figure S4

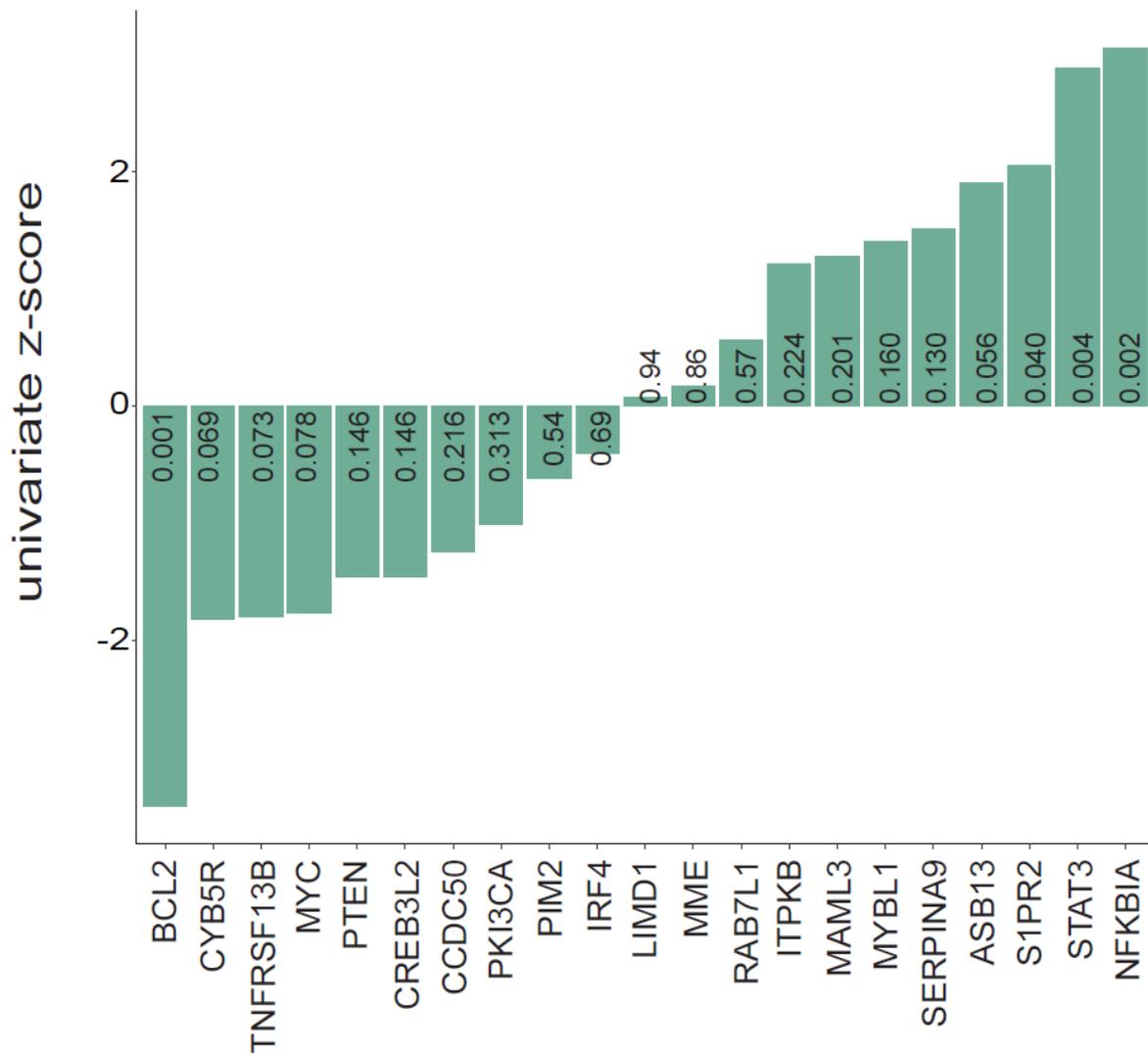
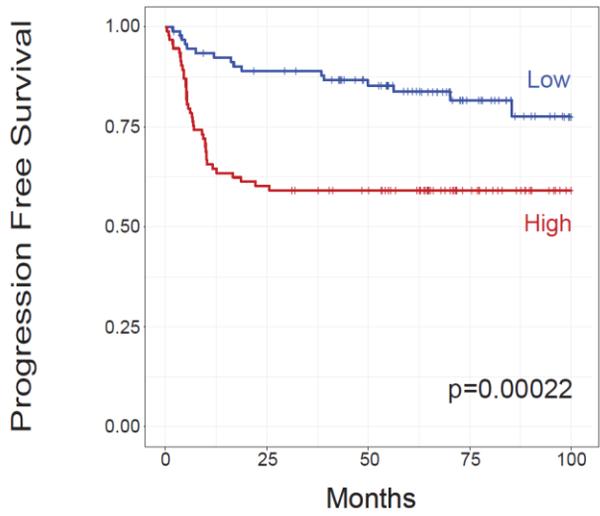


Figure S5

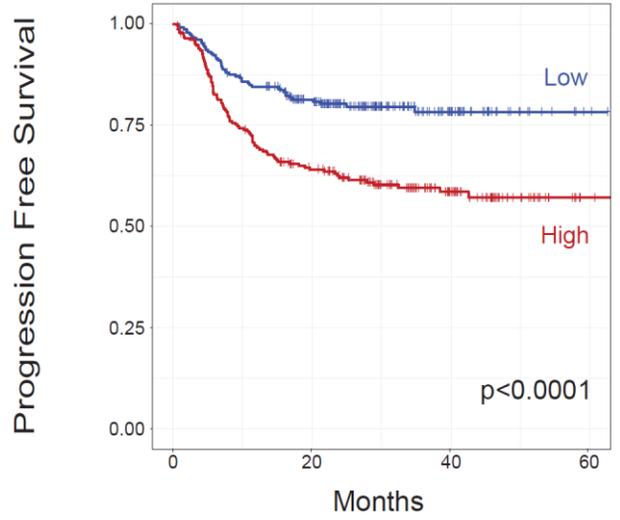
A



Groups

MBN Low	93	79	64	31	9
MBN High	93	56	49	21	1

B

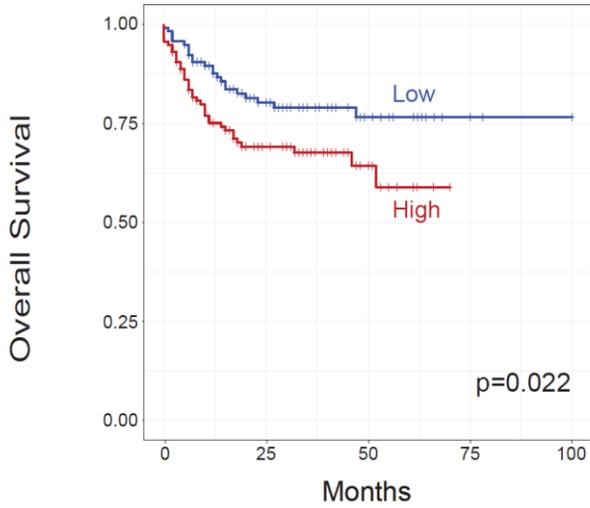


Groups

MBN Low	235	158	39	2
MBN High	234	135	52	2

Figure S6

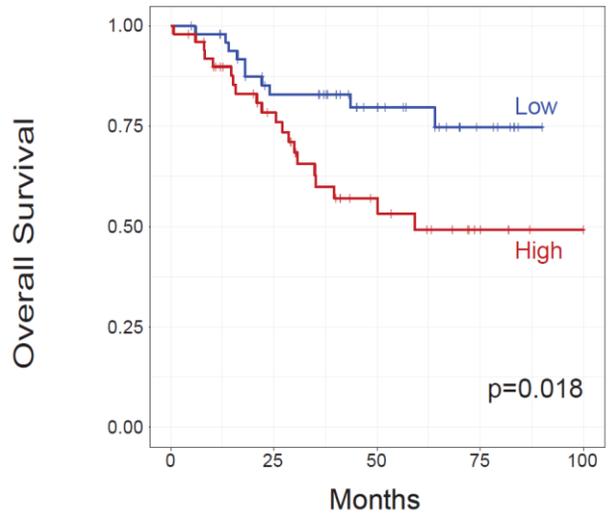
A



Groups

MBN Low	117	66	25	4	2
MBN High	116	54	14	0	0

B

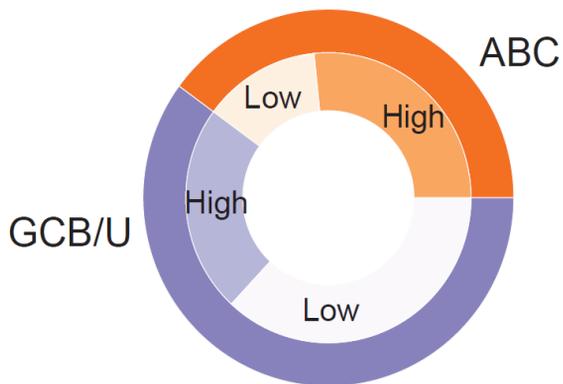


Groups

MBN Low	51	36	21	8	0
MBN High	51	32	15	6	2

C

LENZ COHORT n = 233



D

REAL- LIFE n = 102

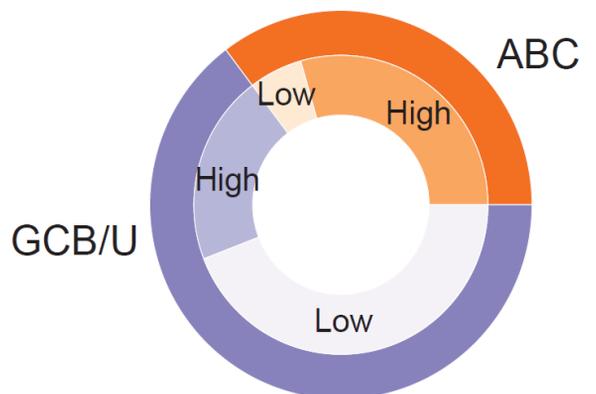
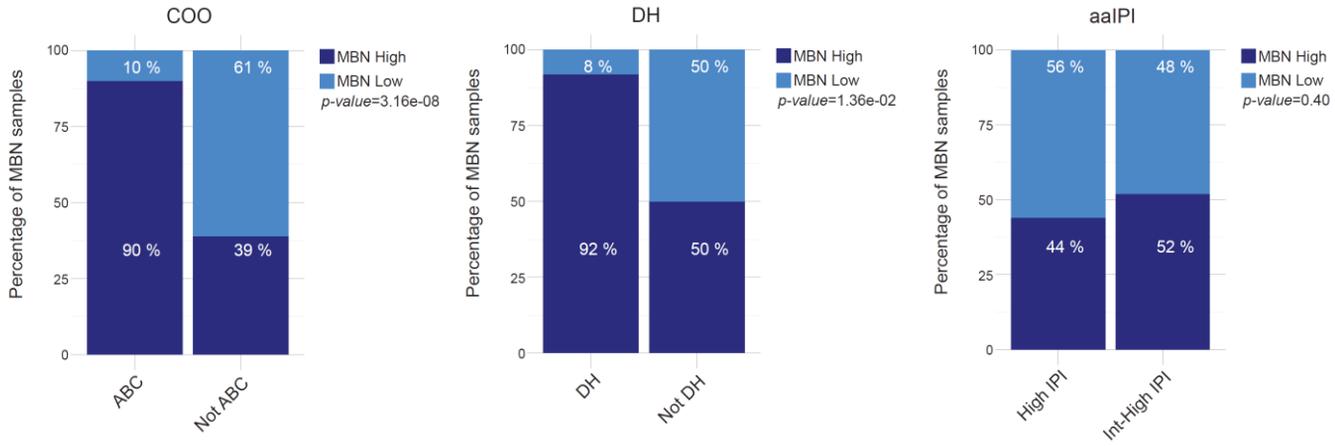


Figure S7

A



B

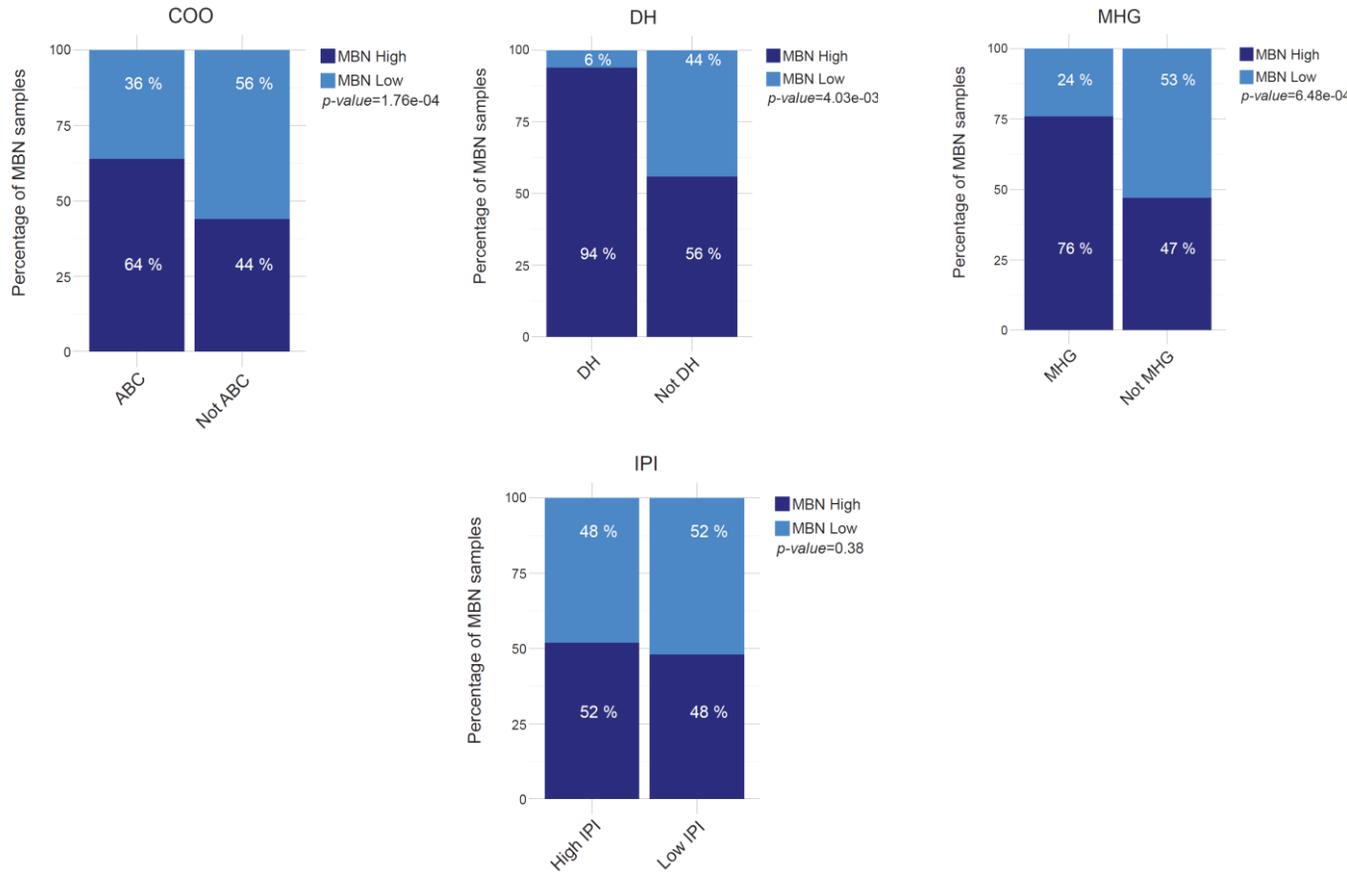
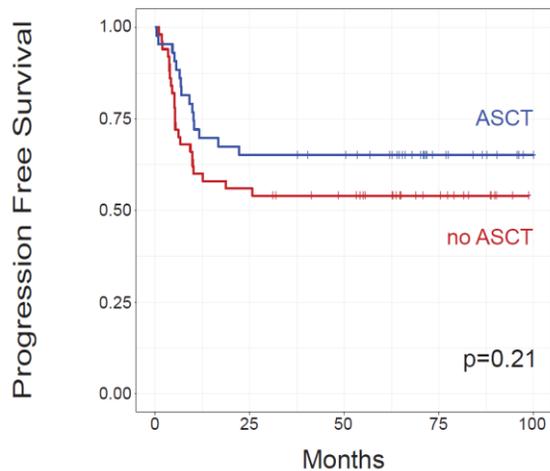


Figure S8

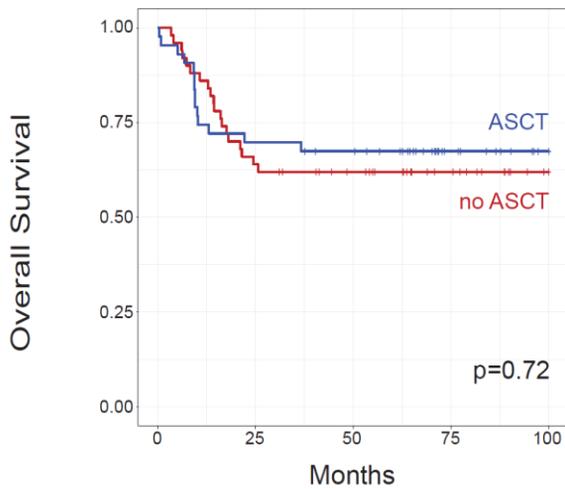
A



Groups

no ASCT	50	28	23	11	0
ASCT	43	28	26	10	1

B



Groups

no ASCT	50	32	25	13	2
ASCT	43	30	27	10	1