haematologica
Journal of The Ferrata Storti Foundation

# A 3-gene signature based on MYC, BCL-2 and NFKBIA improves risk stratification in diffuse large B-cell lymphoma

by Enrico Derenzini, Saveria Mazzara, Federica Melle, Giovanna Motta, Marco Fabbri, Riccardo Bruna, Claudio Agostinelli, Alessandra Cesano, Chiara Antonia Corsini, Ning Chen, Simona Righi, Elena Sabattini, Annalisa Chiappella, Angelica Calleri, Stefano Fiori, Valentina Tabanelli, Antonello Cabras, Giancarlo Pruneri, Umberto Vitolo, Alessandro Massimo Gianni, Alessandro Rambaldi, Paolo Corradini, Pier Luigi Zinzani, Corrado Tarella, and Stefano Pileri

Haematologica 2020 [Epub ahead of print]

*Citation: Enrico Derenzini, Saveria Mazzara, Federica Melle, Giovanna Motta, Marco Fabbri, Riccardo Bruna, Claudio Agostinelli, Alessandra Cesano, Chiara Antonia Corsini, Ning Chen, Simona Righi, Elena Sabattini, Annalisa Chiappella, Angelica Calleri, Stefano Fiori, Valentina Tabanelli, Antonello Cabras, Giancarlo Pruneri, Umberto Vitolo, Alessandro Massimo Gianni, Alessandro Rambaldi, Paolo Corradini, Pier Luigi Zinzani, Corrado Tarella, and Stefano Pileri. A 3-gene signature based on MYC, BCL-2 and NFKBIA improves risk stratification in diffuse large B-cell lymphoma.*
*Haematologica. 2020; 105:xxx*
*doi:10.3324/haematol.2019.236455*

**Title: A 3-gene signature based on *MYC, BCL-2* and *NFKBIA* improves risk stratification in diffuse large B-cell lymphoma.**

**Running head: A 3-gene signature in DLBCL**

**Authors: 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], Stefano Pileri[3].**

**Institutions:**

[1] Onco-Hematology Division, European Institute of Oncology IRCCS, Milan, Italy

[2] Department of Health Sciences, University of Milan, Milan, Italy

[3] Division of Diagnostic Haematopathology, European Institute of Oncology IRCCS, Milan, Italy

[4] Hematopathology Unit, Department of Experimental, Diagnostic, and Specialty Medicine (DIMES), Bologna University School of Medicine, Bologna, Italy

[5] NanoString Technologies Inc, Seattle, WA, USA

[6] Laboratory of Hematology-Oncology, European Institute of Oncology IRCCS, Milan, Italy

[7] Division of Hematology, Fondazione IRCCS Istituto Nazionale dei Tumori, University of Milan, Italy

[8] Department of Pathology, Fondazione IRCCS Istituto Nazionale dei Tumori di Milano, Milan, Italy

[9] Multidisciplinary Oncology Outpatient Clinic, Candiolo Cancer Institute, FPO-IRCCS, Candiolo, Italy

[10] Hematology and Bone marrow Transplant Unit, ASST-Papa Giovanni XXIII, Bergamo, Italy

[11] Institute of Hematology and Medical Oncology "L. e A. Seragnoli", Department of Experimental, Diagnostic, and Specialty Medicine (DIMES), Bologna University School of Medicine, Bologna, Italy

**Correspondence:**

**Dr. Enrico Derenzini, enrico.derenzini@ieo.it, Tel: +390257489944; Prof. Stefano Pileri, stefano.pileri@ieo.it, Tel: +390257489521**

**European Institute of Oncology IRCCS, Via Ripamonti 435, 20141, Milan, Italy.**

**Word count**

Abstract: 249

Main text: 3571

N° tables 2; N° figures 5

Supplemental files: 1

**Trial registration**

The R-HDS0305 and DLCL04 trials are registered at http://www.clinicaltrials.gov as NCT00355199 and NCT00499018 respectively.

**Aknowledgements**

The authors wish to thank Pier Luigi Antoniotti, Sebastiano Spagnolo, Marco Giuffrida and Virginia Maltoni for technical assistance. This study was funded by the AIRC 5x1000 grant to SP (n. 21198) and Italian Ministry of Health with Ricerca Corrente.

**ABSTRACT**

Recent randomized trials focused on gene expression-based determination of the cell of origin in diffuse large B-cell lymphoma could not show significant improvements by adding novel agents to standard chemoimmunotherapy. The aim of this study was the identification of a gene signature able to refine current prognostication algorithms and applicable to clinical practice. Here we used a targeted gene expression profiling panel combining the Lymph2Cx signature for cell of origin classification with additional targets including *MYC*, *BCL-2* and *NFKBIA*, in 186 patients from 2 randomized trials (discovery cohort) (NCT00355199 and NCT00499018). Data were validated in 3 independent series (2 large public datasets and a real-life cohort). By integrating the cell of origin, *MYC/BCL-2* double expressor status and *NFKBIA* expression, we defined a 3-gene signature combining *MYC*, *BCL-2* and *NFKBIA* (MBN-signature), which outperformed the *MYC/BCL-2* double expressor status in multivariate analysis, and allowed further risk stratification within the germinal center B-cell/unclassified subset. The high-risk (MBN Sig-high) subgroup identified the vast majority of double hit cases and a significant fraction of Activated B-Cell-derived diffuse large B-cell lymphomas. These results were validated in 3 independent series including a cohort from the REMoDL-B trial, where, in an exploratory ad hoc analysis, the addition of bortezomib in the MBN Sig-high subgroup provided a progression free survival advantage compared with standard chemoimmunotherapy. These data indicate that a simple 3-gene signature based on *MYC*, *BCL-2* and *NFKBIA* could refine the prognostic stratification in diffuse large B-cell lymphoma, and might be the basis for future precision-therapy approaches.

## INTRODUCTION

The biologic complexity of diffuse large B-cell lymphoma (DLBCL) was first dissected in early 2000s by gene expression profiling (GEP) studies, which subdivided DLBCL into two groups based on GEP signatures reminiscent of the respective cell of origin (COO). These studies showed that DLBCLs with a gene signature related to activated B-lymphocytes (ABC subgroup) had a significantly worse response to anthracycline-based therapies compared to those histogenetically related to germinal center B-cells (GCB subtype), and were addicted to nuclear factor kappa-B (NF-kB) signaling[1-3]. Since immunohistochemical algorithms failed to reproduce the results of GEP[4-10], the Lymphoma Leukemia Molecular Profiling Project (LLMPP) proposed a targeted GEP (T-GEP) panel (Lymph2Cx) desumed from previous studies on fresh/frozen tissue (FFT) [11,12]. This assay was applied on the NanoString platform to formalin-fixed, paraffin embedded (FFPE) tissue from DLBCLs patients treated with R-CHOP[11,12], identifying three subgroups: GCB, ABC and unclassified, the latter representing about 15% of all cases and prognostically closer to the GCB[11,13]. The reproducibility of this assay was confirmed in several studies[13-15]. However, recent results from 3 independent phase 3 randomized trials[16-18] based on COO classification were largely negative. Although these unsatisfactory results could be explained by several reasons, including unexpected toxicities and suboptimal efficacy of these drugs *in vivo*, these data also indicate that the clinical development of predictive T-GEP signatures able to complement the COO for precision therapy approaches is an urgent unmet need. Besides the COO, current evidence indicates a negative prognostic value of double MYC and BCL-2 protein overexpression determined by immunohistochemistry (IHC)[19-21]. Furthermore those DLBCLs with concurrent *MYC* and *BCL-2* and/or *BCL-6* genomic rearrangements are characterized by an even worse prognosis, being now classified as a separate entity, high-grade B-cell lymphoma (HG-BCL) with double/triple hits (w DH/TH)[19,20,22]. Recently large genomic studies integrating DNA and RNA sequencing data identified additional DLBCL subgroups beyond the COO and MYC/BCL-2 double expressor (DE) status[23-25], based on the mutational landscape, GEP signatures, copy number changes, and

differences in outcome. Furthermore, recent studies identified GEP signatures able to define high-risk populations within the GCB/unclassified (GCB/U) subgroup[26,27]. However, given their complexity, large-scale application of these prognostication algorithms could be difficult in daily clinical practice. The aim of this study was the implementation of a simple T-GEP panel able to complement and improve COO-based prognostic stratification for routine clinical application. We designed a panel of genes corresponding to those of the Lymph2Cx assay for COO determination plus additional candidates selected because of their potential prognostic and/or therapeutic interest including *MYC*, *BCL-2* and central nodes of NF-kB, Janus kinase (JAK)/signal transducer and activator of transcription (STAT), and phosphatidylinositol-3 kinase (PI3K) signaling[3,28-33]. This panel of genes was applied to 186 DLBCLs enrolled in two recently reported large italian trials (DLCL04 and R-HDS0305)[34,35]. We found that a 3-gene signature based on *MYC*, *BCL-2* and *NFKBIA* (MBN signature), identified a significant fraction of ABC cases and a subgroup of GCB/U cases (roughly 30%) enriched in HG-BCL w/DH, at increased risk of treatment failure. These data were validated in a real-life cohort and *in silico* in 2 large independent series, including one cohort of patients enrolled in the REMoDL-B trial[18,27], where the addition of bortezomib to chemoimmunotherapy provided a significant advantage for high-risk patients identified by the MBN signature.

## METHODS

### Study design

Patients considered in this study had been enrolled in 2 prospective randomized phase 3 clinical trials investigating the role of first line autologous stem cell transplant (ASCT) consolidation in intermediate/high-risk DLBCL[34,35]. Only cases of DLBCL not-otherwise specified (NOS) (including those originally diagnosed as DLBCL and nowadays included in the HG-BCL

provisional category[22]) were selected for the present study (Figure 1). Patients' characteristics and study algorithm are summarized in Table 1 and Figure 1.

Results were validated in 3 independent cohorts, (two *in silico* validation datasets and one "real-life" cohort): a dataset from Sha and coworkers [n=928 patients: 469 treated with R-CHOP and 459 with R-CHOP plus bortezomib (RB-CHOP)][27]; a public dataset from Lenz et al.[36] (n=233 patients treated with R-CHOP); a "real-life" cohort including 102 consecutive DLBCL-NOS cases with available FFPE tissue, treated with R-CHOP/R-CHOP-like regimens in Bologna (S.Orsola-Malpighi Hospital), and in Milan (European Institute of Oncology) from 2007 to 2018.

This study was approved by the Institutional Review Boards and Ethics Committees of the participating centers, in accordance with the Declaration of Helsinki.

**Procedures**

Gene expression was measured on the NanoString nCounter Analysis System (NanoString Technologies, Seattle, USA). The T-GEP panel contains 26 genes: 15 genes for COO subtyping[11]; 5 housekeeping genes (*UBXN4,ISY1,R3HDM1,WDR55,TRIM56*); and 6 additional genes (*MYC,BCL-2,STAT3,NFKBIA,PTEN,PIK3CA*). Besides *MYC* and *BCL-2*, the additional genes were selected based on their known functions in key pathways involved in DLBCL lymphomagenesis and potential druggability.

**Statistical analyses**

Survival data were analyzed retrospectively. We used Kaplan-Meier method[37] for overall survival (OS) and progression free survival (PFS) analyses. Multivariate and univariate analyses were constructed with the Cox proportional hazards regression model. A p value $\leq 0.05$ was considered as

statistically significant. The Recursive Partitioning Analysis (RPA)[38] was applied to classify patients into more homogenous prognostic groups based on survival. All analyses were performed using R 3.5.0 software[39]. Correlations and differences in patient characteristics were analyzed with the $\chi^2$ and Fisher's exact test.

**Development of the 3-gene prognostic signature (MBN signature)**

An expression ratio-based test was developed by selecting those genes significantly deregulated in the high risk subgroups identified by the RPA shown in Figure 2A and whose normalized mRNA levels were significantly associated with OS. We defined high and low *MYC* and *BCL-2* expressors based on the median normalized *MYC* and *BCL-2* mRNA levels. The high-risk groups included the ABCs and double expressor GCB/unclassified (GCB/U) DLBCL (hereafter defined as DEXP_mRNA); the low risk group was constituted by the non-DEXP_mRNA GCB/U subset. Since the expression levels of *MYC* and *BCL-2* on the one hand and *NFKBIA* on the other had opposing patterns being inversely associated with OS (with higher *MYC/BCL-2* and lower *NFKBIA* levels associated with worse outcome), we combined the expression levels of the three genes in a synthetic predictor called MBN-signature (MBN-Sig) and defined as:

$$MBN - Sig = \frac{MYC + BCL2}{NFKBIA}$$

Detailed information on study cohorts (Table S1), T-GEP procedures with list of genes and target sequences, Fluorescence in situ hybridization (FISH), Immunohistochemistry methods and antibodies (Table S2), and Random Forest classifier are described in supplement.

## RESULTS

### Univariate analyses and a decision-tree classification model integrating the COO and *MYC/BCL-2* status

Given their established clinical relevance, we first investigated the prognostic significance of T-GEP-based COO classification and *MYC/BCL-2* status in the R-HDS0305 and DLCL04 trials[34,35] (discovery cohort). Patient's characteristics are summarized in Table 1. In line with previous findings[11,12] COO classification by T-GEP clearly outperformed the immunohistochemical Hans algorhitm for survival prediction and retained its prognostic significance in the presence or absence of ASCT consolidation (Figure S1A-D). In order to investigate the prognostic impact of concurrent overexpression of *MYC* and *BCL-2*, we defined high and low expressors based on the median normalized *MYC* and *BCL-2* mRNA levels, which correlated well with the respective protein levels assessed by immunohistochemistry (Figure S2A). *MYC/BCL-2* mRNA double expressors (defined as DEXP_mRNA) patients showed a worse outcome compared to non-DEXP_mRNA cases (Figure S2B). Although DEXP_mRNA cases were more prevalent in the ABC compared to the GCB/unclassified (GCB/U) subgroup[9] (Table S3), the prognostic relevance of the *MYC/BCL-2* DEXP_mRNA status was particularly evident in the GCB/U subset (Figure S2C-F). Focusing the analysis on the additional genes (*STAT3, NFKBIA, PTEN, PIK3CA*), which were selected based on their biologic relevance in potentially druggable pathways, only *NFKBIA* and *STAT3* mRNA levels were significantly associated with patient's outcome, with low *STAT3* and low *NFKBIA* expression predicting worse prognosis (Figure S3A,B). In univariate analyses only the age adjusted International Prognostic Index (aaIPI) score (intermediate-high vs high), the COO classification, *MYC/BCL-2*-DE status, *NFKBIA* and *STAT3* levels determined by T-GEP, were significantly associated with OS (Table 2). As observed in the original studies[34,35], first-line ASCT consolidation was not associated with patient's outcome.

In line with the data presented above (Figure S1 and S2), a recursive partitioning analysis integrating the COO with *MYC/BCL-2* status identified 3 main patient subgroups: 2 high risk subsets with similar outcome [ABC (n=40) and *MYC/BCL-2* DEXP_mRNA GCB/U (n=27)], and a low-risk subgroup including non-DEXP_mRNA GCB/U DLBCLs, (n=119) (Figure 2A). Evaluating the relative expression of the additional genes included in the panel across the 3 groups identified by the recursive partitioning analysis (non-DEXP_mRNA GCB/U, DEXP_mRNA GCB/U and ABC DLBCL patients) (Figure 2B), we found that only *MYC*, *BCL-2* and *NFKBIA* were significantly deregulated in both the high-risk ABC and *MYC/BCL-2* DEXP_mRNA GCB/U subgroups, which were characterized by similarly increased *MYC* and *BCL-2* and lower *NFKBIA* mRNA levels compared to the low risk non-DEXP_mRNA GCB/U subset. The *NFKBIA* gene, a frequent target of deletions and mutations in DLBCL[23], encodes for the IkB-α protein, which is a central node of the NF-kB pathway and inhibits nuclear translocation and activity of the NF-kB transcription factors[40]. *STAT3* levels were similar in the high risk ABC and low risk non-DE GCB/U cases being significantly downregulated only in the DEXP_mRNA GCB/U subset. *PIK3CA* and *PTEN* levels did not vary significantly across different groups (Figure 2B).

**Development of a 3-gene prognostic signature combining *MYC*, *BCL-2* and *NFKBIA***

In an effort to build a GEP signature aimed at refining current prognostication algorithms and suitable for clinical practice, we considered only those genes whose expression was significantly associated with OS and differentially represented in both high risk (ABC and DEXP_mRNA GCB/U) vs. low risk (non-DEXP mRNA GCB/U) patient subsets. Using these criteria, we constructed a prognostic signature considering 3 genes (*MYC*, *BCL-2* and *NFKBIA*), which combines the *MYC/BCL-2* DEXP_mRNA status with *NFKBIA* expression (hereafter called MBN signature, see methods). Besides *MYC* and *BCL-2*, (defining the DEXP_mRNA status), *NFKBIA* emerged as the best survival predictor by gene ranking according to the predictive power (univariate

z score) (Figure S4). With this strategy, patients were divided in 2 risk categories characterized by different outcome: low risk patients (MBN-Sig low) had a very favorable prognosis [(91% 5-year OS; 84% 5-year progression free survival (PFS)], whereas high-risk patients (MBN-Sig high) had a significantly worse prognosis (64% 5-year OS; 59% 5-year PFS) (Figure 3A and S5A). Importantly the MBN signature retained its significance and outperformed the *MYC/BCL-2* DEXP_mRNA status in multivariate analysis (Figure 3B, Table S4). In fact, only the COO, the aaIPI score and the MBN signature were significantly associated with outcome in multivariate analyses (Figure 3B). These findings were confirmed *in silico* in a large independent validation cohort of 469 patients[27] treated with R-CHOP (88% 5-year OS and 78% PFS for MBN-Sig low vs 72% OS and 57% PFS for MBN-Sig high patients) (Figure 3C, D, Figure S5B, and Table S5). The prognostic value of the MBN signature was further tested in a publicly available data set including 233 patients (from Lenz at al, 2008)[36] treated with R-CHOP/R-CHOP-like regimens and in a real-life cohort (n=102 patients) with similar results. (Figure S6A and S6B). The MBN signature was able to identify a significant fraction of ABC-derived cases and about a third of GCB/U cases (Figure 3A, C and Figure S6C, D).

**Real life applicability of the MBN signature**

In order to provide a risk stratification tool applicable to routine clinical practice in a prospective manner, we constructed a random forest (RF) model with the expression of genes characterizing the MBN signature. First, the classifier was trained on the discovery cohort splitting it into training (80%) and test (20%) dataset; in this case, the accuracy of the three-gene model was 93% in the training and 94% in test set. In order to confirm the reliability of this three-gene model, we further tested it in an independent dataset (validation set) consisting of the real-life cohort (n=102 cases). Of note, these cases were profiled with the same T-GEP panel and methods used in the discovery cohort, mitigating batch effects phenomena. As result, the three-gene model accurately classified

85% (87 out of 102) cases as either MBN-Sig high or MBN-Sig low subgroups (Figure 4A). As reported in Figure 4B, the model effectively identified MBN-high and low categories with sensitivity (SE) and specificity (SP) of 94% and 76% respectively. Receiver operating characteristic (ROC) curve analysis revealed that the area under the curve (AUC) was 0.94 in the validation set (Figure 4C). Furthermore, this strategy produced a very efficient survival prediction, which as expected showed a worse outcome for the MBN-high subset (Figure 4D) and mirrored the OS curve based on the median MBN value depicted in Figure S6B.

**Correlation of the MBN signature with FISH status and clinical variables**

Focusing the analyses on our discovery cohort of 186 patients (DLCL04 and R-HDS0305 trials)[34,35], we observed that the MBN signature significantly stratified the prognosis GCB/U patients (Figure 5A). Since the MBN signature effectively stratified GCB/U DLBCL patients, we investigated correlations between the MBN-signature, FISH status and clinical variables in our discovery cohort. As shown in Figure 5B, we observed a significantly higher frequency of *MYC* and *BCL-2* rearrangements in the MBN-Sig high subgroup compared to the GCB/U MBN-Sig low subset. According to these observations, there was a significant enrichement of HG-BCL w/DH in the MBN-Sig high subgroup compared to the MBN-Sig low subset (Figure 5B and Figure S7A). No differences in the number of cases with missing FISH analyses were observed between groups (data not shown). In line with the literature[23,26,27], all these cases, except one, were GCB-derived (data not shown). As previously shown in Figure 3, ABC-derived DLBCL were significantly more represented in the MBN-Sig high subgroup (Figure 5B and Figure S7A). Finally, no significant differences in aaIPI score (intermediate high vs. high) were observed between groups (Figure 5B and Figure S7A). These findings were validated *in silico* in the larger cohort from Sha et al[27] (Figure 5C,D). As observed in the discovery cohort, the MBN signature stratified the prognosis of GCB/U patients (Figure 5C) and identified the vast majority of DH cases. Again ABC-derived

DLBCLs were more represented in the MBN-Sig high subgroup (Figure 5D and Figure S7B). In this study, the application of a gene expression classifier identified a molecular high grade (MHG) subgroup strongly enriched in DH lymphomas and comprising 9% of the total patient population[27]. To evaluate how our MBN signature performed in the same patient population, we compared the MBN signature with MHG signature and with FISH status (Figure 5C). Notably the MBN-high subgroup was significantly enriched in MHG cases, identifying 76% of MHG DLBCLs and the vast majority of DH (Figure 5D and Figure S7B). Also in this cohort there were no differences in IPI score between groups (Figure 5D and Figure S7B).

**Rationale for a precision therapy approach in MBN Sig-high DLBCL**

Since the MBN-Sig high subgroup is characterized by relatively higher *MYC* and *BCL-2* expression and lower *NFKBIA* levels indicative of constitutive NF-kB activity, we next investigated the effect of differential therapeutic strategies in this high-risk patient subset. We first analyzed the impact of ASCT vs. standard chemoimmunotherapy in the discovery cohort. ASCT consolidation did not provide any significant PFS or OS advantage compared to standard chemoimmunotherapy in the MBN-Sig high subgroup (Figure S8A, B). The aberrant activation of NF-kB observed in lymphoma is associated with decreased abundance of IkB-α (which is encoded by the *NFKBIA* gene)[41,42]. Since bortezomib is known to increase IkB-α levels by blocking its ubiquitination and therefore inhibiting NF-kB activity[43-45], we next interrogated the Sha dataset[18,27] performing an exploratory ad-hoc analysis to investigate the impact of the addition of bortezomib to standard R-CHOP (RB-CHOP) in the MBN-Sig high subset (characterized by decreased *NFKBIA* levels)[18,27]. Interestingly, RB-CHOP determined a significant PFS advantage in the MBN-Sig high population (p=0.012) (Figure 5E), which translated in an increased OS rate (p=0.052) (Figure 5F).

## DISCUSSION

In this study we applied a customized T-GEP panel (including the Lymph2Cx signature for COO classification and additional genes of potential prognostic and therapeutic interest) to 2 randomized trials[34,35] (n=186 patients) performed in the Rituximab era. The aims of this study were the integration of the COO with additional GEP-based variables, and the identification of a gene signature applicable to routine clinical practice, able to refine current prognostication algorithms. The genes of the T-GEP panel were selected considering the relevance of the respective signaling pathways in B-cell lymphomagenesis, but more importantly based on their potential druggability.

Our study confirmed the prognostic value of GEP-based COO determination, which clearly outperformed the IHC-based Hans algorithm (the ABC DLBCL subgroups having a significantly inferior OS in all case series evaluated here) (Figure S1). The COO retained its prognostic value in patients undergoing ASCT consolidation, suggesting that therapy intensification is not able to overcome the negative prognostic value of the COO. A recursive partitioning analysis integrating COO with *MYC/BCL-2* DEXP_mRNA status identified 3 main subgroups (a low risk non-DEXP_mRNA GCB/U subset and 2 high-risk groups including DEXP_mRNA GCB/U and ABC-DLBCLs) (Figure 2A). The observation of lower *NFKBIA* levels in the ABC and DEXP_mRNA GCB/U subgroups (overexpressing *MYC* and *BCL-2* to a similar extent) (Figure 2B) suggests that, despite known biologic differences, these DLBCL subsets could share similar oncogenic dependencies on *MYC*, *BCL-2* and the NF-kB pathway (being *NFKBIA* a negative regulator of NF-kB signaling). This observations prompted us to design a 3-gene prognostic signature integrating *MYC*, *BCL-2* and *NFKBIA*, which we called the MBN signature. The signature was first tested in our discovery cohort of 186 patients, identifying 2 subgroups characterized by different outcome (Figure 3), and was then applied to 3 independent datasets (469 patients treated with R-CHOP in the Sha cohort[27], 233 patients from the Lenz cohort[36], and 102 patients treated in real-life clinical practice with R-CHOP/R-CHOP-like regimens) confirming its high prognostic significance (total

number of tested cases 990). Since the discovery cohort had some unique characteristics (such as lack of low aa-IPI cases, a relatively low fraction of ABC cases and no uniform first-line treatment), the extensive validation performed in 3 additional cohorts treated with R-CHOP/R-CHOP-like regimens confirms that the key findings of the present study are indeed applicable to an unselected DLBCL population. Importantly, the MBN signature defined a high-risk group including a significant fraction of ABC cases (in line with data shown in table S3 demonstrating a higher incidence of *MYC/BCL-2* DEXP_mRNA and low *NFKBIA* expressors in the ABC subgroup), and about 30% of GCB/U cases (Figure 3). Therefore the MBN-signature could potentially identify an increased proportion of patients at high risk of treatment failure, compared to standard risk stratifications (COO or DE status). The MBN-signature was an independent prognostic predictor, outperforming the *MYC/BCL-2*-DEXP_mRNA status in multivariate analyses (Figure 3), thus confirming the added value of the third gene (*NFKBIA*) for prognostic stratification. The possible clinical applicability of the MBN signature was tested in the real-life cohort using a random forest prediction model built on the discovery cohort, providing a reliable tool for prospective risk stratification (Figure 4). Importantly, the integration of the MBN signature with the COO allowed the identification of 2 risk categories whithin the GCB/U subset. These findings, which were validated in independent cohorts, could have immediate implications (Figure 5A,C and S6A-D). Two recently published studies confirmed the heterogeneity of the GCB subgroup and identified gene signatures allowing better risk stratification of this patient subset[26,27]. These signatures were able to identify a proportion of HG-BCLs with DH/TH and a further group lacking *MYC/BCL-2* rearrangements but characterized by similar clinical features. However, the fact that these signatures are composed by several genes encompassing multiple pathways, could make their successful translation to clinical practice and precision therapy approaches quite challenging.

Our data are in line with these findings confirming that the GCB/U DLBCL subset represents indeed a rather heterogeneous disease category. The MBN signature could identify the majority of tumors with high-grade molecular features (HG-BCLs with DH/TH) in the discovery cohort and

Sha's cohort (Figure 5B,D, and Figure S7A,B). Moreover, by applying the MBN signature to the Sha validation cohort we observed that the MBN-Sig high subroup was significantly enriched in MHG DLBCL cases (Figure 5D and Figure S7B). Taken together, these data indicate that a simple 3-gene signature could efficiently identify high risk GCB/U DLBCL cases. Furthermore, the MBN-signature is based on potentially druggable targets or pathways. For example, *NFKBIA* (encoding for IkB- α) could be targeted by proteasome inhibitors[43-45] and by bromodomain and extraterminal protein (BET) inhibitors, which are able to downregulate MYC while increasing IkB-α levels[46-48]. Our analysis on the impact of bortezomib in the MBN-high subgroup of the Sha cohort[27] (from the REMoDL-B trial) seems to confirm a potential druggability of the MBN signature: in fact treatment with RB-CHOP (R-CHOP plus bortezomib) was associated with a significantly prolonged PFS which translated in increased OS rates in the MBN-high subgroup, as compared to standard R-CHOP (Figure 5E,F). Proteasome inhibitors, BET inhibitors and selective BCL-2 inhibitors could be the basis for rationally-designed combinations for the MBN-Sig high DLBCL subgroup. Alternative strategies to target NF-kB include lenalidomide and B-cell receptor signaling inhibitors, all of which are under clinical investigation in DLBCL. Three COO-based phase 3 trials testing R-CHOP + Ibrutinib (Phoenix trial[16]) or Lenalidomide (ROBUST trial[17]) or bortezomib (REMoDL-B trial[18]) did not meet their primary endpoints. Although several factors concurred to these negative results, the development of alternative and druggable molecular signatures represents an unmet need and could be of primary importance for the design of future precision medicine clinical trials.

The results of our study indicate that a simple and cost-effective 3-gene assay (MBN-signature) could refine current prognostic stratification algorithms providing the rationale for the implementation of precision medicine trials in the MBN-Sig high subset.

**REFERENCES**


1- Alizadeh AA, Eisen MB, Davis RE, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature. 2000;403(6769):503-511.

2- Shipp MA, Ross KN, Tamayo P, et al. Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. Nat Med. 2002;8(1):68-74.

3- Davis RE, Brown KD, Siebenlist U, Staudt LM. Constitutive nuclear factor kappaB activity is required for survival of activated B cell-like diffuse large B cell lymphoma cells. J Exp Med. 2001;194(12):1861-1874.

4- Hans CP, Weisenburger DD, Greiner TC, et al. Confirmation of the molecular classification of diffuse large B-cell lymphoma by immunohistochemistry using a tissue microarray. Blood. 2004;103(1):275-282.

5- Meyer PN, Fu K, Greiner TC, et al. Immunohistochemical Methods for Predicting Cell of Origin and Survival in Patients With Diffuse Large B-Cell Lymphoma Treated With Rituximab. J Clin Oncol. 2011;29(2):200-207.

6- Lawrie CH, Ballabio E, Soilleux E, et al. Inter- and intra-observational variability in immunohistochemistry: a multicentre analysis of diffuse large B-cell lymphoma staining. Histopathology. 2012;61(1):18-25.

7- Castillo JJ, Beltran BE, Song MK, et al. The Hans algorithm is not prognostic in patients with diffuse large B-cell lymphoma treated with R-CHOP. Leuk Res. 2012;36(4):413-417.

8- Coutinho R, Clear AJ, Owen A, et al. Poor concordance among nine immunohistochemistry classifiers of cell-of-origin for diffuse large B-cell lymphoma: implications for therapeutic strategies. Clin Cancer Res. 2013;19(24):6686-6695.

9- Hu S, Xu-Monette ZY, Tzankov A, et al. MYC/BCL2 protein coexpression contributes to the inferior survival of activated B-cell subtype of diffuse large B-cell lymphoma and demonstrates high-risk gene expression signatures: a report from The International DLBCL Rituximab-CHOP Consortium Program. Blood. 2013;121(20):4021-4031.

10- Reinke S, Richter J, Fend F, et al. Round-robin test for the cell-of-origin classification of diffuse large B-cell lymphoma-a feasibility study using full slide staining. Virchows Arch. 2018;473(3):341-349.

11- Scott DW, Wright GW, Williams PM, 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):1214-1217.

12- Scott DW, Mottok A, Ennishi D, et al. Prognostic Significance of Diffuse Large B-Cell Lymphoma Cell of Origin Determined by Digital Gene Expression in Formalin-Fixed Paraffin-Embedded Tissue Biopsies. J Clin Oncol. 2015;33(26):2848-2856.

13- Painter D, Barrans S, Lacy S, et al. Cell-of-origin in diffuse large B-cell lymphomafindings from the UK's population-based Haematological Malignancy Research Network. Br J Haematol. 2019;185(4):752–806.

14- Veldman-Jones MH, Lai Z, Wappett M, et al. Reproducible, Quantitative, and Flexible Molecular Subtyping of Clinical DLBCL Samples Using the NanoString nCounter System. Clin Cancer Res. 2015;21(10):2367-2378.

15- Rimsza LM, Wright G, Schwartz M, et al. Accurate classification of diffuse large B-cell lymphoma into germinal center and activated B-cell subtypes using a nuclease protection assay on formalin-fixed, paraffin-embedded tissues. Clin Cancer Res. 2011;17(11):3727-3732.

16- Younes A, Sehn LH, Johnson P, et al. Randomized Phase III Trial of Ibrutinib and Rituximab Plus Cyclophosphamide, Doxorubicin, Vincristine, and Prednisone in Non-Germinal Center B-Cell Diffuse Large B-Cell Lymphoma. J Clin Oncol. 2019;37(15):1285-1295.

17- Vitolo U, Witzig T, Gascoyne R, et al. ROBUST: First report of phase III randomized study of lenalidomide/R-CHOP ($R^2$-CHOP) vs placebo/R-CHOP in previously untreated ABC-type diffuse large B-cell lymphoma. Hematol Oncol. 2019;37(S2):36-37.

18- Davies A, Cummin TE, Barrans S, et al. Gene-expression profiling of bortezomib added to standard chemoimmunotherapy for diffuse large B-cell lymphoma (REMoDL-B): an open-label, randomised, phase 3 trial. Lancet Oncol. 2019;20(5):649-662.

19- Green TM, Young KH, Visco C, et al. Immunohistochemical double-hit score is a strong predictor of outcome in patients with diffuse large B-cell lymphoma treated with rituximab plus cyclophosphamide, doxorubicin, vincristine, and prednisone. J Clin Oncol. 2012;30(28):3460-3467.

20- Johnson NA, Slack GW, Savage KJ, et al. Concurrent expression of MYC and BCL2 in diffuse large B-cell lymphoma treated with rituximab plus cyclophosphamide, doxorubicin, vincristine, and prednisone. J Clin Oncol. 2012;30(28):3452-3459.

21- Staiger AM, Ziepert M, Horn H, et al. Clinical Impact of the Cell-of-Origin Classification and the MYC/ BCL2 Dual Expresser Status in Diffuse Large B-Cell Lymphoma Treated Within Prospective Clinical Trials of the German High-Grade Non-Hodgkin's Lymphoma Study Group. J Clin Oncol. 2017;35(22):2515-2526.

22- Swerdlow SH, Campo E, Harris NL, et al. WHO Classification of Tumour of Haematopoietic and Lymphoid Tissues, Revised 4$^{th}$ edition. 2017. IARC Press, Lyon.

23- Reddy A, Zhang J, Davis NS, et al. Genetic and Functional Drivers of Diffuse Large B Cell Lymphoma. Cell. 2017;171(2):481-494.

24- Schmitz R, Wright GW, Huang DW, et al. Genetics and Pathogenesis of Diffuse Large B-Cell Lymphoma. N Engl J Med. 2018;378(15):1396-1407.

25- Chapuy B, Stewart C, Dunford AJ, et al. Molecular subtypes of diffuse large B cell lymphoma are associated with distinct pathogenic mechanisms and outcomes. Nat Med. 2018;24(8):679-690.

26- 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.

27- Sha C, Barrans S, Cucco F, et al. Molecular High-Grade B-Cell Lymphoma: Defining a Poor-Risk Group That Requires Different Approaches to Therapy. J Clin Oncol. 2019;37(3):202-212.

28- Pasqualucci L, Dalla-Favera R. Genetics of diffuse large B-cell lymphoma. Blood. 2018;131(21):2307-2319.

29- Roschewski M, Staudt LM, Wilson WH. Diffuse large B-cell lymphoma-treatment approaches in the molecular era. Nat Rev Clin Oncol. 2014;11(1):12-23.

30- Paul J, Soujon M, Wengner AM, et al. Simultaneous Inhibition of PI3Kδ and PI3Kα Induces ABC-DLBCL Regression by Blocking BCR-Dependent and -Independent Activation of NF-κB and AKT. Cancer Cell. 2017;31(1):64-78.

31- Pfeifer M, Grau M, Lenze D, et al. PTEN loss defines a PI3K/AKT pathway-dependent germinal center subtype of diffuse large B-cell lymphoma. Proc Natl Acad Sci U S A. 2013;110(30):12420-12425.

32- Compagno M, Lim WK, Grunn A, et al. Mutations of multiple genes cause deregulation of NF-kappaB in diffuse large B-cell lymphoma. Nature. 2009;459(7247):717-721.

33- Pan YR, Chen CC, Chan YT, et al. STAT3-coordinated migration facilitates the dissemination of diffuse large B-cell lymphomas. Nat Commun. 2018;9(1):3696.

34- Cortelazzo S, Tarella C, Gianni AM, 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.

35- Chiappella A, Martelli M, Angelucci E, 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.

36- Lenz G, Wright G, Dave SS, et al. Stromal gene signatures in large-B-cell lymphomas. N Engl J Med 2008;359(22):2313-2323.

37- Kaplan EL, Meier P. Nonparametric estimations from incomplete observations. J Am Stat Assoc.1958;53(282):457-481.

38- Curran WJ Jr, Scott CB, Horton J, et al. Recursive partitioning analysis of prognostic factors in three Radiation Therapy Oncology Group malignant glioma trials. J Natl Cancer Inst. 1993;85(9):704-710.

39- R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2014. URL http://www.R-project.org/.

40- Oeckinghaus A, Ghosh S. The NF-κB Family of Transcription Factors and Its Regulation. Cold Spring Harbor Perspectives in Biology. 2009;1(4):a000034.

41- Jost PJ, Ruland J. Aberrant NF-kappaB signaling in lymphoma: mechanisms, consequences, and therapeutic implications. Blood. 2007;109(7):2700-2707.

42- Packham G. The role of NF-kappaB in lymphoid malignancies. Br J Haematol. 2008;143(1):3-15.

43- McConkey DJ, Zhu K. Mechanisms of proteasome inhibitor action and resistance in cancer. Drug Resist Updat. 2008;11(4-5):164-179.

44- Mujtaba T, Dou QP. Advances in the understanding of mechanisms and therapeutic use of bortezomib. Discov Med. 2011;12(67):471-480.

45- Bu R, Hussain AR, Al-Obaisi KA, Ahmed M, Uddin S, Al-Kuraya KS. Bortezomib inhibits proteasomal degradation of IκBα and induces mitochondrial dependent apoptosis in activated B-cell diffuse large B-cell lymphoma. Leuk Lymphoma. 2014;55(2):415-424.

46- Delmore JE, Issa GC, Lemieux ME, et al. BET bromodomain inhibition as a therapeutic strategy to target c-Myc. Cell. 2011;146(6): 904-917.

47- Derenzini E, Mondello P, Erazo T, et al. BET Inhibition-Induced GSK3β Feedback Enhances Lymphoma Vulnerability to PI3K Inhibitors. Cell Rep. 2018;24(8):2155-2166.

48- Ceribelli M, Kelly PN, Shaffer AL, et al. Blockade of oncogenic IκB kinase activity in diffuse large B-cell lymphoma by bromodomain and extraterminal domain protein inhibitors. Proc Natl Acad Sci U S A. 2014;111(31):11365-11370.

**TABLES**

**Table 1. Patients characteristics.**

| **Variable** | **Clinical trials** | | | |
|---|---|---|---|---|
| **Trial name** | **RHDS0305** | **DLCL04** | **P-value*** | **RHDS0305 + DLCL04** |
| **N° of patients** | 87 | 99 | - | 186 |
| **Immuno-CHT alone** | 49 (56%) | 56 (57%) | - | 105 (56%) |
| **Immuno-CHT + ASCT** | 38 (44%) | 43 (43%) | - | 81 (44%) |
| **Median age, y (range)** | 53 (21-65) | 52(18-65) | - | 52 (18-65) |
| **COO NanoString** | | | | |
| ABC | 15 (17%) | 25 (25%) | ns | 40 (21%) |
| GCB | 53 (61%) | 58 (59%) | ns | 111 (60%) |
| Unclassified | 19 (22%) | 16 (16%) | | 35 (19%) |
| **COO Hans IHC** | | | | |
| Non-GCB | 58 (67%) | 60 (61%) | | 118 (63%) |
| GCB-like | 29 (33%) | 39 (39%) | ns | 68 (37%) |
| **Stage (Ann Arbor)** | II-IV | III-IV | - | II-IV |
| **aaIPI score** | | | | |
| Low-Low Intermediate (0-1) | - | - | - | - |
| Intermediate-high (2) | 55 (63%) | 81 (82%) | ns | 136 (73%) |
| High (3) | 32 (37%) | 18 (18%) | 0.005 | 50 (27%) |

*Two-sided Fisher's exact test

Abbreviations: N° (number), Immuno-CHT (immunochemotherapy), ASCT (autologous stem cell transplantation), y (years), COO (cell of origin), IHC (immunohistochemistry), aaIPI (age adjusted international prognostic index), ns (not significant).

**Table 2. Univariate analysis for OS.**

| Variable | Hazard Ratio | 95% CI | p-value |
|---|---|---|---|
| **aaIPI** | | | |
| Intermediate-High | Ref | | |
| High | 2.04 | 1.10 - 3.78 | 0.023 |
| **COO Nano** | | | |
| GCB | Ref | | |
| ABC | 3.39 | 1.76 - 6.52 | <0.001 |
| Unclassified | 1.17 | 0.47 - 2.96 | 0.736 |
| **ASCT** | | | |
| No | Ref | | |
| Yes | 0.88 | 0.47 – 1.63 | 0.68 |
| ***MYC-BCL-2* DEXPmRNA** | | | |
| No | Ref | | |
| Yes | 2.32 | 1.26 – 4.29 | 0.007 |
| ***STAT3*** | | | |
| Low | Ref | | |
| High | 0.37 | 0.19 – 0.7 | 0.004 |
| ***NFKBIA*** | | | |
| Low | Ref | | |
| High | 0.34 | 0.17 – 0.68 | 0.002 |

Abbreviations: Signif (significance), ref (reference), IPI (international prognostic index), COO_Nano (COO defined by NanoString), ASCT (autologous stem cell transplant),

MYC/BCL-2 DEXP_mRNA (double expressor status defined by NanoString). * <.05; ** < .01.

## FIGURE LEGENDS

**Figure 1.** Study algorithm.

On the left, the discovery cohort is represented. 224 DLBCL patients enrolled in the DLCL04 (n= 130) and R-HDS0305 (n= 94) trials with available FFPE tissue were initially considered in this analysis. T-GEP success rate was 92.4% (n=207), with 17 cases not yelding enough high-quality mRNA to undergo successful GEP assessment. Only cases originally diagnosed as DLBCL non-otherwise specified (NOS) were considered. Therefore 21 cases classified in different DLBCL categories were excluded. 99 NOS-DLBCL FFPE patient samples from the DLCL04 trial and 87 samples from the R-HDS0305 trial were finally included in this study. On the right, the three validation cohorts: a cohort of 928 patients from Sha and coworkers[27] (469 treated with R-CHOP; 459 with RB-CHOP), a public gene expression dataset (Affymetrix Human Genome U133 Plus 2.0 Array), GSE10846, (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE10846), including 233 patients treated with R-CHOP regimen (Lenz et al 2008)[36]; an additional validation cohort including 102 consecutive DLBCL NOS cases with available FFPE tissue, treated with R-CHOP/R-CHOP-like regimens.

**Figure 2.** Integrating COO with *MYC/BCL-2* DEXP_mRNA status for prognostication in DLBCL.

A) Recursive partitioning analysis integrating COO classification and DEXP_mRNA status, allowing segregation of patients in 3 main prognostic subgroups (a low risk non-DEXP-mRNA GCB/U subset, and 2 high risk groups: *MYC/BCL-2*-DEXP-mRNA GCB/U and ABC).

B) Box plot graphs indicating the expression levels of the additional targets included in the panel (*MYC, BCL-2, NFKBIA, STAT3, PIK3CA, PTEN)* in the 3 main patients subgroups identified by the recursive partitioning analysis (non-DEXP-mRNA GCB/U, *MYC/BCL-2*

DEXP_mRNA GCB/U, ABC-derived DLBCL). P value was calculated with Student T test by comparing non DEXP_mRNA GCB/U group, selected as a reference, vs other groups.

**Figure 3**. Survival curves according to MBN-signature (MBN-Sig) and multivariate analyses for OS.

A) OS of the discovery cohort (R-HDS0305+DLCL04; n=186 patients) according to the MBN-signature showing significant differences in outcome between MBN-Sig low vs. MBN-Sig high patient subsets. P values were calculated with the log rank test. Frequencies of MBN-Sig high vs. low cases in ABC and GCB/U subsets in the discovery cohort are represented in the pie chart.

B) Forest plot depicting multivariate analyses for OS (discovery cohort). Only factors significantly asociated with OS in univariate analyses were considered. According to this analysis only the COO as determined by NanoString-based T-GEP (COO_Nano), the MBN-signature and the aaIPI (age adjusted international prognostic index) score retained statistical significance for OS, whereas *MYC/BCL-2* DEXP-mRNA status, *STAT3* and *NFKBIA* levels determined by T-GEP were not significantly associated with OS. HR (hazard ratio).

C) OS of the 469 patients treated with R-CHOP in the Sha's cohort according to the MBN-signature showing significant differences in outcome between MBN-Sig low vs MBN-Sig high patient subsets. P values were calculated with the log rank test. Frequencies of MBN-Sig high vs. low cases in ABC and GCB/U subsets in the Sha's cohort are represented in the pie chart.

D) Forest plot depicting multivariate analyses for OS (Sha's dataset), confirming the significant independent association with OS of the MBN signature in this large validation cohort.

**Figure 4.** Real-life applicability of the MBN signature

A) Heatmap representing the 3 informative genes of the MBN-signature (MBN-Sig) shown as rows and DLBCL tissue samples shown as columns in the real-life cohort of 102 patients, with the actual MBN signature and the predicted MBN signature class based on the application of a random forest (RF) model built on the discovery cohort on the top of the heatmap.
B) Violin plot showing the fractions of false predictions (false positive, FP, and false negative, FN) as well as true predictions (true positive, TP, and true negative, TN) in the real-life cohort by applying a 3-gene random forest model.
C) ROC curve of the real-life cohort using RF classifier
D) OS curve of the real-life cohort (n=102) based on the predicted MBN signature class. P value was calculated with the log rank test.

**Figure 5.** The MBN signature (MBN-Sig) identifies prognostically distinct subgroups including ABC and a fraction of GCB/U DLBCL enriched in DH DLBCL cases, providing opportunities for precision therapies.

A) OS of the GCB/U subset in the discovery cohort (n=146 patients) according to the integration of the Lymph2Cx with the MBN-signature, distinguishing 2 risk categories according to the MBN signature signature. P value was calculated with the log rank test.
B) Heatmap representing the 3 informative genes of the MBN-signature (MBN-Sig) shown as rows and DLBCL tissue samples shown as columns, in the discovery cohort (n=186 patients). NA (not available), DH (double hit), DE [double expressor (based on DEXP_mRNA status)], aaIPI (age adjusted international prognostic index), COO (cell of origin).

C) OS of the GCB/U subset in the Sha's validation cohort (n=340 patients) according to the MBN-signature, showing superimposable results compared to figure 4C. P value was calculated with the log rank test.

D) Heatmap representing the 3 informative genes of the MBN-signature (MBN-Sig) shown as rows and DLBCL tissue samples shown as columns, in the Sha's cohort (n=469 patients treated with R-CHOP). NA (not available), DH (double hit), DE [double expressor (based on DEXP_mRNA status)], IPI (international prognostic index), COO (cell of origin), MHG (molecular high grade).

E) PFS of patients treated with R-CHOP vs RB-CHOP in the MBN-Sig high subgroup (Full Sha's cohort, n=928 patients; MBN-Sig high n=464 patients). P value was calculated with the log rank test.

F) OS of patients treated with R-CHOP vs RB-CHOP in the MBN-Sig high subgroup (Full Sha's cohort, n=928 patients; MBN-Sig high n=464 patients). P value was calculated with the log rank test.

Figure 1

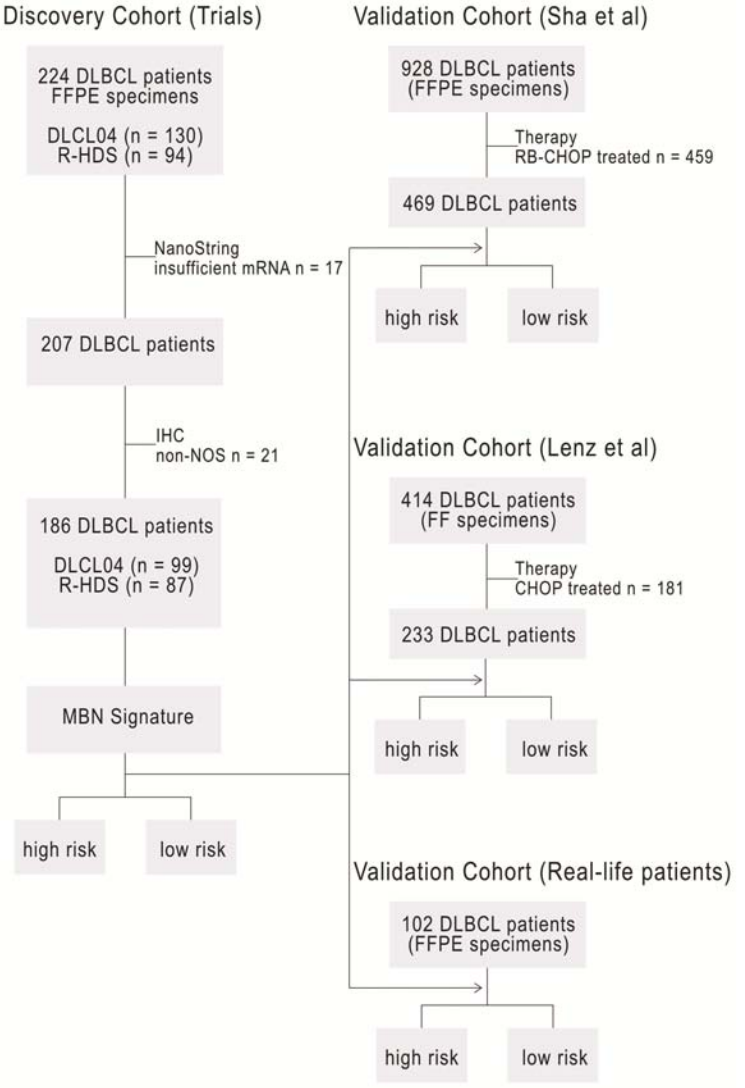

Discovery Cohort (Trials)
224 DLBCL patients
FFPE specimens
DLCL04 (n = 130)
R-HDS (n = 94)
NanoString
insufficient mRNA n = 17
207 DLBCL patients
IHC
non-NOS n = 21
186 DLBCL patients
DLCL04 (n = 99)
R-HDS (n = 87)
MBN Signature
high risk
low risk
Validation Cohort (Sha et al)
928 DLBCL patients
(FFPE specimens)
Therapy
RB-CHOP treated n = 459
469 DLBCL patients
high risk
low risk
Validation Cohort (Lenz et al)
414 DLBCL patients
(FF specimens)
Therapy
CHOP treated n = 181
233 DLBCL patients
high risk
low risk
Validation Cohort (Real-life patients)
102 DLBCL patients
(FFPE specimens)
high risk
low risk

Figure 2

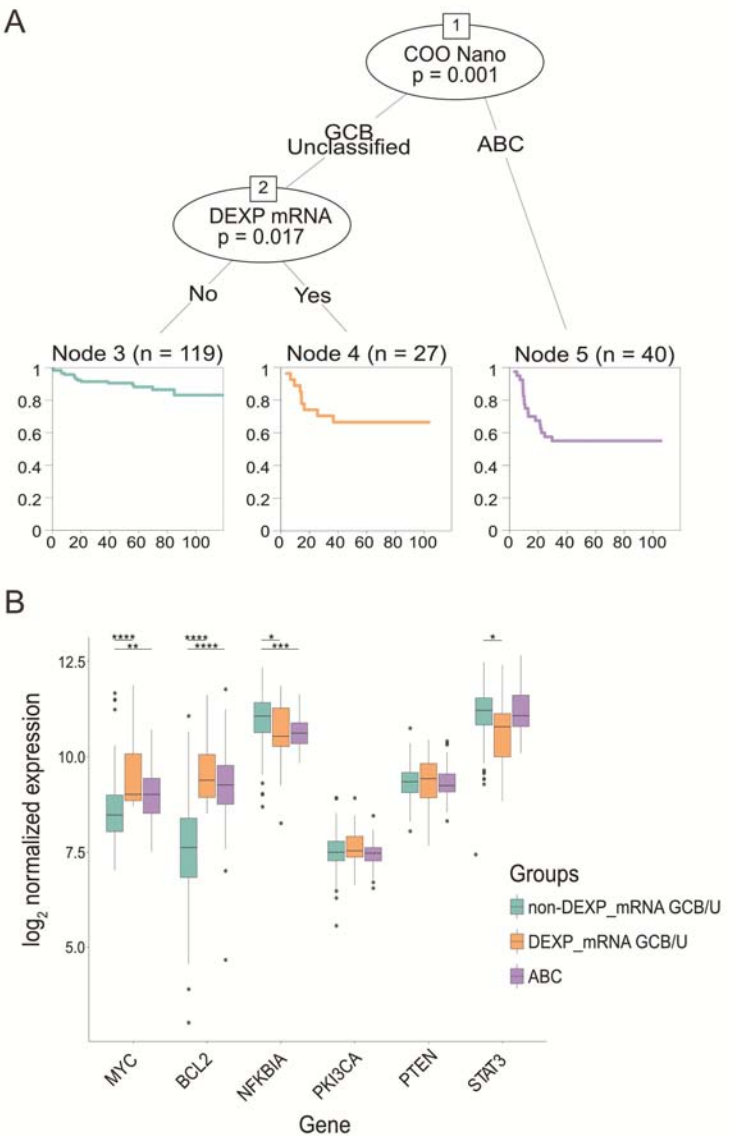
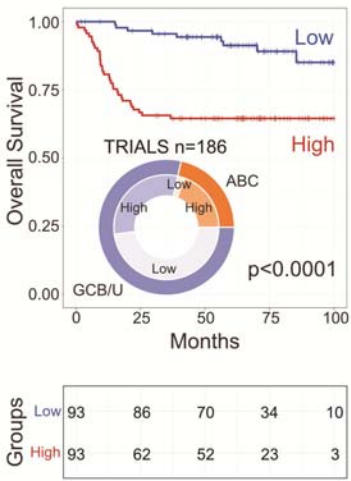

A
1
COO Nano
p = 0.001
GCB
Unclassified
ABC
2
DEXP mRNA
p = 0.017
No
Yes
Node 3 (n = 119)
Node 4 (n = 27)
Node 5 (n = 40)
1
0.8
0.6
0.4
0.2
0
0 20 40 60 80 100
B
log2 normalized expression
12.5
10.0
7.5
5.0
Groups
non-DEXP_mRNA GCB/U
DEXP_mRNA GCB/U
ABC
MYC
BCL2
NFKBIA
PIK3CA
PTEN
STAT3
Gene

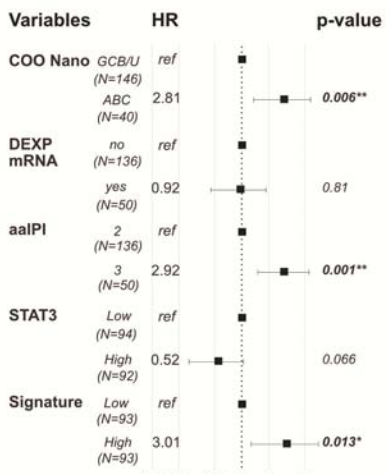
Figure 3

A

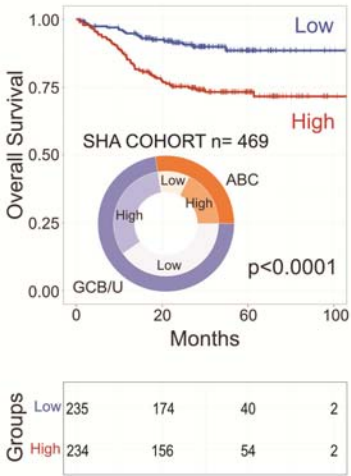



| Groups | | | | | |
|---|---|---|---|---|---|
| Low | 93 | 86 | 70 | 34 | 10 |
| High | 93 | 62 | 52 | 23 | 3 |

B




C




| Groups | | | | | |
|---|---|---|---|---|---|
| Low | 235 | 174 | 40 | 2 | |
| High | 234 | 156 | 54 | 2 | |

D

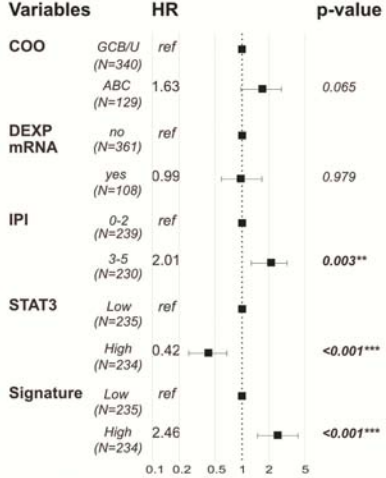


Figure 4

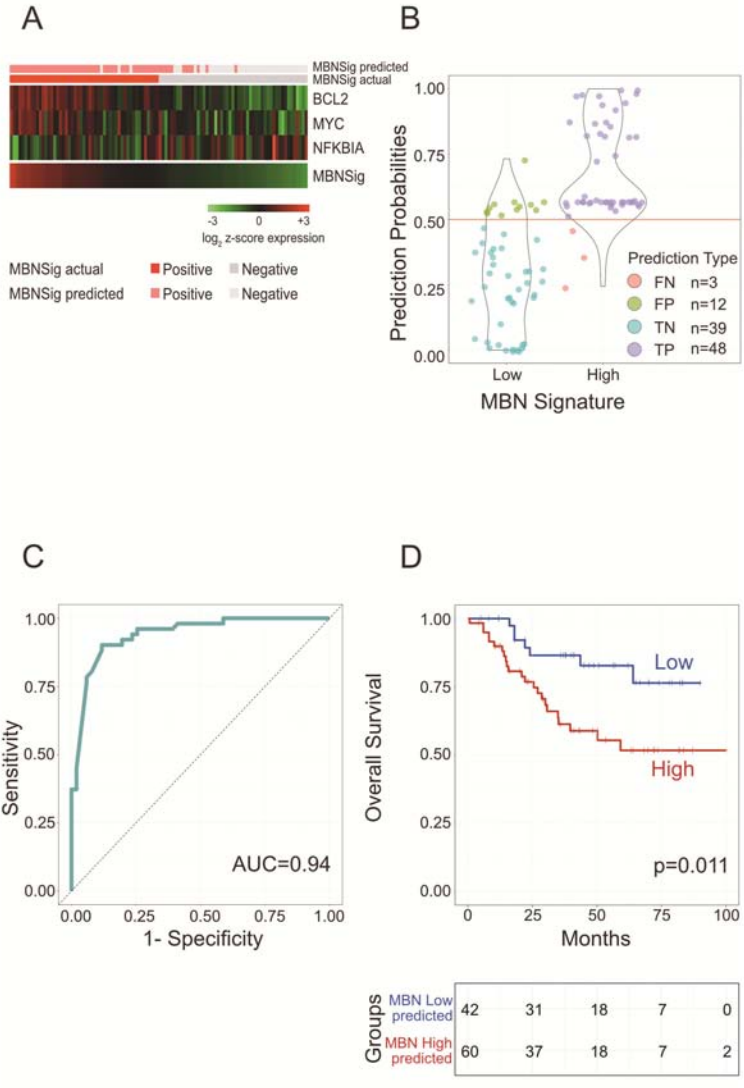
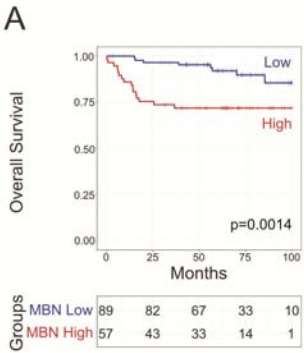
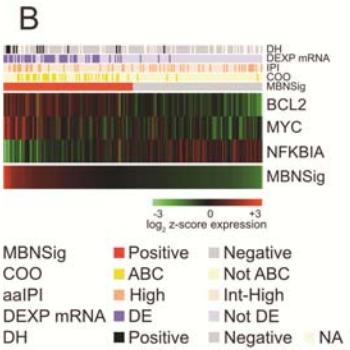
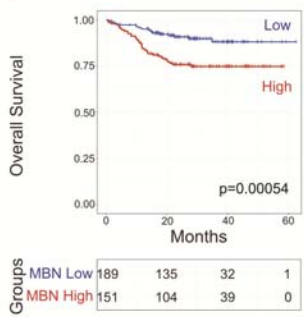
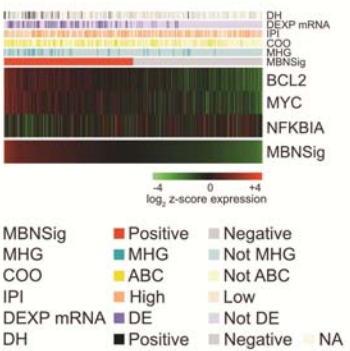

A
MBNSig predicted
MBNSig actual
BCL2
MYC
NFKBIA
MBNSig
-3 0 +3
log2 z-score expression
MBNSig actual Positive Negative
MBNSig predicted Positive Negative
B
Prediction Probabilities
1.00
0.75
0.50
0.25
0.00
Prediction Type
FN n=3
FP n=12
TN n=39
TP n=48
Low
High
MBN Signature
C
Sensitivity
1.00
0.75
0.50
0.25
0.00
AUC=0.94
0.00 0.25 0.50 0.75 1.00
1- Specificity
D
Overall Survival
1.00
0.75
0.50
0.25
0.00
Low
High
p=0.011
0 25 50 75 100
Months
Groups
MBN Low predicted 42 31 18 7 0
MBN High predicted 60 37 18 7 2

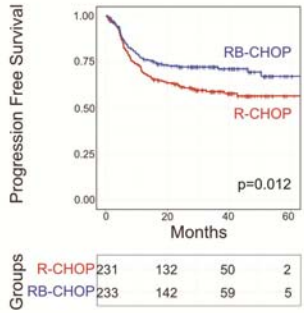
Figure 5

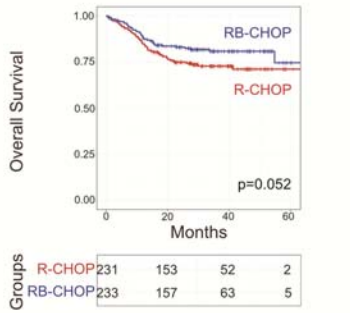

A
Overall Survival
1.00
0.75
0.50
0.25
0.00
Low
High
p=0.0014
0
25
50
75
100
Months
Groups
MBN Low 89 82 67 33 10
MBN High 57 43 33 14 1


B
DH
DEXP mRNA
IPI
COO
MBNSig
BCL2
MYC
NFKBIA
MBNSig
-3
0
+3
log2 z-score expression
MBNSig Positive Negative
COO ABC Not ABC
aaIPI High Int-High
DEXP mRNA DE Not DE
DH Positive Negative NA


C
Overall Survival
1.00
0.75
0.50
0.25
0.00
Low
High
p=0.00054
0
20
40
60
Months
Groups
MBN Low 189 135 32 1
MBN High 151 104 39 0


D
DH
DEXP mRNA
IPI
COO
MHG
MBNSig
BCL2
MYC
NFKBIA
MBNSig
-4
0
+4
log2 z-score expression
MBNSig Positive Negative
MHG MHG Not MHG
COO ABC Not ABC
IPI High Low
DEXP mRNA DE Not DE
DH Positive Negative NA


E
Progression Free Survival
1.00
0.75
0.50
0.25
0.00
RB-CHOP
R-CHOP
p=0.012
0
20
40
60
Months
Groups
R-CHOP 231 132 50 2
RB-CHOP 233 142 59 5


F
Overall Survival
1.00
0.75
0.50
0.25
0.00
RB-CHOP
R-CHOP
p=0.052
0
20
40
60
Months
Groups
R-CHOP 231 153 52 2
RB-CHOP 233 157 63 5

# 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 trial[1] and in the DLCL04 trial[2]. 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, GSE10846[36], 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 4$^{th}$ 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-μm-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 ≥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 sample[4]. 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 $\chi^2$ and Fisher's exact test. A p value $\leq 0.05$ 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 eefficiency 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 | CATCGCGACGGCCAAAAGGAGCGGCGCGGTCTTCGTGGTGTTCGTGGCAGGTGATGATGAACAGTCTACACAGATGGCTGCAAGTTGGGAAGATGATAAA |
| 2 | ISY1 | NM_020701.2 | HOUSEKEEPING | GGCAAAACATCAGTGTCTGTGGGTAGTTGGAATCTTCAGTTCCTGTGAGCGTCGGCGTCTTCTGGGCCTGTGGAGTTTCTTGGACAGGGGCCGCGGGGCT |
| 3 | R3HDM1 | NM_015361.2 | HOUSEKEEPING | CCTGTGTTCCCAAGAGAATTACATTATTGACAAAAGACTCCAAGACGAGGATGCCAGTAGTACCCAGCAGAGGCGCCAGATATTTAGAGTTAATAAAGAT |
| 4 | WDR55 | NM_017706.4 | HOUSEKEEPING | CTACCTCTTCAATTGGAATGGCTTTGGGGCCACAAGTGACCGCTTTGCCCTGAGAGCTGAATCTATCGACTGCATGGTTCCAGTCACCGAGAGTCTGCTG |
| 5 | TRIM56 | NM_030961.1 | HOUSEKEEPING | GTGGAGGCCGAGGACATTTTCCTGAAGGGCAGGGGTTGGCAACTTTTCAACATGGAGTGCCAAACTGCTAACCCGTCTTCTAGTGTGTGAGAATAGGGAC |
| 6 | MYC | NM_002467.3 | ENDOGENOUS | TCGGACACCGAGGAGAATGTCAAGAGGCGAACACACAACGTCTTGGAGCGCCAGAGGAGGAACGAGCTAAAACGGAGCTTTTTTGCCCTGCGTGACCAGA |
| 7 | PKI3CA | NM_006218.2 | ENDOGENOUS | CCTCAGGCTTGAAGAGTGTCGAATTATGTCCTCTGCAAAAAGGCCACTGTGGTTGAATTGGGAGAACCCAGACATCATGTCAGAGTTACTGTTTCAGAAC |

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
CAGTGTGAAGGTGTGTCAAGGTGTATATTCTCGTGGCTCGGCATTC

16 BCL2 NM_000657.2 ENDOGENOUS
GTGAAGCAGAAGTCTGGGAATCGATCTGGAAATCCTCCTAATTTTTACTCCCTCT
CCCCGCGACTCCTGATTCATTGGGAAGTTTCAAATCAGCTATAAC

17 CYB5R NM_016229.3 ENDOGENOUS
CCATGTCTTAGGGCTTCCTGTAGGTAACTATGTCCAGCTCTTGGCAAAAATCGAT
AATGAATTGGTGGTCAGGGCTTACACCCCTGTCTCCAGTGATGAT

18 MAML3 NM_018717.4 ENDOGENOUS
TGGAAGCCATCAACAATTTGCCCAGTAACATGCCACTGCCTTCAGCTTCTCCTCT
TCACCAACTTGACCTGAAACCTTCTTTGCCCTTGCAGAACAGTGG

19 SERPINA9 NM_001042518.1 ENDOGENOUS
CCACTAAATCCTAGGTGGGAAATGGCCTGTTAACTGATGGCACATTGCTAATGC
ACAAGAAATAACAAACCACATCCCTCTTTCTGTTCTGAGGGTGCAT

20 MYBL1 XM_034274.14 ENDOGENOUS
CTCCTTTTAAGAATGCGCTTGCTGCTCAGGAGAAAAAATATGGACCTCTTAAAA
TTGTGTCCCAGCCACTTGCTTTCTTGGAAGAAGATATTCGGGAAGT

21 RAB7L1 NM_001135664.1 ENDOGENOUS
CATTTGAATTGTCTCCTGACTACTGTCCAGTAAGGAGGCCCATTGTCACTTAGAA
AAGACACCTGGAACCCATGTGCATTTCTGCATCTCCTGGATTAGC

22 LIMD1 NM_014240.2 ENDOGENOUS
AAGGCAAGTCTCAGGAACCCATGCAGGTACATCGCTTGCACCTGTTTTTAGCTT
ATTTAATGACGGGCTTTTGGGAAGAGCTGCCCGCATACTGAGAGAC

23 ITPKB NM_002221.3 ENDOGENOUS
GGTTTGCGCCTCTGGGCATGTAGTCTACACAGGACCTGAGAATCTGAGAAACTG
CAGCCGCACGGTTGTTTATGGAGCTTTGGGCGGGGGCTGAGCCCGC

24 PTEN NM_000314.4 ENDOGENOUS
TCTTGACCAATGGCTAAGTGAAGATGACAATCATGTTGCAGCAATTCACTGTAA
AGCTGGAAAGGGACGAACTGGTGTAATGATATGTGCATATTTATTA

25 CREB3L2 NM_001253775.1 ENDOGENOUS
CGCACTTCTCAGAACTTCTGGATGAGTTTTCCCAGAACGTCTTGGGTCAGCTCCT
GAATGATCCTTTCCTCTCAGAGAAGAGTGTGTCAATGGAGGTGGA

26 CCDC50 NM_174908.3 ENDOGENOUS
AGGACATAGCTCGCCTTTTGCAAGAAAAGGAGTTACAGGAAGAGAAAAAGAGA
AAGAAACACTTTCCAGAGTTCCCTGCAACCCGTGCTTATGCAGATAG

## 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 $\chi^2$ 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 $\chi^2$ 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** | *Sha cohort* | *Lenz cohort* | *Real-life* |
| **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 | Diluition |
|---|---|---|---|
| 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** |
|---|---|---|---|
| ***MYC*** | | | |
| Low | 15 | 78 | 0.1 |
| High | 25 | 68 | |
| ***BCL2*** | | | |
| Low | 6 | 87 | <0.001 |
| High | 34 | 59 | |
| ***MYC-BCL-2*** **DEXPmRNA** | | | <0.001 |
| No | 17 | 119 | |
| Yes | 23 | 27 | |
| ***STAT3*** | | | |
| Low | 21 | 73 | 0.9 |
| High | 19 | 73 | |
| ***NFKBIA*** | | | |
| Low | 30 | 66 | 0.001 |
| High | 10 | 80 | |
| ***PIK3CA*** | | | |
| Low | 22 | 71 | 0.59 |
| High | 18 | 75 | |
| ***PTEN*** | | | |
| 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 | ref |
| ABC | 2.81 (1.35-5.83) |
| DEXP mRNA | |
| no | ref |
| yes | 0.92 (0.45-1.86) |
| aaIPI | |
| 2 | ref |
| 3 | 2.92 (1.52-5.60) |
| STAT3 | |
| Low | ref |
| High | 0.52 (0.26-1.04) |
| MBN Signature | |
| Low | ref |
| High | 3.01 (1.27-7.15) |

Table S5. Multivariate Cox regression analysis (Sha cohort)

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

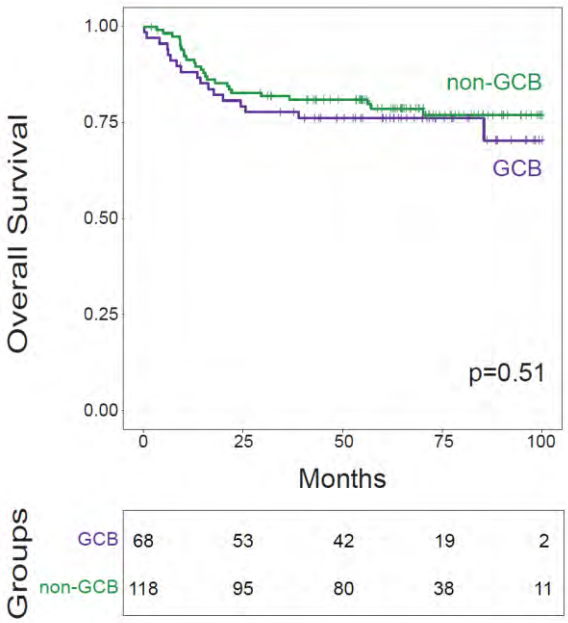
2- Chiappella A, Martelli M, Angelucci E, Brusamolino E, Evangelista A, Carella AM et al. Rituximab-dose-dense chemotherapy with or without high-dose chemotherapy plus autologous stem-cell transplantation in high-risk diffuse large B-cell lymphoma (DLCL04): final results of a multicentre, open-label, 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.

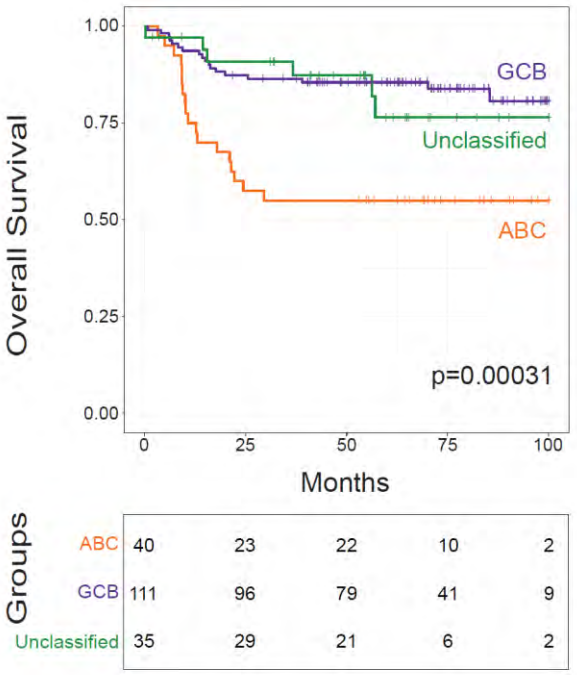
4- Scott DW, Wright GW, Williams PM, Lih CJ, Walsh W, Jaffe ES et al. Determining cell-of-origin subtypes of diffuse large B-cell lymphoma using gene expression in formalin-fixed paraffin-embedded tissue. Blood. 2014; 123(8): 1214-7.

Figure S1

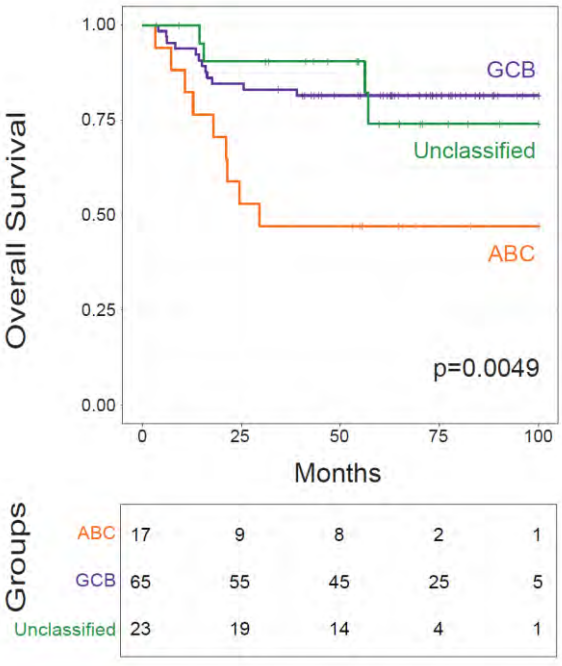
A




B




C




D

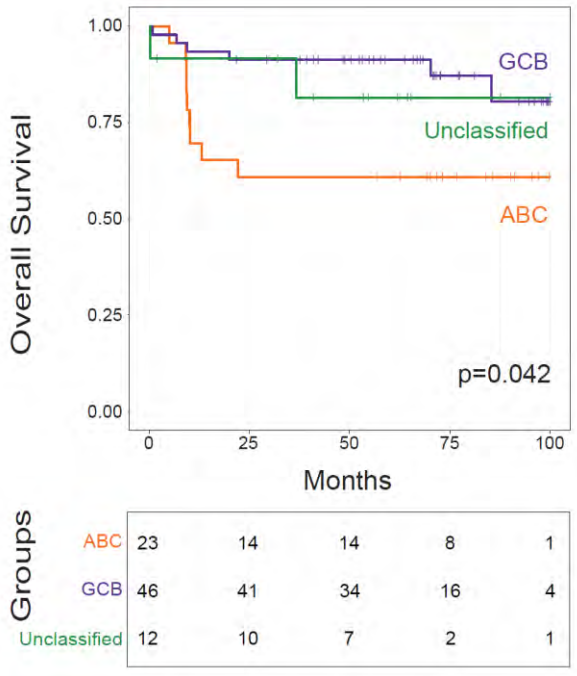


Figure S2

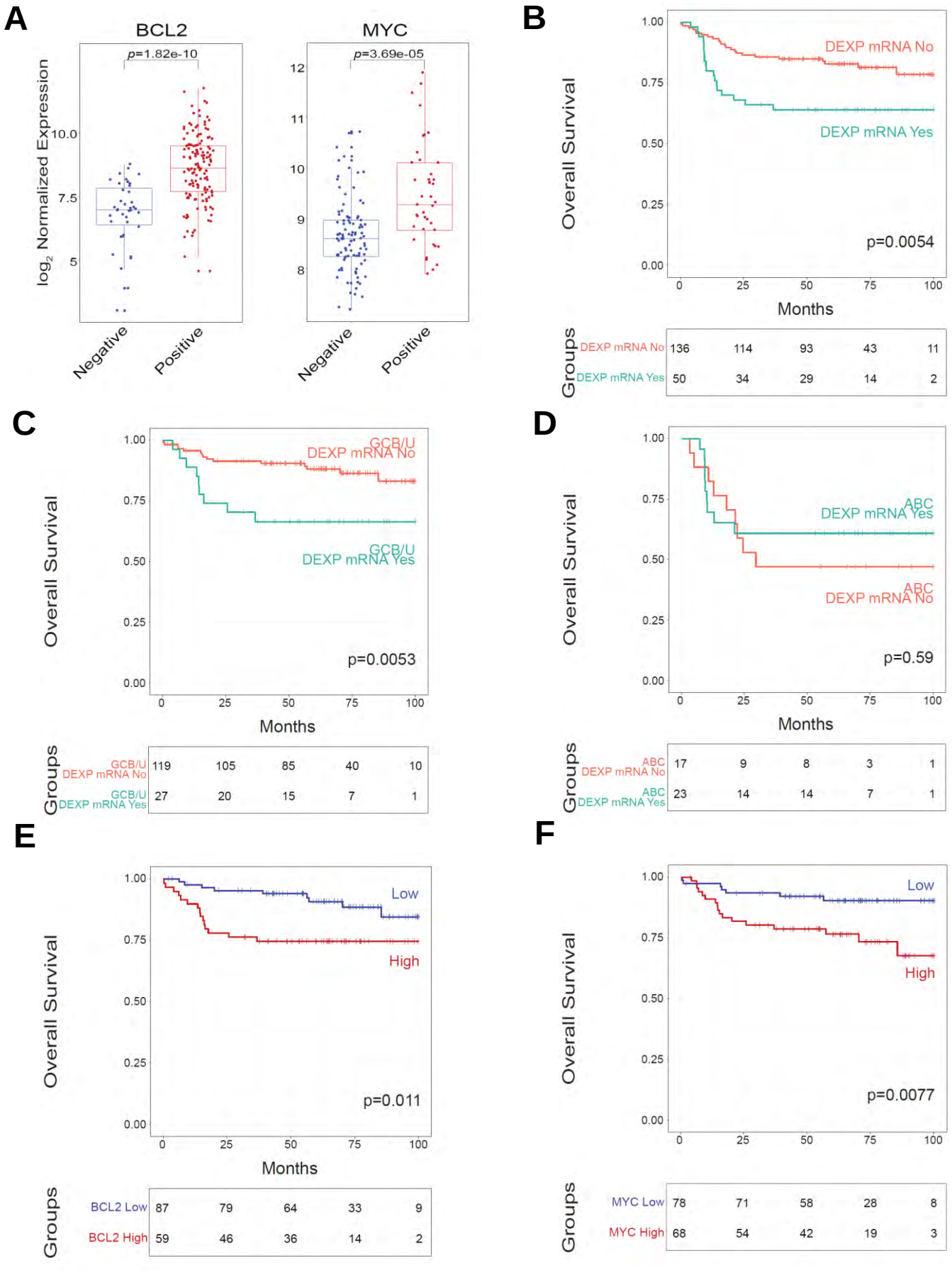

A
BCL2
p=1.82e-10
MYC
p=3.69e-05
log2 Normalized Expression
Negative
Positive
B
DEXP mRNA No
DEXP mRNA Yes
p=0.0054
Overall Survival
Months
Groups
DEXP mRNA No 136 114 93 43 11
DEXP mRNA Yes 50 34 29 14 2
C
GCB/U DEXP mRNA No
GCB/U DEXP mRNA Yes
p=0.0053
Overall Survival
Months
Groups
GCB/U DEXP mRNA No 119 105 85 40 10
GCB/U DEXP mRNA Yes 27 20 15 7 1
D
ABC DEXP mRNA Yes
ABC DEXP mRNA No
p=0.59
Overall Survival
Months
Groups
ABC DEXP mRNA No 17 9 8 3 1
ABC DEXP mRNA Yes 23 14 14 7 1
E
Low
High
p=0.011
Overall Survival
Months
Groups
BCL2 Low 87 79 64 33 9
BCL2 High 59 46 36 14 2
F
Low
High
p=0.0077
Overall Survival
Groups
MYC Low 78 71 58 28 8
MYC High 68 54 42 19 3

Figure S3

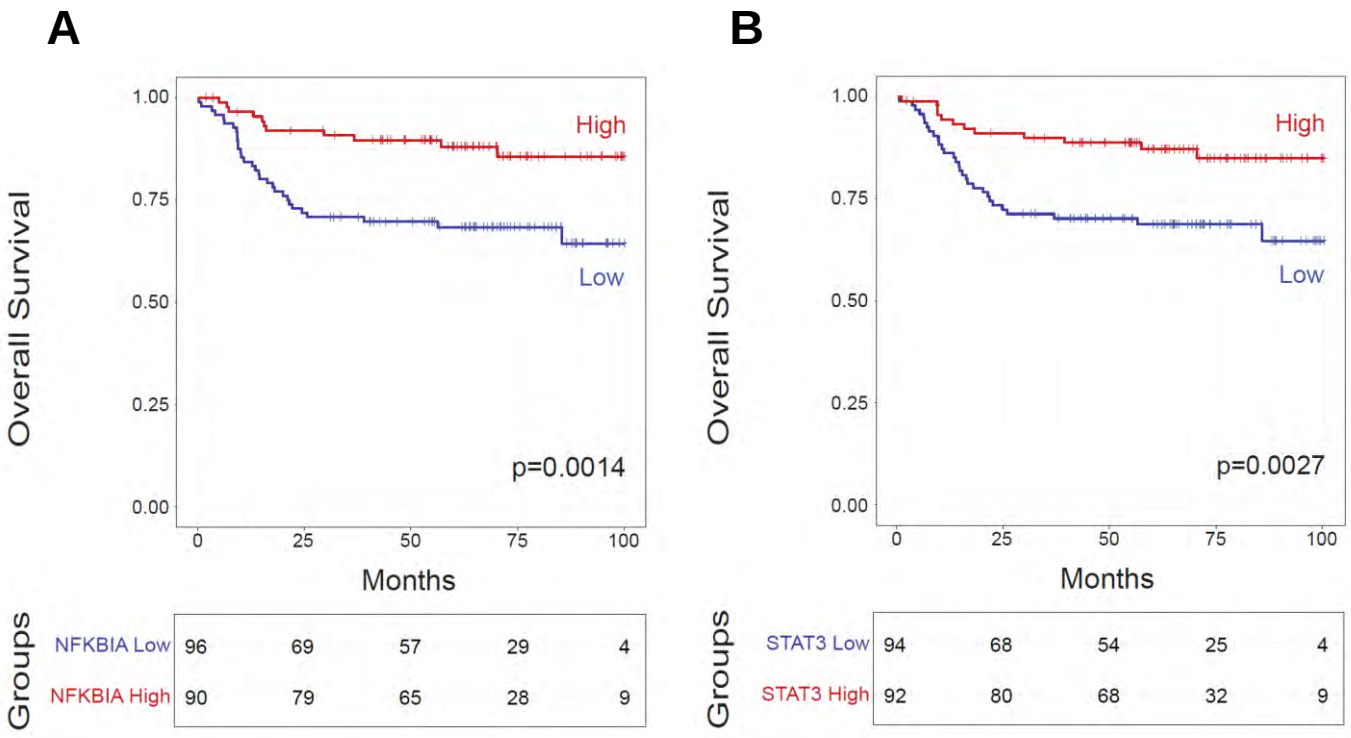

A
Overall Survival
1.00
0.75
0.50
0.25
0.00
High
Low
p=0.0014
0
25
50
75
100
Months
Groups
NFKBIA Low 96 69 57 29 4
NFKBIA High 90 79 65 28 9
B
Overall Survival
1.00
0.75
0.50
0.25
0.00
High
Low
p=0.0027
0
25
50
75
100
Months
Groups
STAT3 Low 94 68 54 25 4
STAT3 High 92 80 68 32 9

Figure S4

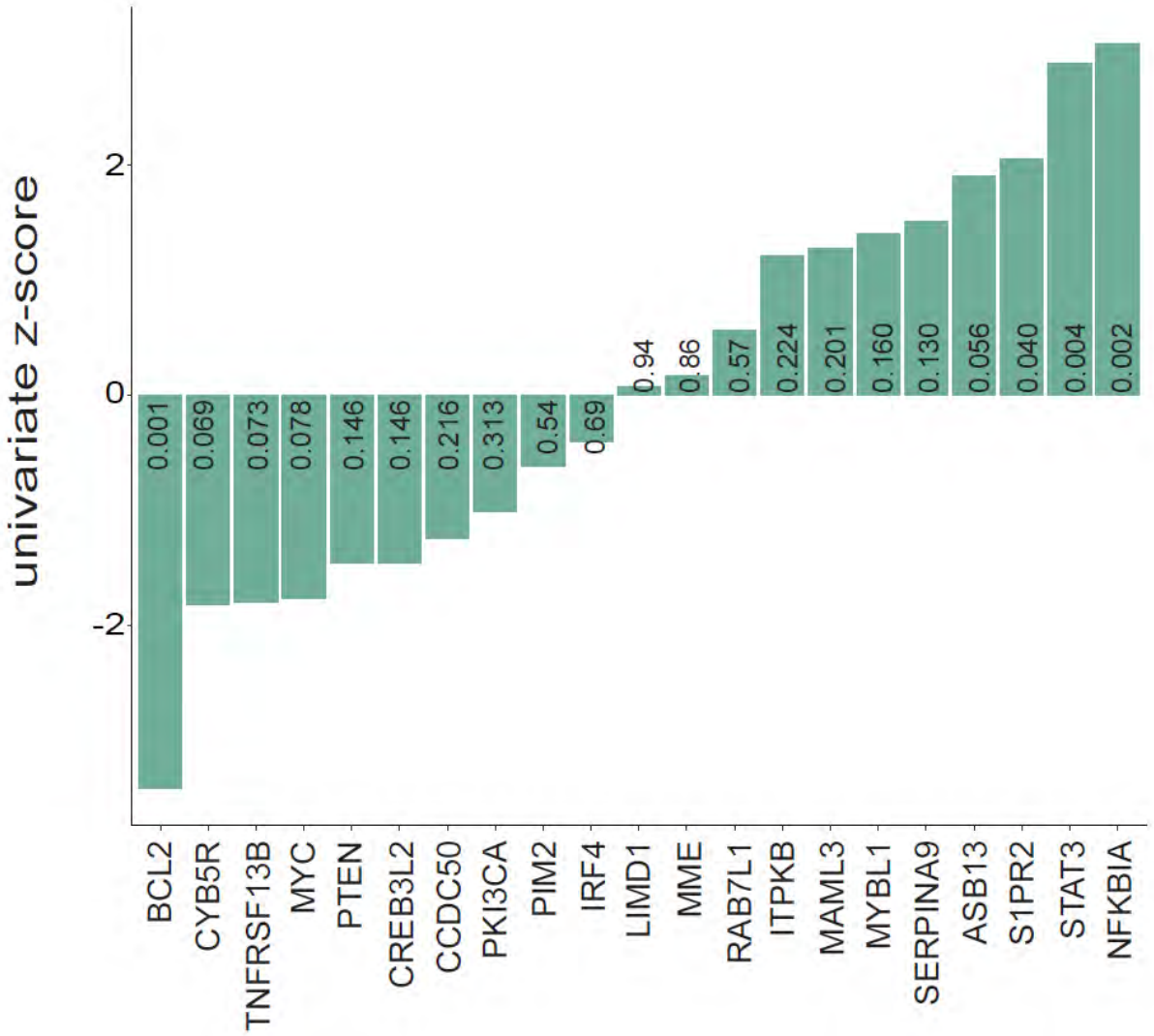
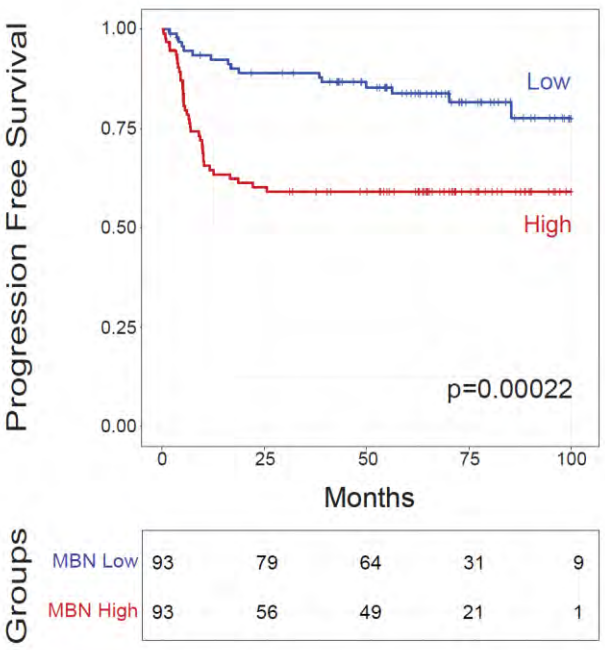

univariate z-score
2
0
-2
BCL2
CYB5R
TNFRSF13B
MYC
PTEN
CREB3L2
CCDC50
PKI3CA
PIM2
IRF4
LIMD1
MME
RAB7L1
ITPKB
MAML3
MYBL1
SERPINA9
ASB13
S1PR2
STAT3
NFKBIA
0.001
0.069
0.073
0.078
0.146
0.146
0.216
0.313
0.54
0.69
0.94
0.86
0.57
0.224
0.201
0.160
0.130
0.056
0.040
0.004
0.002

Figure S5

**A**


Progression Free Survival
1.00
0.75
0.50
0.25
0.00
Low
High
p=0.00022
0
25
50
75
100
Months
Groups
MBN Low 93 79 64 31 9
MBN High 93 56 49 21 1

**B**

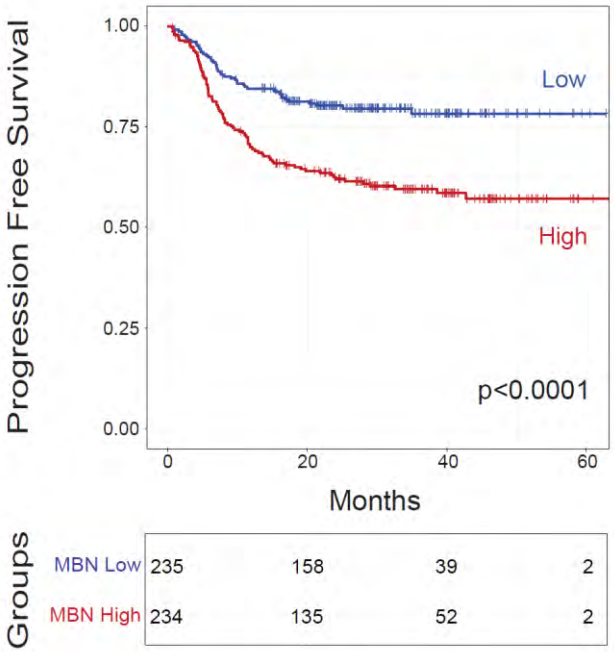
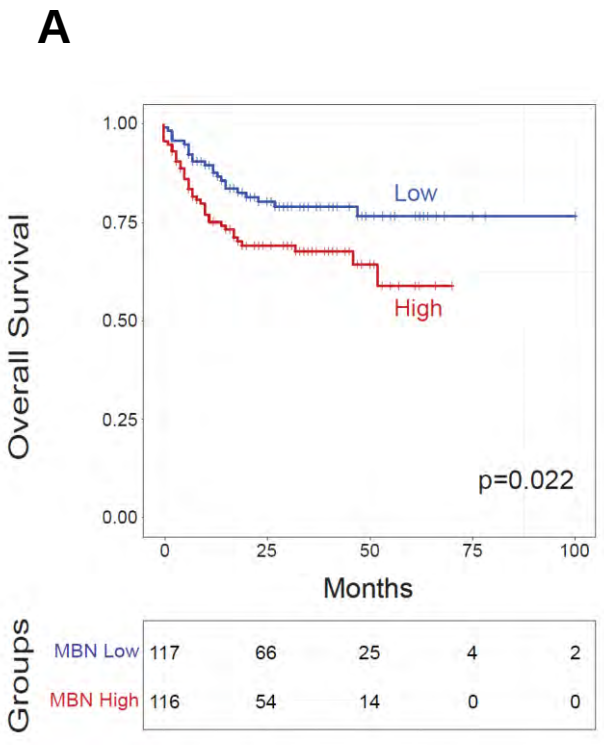
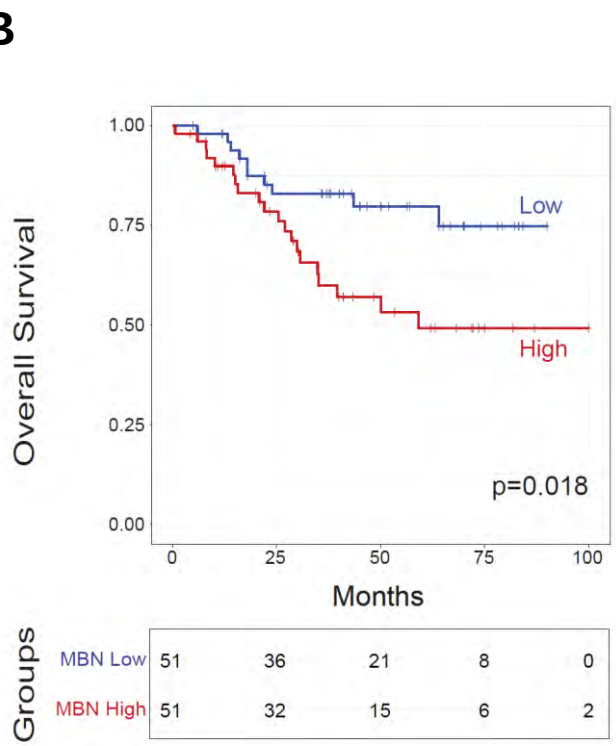
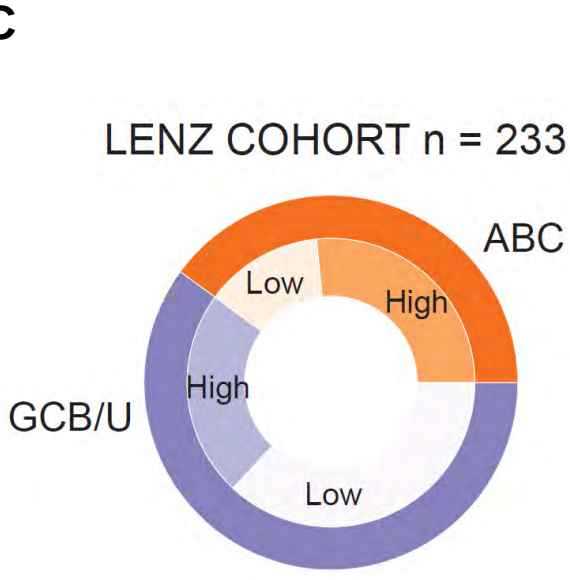

Progression Free Survival
1.00
0.75
0.50
0.25
0.00
Low
High
p<0.0001
0
20
40
60
Months
Groups
MBN Low 235 158 39 2
MBN High 234 135 52 2

Figure S6


A
1.00
0.75
0.50
0.25
0.00
Overall Survival
Low
High
p=0.022
0
25
50
75
100
Months
Groups
MBN Low 117 66 25 4 2
MBN High 116 54 14 0 0


B
1.00
0.75
0.50
0.25
0.00
Overall Survival
Low
High
p=0.018
0
25
50
75
100
Months
Groups
MBN Low 51 36 21 8 0
MBN High 51 32 15 6 2


C
LENZ COHORT n = 233
ABC
Low
High
GCB/U
High
Low

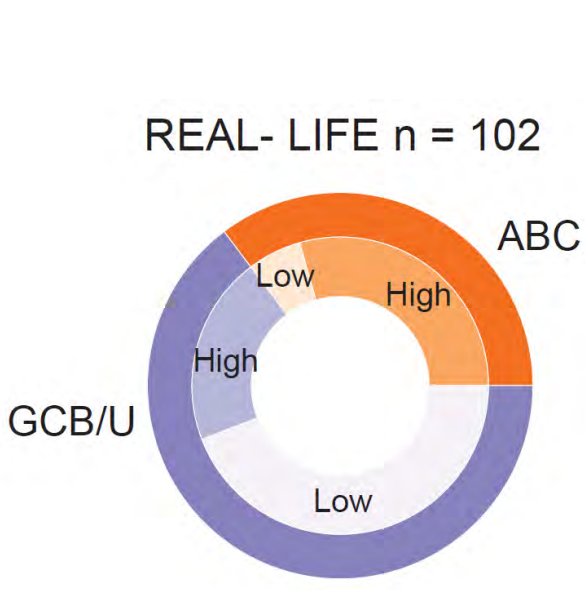

D
REAL- LIFE n = 102
ABC
Low
High
GCB/U
High
Low

Figure S7

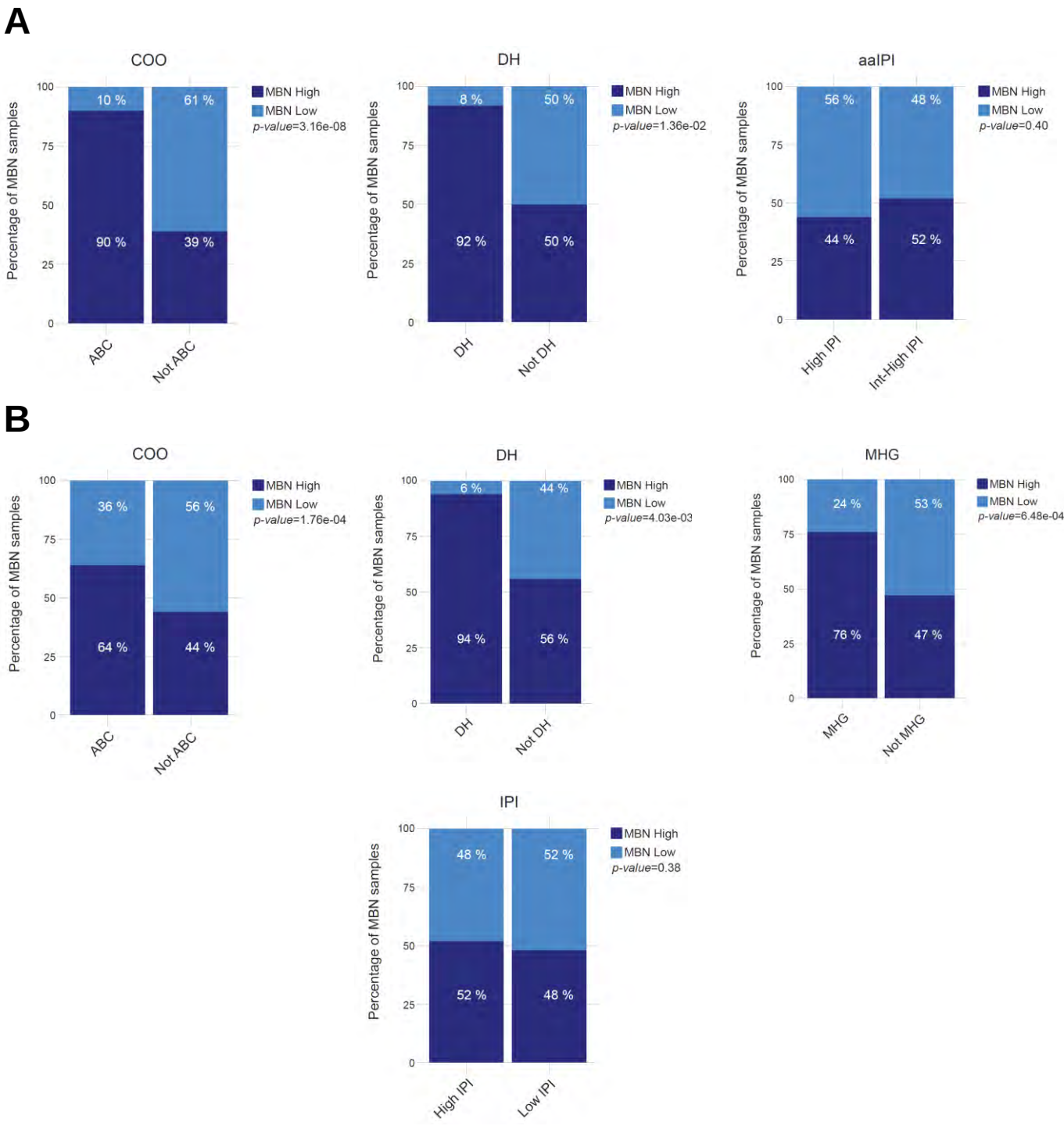
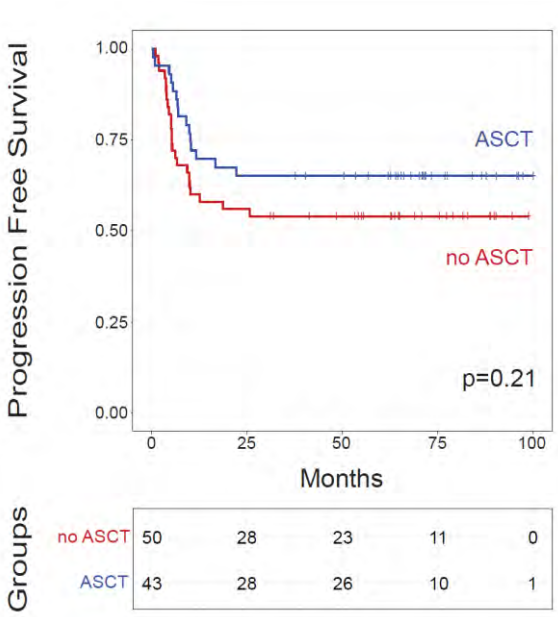
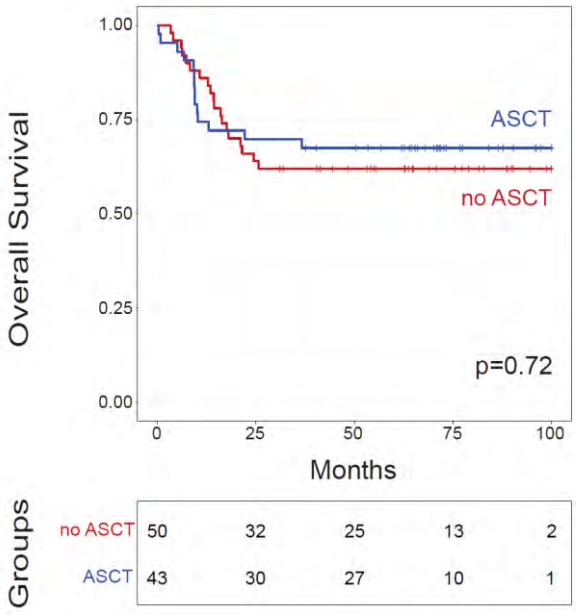

A
COO
Percentage of MBN samples
100
75
50
25
0
10 %
61 %
90 %
39 %
ABC
Not ABC
MBN High
MBN Low
p-value=3.16e-08
DH
8 %
50 %
92 %
50 %
DH
Not DH
MBN High
MBN Low
p-value=1.36e-02
aaIPI
56 %
48 %
44 %
52 %
High IPI
Int-High IPI
MBN High
MBN Low
p-value=0.40
B
COO
36 %
56 %
64 %
44 %
ABC
Not ABC
MBN High
MBN Low
p-value=1.76e-04
DH
6 %
44 %
94 %
56 %
DH
Not DH
MBN High
MBN Low
p-value=4.03e-03
MHG
24 %
53 %
76 %
47 %
MHG
Not MHG
MBN High
MBN Low
p-value=6.48e-04
IPI
48 %
52 %
52 %
48 %
High IPI
Low IPI
MBN High
MBN Low
p-value=0.38

Figure S8

**A**


Progression Free Survival
1.00
0.75
0.50
0.25
0.00
ASCT
no ASCT
p=0.21
0
25
50
75
100
Months
Groups
no ASCT 50 28 23 11 0
ASCT 43 28 26 10 1

**B**


Overall Survival
1.00
0.75
0.50
0.25
0.00
ASCT
no ASCT
p=0.72
0
25
50
75
100
Months
Groups
no ASCT 50 32 25 13 2
ASCT 43 30 27 10 1