haematologica
Journal of The Ferrata Storti Foundation

Prognostic impact of somatic mutations in diffuse large B-cell lymphoma and relationship to cell-of-origin: data from the phase III GOYA study

by Christopher R. Bolen, Magdalena Klanova, Marek Trneny, Laurie H. Sehn, Jie He, Jing Tong, Joseph N. Paulson, Eugene Kim, Umberto Vitolo, Alice Di Rocco, Günter Fingerle-Rowson, Tina Nielsen, Georg Lenz, and Mikkel Z. Oestergaard

Haematologica 2019 [Epub ahead of print]

Citation: Christopher R. Bolen, Magdalena Klanova, Marek Trneny, Laurie H. Sehn, Jie He, Jing Tong, Joseph N. Paulson, Eugene Kim, Umberto Vitolo, Alice Di Rocco, Günter Fingerle-Rowson, Tina Nielsen, Georg Lenz, and Mikkel Z. Oestergaard. Prognostic impact of somatic mutations in diffuse large B-cell lymphoma and relationship to cell-of-origin: data from the phase III GOYA study. doi:10.3324/haematol.2018.227892

Prognostic impact of somatic mutations in diffuse large B-cell lymphoma and relationship to cell-of-origin: data from the phase III GOYA study

Christopher R. Bolen,[1,*] Magdalena Klanova,[2-4,*] Marek Trneny,[2] Laurie H. Sehn,[5] Jie He,[6] Jing Tong,[6] Joseph N. Paulson,[7] Eugene Kim,[7] Umberto Vitolo,[8] Alice Di Rocco,[9] Günter Fingerle-Rowson,[4] Tina Nielsen,[4] Georg Lenz,[10] and Mikkel Z. Oestergaard[11]

[1]Bioinformatics & Computational Biology, Genentech Inc., South San Francisco, CA, USA

[2]1st Department of Medicine, Charles University General Hospital, Prague, Czech Republic

[3]Institute of Pathological Physiology, 1st Faculty of Medicine, Charles University, Prague, Czech Republic

[4]Pharma Development Clinical Oncology, F. Hoffmann-La Roche Ltd, Basel, Switzerland

[5]British Columbia Cancer Centre for Lymphoid Cancer, Vancouver, Canada

[6]Foundation Medicine, Inc., Cambridge, MA, USA

[7]Department of Biostatistics, Product Development, Genentech Inc., South San Francisco, CA, USA

[8]A.O. Universitaria Città della Salute e della Scienza di Torino, Ematologia, Torino, Italy

[9]Department of Cellular Biotechnologies and Hematology, Sapienza University, Rome, Italy

[10]Department of Medicine A, Hematology, Oncology and Pneumology, University

Hospital Münster, Münster, Germany

[11]Oncology Biomarker Development, F. Hoffmann-La Roche Ltd, Basel, Switzerland

**CRB and MK contributed equally to this study.*

Running title: Somatic mutations in diffuse large B-cell lymphoma

Correspondence: Bioinformatics & Computational Biology, Genentech Inc., One DNA Way, MS 444A, South San Francisco, CA 94080, USA; phone: +1 650-467-1878; e-mail: bolen.christopher@gene.com

Abstract word count: 213/250 max

Word count: 3925/4000 max

Figures/tables: 7/8 max

Supplemental files: 1

Acknowledgments

The authors would like to thank the GOYA study team investigators, coordinators, nurses and patients. GOYA was supported by F. Hoffmann-La Roche Ltd, with scientific support from the Fondazione Italiana Linfomi. Editorial support was provided by Louise Profit, PhD (Gardiner-Caldwell Communications Ltd, Macclesfield, UK), and was funded by F. Hoffmann-La Roche Ltd.

References: 28

Article topic: non-Hodgkin lymphoma

Key words: Diffuse large B-cell lymphoma (DLBCL), obinutuzumab, prognostic biomarkers, next-generation sequencing (NGS)

ABSTRACT

Diffuse large B-cell lymphoma represents a biologically and clinically heterogeneous diagnostic category with well-defined cell-of-origin subtypes. Using data from the GOYA study (NCT01287741), we characterized the mutational profile of diffuse large B-cell lymphoma and evaluated the prognostic impact of somatic mutations in relation to cell-of-origin. Targeted DNA next-generation sequencing was performed in 499 formalin-fixed paraffin-embedded tissue biopsies from previously untreated patients. Prevalence of genetic alterations/mutations was examined. Multivariate Cox regression was used to evaluate the prognostic effect of individual genomic alterations. Of 465 genes analyzed, 59 were identified with mutations occurring in at least 10 of 499 patients (≥2% prevalence); 334 additional genes had mutations occurring in ≥1 patient. Single nucleotide variants were the most common mutation type. On multivariate analysis, *BCL2* alterations were most strongly associated with shorter progression-free survival (multivariate hazard ratio: 2.6; 95% confidence interval: 1.6 to 4.2). *BCL2* alterations were detected in 102 of 499 patients; 92 had *BCL2* translocations, 90% of whom had germinal center B-cell-like diffuse large B-cell lymphoma. *BCL2* alterations were also significantly correlated with *BCL2* gene and protein expression levels. Validation of published mutational subsets revealed consistent patterns of co-occurrence, but no consistent prognostic differences between subsets. Our data confirm the molecular heterogeneity of diffuse large B-cell lymphoma, with potential treatment targets occurring in distinct cell-of-origin subtypes. **clinicaltrials.gov identifier:** NCT01287741.

Introduction

Diffuse large B-cell lymphoma (DLBCL) represents a biologically and clinically heterogeneous diagnostic category. Distinct DLBCL cell-of-origin (COO) subtypes, arising from different stages of normal B-cell development and with different prognostic outcomes, were identified almost 2 decades ago.[1-3] Several studies have since described the landscape of recurrent somatic mutations in DLBCL and demonstrated the molecular uniqueness of the distinct COO subtypes, and recent studies have suggested clinically relevant genetic subgroups exist within each subtype.[4-9] While germinal center B-cell-like (GCB) DLBCL is characterized by frequent translocations of the *BCL2* gene, a key regulator of the intrinsic apoptotic pathway, or mutations of the epigenetic modifiers, *CREBBP* and *EZH2*, these abnormalities are rare in activated B-cell-like (ABC) DLBCL.[10] In contrast, mutations in genes encoding proteins implicated in B-cell receptor signaling and the NFκB pathway, such as *CD79b or MYD88*, or genes involved in regulation of the cell cycle such as *CDKN2A,* contribute to the molecular pathogenesis of ABC DLBCL.[11-14]

While the prognostic impact of the distinct COO subtypes has been confirmed in several studies,[2,3,15,16] the influence of key genomic alterations on the clinical outcomes of DLBCL patients is less clear, particularly their added clinical prognostic value over the International Prognostic Index (IPI) and COO. Mutations of several genes, such as *TP53*, *MYD88* or *CDKN2A,* have been shown to be associated with poor prognosis in DLBCL patients.[11,17-19] Many of these alterations, such as loss of *CDKN2A* or mutations of *MYD88,* are significantly enriched within the prognostically inferior ABC subtype and their independent prognostic role needs to be confirmed.

A recent observational study by Reddy *et al.*[19] retrospectively explored 150 genetic drivers of DLBCL in 1001 patients and developed a genomic risk model

comprising genetic alterations, COO DLBCL subtype, IPI score and dual *MYC* and *BCL2* expression, which had greater prognostic ability for overall survival than molecular or clinical factors (COO, *MYC/BCL2* expression, IPI) alone.[19] Additionally, the studies by Schmitz and colleagues (2018) and Chapuy *et al.* (2018) helped elucidate some of the reported clinical and genetic heterogeneity in transcriptionally defined COO subsets of front-line DLBCL.[8,9] Using a set of common genetic alterations, both studies identified distinct molecular subtypes and evaluated their clinical prognostic outcome. Both studies identified a number of common mutational profiles, including two distinct subsets of ABCs—one enriched for mutations in *MYD88* and *CD79B*, and another for *BCL6* and *NOTCH* mutations—and a GCB subset enriched for *BCL2* translocations and mutations in *CREBBP* and *EZH2*. Importantly, these clusters had distinct prognostic profiles, many reflecting the established prognostic impact of the dominant mutations in each group (e.g. worse prognosis for the *BCL2* and *MYD88* subsets).[9]

Here, we perform an integrated analysis to evaluate if somatic mutations in DLBCL provide clinical prognostic value over established clinical and biological risk factors, including COO and IPI. Using data from the phase III GOYA study, the largest (n=1418) randomized clinical trial in patients with previously untreated DLBCL to date, we analyzed the mutational profile of DLBCL using a well-established, highly validated targeted next-generation sequencing (NGS) platform, and evaluated the prognostic impact of somatic mutations and their relationship with COO. A previous exploratory analysis in the GOYA study showed that patients with GCB DLBCL achieved a better outcome, in terms of progression-free survival (PFS), than those with the ABC subtype, irrespective of treatment.[3]

Methods

Patient treatment and assessments

The GOYA study design has previously been described.[3] Included patients had previously untreated, histologically documented, CD20-positive DLBCL; inclusion criteria are further described in the *Online Supplementary Methods*.

The study was conducted in accordance with the European Clinical Trial Directive (for European centers), Declaration of Helsinki and the International Conference on Harmonisation guidelines for Good Clinical Practice. The protocol was approved by the ethics committees of participating centers and registered at ClinicalTrials.gov (NCT01287741). All patients provided written informed consent.

Staging investigations included computed tomography scanning and bone marrow biopsy. Tumor response and progression were assessed by the investigator using regular clinical and laboratory examinations and computed tomography scans. Response was evaluated according to the Revised Response Criteria for Malignant Lymphoma[20] 4–8 weeks after last study treatment, or at early discontinuation.

COO analysis

COO classification was based on gene expression profiling using the NanoString Lymphoma Subtyping Research-Use-Only assay (NanoString Technologies, Inc., Seattle, WA). COO data were available in 933 patients (reasons for non-availability: restricted Chinese export license [n=252], *CD20*+ DLBCL not confirmed by central pathology [n=102] and missing/inadequate tissue [n=131]).

Immunohistochemical analyses

Pretreatment tumor samples were analyzed by a central laboratory using the Ventana BCL2 (124) and MYC (Y69) investigational use only immunohistochemical assays. The prespecified scoring algorithm incorporated percentage of tumor cells stained and their intensity: BCL2 immunohistochemistry-positive was defined as moderate/strong cytoplasmic staining in ≥50% of tumor cells and MYC immunohistochemistry-positive was defined as nuclear staining at any intensity in ≥40% of tumor cells.

Targeted NGS

Genomic DNA was extracted from diagnostic formalin-fixed, paraffin-embedded tissue sections containing ≥20% tumor cells. Samples were submitted to a central laboratory for NGS-based genomic profiling and processed as previously described.[21,22] Adaptor-ligated DNA underwent hybrid capture for all coding exons of 465 cancer-related genes (FoundationOne Heme™ platform, Foundation Medicine Incorporated [FMI], MA) (*Online Supplementary Methods*). NGS data were available for 499 of the 1418 patients included in the intent-to-treat (ITT) population of the GOYA study; both NGS and COO were available in 482 patients. Information about known drug targets and ongoing clinical trials targeting individual mutations was queried on March 23, 2018 through an FMI internal database populated using data from clinicaltrials.gov and other publicly available sources.

Validation of mutational models

We sought to confirm the prognostic value of the mutational genomic risk model generated by Reddy *et al.*,[19] Chapuy *et al.*[9] and Schmitz *et al.*,[8] as described in the *Online Supplementary Methods*.

Statistical analysis

Only genetic alterations with known somatic and functional status were included in the statistical analysis.[21] Univariate and multivariate Cox regression analyses were used to evaluate the prognostic effect of a genetic alteration if there were ≥10 progression events in mutated patients or ≥40 patients in total with the mutation. Multivariate Cox regression analysis was performed to control for COO, IPI, treatment arm, number of planned chemotherapy cycles and geographic region. Multiple testing adjustment was done by estimating false discovery rates (FDRs) using the Benjamini-Hochberg procedure (significance <5% FDR).

Results

Baseline disease characteristics were similar between patients with NGS available and the overall GOYA ITT population, except for race (*Online Supplementary Table S1*) and geographic region (data not shown), due to lack of access to samples from China.

Genomic alterations detectable by targeted NGS

Of 465 sequenced genes, 59 (13%) were identified as functionally altered (i.e. having mutations that significantly alter the function of a gene in a manner that has been previously reported to drive cancer progression) in at least 10 of 499 patient samples (≥2% prevalence), and 334 additional genes with alterations were identified in ≥1 patient; 3% of patients had no identified mutation. The median number of gene alterations per patient was 6 (range 0-17). The median number of single nucleotide variants (SNVs) and copy number abnormalities (CNAs) per patient were 4 (range,

0-16) and 0 (range, 0-10), respectively. Ninety-seven percent of cases harbored ≥1 alteration and 93% of cases harbored multiple (≥2) alterations.

The most frequently (≥2% cases) observed gene alterations (SNVs, amplifications and deletions) are shown in Figure 1A. SNVs were the most common mutation type, while CNAs were specific to a few genes, including *CDKN2A/B* and *REL*. Of the 31 analyzed gene rearrangements, *BCL2*, *MYC* and *BCL6* were the most frequently rearranged; for these genes, the most frequently observed translocation partner was the immunoglobulin heavy chain locus, found in 92/92 (100%), 29/32 (90.6%) and 57/100 (57.0%) cases where the rearrangement partner could be determined, respectively (*Online Supplementary Table S2*).

Frequencies of genomic alterations among COO subsets

Of the patients for whom both COO and NGS were available (n=482), 272 (56%), 78 (16%) and 132 (27%) were classified as GCB, unclassified and ABC DLBCL, respectively (*Online Supplementary Table S1*). This was similar to findings for the overall COO population (n=933; GCB, n=540 [58%]; unclassified, n=150 [16%]; ABC, n=243 [26%]). Within the GCB subtype, the most prevalent mutated genes were *BCL2* (88/272 [32%]), MLL2 (*KMT2D)* (82/272 [30%]) and *CREBBP* (60/272 [22%]); loss of *CDKN2A* (64/132 [49%]) and *CDKN2B* (40/132 [30%]) and mutations of *MYD88* (45/132 [34%]) were most frequently observed in the ABC subtype (Table 1 and *Online Supplementary Table S3*). Fifteen genes were found to be significantly differentially mutated between the GCB and ABC subtypes at FDR <0.05 (Figure 1B). Alterations of *BCL2*, *CREBBP*, *TNFRSF14*, *EZH2*, *REL, BCL7A* and *SGK1* were more frequently observed in GCB DLBCL whereas *BCOR*, *ETV6*, *PRDM1*, *PIM1*, *CD79b*, *CDKN2B*, *MYD88* and *CDKN2A* were more frequently

mutated in ABC DLBCL (Figure 1B). In the case of *BCL2* and *CDKN2A*, specific types of alterations displayed different frequencies between the GCB and ABC subtypes (Figure 1C). While *BCL2* translocations and SNVs were more frequently found in the GCB subtype, high-level *BCL2* amplifications (≥6 copies) were enriched within the ABC subtype (ABC, 9/132 [6.8%]; GCB, 4/272 [1.5%]; Fisher's exact test *P*=0.012). An analysis of low-level BCL2 amplifications (≥1 copy above median ploidy and ≥3 copies) confirmed the enrichment in ABC DLBCL samples (ABC, 83/132 [62.9%]; GCB, 45/272 [16.5%]; Fisher's exact test *P*<0.001). The enrichment of *CDKN2A* alterations within the ABC subtype was pronounced only for *CDKN2A* deletions; SNVs occurred to the same degree in all COO subtypes.

Correlation of individual alterations with clinical outcomes

Alterations of 23 genes (fulfilling the predefined criteria based on their prevalence in the analyzed cohort) were evaluated for association with PFS on univariate and multivariate analyses. Prognostic trends were observed among a number of previously studied biomarkers, including *BCL2*, *CREBBP*, *REL*, *TP53* and *CDKN2A* (all *P*<0.05 [unadjusted]). However, alterations of *BCL2* (including translocations, SNVs and high-level amplifications) were the most strongly associated with PFS (hazard ratio [HR]: 2.6; 95% confidence interval [CI]: 1.6 to 4.2; FDR, 0.0037) independent of COO, IPI, treatment arm, number of planned chemotherapy cycles and geographic region (Table 2). None of the 23 biomarkers showed significant differences in prognostic impact between treatment arms. The *BCL2* prognostic effect was observed for both *BCL2* SNVs (HR: 2.6; 95% CI: 1.5 to 4.7; FDR, 0.022) and translocations (HR: 2.5; 95% CI: 1.4 to 4.2; FDR, 0.0028) (Table 2 and Figure 2) individually. The prognostic role of high-level *BCL2*

amplification was not tested separately due to the low prevalence of this alteration in the current study. No association was found between survival and low-level *BCL2* amplifications (HR: 1.2; 95% CI: 0.8 to 1.9; FDR, 0.58). *BCL2* alterations were detected in 20% (102/499) of patients, with 92 of 102 patients having a *BCL2* translocation, 90% (83/92) of whom were GCB patients, with only one translocated ABC patient. Of 39 patients with *BCL2* SNVs, 80% (31/39) and 15% (6/39) were in the GCB and ABC subgroups, respectively. The majority of patients with *BCL2* SNVs harbored *BCL2* translocations (74% [29/39]) (Figure 3), but *BCL2* SNVs were still associated with worse prognosis among patients without a *BCL2* translocation (HR: 2.8; 95% CI: 1.0 to 7.9; *P*=0.047). *BCL2* mutations were also significantly correlated with *BCL2* gene and protein expression levels (*Online Supplementary Figure S1*).

Alterations of *CREBBP* (HR: 2.1; 95% CI: 1.3 to 3.4; FDR, 0.054) and *TP53* (HR: 1.6; 95% CI: 1.1 to 2.5; FDR, 0.22) were also associated with PFS on multivariate analysis, but did not fulfill the predefined criteria for significance (FDR <0.05). Alterations of *CREBBP* were detected in 15% (73/499) of patients; 82% (60/73), 8% (6/73) and 7% (5/73) of whom belonged to the GCB, unclassified and ABC subtypes, respectively. Four of the 73 patients harbored two different *CREBBP* mutations. In the majority of cases, *CREBBP* alterations were SNVs (97% [71/73]), with only two cases of *CREBBP* deletion. Alterations of *TP53* were found in 18% (92/499) of patients, of whom 58% (53/92), 15% (14/92) and 22% (20/92) had the GCB, unclassified and ABC DLBCL subtype, respectively. Overall, 105 *TP53* alterations were observed in 92 patients, with 13/92 patients harboring two simultaneous *TP53* mutations. SNVs were the most frequently observed *TP53* alterations (98% [103/105]), while *TP53* deletions and rearrangements were observed in two cases, and one case, respectively.

CDKN2A alterations were associated with shorter PFS on univariate analysis (HR: 1.7; 95% CI: 1.2 to 2.5; FDR, 0.13). This effect was driven by *CDKN2A* deletions (HR: 1.6; 95% CI: 1.1 to 2.4; FDR, 0.058). No significant association with PFS was observed on multivariate analysis for all *CDKN2A* alterations, or for *CDKN2A* deletions only (Table 2). *CDKN2A* alterations were observed in 23% (113/499) of DLBCL patients. Of all cases with any *CDKN2A* alteration, 25% (28/113), 15% (17/113) and 57% (64/113) belonged to the GCB, unclassified and ABC subtypes, respectively. The majority of the *CDKN2A* alterations were homozygous gene deletions, which were enriched within the ABC subtype. Patients with *CDKN2A* deletions had adverse clinical disease characteristics (IPI, extranodal sites, age, and serum LDH) compared with patients without a *CDKN2A* deletion, both in the total FMI evaluable patients and among the ABC subtype (*Online Supplementary Table S4*).

In a survival analysis according to COO subtype, *BCL2* translocations (HR: 2.3; 95% CI: 1.3 to 4.2; *P*=0.0049; FDR, 0.017) were significantly associated with shorter PFS independent of clinical factors in the GCB subtype, while none of the identified genetic alterations were significantly prognostic within the ABC subtype (*Online Supplementary Table S5*).

Correlation of combined genomic risk model with clinical outcomes

We evaluated the performance of a combined genomic risk model for predicting clinical outcomes using a single comprehensive NGS assay. When applying a modified mutational model generated by Reddy *et al.*,[19] the risk scores ranged from -3 to 7, with most patients centered at 0 (Figure 4A). Low-risk was defined by a score <0 (n=112), low-intermediate-risk with a score 0 (n=215), high-intermediate-risk

patients had a score >0 and <3 (n=107) and high-risk had a score ≥3 (n=29). This genomic scoring system provided clear separation between the low/low-intermediate and high/high-intermediate groups (Figure 4B). Using a simple dichotomization of the score into low- and high-risk subgroups, the overall univariate HR for the prognostic score was 0.61; 95% CI: 0.42 to 0.88; *P*=0.0087. The risk groups were highly correlated with COO subtypes, and after correcting for COO, the model was no longer significant in the entire cohort (HR: 0.77; 95% CI: 0.49 to 1.2; *P*=0.27). When tested within COO subtypes, no significant prognostic signal was found, although there was a trend for added prognostic information among the GCB subset (HR: 0.5; 95% CI: 0.24 to 1.04; *P*=0.06) but not the ABC subset (HR: 1.2; 95% CI: 0.66 to 2.32; *P*=0.5).

Validation of new molecular classifications

Although no publicly available tool exists for classifying samples into molecular subtypes as defined by Schmitz *et al.*[8] and Chapuy *et al.*,[9] we sought to validate these classifications using an approximation of their clusters. For Schmitz *et al.*, we approximated the EZB, BN2, N1 and MCD clusters using each cluster's founder alterations (*EZH2* or *BCL2*; *BCL6* or *NOTCH2*; *NOTCH1*; and *MYD88*, *L265P* or *CD79B*, respectively; see Methods). Prevalence of these four clusters was consistent with those reported by Schmitz *et al.* (Figure 5A); however, we observed no difference in prognosis among any of the four mutational subgroups (log-rank *P*=0.94), although the mutational subsets did perform worse than the unclassified "other GCB" subset (pooled mutational clusters *vs.* other GCB *P*=0.021; EZB *vs.* other GCB *P*=0.023; Figure 5B).

To recreate the Chapuy classifications, we applied the non-negative matrix factorization (NMF) clustering algorithm to the set of mutations overlapping with those reported by Chapuy *et al.* This resulted in five clusters (plus an unmutated cluster – labeled C0) sharing very similar mutational profiles and distribution of COO subsets with Chapuy *et al.*'s clusters (Figure 5C and *Online Supplementary Figure S2*), with the notable exception that *CDKN2A/2B* (9p21) deletions significantly co-occurred with *MYD88* and *CD79B* alterations, rather than with *TP53* alterations as observed in Chapuy *et al.* We observed similar prognostic trends among these subsets, with our clusters G2, G3 and G5 (equivalent to Chapuy C2, C3 and C5) showing significantly worse prognosis when compared with clusters G0, G1 and G4 (Chapuy C0, C1 and C4, respectively) (Figure 5D; HR: 1.8; 95% CI: 1.2 to 2.6; *P*=0.0033).

Discussion

In the current study, we analyzed the mutational profile and prognostic impact of genomic alterations in newly diagnosed DLBCL patients who were uniformly treated with anti-CD20-based immunochemotherapy (obinutuzumab or rituximab plus cyclophosphamide, doxorubicin, vincristine and prednisone [G-/R-CHOP]) in the phase III GOYA trial. Using a well-established, highly validated targeted NGS platform, we analyzed SNVs and CNAs in 465 cancer-related genes and 31 select gene rearrangements in 499 patients, so far the largest prospectively collected dataset in DLBCL. These data serve as a valuable resource for understanding the clinical relevance of mutations as measured by this platform.

Alteration of the *BCL2* gene was the only genetic abnormality significantly associated with shorter PFS independently of molecular or clinical factors (treatment

arm, COO, IPI, number of planned chemotherapy cycles and geographic region). This effect was observed for both *BCL2* translocations and SNVs. The co-occurrence of *BCL2* SNVs with *BCL2* translocations, possibly as a consequence of aberrant somatic hypermutation,[23] may partially explain the negative prognostic impact of *BCL2* SNVs, although the negative prognostic effect of BCL2 SNVs among patients without BCL2 translocations may point to an independent biological role for these alterations. *BCL2* translocations were significantly enriched within the GCB subtype and were associated with shorter PFS within this subtype. *BCL2* translocations were associated with high levels of BCL2 mRNA and protein expression, both of which have been shown to be associated with an adverse prognosis in DLBCL independent of COO and IPI, including in the GOYA study.[24] Our data suggest that pharmacological inhibition of the BCL2 protein could be a promising treatment strategy in a subset of DLBCL patients. Venetoclax, a highly specific BCL2 inhibitor,[25] is currently being tested in clinical trials in patients with newly diagnosed DLBCL; however, the subpopulation of DLBCL patients who could benefit from venetoclax needs to be defined.

Given the molecular uniqueness and prognostic value of the particular COO subtypes, we aimed to analyze the prognostic impact of genetic alterations within these subtypes. The only genetic alteration significantly associated with shorter PFS within the GCB subtype was *BCL2* translocation. None of the tested genetic alterations were significantly associated with outcome within the prognostically-inferior ABC subtype, supporting the strong prognostic significance of COO assessed by gene expression profiling.

In this study, we observed prognostic trends in several genes, including *TP53*, *CREBBP* and *CDKN2A*, but none met our thresholds for significance. There are

several potential explanations for this observation. First, in the current study we used robust statistical methods with strict predefined criteria for significance to test the association of particular gene alterations with clinical outcomes. Second, only truncating/frameshift mutations and previously reported loss-of-function mutations were included in this study. Alteration of several genes, such as *CREBBP* and *TP53*, were associated with shorter PFS in our study, in the absence of multiple testing correction.

When validating the genomic risk model from Reddy *et al.*,[19] although the model was prognostic in our population when stratified into high- and low-risk groups (HR: 0.61; 95% CI: 0.42 to 0.88; $P<0.01$), when corrected for COO, the model was no longer significant (HR: 0.77; 95% CI: 0.49 to 1.2; $P=0.27$), indicating that it provided little additional benefit over the most commonly used gene expression profiling and FISH assays, and that COO evaluation in combination with *BCL2* and *MYC* translocation status may be a simpler approach with similar overall prognostic relevance, although other genomic features such as TP53 or CREBBP may provide additional information that is worth considering. However, it should be noted that we were unable to apply the Reddy *et al.* model in its entirety due to some differences in gene availability on the FMI platform, and that Reddy *et al.* evaluated the model in terms of overall survival, whereas our study evaluated it in terms of PFS.

The current study also demonstrated the molecular heterogeneity of DLBCL, with the majority of the observed genetic alterations shared by COO subtypes; however, the frequency of mutations in 15 genes was enriched between GCB and ABC subtypes. In addition, approximating the molecular clusters described by Schmitz *et al.* and Chapuy *et al.* revealed a consistent set of molecular subgroups, with some specific to either GCB (EZB-like, G3), ABC (MCD- or N1-like, G5) or

Unclassified (BN2-like) COO subtypes, and others appearing to be independent of the tumor cell of origin. Among the clusters defined by NMF, we observed a significantly worse prognosis for clusters G2, G3 and G5, consistent with Chapuy's C2, C3 and C5 clusters. This is most likely driven by the enrichment of individual prognostic alterations among these subgroups (*BCL2* and *CREBBP* in G3; *TP53* and *REL* in G2), or by enrichment for the ABC subset (G5). By contrast, our approximation of the Schmitz clusters identified four sets of clusters with approximately equivalent prognosis, suggesting that the founder alterations used to define these clusters are not sufficient to identify patients with worse prognosis. Although we cannot directly recapitulate the clusters defined by Schmitz and Chapuy, both due to limitations of the FMI panel and because algorithms for classifying DLBCL samples are not publicly available, our results here show that we can successfully capture the molecular heterogeneity of DLBCL using this targeted mutational panel.

Since 2011, several studies have characterized the landscape of somatic mutations in DLBCL by whole exome NGS technologies[5-7,26] or FMI's targeted exome-sequencing platform,[4] and have identified recurrent genetic alterations. Our study identified a relatively lower number of genetic alterations compared with whole-exome studies, but it was relatively consistent with the frequencies of mutations identified by Intlekofer *et al.*[4] This is most likely because both our study and the study by Intlekofer *et al.* focused on mutations with known or likely somatic and functional status. FMI may also lack some alterations of potential relevance in DLBCL, including alterations in the human leukocyte antigen genes, potentially limiting the scope of this analysis. In contrast, the relatively low prevalence of *MYC* translocations in this dataset may be reflective of an accrual bias during patient

recruitment. Patients with these alterations, particularly in combination with *BCL2* translocations (double-hit lymphoma) have been well characterized as having particularly aggressive disease and are generally more difficult to recruit for clinical trials. These patients may also benefit from more aggressive chemotherapy than G-/R-CHOP, which could also explain why these patients were not enrolled in GOYA.

Our data show that DLBCL contains mutations in a variety of potentially targetable pathways. In total, a majority (59%) of patients harbor ≥1 alteration in genes that would be eligible for potential targeted therapies approved in other indications (e.g. venetoclax for *BCL2* translocations/amplifications, everolimus for *PTEN* loss, and ruxolitinib and tofacitinib for *JAK2* mutations) and over 70% of patients would potentially qualify to be enrolled in ongoing clinical trials based on genomic information, according to FMI's clinical trial database. Genes enriched between GCB and ABC subtypes also included previously reported driver mutations and gene alterations that can be targeted by novel therapies, such as the gain of function mutation of *EZH2* in the GCB DLBCL subtype,[27] and the BCL2 translocations and amplifications.[28] These mutations, along with COO subtype information, would be useful for the design of clinical trials involving combinations of novel targeted therapies.

In conclusion, using the largest prospective dataset in previously untreated DLBCL to date, we demonstrated the molecular heterogeneity of DLBCL, with potential treatment targets harbored by the distinct COO subtypes. Only alterations in *BCL2* were significantly associated with clinical outcome independently of COO and clinical factors, thereby demonstrating the strong prognostic value of COO for clinical outcome in DLBCL.

Data sharing

Qualified researchers may request access to individual patient level data through the clinical study data request platform (www.clinicalstudydatarequest.com). Further details on Roche's criteria for eligible studies are available here (https://clinicalstudydatarequest.com/Study-Sponsors/Study-Sponsors-Roche.aspx). For further details on Roche's Global Policy on the Sharing of Clinical Information and how to request access to related clinical study documents, see here (https://www.roche.com/research_and_development/who_we_are_how_we_work/clinical_trials/our_commitment_to_data_sharing.htm).

Authorship

Contribution:

CRB, MZO and MK contributed to the design of the research, performed the research and the analysis and interpretation of the data; GFR and TN contributed to the design of the research and the analysis and interpretation of the data; MT, LHS, EK, JNP, UV, ADiR and GL contributed to the analysis and interpretation of the data; CB performed the statistical analysis; MK wrote the manuscript; all authors reviewed the manuscript, approved for submission, and take responsibility for the content.

Conflict-of-interest disclosure

CRB, JNP and EK report employment and equity ownership for Genentech. MK reports employment for F. Hoffmann-La Roche Ltd. MT reports consultancy and honoraria for F. Hoffmann-La Roche Ltd, Celgene, Janssen, AbbVie, BMS, Takeda and Gilead, and research funding for F. Hoffmann-La Roche Ltd and Celgene. LHS reports consultancy and honoraria for AbbVie, Amgen, Celgene, F. Hoffmann-La

Roche Ltd/Genentech, Janssen and Seattle Genetics. UV reports honoraria for Celgene, Janssen, Gilead, F. Hoffmann-La Roche Ltd, Sandoz and Takeda, advisory board membership for Celgene, Janssen and F. Hoffmann-La Roche Ltd, and research funding for F. Hoffmann-La Roche Ltd. TN, GFR and MZO report employment and equity ownership in F. Hoffmann-La Roche Ltd. GL reports consultancy and honoraria for Amgen, Bayer, BMS, Celgene, Gilead, F. Hoffmann-La Roche Ltd, Janssen, Morphosys, NanoString, Novartis and Takeda, and research funding for Agios, AstraZeneca, Bayer, Celgene, F. Hoffmann-La Roche Ltd, Janssen and Novartis. MT reports consultancy for Celgene, F. Hoffmann-La Roche Ltd, Janssen, Sandoz, Takeda and Mundipharma, advisory board membership for F. Hoffmann-La Roche Ltd, Janssen and Sandoz, and speakers' bureaus for F. Hoffmann-La Roche Ltd and Takeda.

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. Rosenwald A, Wright G, Chan WC, et al. The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. N Engl J Med. 2002;346(25):1937-1947.
3. Vitolo U, Trneny M, Belada D, et al. Obinutuzumab or rituximab plus cyclophosphamide, doxorubicin, vincristine, and prednisone in previously untreated diffuse large B-cell lymphoma. J Clin Oncol. 2017:35(31):3529-3537.
4. Intlekofer AM, Joffe E, Batlevi CL, et al. Integrated DNA/RNA targeted genomic profiling of diffuse large B-cell lymphoma using a clinical assay. Blood Cancer J. 2018;8(6):60.
5. Morin RD, Mendez-Lago M, Mungall AJ, et al. Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. Nature. 2011;476(7360):298-303.
6. Pasqualucci L, Trifonov V, Fabbri G, et al. Analysis of the coding genome of diffuse large B-cell lymphoma. Nat Genet. 2011;43(9):830-837.
7. Zhang J, Grubor V, Love CL, et al. Genetic heterogeneity of diffuse large B-cell lymphoma. Proc Natl Acad Sci U S A. 2013;110(4):1398-1403.
8. Schmitz R, Wright GW, Huang, et al. Genetics and pathogenesis of diffuse large-B-cell lymphoma. N Engl J Med. 2018;378(15):1396-1407.
9. 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(5):679-690.

10. Lunning MA, Green MR. Mutation of chromatin modifiers; an emerging hallmark of germinal center B-cell lymphomas. Blood Cancer J. 2015;5:e361.
11. Jardin F, Jais JP, Molina TJ, et al. Diffuse large B-cell lymphomas with CDKN2A deletion have a distinct gene expression signature and a poor prognosis under R-CHOP treatment: a GELA study. Blood. 2010;116(7):1092-1104.
12. 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.
13. Davis RE, Ngo VN, Lenz G, et al. Chronic active B-cell-receptor signalling in diffuse large B-cell lymphoma. Nature. 2010;463(7277):88-92.
14. Young RM, Shaffer AL, Phelan JD, Staudt LM. B-cell receptor signaling in diffuse large B-cell lymphoma. Semin Hematol. 2015;52(2):77-85.
15. Lenz G, Wright G, Dave SS, et al. Stromal gene signatures in large-B-cell lymphomas. N Eng J Med. 2008;359(22):2313-2323.
16. 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.
17. Xu-Monette ZY, Wu L, Visco C, et al. Mutational profile and prognostic significance of TP53 in diffuse large B-cell lymphoma patients treated with R-CHOP: report from an International DLBCL Rituximab-CHOP Consortium Program Study. Blood. 2012;120(19):3986-3996.
18. Fernandez-Rodriguez C, Bellosillo B, Garcia-Garcia M, et al. MYD88 (L265P) mutation is an independent prognostic factor for outcome in patients with diffuse large B-cell lymphoma. Leukemia. 2014;28(10):2104-2106.

19. Reddy A, Zhang J, Davis NS, et al. Genetic and functional drivers of diffuse large B cell lymphoma. Cell. 2017;171(2):481-494.e15.
20. Cheson BD, Pfistner B, Juweid ME, et al. Revised response criteria for malignant lymphoma. J Clin Oncol. 2007;25(5):579-586.
21. Frampton GM, Fichtenholtz A, Otto GA, et al. Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing. Nat Biotechnol. 2013;31(11):1023-1031.
22. He J, Abdel-Wahab O, Nahas MK, et al. Integrated genomic DNA/RNA profiling of hematologic malignancies in the clinical setting. Blood. 2016;127(24):3004-3014.
23. Schuetz JM, Johnson NA, Morin RD, et al. BCL2 mutations in diffuse large B-cell lymphoma. Leukemia. 2012;26(6):1383-1390.
24. Sehn LH, Oestergaard MZ, Trněný M, et al. Prognostic impact of BCL2 and MYC expression and translocation in untreated DLBCL: results from the phase III GOYA study. Hematol Oncol. 2017;35(S2):131-133.
25. Souers AJ, Leverson JD, Boghaert ER, et al. ABT-199, a potent and selective BCL-2 inhibitor, achieves antitumor activity while sparing platelets. Nat Med. 2013;19(2):202-208.
26. Lohr JG, Stojanov P, Lawrence MS, et al. Discovery and prioritization of somatic mutations in diffuse large B-cell lymphoma (DLBCL) by whole-exome sequencing. Proc Natl Acad Sci U S A. 2012;109(10):3879-3884.
27. McCabe MT, Ott HM, Ganji G, et al. EZH2 inhibition as a therapeutic strategy for lymphoma with EZH2-activating mutations. Nature. 2012;492(7427):108-112.

28. Roberts AW, Huang D. Targeting BCL2 with BH3 mimetics: basic science and clinical application of venetoclax in chronic lymphocytic leukemia and related B cell malignancies. Clin Pharmacol Ther. 2017;101(1):89-98.

Table 1. Prevalence of most frequent* gene mutations according to DLBCL COO subtype.

	GCB, n=272 (%)	Unclassified, n=78 (%)	ABC, n=132 (%)
BCL2	32.4	5.1	4.5
KMT2D	30.1	21.8	28.8
CREBBP	22.1	7.7	3.8
TP53	19.5	17.9	15.2
BCL6	18.8	35.9	22.0
B2M	17.6	12.8	12.9
TNFRSF14	17.3	1.3	0.0
EZH2	16.2	6.4	0.8
TNFAIP3	15.4	11.5	9.1
REL	13.2	5.1	0.8
BCL7A	10.7	2.6	2.3
CDKN2A	10.3	21.8	48.5
MYD88	8.8	15.4	34.1
CD58	8.5	10.3	6.8
TMEM30A	8.1	11.5	8.3
CD70	7.7	17.9	6.1
PIM1	7.0	5.1	24.2
CDKN2B	5.1	11.5	30.3
NOTCH2	4.0	10.3	6.8
CD79B	2.2	9.0	25.0
PRDM1	1.5	3.8	19.7
ETV6	0.7	5.1	10.6

Listed in order of frequency in the GCB subgroup. *Gene mutations occurring in ≥10% of patients in any COO subgroup. ABC: activated B-cell-like; COO: cell-of-origin; DLBCL: diffuse large B-cell lymphoma; GCB: germinal center B-cell-like.

Table 2. Results from prognostic evaluation of prioritized candidate genes.

Gene	Univariate HR (95% CI)*	*P*-value	FDR	Multivariate HR (95% CI)†	*P*-value	FDR
BCL2	**1.7 (1.1-2.5)**	**0.012**	**0.14**	**2.6 (1.6-4.2)**	**0.00016**	**0.0037**
BCL2 translocation	**1.6 (1.0-2.4)**	**0.036**	**0.096**	**2.5 (1.4-4.2)**	**0.00095**	**0.0028**
BCL2 SNVs	**2.2 (1.3-3.8)**	**0.0025**	**0.041**	**2.6 (1.5-4.7)**	**0.0014**	**0.022**
CREBBP	1.4 (0.9-2.2)	0.14	0.37	2.1 (1.3-3.4)	0.0047	0.054
REL	1.3 (0.8-2.3)	0.32	0.67	1.9 (1.0-3.4)	0.043	0.25
CD274	1.6 (0.9-3.2)	0.13	0.37	1.7 (0.9-3.3)	0.13	0.54
TP53	1.6 (1.0-2.4)	0.034	0.26	1.6 (1.1-2.5)	0.029	0.22
TP53 SNVs	1.5 (1.0-2.3)	0.044	0.35	1.6 (1.0-2.5)	0.034	0.18
TNFRSF14	1.2 (0.7-2.1)	0.49	0.74	1.4 (0.8-2.7)	0.26	0.54
KMT2D	1.2 (0.8-1.7)	0.46	0.74	1.3 (0.9-1.9)	0.23	0.54
CD58	1.2 (0.7-2.1)	0.59	0.79	1.3 (0.7-2.4)	0.38	0.62
MYC	1.6 (0.9-2.8)	0.15	0.37	1.2 (0.6-2.2)	0.60	0.72
MYC translocation	1.8 (0.9-3.2)	0.064	0.096	1.4 (0.7-2.5)	0.30	0.30
ARID1A	1.2 (0.6-2.2)	0.66	0.79	1.2 (0.6-2.4)	0.55	0.70
CDKN2A	1.7 (1.2-2.5)	0.0056	0.13	1.2 (0.8-1.9)	0.46	0.70
CDKN2A deletion	1.6 (1.1-2.4)	0.014	0.058	1.1 (0.7-1.7)	0.85	0.99
CDKN2B	1.5 (1.0-2.4)	0.077	0.35	1.1 (0.7-1.7)	0.82	0.85
BCL7A	1.1 (0.6-2.1)	0.81	0.88	1.1 (0.6-2.3)	0.68	0.75
TNFAIP3	0.9 (0.5-1.5)	0.63	0.79	1.0 (0.6-1.8)	0.85	0.85
MYD88	1.2 (0.8-1.9)	0.44	0.74	0.9 (0.5-1.4)	0.52	0.70
B2M	0.8 (0.5-1.4)	0.52	0.74	0.9 (0.5-1.5)	0.63	0.72
EZH2	0.5 (0.3-1.2)	0.12	0.37	0.8 (0.4-1.7)	0.50	0.70
BCL6	1.0 (0.7-1.6)	0.86	0.9	0.8 (0.5-1.2)	0.27	0.54
PIM1	0.8 (0.5-1.4)	0.48	0.74	0.7 (0.4-1.2)	0.21	0.54
CD79B	0.9 (0.5-1.6)	0.77	0.88	0.7 (0.4-1.3)	0.28	0.54
CD70	1.0 (0.5-1.9)	0.93	0.93	0.7 (0.4-1.4)	0.38	0.62
CARD11	0.5 (0.2-1.1)	0.076	0.35	0.6 (0.3-1.4)	0.22	0.54
TMEM30A	0.6 (0.3-1.3)	0.19	0.43	0.6 (0.3-1.4)	0.25	0.54

Listed in order of multivariate HR. Significant alterations on multivariate analysis (FDR <0.05) shown in bold. *Adjusted for treatment only. †Adjusted for treatment arm, IPI, COO, number of planned chemotherapy cycles and geographic region. CI: confidence interval; COO: cell-of-origin; FDR: false discovery rate; HR: hazard ratio; IPI: International Prognostic Index.

Figure legends

Figure 1. Frequently observed gene alterations in patients with DLBCL in GOYA. (A) Most frequently (≥2% of cases) observed gene alterations (SNVs, amplifications and deletions). (B) Genes with significant differences in mutation rates* between the ABC and GCB DLBCL subtypes. (C) Frequency of *BCL2* and *CDKN2A* alterations in the ABC and GCB DLBCL subtypes. *FDR <0.05. ABC: activated B-cell-like; CNA: copy number abnormality; DLBCL: diffuse large B-cell lymphoma; FDR: false discovery rate; GCB: germinal center B-cell-like; SNV: single nucleotide variant; trans: translocation.

Figure 2. Association between *BCL2* gene alterations and PFS in DLBCL. (A) All *BCL2* alterations. (B) *BCL2* SNVs. (C) *BCL2* translocations. CI: confidence interval; DLBCL: diffuse large B-cell lymphoma; FDR: false discovery rate; HR: hazard ratio; MUT: mutant; PFS: progression-free survival; SNV: single nucleotide variant; WT: wild-type.

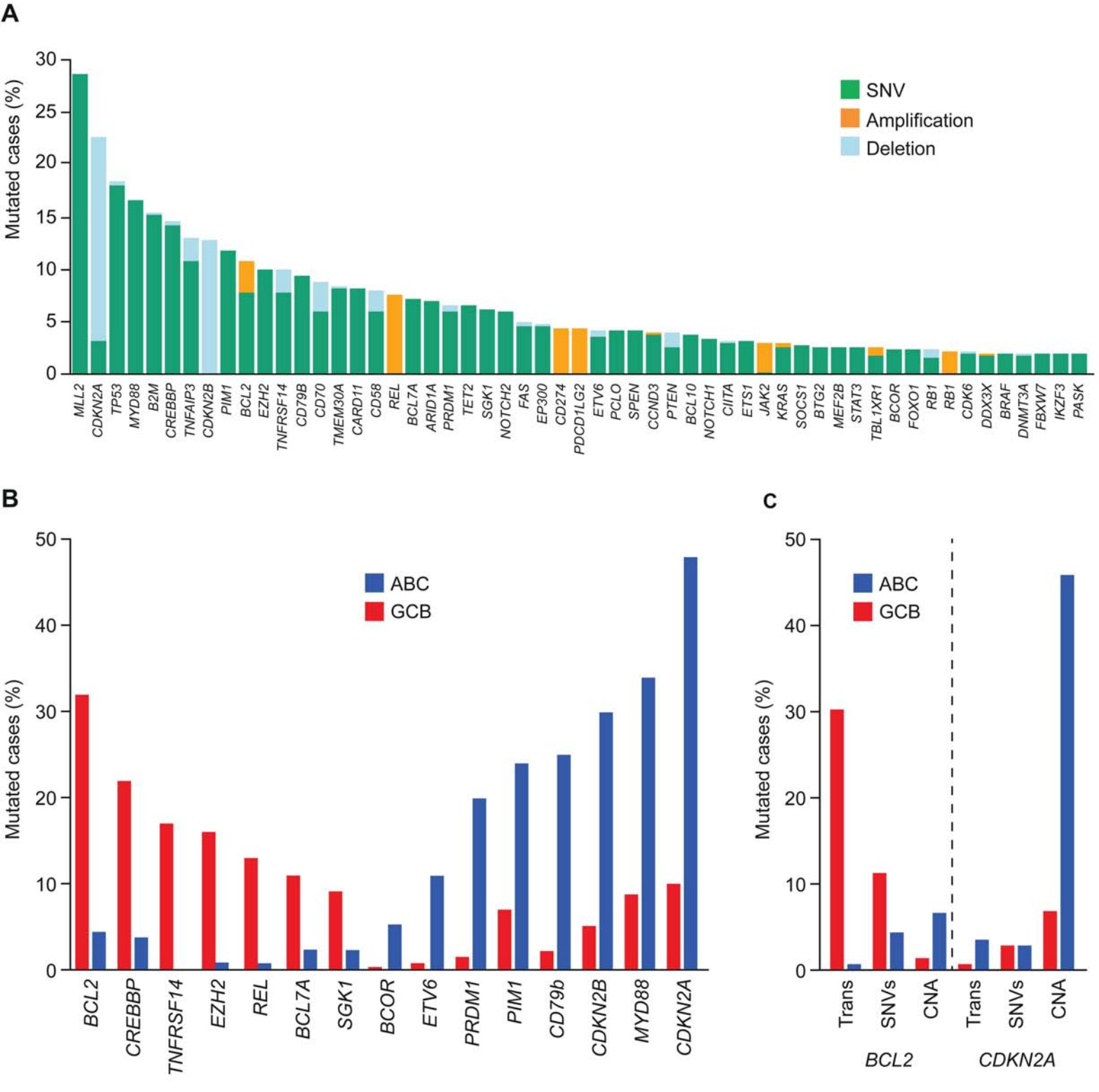
Figure 3. *BCL2* alterations according to COO subtype. *GCB, 31%; unclassified, 5.1%; ABC, 0.8%. †GCB, 11%; unclassified, 0%; ABC, 4.5%. ‡GCB, 1.5%; unclassified, 2.6%; ABC, 6.8%. ABC: activated B-cell-like; amp: amplification; COO: cell-of-origin; GCB: germinal center B-cell-like; NA: not available; SNV: single nucleotide variant; trans: translocation.

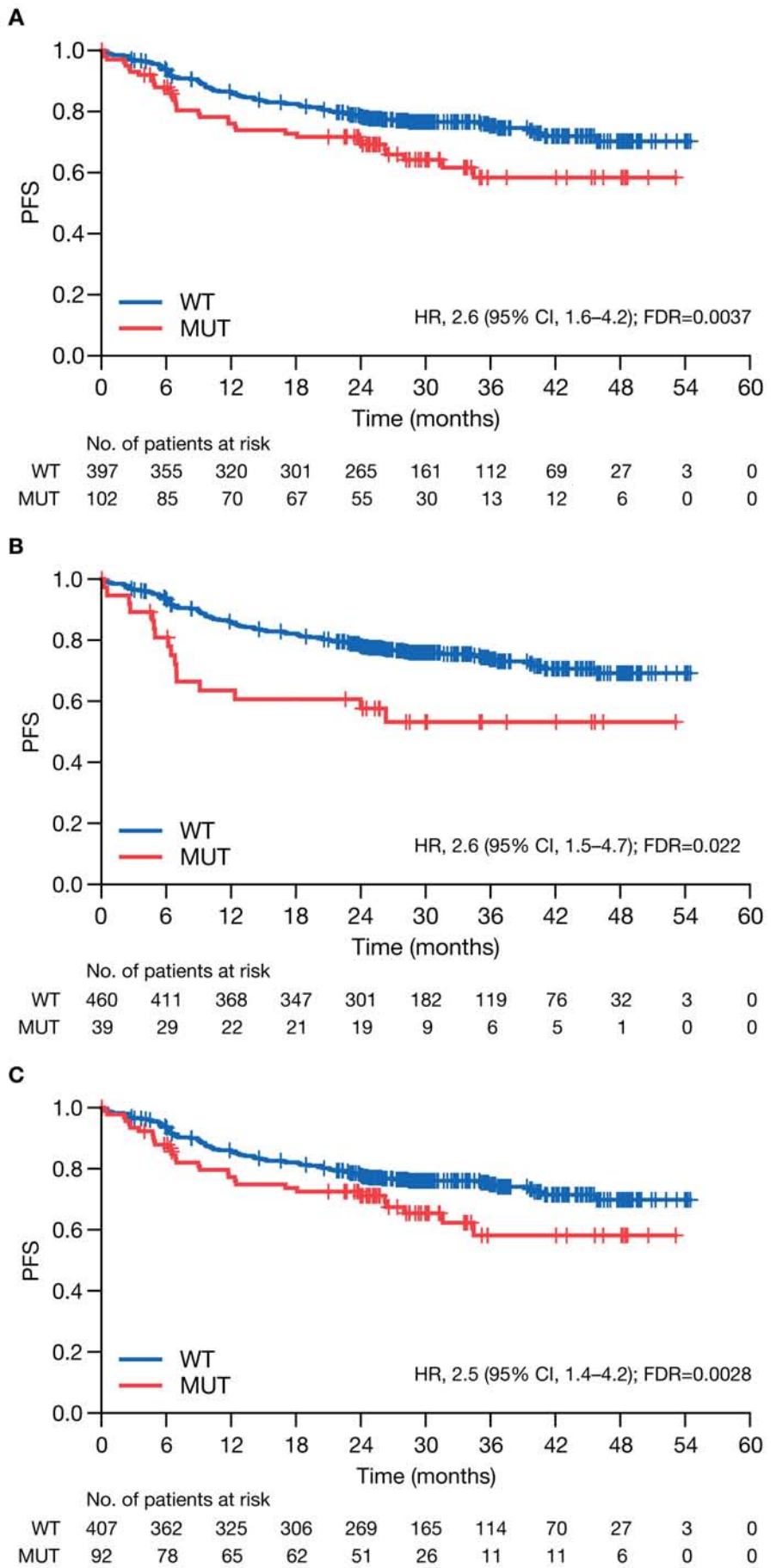
Figure 4. (A) Distribution of risk scores using the applied Reddy *et al.* prognostic model, and (B) PFS by risk group (n=443). int: intermediate; PFS: progression-free survival.

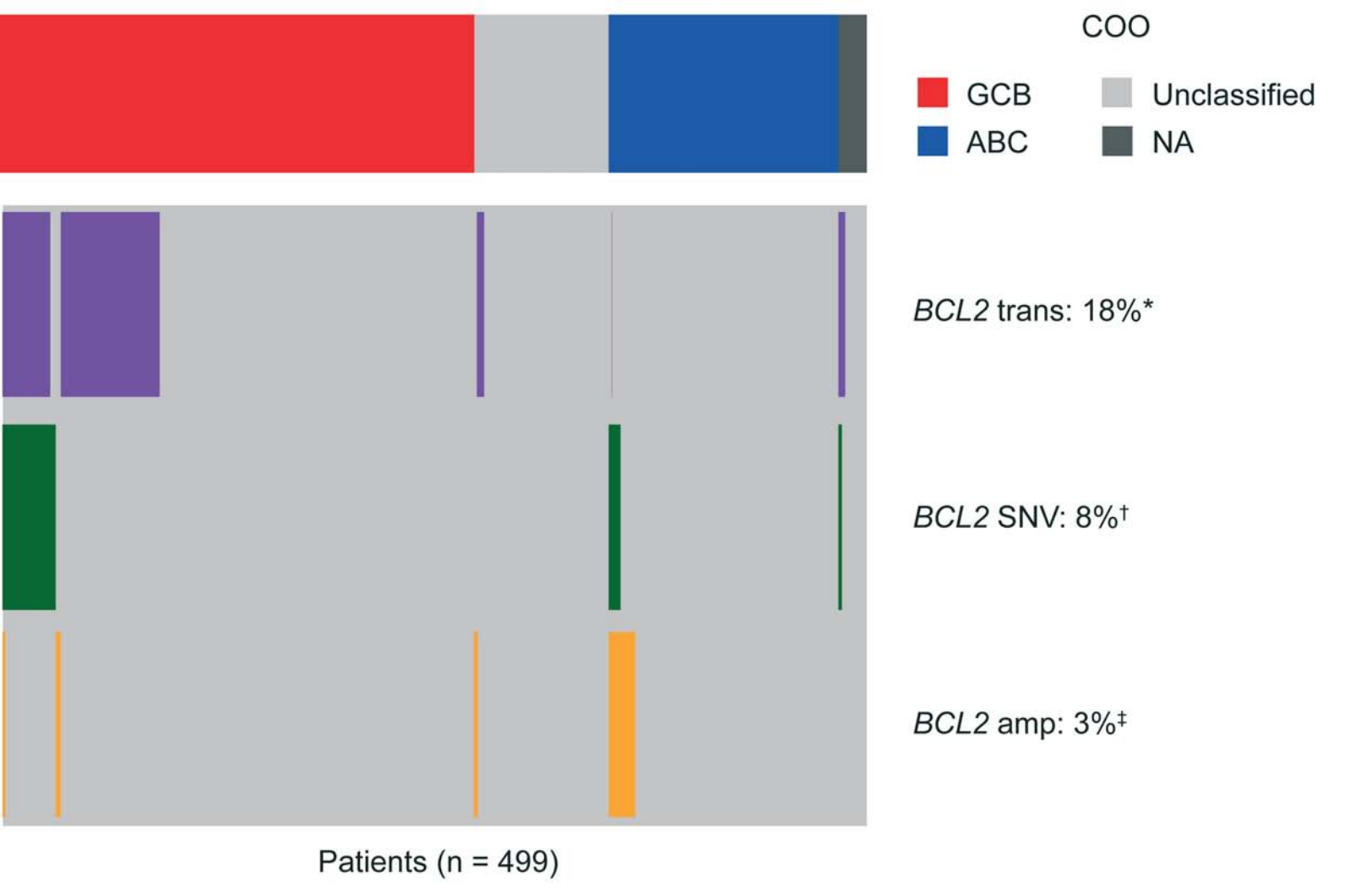
Figure 5. DLBCL mutational subset validation. (A) Prevalence and (B) association of Schmitz *et al.* classifications with PFS. Schmitz clusters were approximated using the seed mutations: EZB - *EZH2* or *BCL2*; BN2 - *BCL6 or*

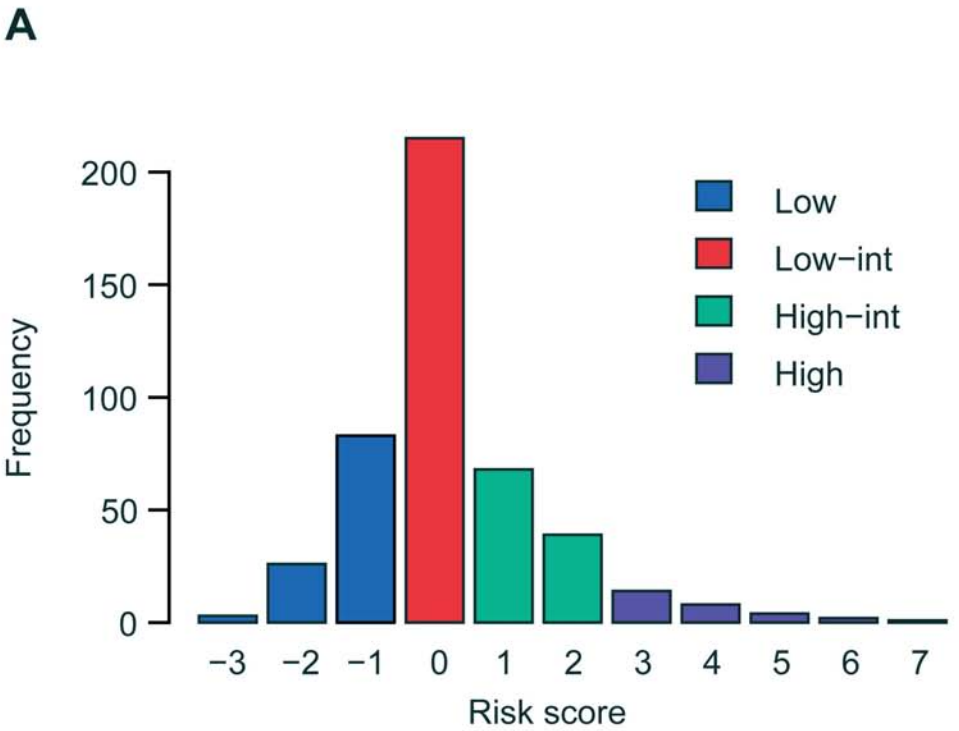
NOTCH2; N1 - *NOTCH1*; MCD - *MYD88*, *L265P* or *CD79B*; Multi – multiple seed mutations from more than one cluster. (C) Chapuy *et al.* clusters were approximated by application of non-negative matrix factorization (NMF) to the GOYA FMI dataset and selecting five clusters (G1-G5). Mutations with significant enrichment in one or more clusters are shown. (D) Association between NMF clusters and PFS. ABC: activated B-cell-like; alt: alteration; CNA: copy number abnormality; COO: cell-of-origin; DLBCL: diffuse large B-cell lymphoma; FMI: Foundation Medicine Incorporated; GCB: germinal center B-cell-like; HR: hazard ratio; PFS: progression-free survival; SNV: single nucleotide variant.

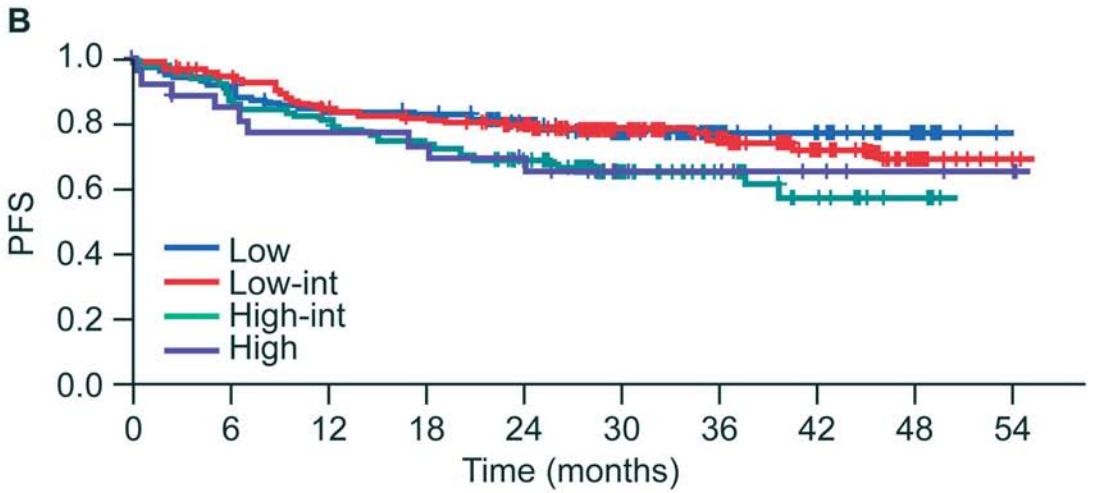
Figure 1


A
Mutated cases (%)
SNV
Amplification
Deletion
MLL2
CDKN2A
TP53
MYD88
B2M
CREBBP
TNFAIP3
CDKN2B
PIM1
BCL2
EZH2
TNFRSF14
CD79B
CD70
TMEM30A
CARD11
CD58
REL
BCL7A
ARID1A
PRDM1
TET2
SGK1
NOTCH2
FAS
EP300
CD274
PDCD1LG2
ETV6
PCLO
SPEN
CCND3
PTEN
BCL10
NOTCH1
CIITA
ETS1
JAK2
KRAS
SOCS1
BTG2
MEF2B
STAT3
TBL1XR1
BCOR
FOXO1
RB1
RB1
CDK6
DDX3X
BRAF
DNMT3A
FBXW7
IKZF3
PASK
B
Mutated cases (%)
ABC
GCB
BCL2
CREBBP
TNFRSF14
EZH2
REL
BCL7A
SGK1
BCOR
ETV6
PRDM1
PIM1
CD79b
CDKN2B
MYD88
CDKN2A
C
Mutated cases (%)
ABC
GCB
Trans
SNVs
CNA
Trans
SNVs
CNA
BCL2
CDKN2A


A
1.0
0.8
0.6
0.4
0.2
0.0
PFS
WT
MUT
HR, 2.6 (95% CI, 1.6–4.2); FDR=0.0037
0 6 12 18 24 30 36 42 48 54 60
Time (months)
No. of patients at risk
WT 397 355 320 301 265 161 112 69 27 3 0
MUT 102 85 70 67 55 30 13 12 6 0 0
B
1.0
0.8
0.6
0.4
0.2
0.0
PFS
WT
MUT
HR, 2.6 (95% CI, 1.5–4.7); FDR=0.022
0 6 12 18 24 30 36 42 48 54 60
Time (months)
No. of patients at risk
WT 460 411 368 347 301 182 119 76 32 3 0
MUT 39 29 22 21 19 9 6 5 1 0 0
C
1.0
0.8
0.6
0.4
0.2
0.0
PFS
WT
MUT
HR, 2.5 (95% CI, 1.4–4.2); FDR=0.0028
0 6 12 18 24 30 36 42 48 54 60
Time (months)
No. of patients at risk
WT 407 362 325 306 269 165 114 70 27 3 0
MUT 92 78 65 62 51 26 11 11 6 0 0

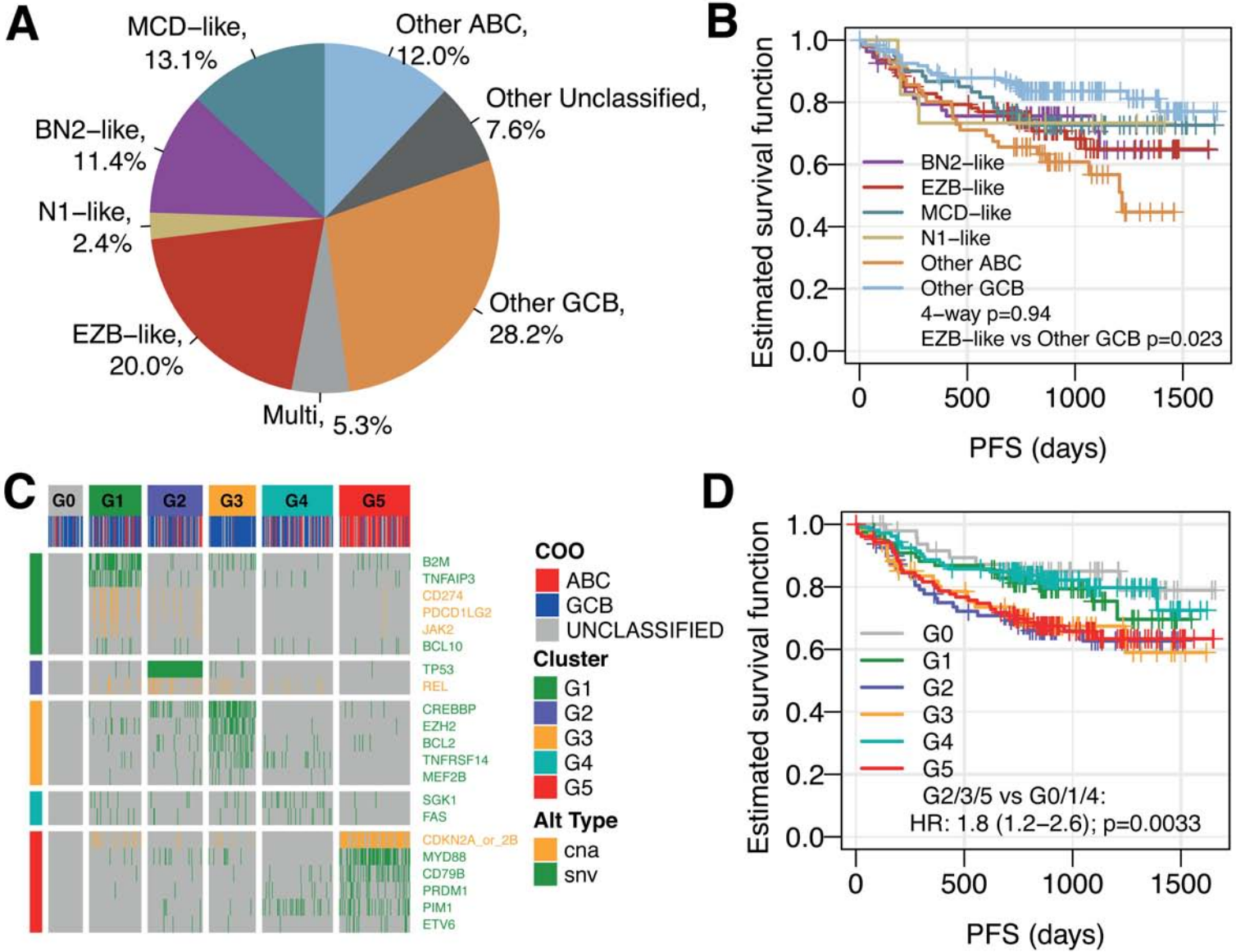

COO
GCB
Unclassified
ABC
NA
BCL2 trans: 18%*
BCL2 SNV: 8%†
BCL2 amp: 3%‡
Patients (n = 499)





No. of patients at risk

Low	112	99	86	84	71	47	31	24	11	0
Low–int	215	194	172	163	143	83	55	35	16	1
High–int	107	94	84	76	65	39	22	11	3	0
High	29	22	20	19	17	10	9	6	3	2


A
MCD–like, 13.1%
Other ABC, 12.0%
Other Unclassified, 7.6%
BN2–like, 11.4%
N1–like, 2.4%
Other GCB, 28.2%
EZB–like, 20.0%
Multi, 5.3%
B
Estimated survival function
1.0
0.8
0.6
0.4
0.2
0.0
BN2–like
EZB–like
MCD–like
N1–like
Other ABC
Other GCB
4–way p=0.94
EZB–like vs Other GCB p=0.023
0
500
1000
1500
PFS (days)
C
G0
G1
G2
G3
G4
G5
B2M
TNFAIP3
CD274
PDCD1LG2
JAK2
BCL10
TP53
REL
CREBBP
EZH2
BCL2
TNFRSF14
MEF2B
SGK1
FAS
CDKN2A_or_2B
MYD88
CD79B
PRDM1
PIM1
ETV6
COO
ABC
GCB
UNCLASSIFIED
Cluster
G1
G2
G3
G4
G5
Alt Type
cna
snv
D
Estimated survival function
1.0
0.8
0.6
0.4
0.2
0.0
G0
G1
G2
G3
G4
G5
G2/3/5 vs G0/1/4:
HR: 1.8 (1.2–2.6); p=0.0033
0
500
1000
1500
PFS (days)

Supplementary methods

GOYA inclusion criteria

Included patients had previously untreated, histologically documented, CD20-positive DLBCL; Eastern Cooperative Oncology Group performance status of 0-2; and IPI score ≥2. Patients with an IPI score of 1 and aged ≤60 years, with or without bulky disease, and those with an IPI score of 0 and bulky disease (i.e. one lesion ≥7.5 cm) were also included. Patients were treated with eight 21-day cycles of G 1000 mg (days 1, 8 and 15, cycle 1; day 1, cycles 2-8) or R 375 mg/m^2 (day 1, cycles 2-8) plus 6-8 cycles of CHOP chemotherapy.

Targeted NGS

Captured libraries were sequenced to a median exon coverage depth of >500x (DNA) using Illumina sequencing, and resultant sequences were analyzed for SNVs (base substitutions, and small insertions and deletions [indels]), CNAs (focal amplifications, and homozygous deletions) and gene fusions/rearrangements, as previously described.[1] Frequent germline variants identified in the 1000 Genomes Project (dbSNP135) were removed. To maximize mutation-detection accuracy (sensitivity and specificity) in impure clinical specimens, the test was previously optimized and validated to detect base substitutions at a ≥5% mutant allele frequency, indels at a ≥10% mutant allele frequency with ≥99% accuracy and fusions occurring within baited introns/exons with >99% sensitivity.[1] Known, confirmed somatic alterations deposited in the Catalogue of Somatic Mutations in Cancer (COSMIC v62) were called at allele frequencies ≥1%.[2] Genes were only considered amplified with a copy number ≥6. A separate analysis was performed to identify “low level” *BCL2* amplifications, where patients were considered to have amplifications

with *BCL2* copy number ≥3, ≥1 copy more than median copy number and an amplification signal significantly higher than noise level.

Validation of mutational model generated by Reddy *et al.*[3]

Whole transcriptome gene expression was analyzed using TruSeq RNA sequencing in tumor tissue from 443 of the 499 DLBCL samples with FMI mutational data, and raw read counts were normalized using Limma-voom;[4] expression of *MYC* and *BCL2* were estimated, and a median cut-point was used to split samples into high- and low-expressers. These data were combined with the COO classification from NanoString, and the known/likely mutation calls from the FMI platform. Not all of the genes used in the Reddy *et al.* model were available on the FMI platform, thus a number of the model coefficients were excluded (*ZFAT*, *KLHL14*, *BIRC6*, *SETD5*, *CHD1*, *ZEB2*, *DDX10* and *ARID5B*). Therefore, rather than retaining the coefficients used in the trained model, we retained the sign of the coefficients and gave each coefficient equal weight. High-risk biomarkers were given a weight of +1, while low-risk biomarkers were given a weight of -1. The score was then calculated as the sum of biomarkers present in each sample.

Validation of complex molecular subtypes

Approximation of the Schmitz *et al.*[5] molecular subtypes was calculated based on presence of alterations in at least one of the clusters' "founder" genes: for EZB – *EZH2* SNV or *BCL2* translocation; BN2 – *BCL6* rearrangement or *NOTCH2* SNV; N1 – *NOTCH1* SNV; MCD – *MYD88*, *L265P* or *CD79B* SNV. Samples with alterations in founder genes from multiple clusters are referred to as "multi", and were not included in any of the individual clusters.

Approximation of the Chapuy *et al.*[6] molecular subtypes was calculated by applying NMF to a subset of the FMI platform. Among SNVs and CNAs that are measured by the FMI platform, all genes reported by Chapuy *et al.* as significantly enriched in at least one molecular subtype were included. Additionally, copy number alterations in *CDKN2A* and *CDKN2B* were collapsed to represent deletions of the 9p21 region. A total of 51 features with at least one alteration were included for modeling. Patients with no alterations were removed, and NMF was applied using 100 runs and five clusters. The resulting clusters were manually examined and labeled to match the clusters described by Chapuy *et al.*

1. Frampton GM, Fichtenholtz A, Otto GA, et al. Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing. Nat Biotechnol. 2013;31(11):1023-1031.
2. Forbes SA, Bindal N, Bamford S, et al. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. Nucleic Acids Res. 2011;39(Database issue):D945-D950.
3. Reddy A, Zhang J, Davis NS, et al. Genetic and functional drivers of diffuse large B cell lymphoma. Cell. 2017;171(2):481-494.e15.
4. Law CW, Chen Y, Shi W, Smyth GK. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol. 2014;15(2):R29.
5. 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.
6. 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(5):679-690.

Supplementary tables and figures

Table S1. Baseline disease characteristics of patients with NGS data available and the overall GOYA study population.

	GOYA ITT population, n=1418	Patients with NGS available, n=499
Median age (range), years	62 (18-86)	64 (18-86)
Male, n (%)	752 (53)	257 (52)
Race, n (%)		
White	856 (60)	393 (79)
Asian	522 (37)	88 (18)
Other	40 (3)	18 (4)
3-year PFS (95% CI), fraction	0.68 (0.66-0.71)	0.72 (0.68-0.77)
3-year OS (95% CI), fraction	0.81 (0.79-0.83)	0.85 (0.81-0.88)
ECOG PS, n (%)		
0-1	1231 (87)	445 (89)
2-3	186 (13)	53 (11)
IPI, n (%)		
Low	283 (20)	102 (20)
Low-int	502 (35)	171 (34)
High-int	413 (29)	146 (29)
High	220 (16)	80 (16)
Ann Arbor stage, n (%)	n=1417	n=499
I	103 (7)	35 (7)
II	238 (17)	83 (17)
III	466 (33)	172 (34)

IV	610 (43)	209 (42)
COO, n (%)	n=933	n=482
GCB	540 (58)	272 (56)
ABC	243 (26)	132 (27)
Unclassified	150 (16)	78 (16)
Bone marrow involvement, n (%)*	153 (11)	67 (14)
Elevated serum LDH†	816 (58)	283 (57)
Number of extranodal sites, n (%)	n=954	n=322
0	14 (1)	7 (2)
1	437 (46)	158 (49)
>1	503 (53)	157 (49)
Bulky disease at baseline, n (%)‡	523 (37)	175 (35)

*Data missing for 14 patients for bone marrow involvement for the GOYA ITT population, and seven patients for the NGS available population. †Data missing for five patients for serum LDH for the GOYA ITT population, and one patient for the NGS available population. ‡Data missing for five patients for bulky disease at baseline for the GOYA ITT population.
ABC: activated B-cell-like; CI: confidence interval; COO: cell-of-origin; ECOG PS: Eastern Cooperative Oncology Group performance status; GCB: germinal center B-cell-like; int: intermediate; IPI: International Prognostic Index; ITT: intent-to-treat; LDH: lactate dehydrogenase; NGS: next-generation sequencing; OS: overall survival; PFS: progression-free survival.

Table S2. Listing of translocation partners for *BCL2*, *MYC* and *BCL6*.

	BCL2	***MYC***	***BCL6***
IGH	92	29	57
Unknown	1	1	9
CIITA	0	0	6
IKZF1	0	0	4
HIST1H2BK	0	0	4
BCL6	0	0	3
IGL	0	0	3
RHOH	0	0	3
HSP90AA1	0	0	2
IGLL5	0	0	2
TFRC	0	0	2
EIF4A2	0	0	2
TMSB4X	0	0	1
TRA2B	0	0	1
LCP1	0	0	1
RPLP0	0	0	1
HMGB2	0	0	1
GAS5	0	0	1
G3BP1	0	0	1
LPP	0	0	1
NR3C1	0	0	1
HSPD1	0	0	1
RPSA	0	0	1

HNRNPC	0	0	1
BIRC3	0	1	0
FLJ21408	0	1	0
DMD	0	1	0

Table S3. Prevalence of gene mutations according to DLBCL COO subtype.

	GCB, n=272 (%)	Unclassified, n=78 (%)	ABC, n=132 (%)
BCL2	32.4	5.1	4.5
KMT2D	30.1	21.8	28.8
CREBBP	22.1	7.7	3.8
TP53	19.5	17.9	15.2
BCL6	18.8	35.9	22.0
B2M	17.6	12.8	12.9
TNFRSF14	17.3	1.3	0.0
EZH2	16.2	6.4	0.8
TNFAIP3	15.4	11.5	9.1
REL	13.2	5.1	0.8
BCL7A	10.7	2.6	2.3
CDKN2A	10.3	21.8	48.5
CARD11	9.9	2.6	6.1
ARID1A	9.6	6.4	3.0
SGK1	9.2	1.3	2.3
MYD88	8.8	15.4	34.1
CD58	8.5	10.3	6.8
TMEM30A	8.1	11.5	8.3
CD70	7.7	17.9	6.1
PIM1	7.0	5.1	24.2
MYC	7.0	5.1	9.1
TET2	7.0	9.0	1.5

CIITA	6.3	5.1	1.5
FAS	5.9	6.4	2.3
CDKN2B	5.1	11.5	30.3
PCLO	5.1	2.6	3.0
PTEN	5.1	2.6	2.3
SOCS1	5.1	1.3	0.8
CD274	4.8	6.4	5.3
PDCD1LG2	4.4	6.4	3.0
NOTCH2	4.0	10.3	6.8
SPEN	4.0	6.4	3.8
STAT3	4.0	2.6	0.0
EP300	3.7	9.0	4.5
MEF2B	3.7	0.0	3.0
RB1	3.7	2.6	0.0
KRAS	3.3	2.6	3.0
DDX3X	3.3	2.6	0.0
FOXO1	3.3	1.3	0.0
ETS1	2.9	2.6	4.5
BTG2	2.9	3.8	0.8
CDK6	2.9	2.6	0.8
BRAF	2.9	1.3	0.8
CCND3	2.6	9.0	4.5
JAK2	2.6	3.8	3.0
FBXO11	2.6	1.3	0.0
HIST1H1E	2.6	1.3	0.0

CD79B	2.2	9.0	25.0
TBL1XR1	2.2	2.6	3.8
IKZF3	2.2	1.3	2.3
BCL10	1.8	6.4	6.8
PASK	1.8	0.0	3.8
XPO1	1.8	0.0	2.3
KLHL6	1.8	0.0	2.3
DNMT3A	1.8	3.8	1.5
FBXW7	1.8	2.6	1.5
CD36	1.8	2.6	0.8
PRDM1	1.5	3.8	19.7
NOTCH1	1.5	6.4	6.1
FOXP1	1.5	0.0	3.8
LRP1B	1.5	0.0	2.3
FANCA	1.5	2.6	0.8
IKZF1	1.1	2.6	3.8
BTG1	1.1	1.3	3.8
GNAS	1.1	1.3	3.0
KDM4C	1.1	2.6	2.3
ETV6	0.7	5.1	10.6
MALT1	0.7	2.6	3.8
BCOR	0.4	3.8	5.3

Listed in order of frequency in the GCB subtype. ABC: activated B-cell-like; COO: cell-of-origin; DLBCL: diffuse large B-cell lymphoma; GCB: germinal center B-cell-like.

Table S4. Patient disease characteristics according to CDKN2A alteration type.

		DLBCL			ABC DLBCL	
		***CDKN2A* del, (n=99) (%)**	***CDKN2A* WT or SNV, (n=400) (%)**		***CDKN2A* del, (n=62) (%)**	***CDKN2A* WT or SNV, (n=70) (%)**
IPI	Low/low-int	**38 (38.4)**	**235 (58.8)**	Low/low-int	**24 (38.7)**	**37 (52.9)**
	High-int	28 (28.3)	118 (29.5)	High-int	18 (29.0)	24 (34.3)
	High	**33 (33.3)**	**47 (11.8)**	High	**20 (32.3)**	**9 (12.9)**
Ann Arbor	I	6 (6.1)	29 (7.3)	I	4 (6.5)	3 (4.3)
	II	14 (14.1)	69 (17.3)	II	8 (12.9)	9 (12.9)
	III	30 (30.3)	142 (35.5)	III	21 (33.9)	30 (42.9)
	IV	49 (49.5)	160 (40.0)	IV	29 (46.8)	28 (40.0)
EN sites	0	29 (29.3)	155 (38.8)	0	18 (29.0)	27 (38.6)
	1	28 (28.3)	130 (32.5)	1	19 (30.6)	22 (31.4)
	>1	**42 (42.4)**	**115 (28.8)**	>1	**25 (40.3)**	**21 (30.0)**
ECOG PS	0	45 (45.5)	216 (54.0)	0	32 (51.6)	38 (54.3)
	1	36 (36.4)	148 (37.0)	1	20 (32.3)	26 (37.1)
	2	18 (18.2)	35 (8.8)	2	10 (16.1)	6 (8.6)
Age	<60	**24 (24.2)**	**155 (38.8)**	<60	**14 (22.6)**	**25 (35.7)**
	>60	**75 (75.8)**	**245 (61.3)**	>60	**48 (77.4)**	**45 (64.3)**
Serum LDH	Normal	**32 (32.3)**	**183 (45.8)**	Normal	18 (29.0)	19 (27.1)
	Elevated	**67 (67.7)**	**216 (54.0)**	Elevated	45 (72.6)	49 (70.0)

Differences ≥10% between *CDKN2A* del and *CDKN2A* WT or SNV are shown in bold. ABC: activated B-cell-like; del: deletion; DLBCL: diffuse large B-cell lymphoma; ECOG PS: Eastern Cooperative Oncology Group performance status; EN: extranodal; int: intermediate; IPI: International Prognostic Index; LDH: lactate dehydrogenase; SNV: single nucleotide variant; WT: wild-type.

Table S5. Multivariate analysis of the association between genetic alterations and PFS in (A) GCB and (B) ABC DLBCL.

A

Gene	HR (95% CI)	FDR
CD274	2.6 (1.0-7.0)	0.24
BCL2	2.3 (1.3-4.1)	0.066
BCL2_SNVs	2.4 (1.2-5.0)	0.11
BCL2_trans	**2.3 (1.3-4.2)**	**0.017**
BCL2_amp (low level)	1.4 (0.8-2.7)	0.34
CDKN2A	2.2 (1.1-4.3)	0.19
CDKN2A_CNAs	1.9 (0.9-4.1)	0.24
CREBBP	2.1 (1.2-3.9)	0.1
ARID1A	2.0 (0.9-4.3)	0.36
MYC	1.7 (0.7-4.1)	0.44
MYC_trans	2.0 (0.8-4.7)	0.27
CD70	1.9 (0.8-4.4)	0.37
CDKN2B	1.8 (0.7-4.3)	0.44
CDKN2B_CNAs	1.9 (0.8-4.6)	0.24
TP53	1.6 (0.9-3.0)	0.43
TP53_SNVs	1.6 (0.8-3.0)	0.42
REL	1.6 (0.8-3.1)	0.43
REL_CNVs	1.4 (0.7-2.8)	0.34
BCL7A	1.6 (0.7-3.7)	0.6
MYD88	1.5 (0.7-3.3)	0.51
KMT2D	1.4 (0.8-2.4)	0.51

TNFRSF14	1.3 (0.7-2.5)	0.57
CD79B	1.3 (0.4-4.5)	0.85
TNFAIP3	1.2 (0.6-2.7)	0.88
B2M	1.1 (0.5-2.2)	0.88
TMEM30A	0.9 (0.3-2.6)	0.88
EZH2	0.9 (0.4-2.2)	0.85
PIM1	0.9 (0.3-2.8)	0.88
BCL6	0.8 (0.4-1.6)	0.49
CD58	0.7 (0.2-1.9)	0.67
CARD11	0.3 (0.1-1.1)	0.26

B

Gene	**HR (95% CI)**	***P* value**	**FDR**
BCL2	4.3 (1.5-13.0)	0.0083	0.11
BCL2_SNVs	4.3 (1.5-13.0)	0.0083	0.058
BCL2_amp (low level)	0.9 (0.5-1.8)	0.86	1.00
TP53	1.6 (0.7-3.4)	0.25	0.49
TNFAIP3	1.1 (0.4-3.2)	0.92	0.99
TNFAIP3_SNVs	1.2 (0.4-3.7)	0.73	0.98
KMT2D	1.1 (0.6-2.3)	0.69	0.81
MYC	0.7 (0.2-2.1)	0.54	0.76
MYC_trans	0.9 (0.3-2.6)	0.83	0.83
B2M	0.9 (0.3-2.4)	0.84	0.98
CDKN2B	0.8 (0.4-1.7)	0.57	0.86

CDKN2A	0.7 (0.4-1.4)	0.35	0.6
CD79B	0.7 (0.3-1.4)	0.28	0.49
PIM1	0.7 (0.4-1.5)	0.41	0.82
MYD88	0.6 (0.3-1.2)	0.13	0.46
BCL6	0.4 (0.2-1.0)	0.058	0.12
TMEM30A	0.3 (0.04-2.1)	0.22	0.68

Genes listed in order of multivariate HR. Significant alterations indicated in bold (FDR <0.05). Only genes with >10 mutated samples shown for the ABC subtype. ABC: activated B-cell-like; amp: amplification; CI: confidence interval; CNA: copy number abnormality; CNA: copy number variation; DLBCL: diffuse large B-cell lymphoma; FDR: false discovery rate; GCB: germinal center B-cell-like; HR: hazard ratio; PFS: progression-free survival; SNV: single nucleotide variant; trans: translocation.

Figure S1. Correlation between (A) all *BCL2* mutations and (B) *BCL2* translocations and protein and gene expression levels. FMI: Foundation Medicine Incorporated; IHC: immunohistochemistry; MUT: mutant; trans: translocation; WT: wild-type.

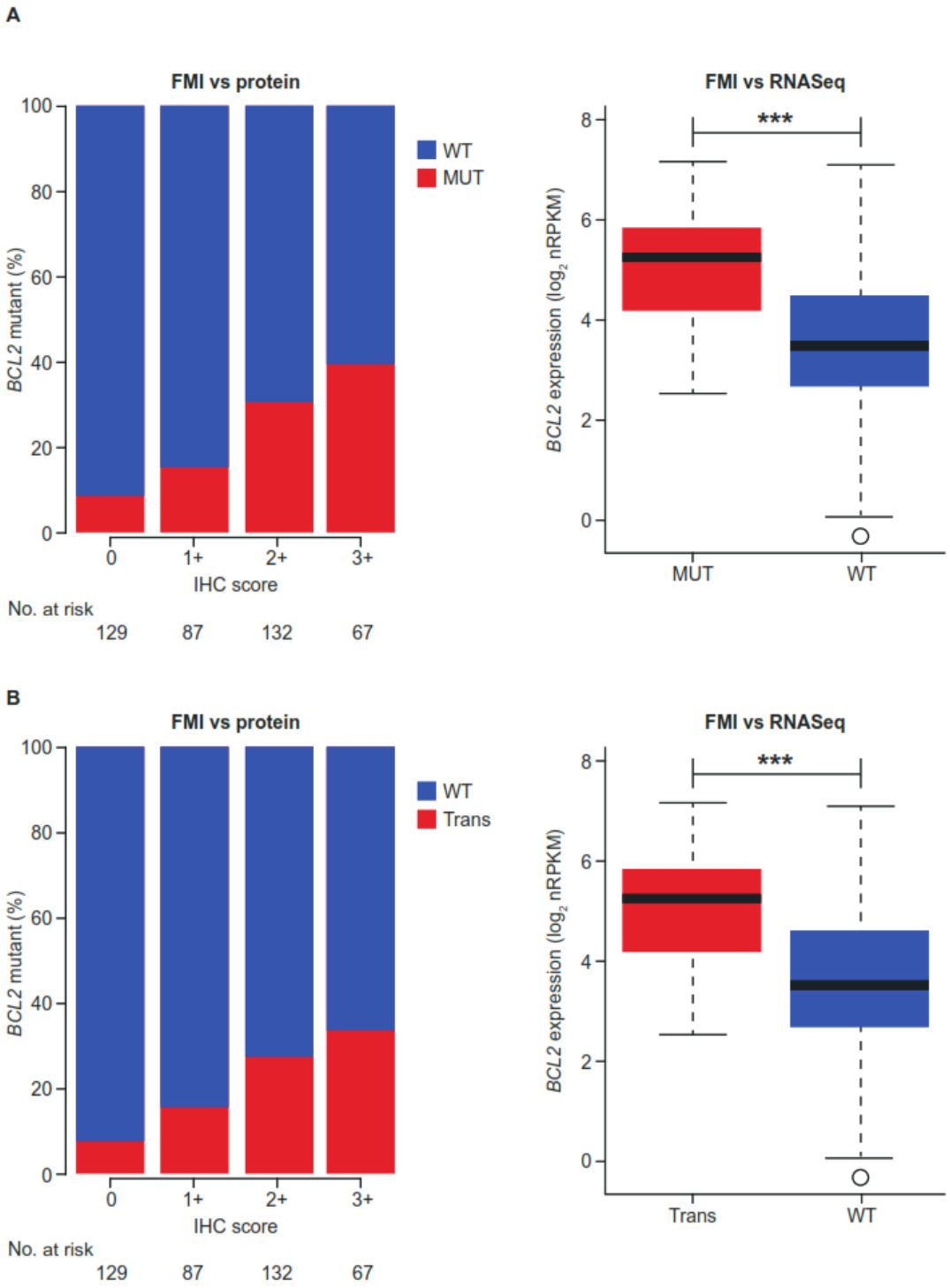


Figure S2. DLBCL mutational subset validation. Chapuy *et al.* clusters were approximated by application of non-negative matrix factorization (NMF) to the GOYA FMI dataset and selecting five clusters (G1-G5). ABC: activated B-cell-like; ALT: alteration; CNA: copy number abnormality; COO: cell of origin; DLBCL: diffuse large B-cell lymphoma; FMI: Foundation Medicine Incorporated; GCB: germinal center B-cell-like; SNV: single nucleotide variant.