SUPPLEMENTARY APPENDIX

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 Trnény,[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.*

 doi:10.3324/haematol.2019.227892

Received: June 14, 2019.

Accepted: November 14, 2019.

Pre-published: November 14, 2019.

Correspondence: *CHRISTOPHER R. BOLEN* - bolen.christopher@gene.com

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.

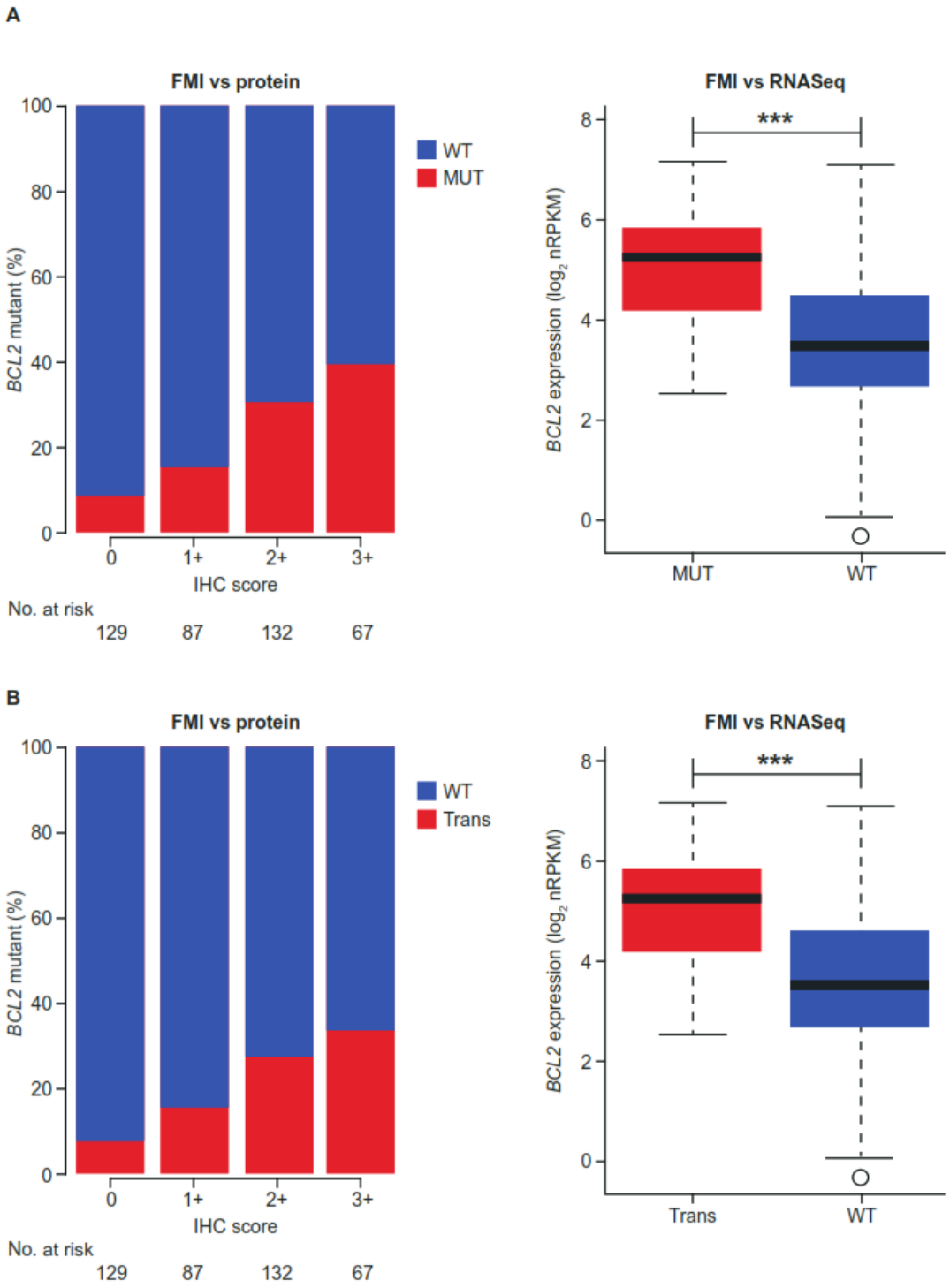

Figure S2. DLBCL mutational subset validation. Chapuy *et al.* clusters were approximated by application of non-negative matrix factorization (NMF) to the GOYA FMI dataset and selecting five clusters (G1-G5). ABC: activated B-cell-like; ALT: alteration; CNA: copy number abnormality; COO: cell of origin;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