Haematologica 2020
Volume 105(9):2298-2307

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,[2,3,4*] Marek Trněný,[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 and 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, British Columbia, 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, Dipartimento di 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 and [11]Oncology Biomarker Development, F. Hoffmann-La Roche Ltd. Basel, Switzerland

**CRB and MK contributed equally as co-first authors.*

ABSTRACT

Diffuse large B-cell lymphoma (DLBCL) represents a biologically and clinically heterogeneous diagnostic category with well-defined cell-of-origin (COO) subtypes. Using data from the GOYA study (*clinicaltrials.gov identifier: NCT01287741*), we characterized the mutational profile of DLBCL and evaluated the prognostic impact of somatic mutations in relation to COO. 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: 1.6-4.2). *BCL2* alterations were detected in 102 of 499 patients; 92 had *BCL2* translocations, 90% of whom had germinal center B-cell-like DLBCL. *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 DLBCL, with potential treatment targets occurring in distinct COO subtypes.

Correspondence:

CHRISTOPHER R. BOLEN
bolen.christopher@gene.com

Received: June 14, 2019.

Accepted: November 14, 2019.

Pre-published: November 14, 2019.

doi:10.3324/haematol.2019.227892

©2020 Ferrata Storti Foundation

Material published in Haematologica is covered by copyright. All rights are reserved to the Ferrata Storti Foundation. Use of published material is allowed under the following terms and conditions: https://creativecommons.org/licenses/by-nc/4.0/legalcode. Copies of published material are allowed for personal or internal use. Sharing published material for non-commercial purposes is subject to the following conditions: https://creativecommons.org/licenses/by-nc/4.0/legalcode, sect. 3. Reproducing and sharing published material for commercial purposes is not allowed without permission in writing from the publisher.

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 two 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 sub-

type.[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 1,001 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 *et al*.[8] and Chapuy *et al*.[9] 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 ABC (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=1,418) 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 been described previously.[3] Patients included in the study had previously untreated, histologically documented, CD20+ DLBCL; details of the inclusion criteria are available in the *Online Supplementary Methods*.

The study was conducted in accordance with the European Clinical Trial Directive (for European centers), the 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 identifier: NCT01287741*. All patients provided written informed consent.

Staging investigations included computed tomography (CT) scanning and bone marrow biopsy. Tumor response and progression were assessed by the investigator using regular clinical and laboratory examinations and CT 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.

Cell-of-origin analysis

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

Immunohistochemical analyses

Pre-treatment tumor samples were analyzed by a central laboratory using the Ventana BCL2 (124) and MYC (Y69) investigational use only immunohistochemical assays. The pre-specified 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 next-generation sequencing

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, USA] (*Online Supplementary Methods*). NGS data were available for 499 of the 1,418 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

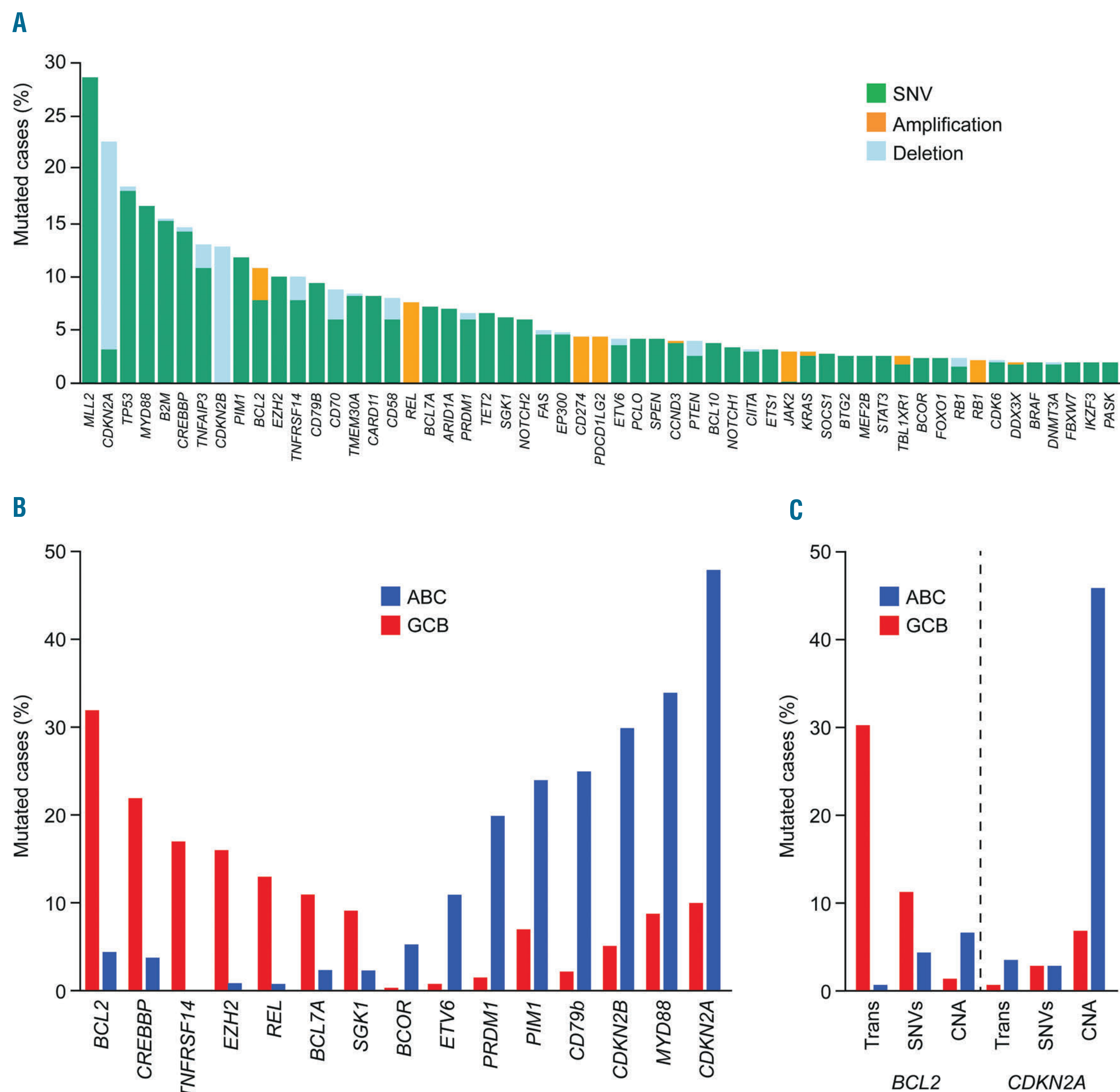

Figure 1. Frequently observed gene alterations in patients with diffuse large B-cell lymphoma (DLBCL) in the GOYA trial (*clinicaltrials.gov identifier: NCT01287741*). (A) Most frequently (≥2% of cases) observed gene alterations: single nucleotide variant (SNV), amplifications and deletions. (B) Genes with significant differences in mutation rates* between the activated B-cell-like (ABC) and germinal center B-cell-like (GCB) DLBCL subtypes. (C) Frequency of *BCL2* and *CDKN2A* alterations in the ABC and GCB DLBCL subtypes. *False discovery rate (FDR) <0.05. CNA: copy number abnormality; trans: translocation.

chemotherapy cycles, and geographic region. Multiple testing adjustment was performed by estimating false discovery rates (FDR) 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 next-generation sequencing

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 (SNV) and copy number abnormalities (CNA) 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 (SNV, amplifications and deletions) are shown in Figure 1A. SNV were the most common mutation type, while CNA 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 of 92 (100%), 29 of 32 (90.6%), and 57 of 100 (57.0%) cases where the rearrangement partner could be determined, respectively (*Online Supplementary Table S2*).

Frequencies of genomic alterations among cell-of-origin 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 of 272 (32%)], MLL2 (*KMT2D*) [82 of 272 (30%)] and *CREBBP* [60/272 (22%)]; loss of *CDKN2A* [64 of 132 (49%)] and *CDKN2B* [40 of 132 (30%)] and mutations of *MYD88* [45 of 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 SNV were more frequently found in the GCB subtype, high-level BCL2 amplifications (≥6 copies) were enriched within the ABC subtype [ABC, 9 of 132 (6.8%); GCB, 4 of 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 of 132 (62.9%); GCB, 45 of 272 (16.5%); Fisher's exact test *P*<0.001]. The enrichment of *CDKN2A* alterations within the ABC subtype was pronounced only

Table 1. Prevalence of most frequent* gene mutations according to diffuse large B-cell lymphoma cell-of-origin (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 germinal center B-cell-like (GCB) subgroup. *Gene mutations occurring in ≥10% of patients in any COO subgroup. n=number; ABC: activated B-cell-like.

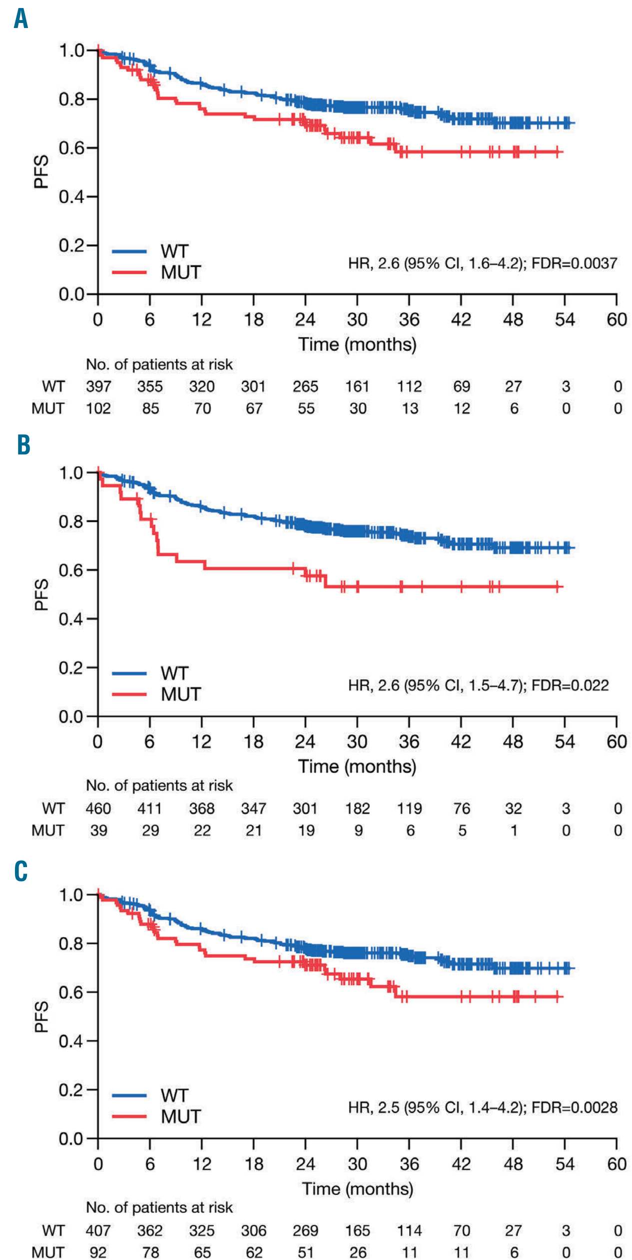

Figure 2. Association between *BCL2* gene alterations and progression-free survival (PFS) in diffuse large B-cell lymphoma (DLBCL). (A) All *BCL2* alterations. (B) *BCL2* single nucleotide variant. (C) *BCL2* translocations. CI: confidence interval; FDR: false discovery rate; HR: hazard ratio; MUT: mutant;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

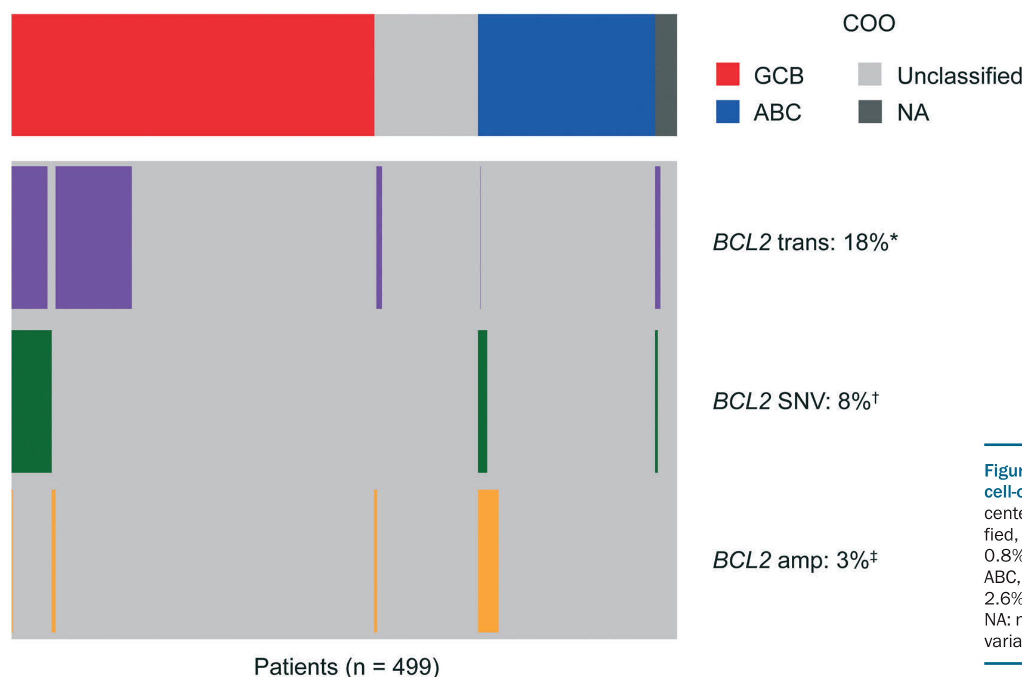

Figure 3. *BCL2* alterations according to cell-of-origin (COO) subtype. *Germinal center B-cell-like (GCB), 31%; unclassified, 5.1%; activated B-cell-like (ABC), 0.8%. †GCB, 11%; unclassified, 0%; ABC, 4.5%. ‡GCB, 1.5%; unclassified, 2.6%; ABC, 6.8%. amp: amplification; NA: not available; SNV: single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

for *CDKN2A* deletions; SNV 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, SNV and high-level amplifications) were the most strongly associated with PFS [hazard ratio (HR): 2.6; 95% confidence interval (CI): 1.6-4.2; 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* SNV (HR: 2.6; 95%CI: 1.5-4.7; FDR, 0.022) and translocations (HR: 2.5; 95%CI: 1.4-4.2; 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: 0.8-1.9; FDR, 0.58). *BCL2* alterations were detected in 20% (102 of 499) of patients, with 92 of 102 patients having a BCL2 translocation, 90% (83 of 92) of whom were GCB patients, with only one translocated ABC patient. Of 39 patients with *BCL2* SNV, 80% (31 of 39) and 15% (6 of 39) were in the GCB and ABC subgroups, respectively. The majority of patients with BCL2 SNV harbored *BCL2* translocations (74%, 29 of 39) (Figure 3), but *BCL2* SNV were still associated with worse prognosis among patients without a *BCL2* translocation (HR: 2.8; 95%CI: 1.0-7.9; $P=0.047$). *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: 1.3-3.4; FDR, 0.054) and *TP53* (HR: 1.6; 95%CI: 1.1-2.5; 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 of 499) of patients; 82% (60 of 73), 8% (6 of 73), and 7% (5 of 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 SNV (97%, 71 of 73), with only two cases of *CREBBP* deletion. Alterations of *TP53* were found in 18% (92 of 499) of patients, of whom 58% (53 of 92), 15% (14 of 92), and 22% (20 of 92) had the GCB, unclassified, and ABC DLBCL subtype, respectively. Overall, 105 TP53 alterations were observed in 92 patients, with 13 of 92 patients harboring two simultaneous *TP53* mutations. SNV were the most frequently observed *TP53* alterations (98%, 103 of 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: 1.2-2.5; FDR, 0.13). This effect was driven by *CDKN2A* deletions (HR: 1.6; 95%CI: 1.1-2.4; 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 of 499) of DLBCL patients. Of all cases with any *CDKN2A* alteration, 25% (28 of 113), 15% (17 of 113), and 57% (64 of 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 lactate dehydrogenase) compared with patients without a *CDKN2A* deletion,

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

Gene	Univariate HR (95%CI)*	*P*	FDR	Multivariate HR (95%CI)†	*P*	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 SNV	**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 SNV	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 hazard ratio (HR). Significant alterations on multivariate analysis [false discovery rate (FDR) <0.05] shown in bold. *Adjusted for treatment only. †Adjusted for treatment arm, International Prognostic Index, cell-of-origin, number of planned chemotherapy cycles, and geographic region. CI: confidence interval.

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: 1.3-4.2; *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: 0.42-0.88; *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: 0.49-1.2; *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: 0.24-1.04; *P*=0.06) but not the ABC subset (HR: 1.2; 95%CI: 0.66-2.32; *P*=0.5).

Validation of new molecular classifications

Although there is no publicly available tool 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.*,[8] 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.*[8] (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.*[9] This resulted in five clusters (plus an unmutated cluster: C0) sharing very similar mutational profiles and distribution of COO subsets with the clusters of Chapuy *et al.*[9] (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.*[9] 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) (HR: 1.8; 95%CI: 1.2-2.6; *P*=0.0033) (Figure 5D) .

Discussion

In this 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 SNV and CNA in 465 cancer-related genes and 31 select gene rearrangements in 499 patients. This is the largest prospectively collected dataset in DLBCL so far. 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 independent 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 SNV. The co-occurrence of *BCL2* SNV with *BCL2* translocations, possibly as a consequence of aberrant somatic hypermutation,[23] may partially explain the negative prognostic

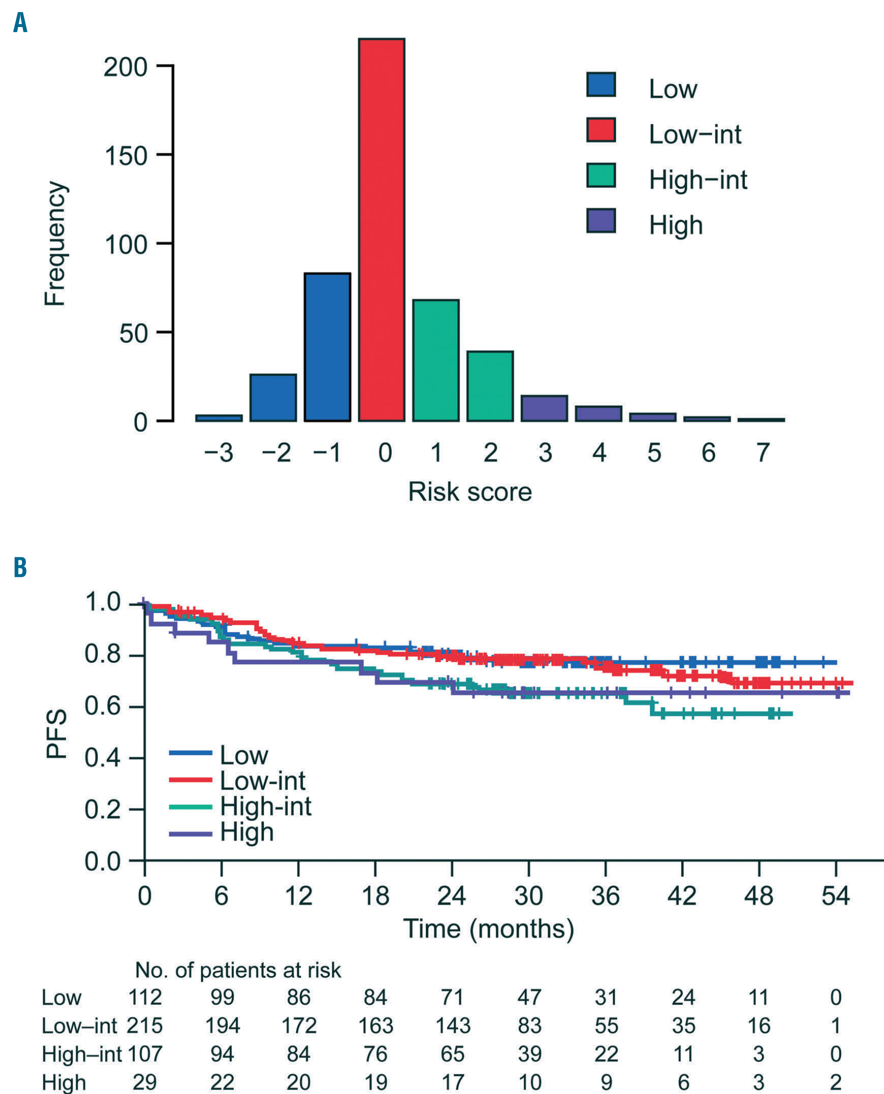

Figure 4. (A) Distribution of risk scores using the applied Reddy *et al.*[19] prognostic model, and (B) progression-free survival (PFS) by risk group (n=443). int: intermediate.

impact of *BCL2* SNV, although the negative prognostic effect of BCL2 SNV 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 pre-defined 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;

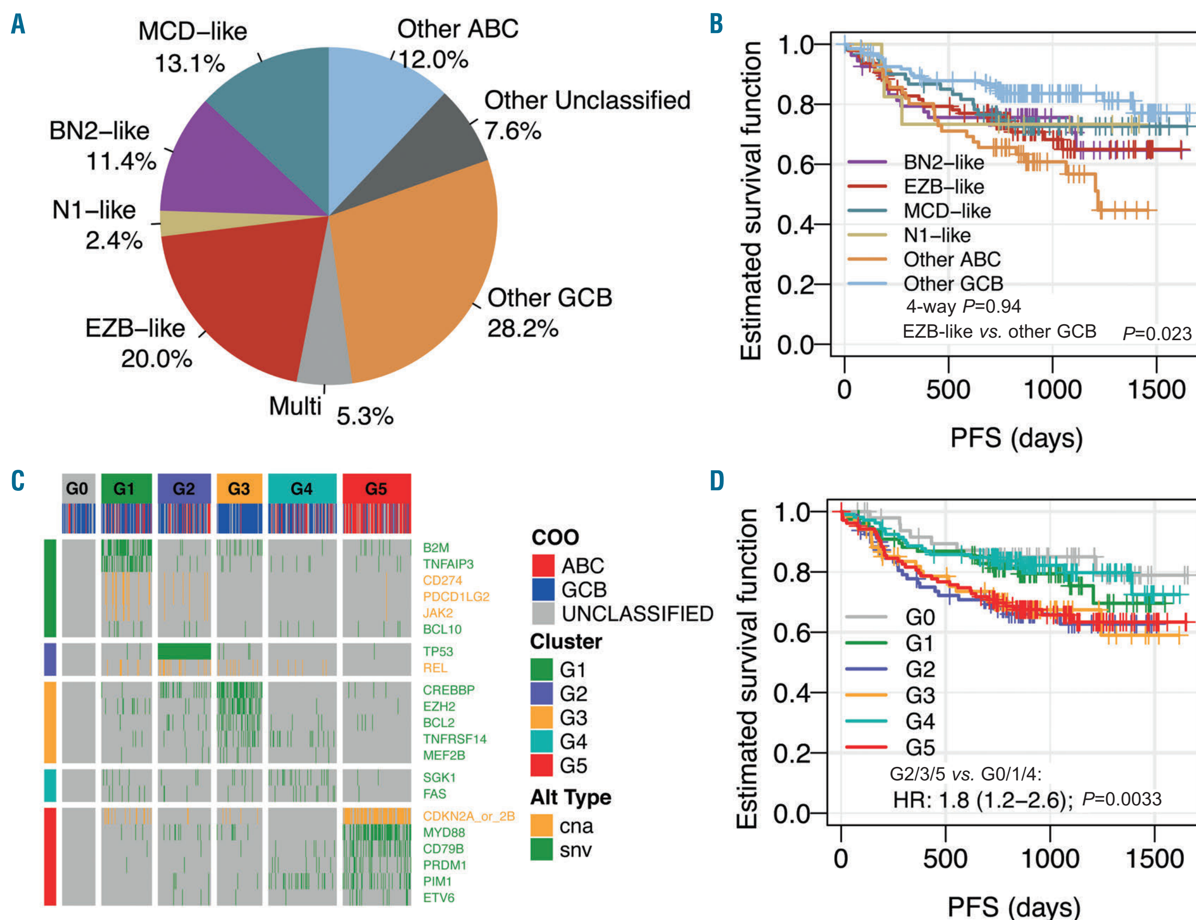

Figure 5. Diffuse large B-cell lymphoma (DLBCL) mutational subset validation. (A) Prevalence and (B) association of Schmitz *et al.*[8] classifications with progression-free survival (PFS). Schmitz clusters were approximated using the seed mutations: EZB - *EZH2* or *BCL2*; BN2 - *BCL6* or *NOTCH2*; N1 - *NOTCH1*; MCD - *MYD88, L265P* or *CD79B*; Multi: multiple seed mutations from more than one cluster. (C) Chapuy *et al.*[9] clusters were approximated by application of non-negative matrix factorization (NMF) to the GOYA Foundation Medicine Incorporated (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; GCB: germinal center B-cell-like; HR: hazard ratio; SNV: single nucleotide variant.

95%CI: 0.42-0.88; *P*<0.01), when corrected for COO, the model was no longer significant (HR: 0.77; 95%CI: 0.49-1.2; *P*=0.27), indicating that it provided little additional benefit over the most commonly used gene expression profiling and fluorescence *in situ* hybridization 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.*[19] model in its entirety due to some differences in gene availability on the FMI platform, and for the fact that Reddy *et al.*[19] 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.*[8] and Chapuy *et al.*[9] 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 COO. 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.[9] 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 *et al.*[8] and Chapuy *et al.*,[9] 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 the FMI 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 the FMI 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 independent 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. Further details on Roche's criteria for eligible studies are available here (https://vivli.org/members/ourmembers/). 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).

Acknowledgments

The authors would like to thank the GOYA study team investigators, coordinators, nurses and patients.

Funding

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

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 DW, 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.