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

Ferrata Storti Foundation

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 Pileri[3]

[1]Onco-Hematology Division, IEO European Institute of Oncology IRCCS, Milan and European Institute of Oncology IRCCS, Milan, Italy; [2]Department of Health Sciences, University of Milan, Milan, Italy; [3]Division of Diagnostic Hematopathology, IEO European Institute of Oncology IRCCS, Milan and 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, IEO European Institute of Oncology IRCCS, Milan and European Institute of Oncology IRCCS, Milan, Italy; [7]Division of Hematology, Fondazione IRCCS Istituto Nazionale dei Tumori, University of Milan, 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 and [11]Institute of Hematology and Medical Oncology "L. e A. Seràgnoli", Department of Experimental, Diagnostic, and Specialty Medicine (DIMES), Bologna University School of Medicine, Bologna, Italy

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 two randomized trials (discovery cohort) (clinicaltrials gov. Identifier: NCT00355199 and NCT00499018). Data were validated in three independent series (two 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 three-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 three 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 three-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 the 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 DLBCL 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 dependent on nuclear factor k-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 DLBCL 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 three independent phase III randomized trials[16-18] based on COO classification were largely negative. Although these unsatisfactory results could be due to 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 DLBCL 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 DLBCL enrolled in two recently reported large Italian trials (DLCL04 and R-HDS0305; clinicaltrials. gov. Identifier: NCT00355199 and NCT00499018).[34,35] We found that a three-gene signature based on *MYC, BCL-2* and *NFKBIA* (MBN signature), identified a significant fraction of ABC cases and a sub-group 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 two 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 two prospective randomized phase III 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 three 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] five housekeeping genes (*UBXN4, ISY1, R3HDM1, WDR55, TRIM56*); and six 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 analysis

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 ≤0.05 was considered statistically significant. 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 χ^2 and Fisher's exact test.

Development of the three-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 ABC 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 one hand and *NFKBIA* on the other hand 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:

$$MBN\text{-}Sig = (MYC + BCL\text{-}2)/NFKBIA$$

Detailed information on study cohorts (*Online Supplementary Table S1*), T-GEP procedures with list of genes and target sequences, fluorescence *in situ* hybridization (FISH), IHC methods and antibodies (*Online Supplementary Table S2*), and random forest (RF) classifier are described in the *Online Supplementary Appendix*.

Results

Univariate analyses and a decision-tree classification model integrating the cell of origin 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 (*Online Supplementary Figure S1A* to *D*). In order to investigate the prognostic impact of concurrent overexpression of *MYC* and *BCL-2*, we defined high and low expressors based on
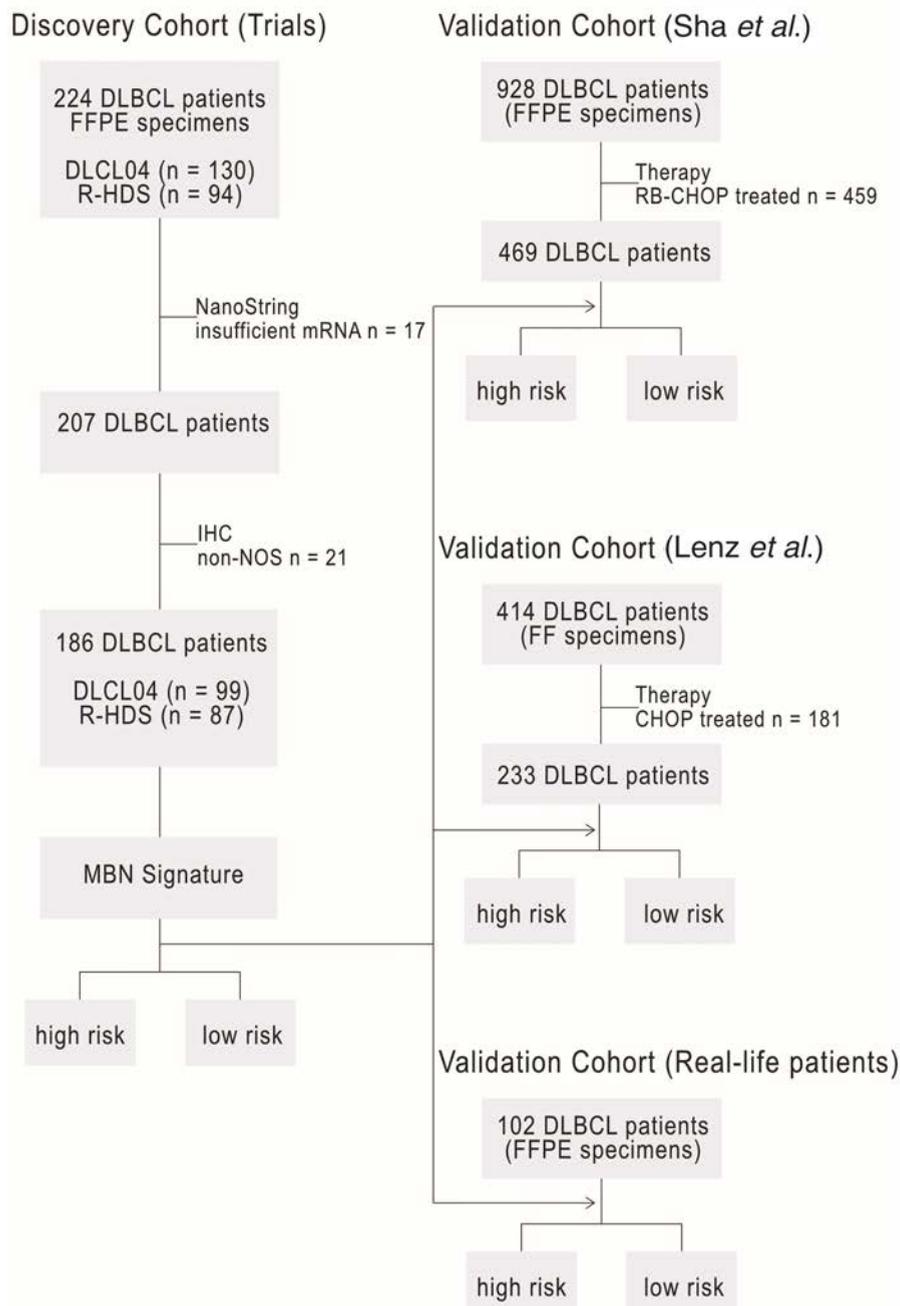
Figure 1. Study algorithm. On the left, the discovery cohort is represented; 224 diffuse large B-cell lymphoma (DLBCL) patients enrolled in the DLCLO4 (n= 130) and R-HDS0305 (n= 94) trials with available formalin-fixed, paraffin embedded (FFPE) tissue were initially considered in this analysis. Targeted gene expression profiling (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 DLCLO4 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, (h t t p s : / / w w w . n c b i . n l m . nih.gov/geo/query/acc.cgi?acc=GSE1 0846), 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. RB-CHOP: R-CHOP plus bortezomib.

the median normalized *MYC* and *BCL-2* mRNA levels, which correlated well with the respective protein levels assessed by IHC (*Online Supplementary Figure S2A*). *MYC/BCL-2* mRNA double expressors (defined as DEXP_mRNA) patients showed a worse outcome compared to non-DEXP_mRNA cases (*Online Supplementary Figure S2B*). Although DEXP_mRNA cases were more prevalent in the ABC compared to the GCB/unclassified (GCB/U) subgroup[9] (*Online Supplementary Table S3*), the prognostic relevance of the MYC/BCL-2 DEXP_mRNA status was particularly evident in the GCB/U subset (*Online Supplementary Figure S2C* to *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 (*Online*
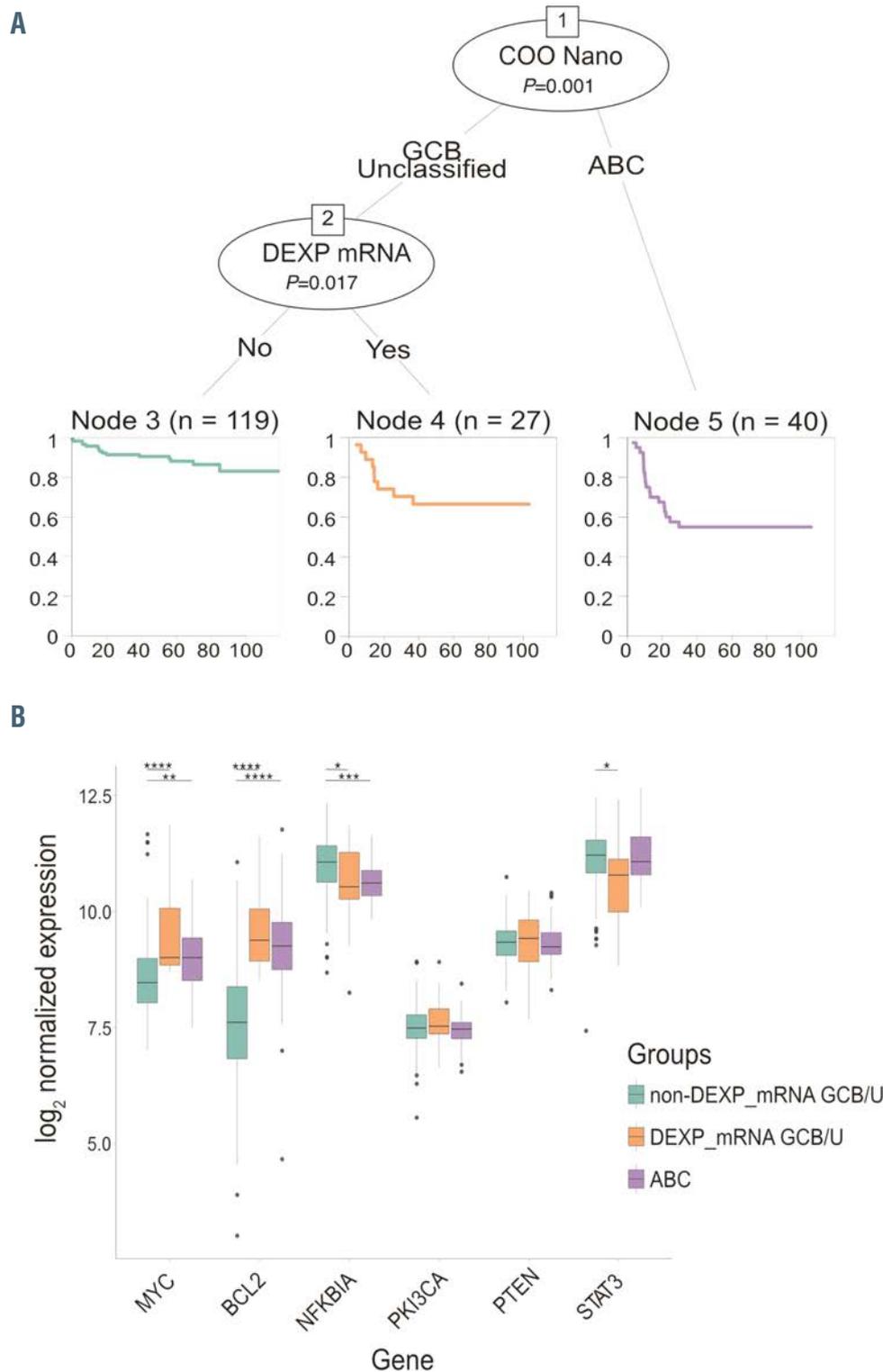

Figure 2. Integrating cell of origin with *MYC/BCL-2* DEXP_mRNA status for prognostication in diffuse large B-cell lymphoma. (A) Recursive partitioning analysis integrating cell of origin (COO) classification and DEXP_mRNA status, allowing segregation of patients in three main prognostic subgroups (a low risk non-DEXP-mRNA GCB/U subset, and two 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 three main patients subgroups identified by the recursive partitioning analysis (non-DEXP-mRNA GCB/U, *MYC/BCL-2* DEXP_mRNA GCB/U, ABC-derived diffuse large B-cell lymphoma [DLBCL]. *P*-value was calculated with Student *t*-test by comparing non DEXP_mRNA GCB/U group, selected as a reference, *versus* other groups. ABC: activated B-cells; GCB: germinal center B cells; GCB/U: GCB unclassified; DEXP_mRNA: double expressor GCB/U DLBCL. COO Nano: COO as determined by T-GEP with NanoString profiling

Supplementary Figures S3A and 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 (*Online Supplementary Figure S1* and *S2*), a recursive partitioning analysis integrating the COO with *MYC/BCL-2* status identified three main patient subgroups: two 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 DLBCL, (n=119) (Figure 2A). Evaluating the relative expression of the additional genes included in the panel across the three 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 three-gene prognostic signature combining *MYC, BCL-2* and *NFKBIA*

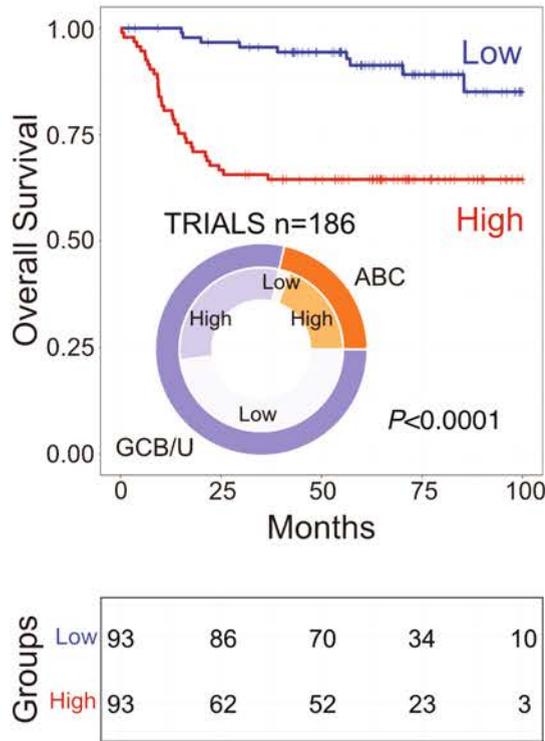In an effort to build a GEP signature aimed at refining current prognostication algorithms and suitable for clinical practice, we considered only those genes whose expression was significantly associated with OS and differentially represented in both high risk (ABC and DEXP_mRNA GCB/U) *versus* low risk (non-DEXP mRNA GCB/U) patient subsets. Using these criteria, we constructed a prognostic signature considering three 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) (*Online Supplementary Figure S4*). With this strategy, patients were divided in two risk categories characterized by different outcome: low risk patients (MBN-Sig low) had a very favorable prognosis (91% 5-year OS; 84% 5-year PFS), whereas high-risk patients (MBN-Sig high) had a significantly worse prognosis (64% 5-year OS; 59% 5-year PFS) (Figure 3A; *Online Supplementary Figure S5A*). Importantly the MBN signature retained its significance and outperformed the *MYC/BCL-2* DEXP_mRNA status in multivariate analysis (Figure 3B; *Online Supplementary 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 and D; *Online Supplementary Figure S5B; Online Supplementary 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 (*Online Supplementary Figure S6A* and *B*). The MBN signature was able to identify a significant fraction of ABC-derived cases and about a third of GCB/U cases (Figure 3A and C; *Online Supplementary Figure S6C* and *D*).
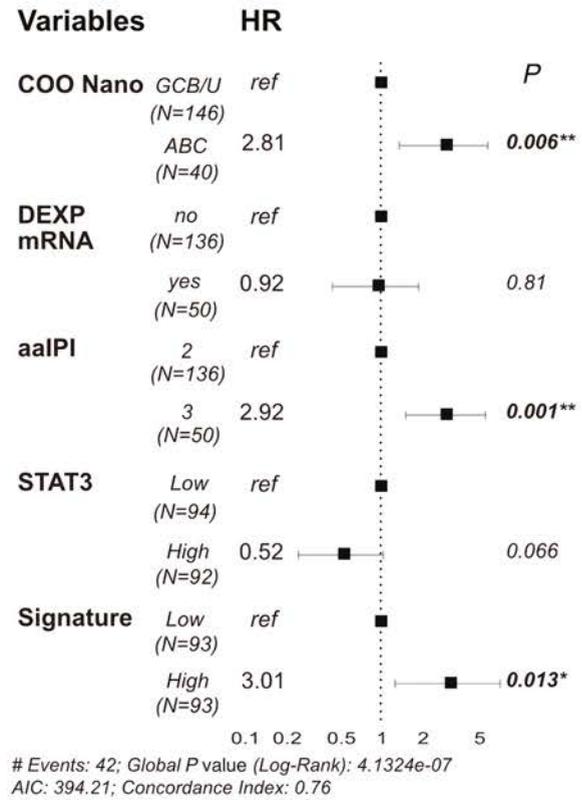
Table 1. Patients characteristics.

Trial name	RHDS0305	DLCL04	P*	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: 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; GCB: germinal center B cells; ABC: activated B-cells; IHC: immunohistochemistry.
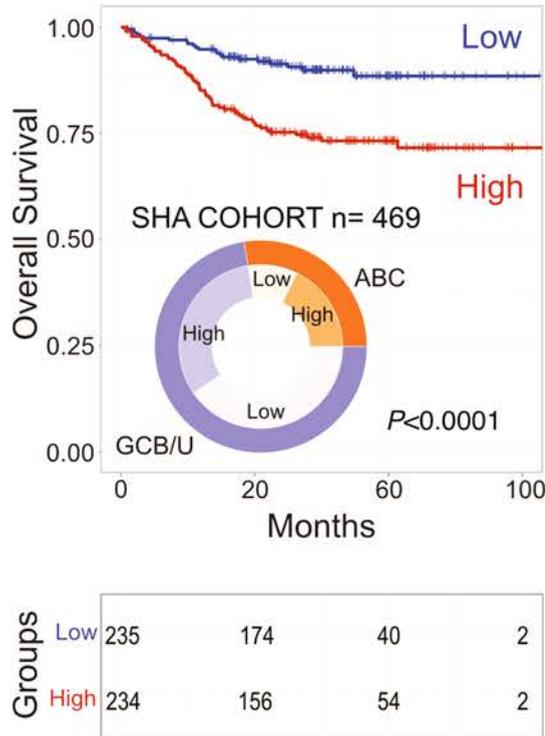
A

B

Variables		HR		P
COO Nano	GCB/U (N=146)	ref		
	ABC (N=40)	2.81		*0.006***
DEXP mRNA	no (N=136)	ref		
	yes (N=50)	0.92		0.81
aaIPI	2 (N=136)	ref		
	3 (N=50)	2.92		*0.001***
STAT3	Low (N=94)	ref		
	High (N=92)	0.52		0.066
Signature	Low (N=93)	ref		
	High (N=93)	3.01		*0.013**

 # Events: 42; Global P value (Log-Rank): 4.1324e-07
AIC: 394.21; Concordance Index: 0.76

C

D

Variables		HR		P
COO	GCB/U (N=340)	ref		
	ABC (N=129)	1.63		0.065
DEXP mRNA	no (N=361)	ref		
	yes (N=108)	0.99		0.979
IPI	0-2 (N=239)	ref		
	3-5 (N=230)	2.01		*0.003***
STAT3	Low (N=235)	ref		
	High (N=234)	0.42		*<0.001****
Signature	Low (N=235)	ref		
	High (N=234)	2.46		*<0.001****

 # Events: 81; Global P value (Log-Rank): 2.9833e-08
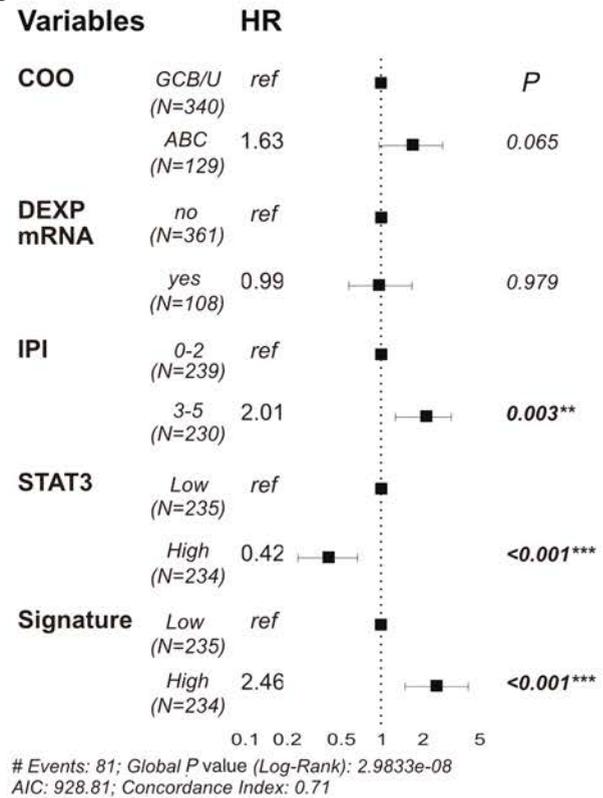*AIC: 928.81; Concordance Index: 0.71*

Figure 3. Legend on next page.

Figure 3. Survival curves according to MBN signature and multivariate analyses for overall survival. (A) Overall survival (OS) of the discovery cohort (R-HDS0305+DLCL04; n=186 patients) according to the MBN signature (MBN-Sig) showing significant differences in outcome between MBN-Sig low *versus* MBN-Sig high patient subsets. *P*-values were calculated with the log rank test. Frequencies of MBN-Sig high *versus* low cases in activated B-cells (ABC) and germinal center B cells/ unclassified (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 cell of origin (COO) as determined by NanoString-based targeted gene expression profiling (T-GEP) (COO_Nano), the MBN-Sig and the age adjusted international prognostic index (aaIPI) 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 *versus* MBN-Sig high patient subsets. *P*-values were calculated with the log rank test. Frequencies of MBN-Sig high *versus* low cases in ABC and GCB/U subsets in the Sha's cohort are represented in the pie chart. (C) Forest plot depicting multivariate analyses for OS (Sha's dataset), confirming the significant independent association with OS of the MBN-Sig in this large validation cohort.

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 an 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 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 the *Online Supplementary Figure S6B*.

Correlation of the MBN signature with fluorescence *in situ* hybridization 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* re-arrangements in the MBN-Sig high subgroup compared to the GCB/U MBN-Sig low subset. According to these observations, there was a significant enrichment of HG-BCL w/DH in the MBN-Sig high subgroup compared to the MBN-Sig low subset (Figure 5B; *Online Supplementary 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; *Online Supplementary Figure S7A*). Finally, no significant differences in the aaIPI score (intermediate high *vs.* high) were observed between groups (Figure 5B; *Online Supplementary Figure S7A*). These findings were validated *in silico* in the larger cohort from Sha *et al.*[27] (Figure 5C and D). As observed in the discovery cohort, the MBN signa-

Table 2. Univariate analysis for overall survival.

	Hazard Ratio	95% CI	*P*
aaIPI			
Intermediate-High	Ref	1.10-3.78	0.023
High	2.04		
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	0.47-1.63	0.68
Yes	0.88		
MYC-BCL-2 DEXPmRNA			
No	Ref	1.26-4.29	0.007
Yes	2.32		
STAT3			
Low	Ref	0.19-0.7	0.004
High	0.37		
NFKBIA			
Low	Ref	0.17-0.68	0.002
High	0.34		

Signif: significance; ref: reference; IPI: international prognostic index; COO: cell of origin; COO_Nano: COO defined by NanoString; ASCT: autologous stem cell transplant; MYC/BCL-2 DEXP_mRNA: double expressor status defined by NanoString: CI: Confidence Interval. *P<0.05; **P<0.01.

ture stratified the prognosis of GCB/U patients (Figure 5C) and identified the vast majority of DH cases. Again ABC-derived DLBCL were more highly represented in the MBN-Sig high subgroup (Figure 5D; *Online Supplementary 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] In order to evaluate how our MBN signature performed in the same patient population, we compared the MBN signature with the MHG signature and with the FISH status (Figure 5C). Notably the MBN-high subgroup was significantly enriched in MHG cases, identifying 76% of MHG DLBCL and the vast majority of DH (Figure 5D; *Online Supplementary Figure S7B*). Also in this cohort there were no differences in IPI score between groups (Figure 5D; *Online Supplementary Figure S7B*).

Rationale for a precision therapy approach in MBN-Sig high-risk diffuse large B-cell lymphoma patients

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 *versus* standard chemoimmunotherapy in the discovery cohort. ASCT consolida-

tion did not provide any significant PFS or OS advantage compared to standard chemoimmunotherapy in the MBN-Sig high subgroup (*Online Supplementary Figure S8A and 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 examined 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
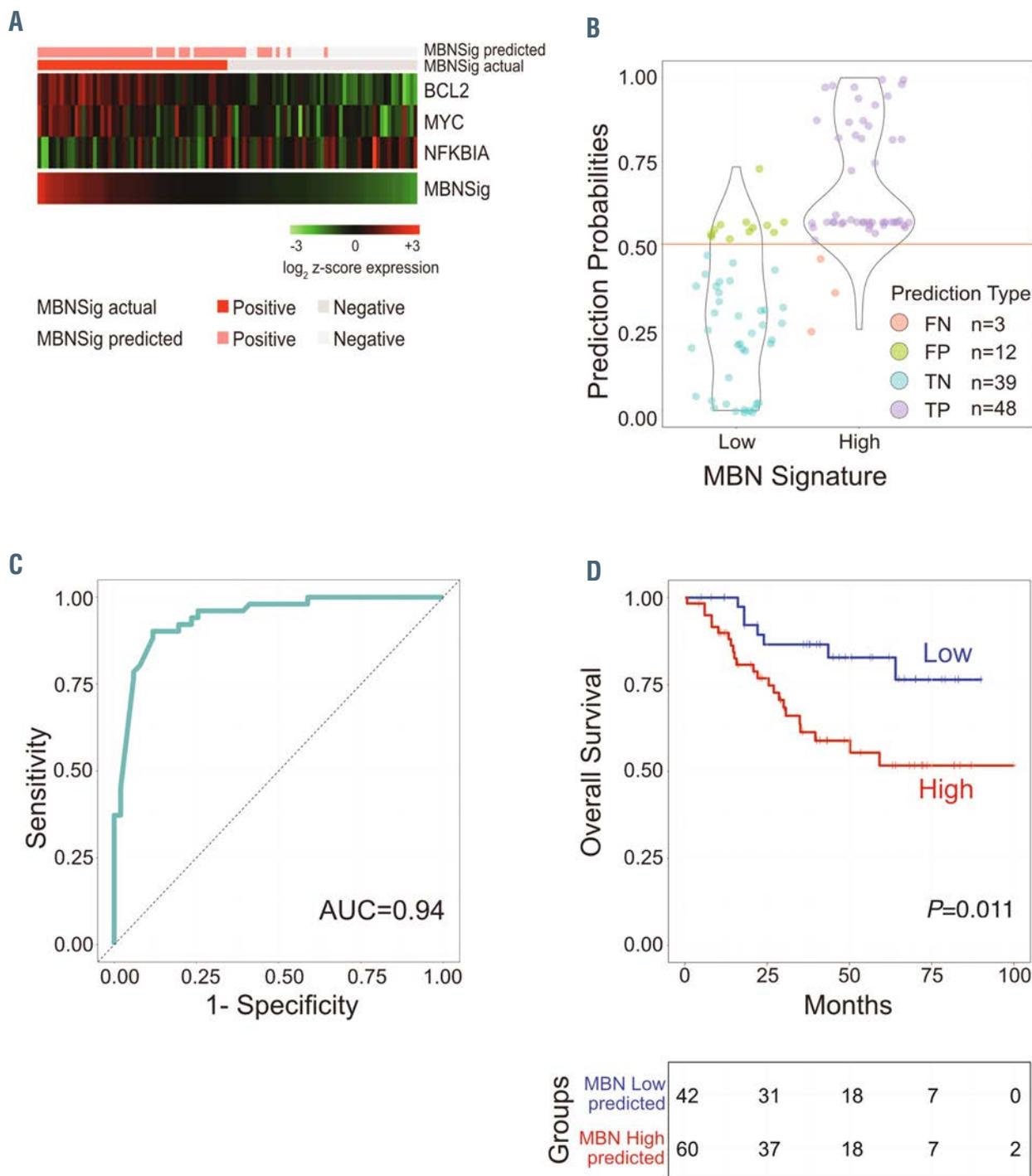
Figure 4. Real-life applicability of the MBN signature. (A) Heatmap representing the three informative genes of the MBN signature (MBN-Sig) shown as rows and diffuse large B-cell lymphoma (DLBCL) tissue samples shown as columns in the real-life cohort of 102 patients, with the actual MBN-Sig and the predicted MBN-Sig 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 three-gene RF model. (C) ROC curve of the real-life cohort using RF classifier. (D) Overall survival (OS) curve of the real-life cohort (n=102) based on the predicted MBN-Sig class. *P*-value was calculated with the log rank test.
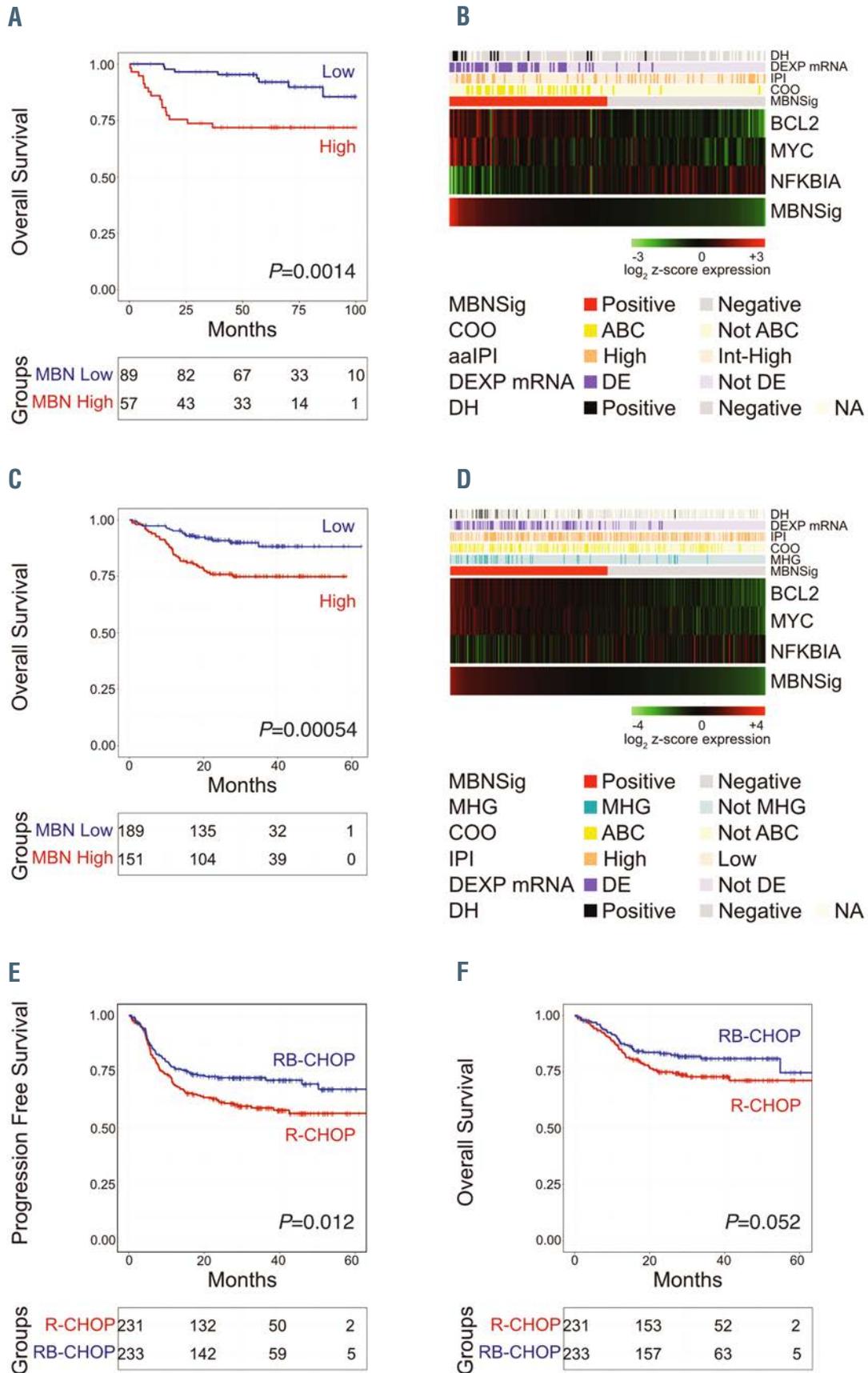
Figure 5. Legend on following page.

(*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 two 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) (*Online Supplementary 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 three main subgroups (a low risk non-DEXP_mRNA GCB/U subset and two 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 three-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 two subgroups characterized by different outcome (Figure 3), and was then applied to three 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 three 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 the *Online Supplementary 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 an RF 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 two risk categories whithin the GCB/U subset. These findings, which were validated in independent cohorts, could have immediate implications (Figure 5A and C; *Online Supplementary Figure S6A to 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-BCL with DH/TH and a further group lacking *MYC/BCL-2* re-arrangements 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-BCL with DH/TH) in the discovery cohort and Sha's cohort (Figure 5B and D; *Online Supplementary Figure S7A and 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; *Online Supplementary Figure S7B*). Taken together, these data indicate that a simple there-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 and 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 III 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 three-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.

Disclosures

ACe and NC are employees and shareholders of NanoString technology; ED has received research funding from TG-Therapeutics, ADC-Therapeutics, Takeda and sits on the Advisory Board for Gilead; ES has received support from Novartis and Eusapharma for educational events; ACh sit on the Advisory Boards with Celgene, Gilead-Kite, Janssen, Iqone, Takeda and has received honoraria for lectures from Celgene, Gilead-Kite, Janssen, Roche, Servier; UV has a consulting or advisory role for Celgene, Gilead and Janssen and is part of the speakers' bureau with Roche, Celgene, Janssen, Gilead Sciences, Abbvie, Sandoz; AR sits on the National or International Advisory Boards for Gilead, Amgen, Novartis, Pfizer, Celgene, Italfarmaco, Sanofi-Aventis, Astellas, Roche, Omeros and has sponsored symposia for Amgen, Novartis, Celgene, Roche; PC has received honoraria for Advisory Board participation or as a lecturer from AbbVie, Amgen, Celgene, Daiichi Sankyo, Gilead, Incyte, Janssen, Kite, KiowaKirin, Novartis, Roche, Sanofi, Servier, Takeda; PLZ has received honoraria for speakers' bureau or Advisory Boards for Verastem, Celltrion, Gilead, Janssen-Cilag, BMS, Servier, Sandoz, MSD, Immune Design, Celgene, Portola, Roche, Eusapharma, Kyowa Kirin, Sanofi; CT sit on the Advisory Board for ADC-Therapeutics; SP sits on the Advisory Boards for Celgene, NanoString, Roche; SM, FM, GM, MF, RB, CA, CC, SR, ACa, SF, VT, ACab, GP, AMG have no conflicts of interest to disclose.

Contributions

ED and SP designed the study, interpreted the data and wrote the manuscrip; SM and MF performed bioinformatics and statistical analyes, and SM helped with manuscript writing; FM and GM performed T-GEP experiments; VT, SF, CA and ACa performed immunohistochemistry; SP, ES, VT, SF and CA evaluated immunohistochemistry data; CC and SR performed FISH analyses; ACe and NC helped designing T-GEP experiments and helped with data interpretation; RB helped with data collection; ACab, GP, CT, AMG, PLZ, AR, PC, UV and ACh helped with data collection and interpretation. All authors critically reviewed the draft and approved the manuscript.

Aknowledgments

Funding

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 (R2 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. taiger 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 4th 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-κB 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.