Next Generation Sequencing in Diffuse Large B Cell Lymphoma Highlights Molecular Divergence and Therapeutic Opportunities: a LYSA Study

Sydney Dubois1, Pierre-Julien Viailly1,2, Sylvain Mareschal1, Elodie Bohers1, Philippe Bertrand1, Philippe Rumin1, Catherine Maingonnat1, Jean-Philippe Jais3, Pauline Peyrouze4, Martin Figeac4, Thierry J Molina5, Fabienne Desmots6, Thierry Fest6, Corinne Haioun7, Thierry Lamy6, Christiane Copie-Bergman8, Josette Brière9,10, Tony Petrella11, Danielle Canioni12, Bettina Fabiani13, Bertrand Coiffier14, Richard Delarue15, Frédéric Peyrade16, André Bosly17, Marc André17, Nicolas Ketterer18, Gilles Salles14, Hervé Tilly1, Karen Leroy19, Fabrice Jardin1

1 Inserm U918, Centre Henri Becquerel, Université de Rouen, IRIB, Rouen, France 2 LITIS EA 4108, Normandie Université, Rouen, France 3 Inserm UMRS 872, AP-HP Hôpital Necker, Paris, France 4 Université de Lille 2, Lille, France 5 Pathology, AP-HP, Hôpital Necker, EA 7324, Université Paris Descartes, France 6 Inserm U917, CHU Pontchaillou, Rennes, France 7 Unité Hémopathies Lymphoïdes, AP-HP Hôpital Henri Mondor, Créteil, France 8 IMRB – Inserm U955, AP-HP Hôpital Henri Mondor, Créteil, France 9 Inserm U728, Université Paris Diderot, Sorbonne Paris Cité 10 Department of Pathology, AP-HP Hôpital Saint-Louis, Paris, France 11 Department of Pathology, Hôpital Maisonneuve-Rosemont, Montréal (QC), H1T2M4, Canada 12 Laboratoire de Pathologie, AP-HP Hôpital Necker, Paris, France 13 Laboratoire de Pathologie, AP-HP Hôpital Saint Antoine, Paris, France 14 CNRS UMR5239, Hospices Civils de Lyon, Lyon, France 15 Department of Hematology, AP-HP Hôpital Necker, Paris, France 16 Department of Hematology, Lacassagne Center, Nice, France 17 CHU Dinant Godinne, UcL Namur, Yvoir, Belgium 18 Department of Oncology, Lausanne Hospital, Lausanne, Switzerland 19 Inserm U955 Team 09, AP-HP Hôpital Henri Mondor, Créteil, France

Running Title: Targeted NGS details DLBCL divergence and actionable targets

Corresponding author: Fabrice Jardin
Address: 1 rue d’Amiens, Centre Henri Becquerel, Rouen, France
Email: fabrice.jardin@chb.unicancer.fr
Phone: +33 (0)232082465
Fax: +33 (0)232082455

Keywords: Diffuse Large B Cell Lymphoma, Primary Mediastinal B-cell Lymphoma, Next Generation Sequencing

Declaration of Interests: The authors disclose no potential conflicts of interest.

Total word count: 3866, Abstract word count: 247, Number of figures and tables: 6
Translational Relevance

This is the first Next Generation Sequencing (NGS) study of such a large, prospective cohort using a targeted gene panel, the Lymphopanel, focusing on genes identified as important for lymphomagenesis or whose potential has been pinpointed in Whole Exome Sequencing studies. Particular attention has been paid to the inclusion of actionable targets in the Lymphopanel, highlighting potential candidate patients for novel personalized therapies. This study further details subtype-enriched gene and pathway mutations in order to optimize DLBCL treatment, and highlights several novel, frequent, and actionable mutations, notably in PMBL. In addition, Lymphopanel NGS has enabled the detection of novel clinical and prognostic correlations, potentially impacting treatment decisions. As benchtop sequencers become increasingly available in academic research and routine clinical settings, our study demonstrates the feasibility and impact of NGS with a consensus gene panel in DLBCL patient management, contributing essential information to today’s precision therapy era treatment decisions.
Abstract

**Purpose:** Next Generation Sequencing (NGS) has detailed the genomic characterization of Diffuse Large B Cell Lymphoma (DLBCL) by identifying recurrent somatic mutations. We set out to design a clinically feasible NGS panel focusing on genes whose mutations hold potential therapeutic impact. Furthermore, for the first time, we evaluated the prognostic value of these mutations in prospective clinical trials.

**Experimental Design:** A Lymphopanel was designed to identify mutations in 34 genes, selected according to literature data and a whole exome sequencing study of relapsed/refractory DLBCL patients. The tumor DNA of 215 patients with CD20+ *de novo* DLBCL in the prospective, multicenter and randomized LNH-03B LYSA clinical trials was sequenced to deep, uniform coverage with the Lymphopanel. Cell of origin molecular classification was obtained through gene expression profiling with HGU133+2.0 Affymetrix GeneChip arrays.

**Results:** The Lymphopanel was informative for 96% of patients. A clear depiction of DLBCL subtype molecular heterogeneity was uncovered with the Lymphopanel, confirming that Activated B Cell-like (ABC), Germinal Center B-cell like (GCB) and Primary Mediastinal B-cell Lymphoma (PMBL) are frequently affected by mutations in NFkB, epigenetic, and JAK-STAT pathways respectively. Novel truncating immunity pathway, *ITPKB*, *MFHAS1* and *XPO1* mutations were identified as highly enriched in PMBL. Notably, *TNFAIP3* and *GNA13* mutations in ABC patients treated with R-CHOP were associated with significantly less favorable prognoses.

**Conclusions:** This study demonstrates the contribution of NGS with a consensus gene panel to personalized therapy in DLBCL, highlighting subtypes’ molecular heterogeneity and identifying somatic mutations with therapeutic and prognostic impact.
Introduction

Diffuse large B-cell lymphoma (DLBCL) is the most common form of adult lymphoma worldwide, accounting for 30-40% of newly diagnosed Non-Hodgkin Lymphoma (NHL)(1). Gene expression profiling (GEP) has made strides in deciphering the molecular heterogeneity of DLBCL, enabling the entity’s subdivision into three main molecular subtypes: Germinal Center B-cell like (GCB), Activated B-Cell like (ABC), and Primary Mediastinal B-cell Lymphoma (PMBL)(2,3). Among other genetic aberrations, the GCB subtype is characterized by t(14;18)(q32;q21) translocations(4) and loss of PTEN(5), while the ABC subtype is characterized by t(3;14)(q27;q32) translocations, deletion of the INK4A-ARF locus(6) and BCL2 amplification(7). PMBL displays strong molecular similarities to classical Hodgkin Lymphoma (cHL), exhibiting frequent amplifications of JAK2 and deletions of SOCS1(8,9).

Arising from B cells at distinct stages of differentiation and maturation, these subtypes are also diverse in clinical presentation, response to immunochemotherapy and outcome, with the ABC subtype having the most unfavorable prognosis(2,7). As targeted therapies become increasingly widespread, it is essential to thoroughly characterize each molecular subtype, in order to ensure optimal care for each patient. Unfortunately, the cell of origin (COO) classification of DLBCL still has relatively little influence on clinical practice, as GEP is poorly transposable to routine diagnosis and surrogate immunohistochemical algorithms remain unreliable(10), although recent techniques offer better bench to bedside translation(11,12).

Next generation sequencing (NGS) technologies, enabling high-throughput DNA sequencing, have emerged over the past decade and have provided new insights into the genomic characterization of DLBCL by identifying recurrent single nucleotide variants (SNVs), which can be enriched in a particular subtype(13–22). Although DLBCL is defined by widespread genetic heterogeneity, several of the pinpointed recurrent SNVs warrant special interest as they occur in actionable targets and/or correlate with antitumoral response. Equally of note, despite their probable physiopathological relevance, neither the prognostic value of these mutations nor their potential to predict resistance to conventional B-cell lymphoma immunochemotherapy treatment has been properly evaluated in prospective clinical trials. Furthermore, mutation feature analyses have identified activation-induced cytidine deaminase (AID)-driven somatic hypermutation (SHM) as a mechanism mediating the development of
some of these recurrent SNVs, rendering the task of distinguishing driver from passenger mutations even more challenging(23).

To more feasibly identify molecular subtypes and reach the goal of precision therapy in DLBCL, a limited and clinically relevant panel of target genes must be designed, with efforts made to thoroughly describe the SNVs they harbor. Furthermore, compact sequencers readily available in academic laboratories must be used in order to generalize this practice and apply it to routine clinical disease management. With this idea in mind, we developed a Lymphopanel NGS assay, designed to identify mutations in 34 genes important for lymphomagenesis based on a literature review of WES studies in DLBCL(13,14,20,24) as well as a WES study of relapsed/refractory DLBCL cases(25). In the current study, we sequenced 215 patients with de novo DLBCL enrolled in the prospective, multicenter and randomized LNH-03B LYSA trials, using the Lymphopanel. Our intent was to determine whether the Lymphopanel was informative enough to identify targetable SNVs and/or pathway mutations that might alter treatment decisions, highlight subtype-specific mutational profiles, distinguish genes impacted by AID, and potentially foretell clinical outcome.

Materials and Methods

Patients

215 adult patients with de novo CD20+ DLBCL enrolled in the prospective, multicenter and randomized LNH-03B LYSA trials with available frozen tumor samples, centralized histopathological review and adequate DNA/RNA quality were selected. The LNH03-B LYSA trials were initiated in 2003 and included 1704 patients overall in six distinct clinical trials (Supplementary Methods). COO molecular classification was obtained with HGU133+2.0 Affymetrix GeneChip arrays (Affymetrix), grouping patients into ABC, GCB, PMBL and “other”, also referred to as unclassified (detailed in Supplementary Methods). Clinical features of the patients are indicated in Table S1. The study was performed with approval of an institutional review board and written informed consent was obtained from all participants at the time of enrollment.

NGS experiments and data analysis

The Lymphopanel was designed to identify mutations in 34 genes important for lymphomagenesis, based on literature data (Table S2)(24) and WES of relapsed/refractory
DLBCL sequencing(25). The design covers 87,703 bases using 872 amplicons. Genes were grouped into 8 specific pathways: Immunity (CIITA, B2M, TNFRSF14 and CD58), NOTCH (NOTCH1 and NOTCH2), Apoptosis/Cell cycle (MFHAS1, XPO1, MYC, CDKN2A/B, FOXO1, TP53, GNA13 and BCL2), NFkB (TNFAIP3, MYD88, PIM1, CARD11, IRF4 and PRDM1), Epigenetic Regulation (EZH2, KMT2D, EP300, MEF2B and CREBBP), MAP Kinase (BRAF), JAK-STAT (SOCSI and STAT6), and BCR (CD79A/B, ITPKB, TCF3 and ID3) (Table S3).

Ion Torrent Personal Genome Machine (PGM) Sequencing and PGM data analysis was performed as previously described(26) and detailed in Supplementary Methods. Variant analysis was performed using an in-house generated bioinformatics pipeline, detailed in Supplementary Methods and Figures S1 and S2. This pipeline ensured excellent filtering of artefacts and polymorphisms, as proven by an independent study of seven non-tumoral samples from DLBCL patients with a mean error rate of non-filtered variants of only 0.7% (Table S4).

Statistical methods

All statistical analyses were performed using R software version 3.1.2(27). Progression-Free survival (PFS) was evaluated from the date of enrolment to the date of disease progression, relapse, re-treatment or death from any cause. Overall survival (OS) was evaluated from the date of enrolment to the date of death from any cause. Multivariate Cox and log-rank tests (“survival” R package version 2.37.7) were used to assess differences in OS and PFS rates calculated by Kaplan-Meir estimates. Ward hierarchical unsupervised clustering (“LPS” R package version 1.0.10) was performed using Jaccard distance. Statistical differences between all other parameters were determined using χ², Mann-Whitney or Fisher’s exact tests when appropriate. P values or False Discovery Rates (FDR) < 0.05 were considered statistically significant.

Results and Discussion

Lymphopanel Variant Features

Lymphopanel NGS was performed on 215 DLBCL patients (81 ABC, 83 GCB, 33 “other” and 18 PMBL). The median overall sequencing depth was 225x [110x-331x]. An initial 13,965 variants were filtered for quality, SNPs and functional relevance, leading to 1064 (7.6%) variants ultimately validated (Table S5 and Figure S1). The Lymphopanel
informativity was of 96%, with 206 patients (78 GCB, 80 ABC, 30 “other” and 18 PMBL) presenting at least one variant. All genes in the Lymphopanel were informative for at least one patient, with the number of variants per gene ranging from one to 124. Gene mutation frequencies in the total cohort and by subtype are presented in Figure 1.

We subdivided the genes targeted by the Lymphopanel into eight specific pathways (Figure 2, Table S3). As expected, we confirmed that the ABC subtype is dominated by NFkB pathway mutations (45% of total variants), followed by epigenetic regulation pathway mutations (20.2% of total variants) (Figure 2A), whereas the GCB subtype is mostly characterized by mutations in the epigenetic regulation pathway (32.3% of total variants) and the apoptosis/cell cycle pathway (26.3% of total variants) (Figure 2B). Mutations in the JAK-STAT, immunity and apoptosis/cell cycle pathways were prevalent in PMBL (29.1%, 20.9% and 20.9% of total variants respectively) (Figure 2C). Interestingly, PMBL presented a significantly higher number of variants per sample than the other subtypes combined (p=2.76x10^-5, Figure S3). The “other” subtype showed a fairly hybrid spectrum of pathway mutations (Figure 2D), perhaps suggesting the involvement of the three main subtypes, rather than the existence of a distinct entity.

We sought to assess the impact of AID involvement in the Lymphopanel genes (detailed in Supplementary Methods and Figure S4). Genes with an AID mutation frequency superior to the average of that of the Lymphopanel are presented in Table S6. Previously identified SHM targets in DLBCL among these genes include MYC, PIM1, IRF4, BCL2, SOCS1 and CIITA(23). Potentially novel SHM targets include MFHAS1, PRDM1, GNA13, MYD88, ITPKB and NOTCH2. There was no significant difference in AID involvement between DLBCL subtypes.

**Lymphopanel NGS identifies mutations with potential treatment impact**

The Lymphopanel included genes whose mutations could serve to stratify patients according to treatment options. This includes actionable mutations, currently targeted by precision therapies, highlighted in Figure 1, as well as mutations whose presence might call into question the use of certain targeted therapy options.

Recent results have shown that doubly mutated MYD88/CD79B patients are significantly more responsive to single therapy Ibrutinib, a BTK inhibitor, highlighting the importance of targeted NGS for DLBCL patients in the clinical setting(28). Furthermore, the presence of CARD11 and TNFAIP3 mutations has been shown to lead to decreased activity of both Ibrutinib and Sotrastaurin, a Protein Kinase C inhibitor(29,30). MYD88 mutations are frequent
in ABC DLBCL, L265P being the most prevalent variant (up to 29% in previous studies)(18). In our cohort, MYD88 mutations were significantly enriched in ABC patients (28.4%, FDR=4.7x10^{-3}), with the L265P mutation present in 21% of ABC patients (Figures 3 and S5A). Of 17 ABC patients with MYD88 L265P mutations, we found ten (58.8%) with associated CD79B mutations, who might respond more favorably to Ibrutinib treatment. CD79B mutations were identified in 10.7% of total patients, slightly less than previously reported, but were significantly enriched in ABC DLBCL (24.7%, FDR=3.2x10^{-5}), corroborating previous studies(13,31). Of note, two (11.8%) MYD88 L265P mutated ABC patients were CARD11 mutated and two were TNFAIP3 mutated, potentially diminishing the effect of Ibrutinib and Sotrastaurin, while six (35.3%) were PRDM1 mutated (Figure S6).

ABC cells have been shown to be addicted to IRF4 activity (32), responsible for Lenalidomide treatment response. 21 IRF4 variants were discovered in 7.9% of patients (13.6% of ABC, 4.8% of GCB, 0% of “other” and 11.1% of PMBL) (Figure 1). Although the impact on expression of these mutations was not analyzed, only one IRF4 variant was a stop-gain mutation, indicating that expression is most likely at least conserved. Most of the non-synonymous SNVs were clustered around amino acid positions 18 (two S18 variants and two G20 variants) and 99 (four C99R variants, SIFT scores=0) (Figure 4). Interestingly, IRF4 mutations were almost equally frequent in PMBL and ABC in our cohort. Although IRF4 addiction has not been demonstrated in PMBL, this predilection for IRF4 mutations, added to the well-known high expression of IRF4 in this subtype, may suggest a potential role for Lenalidomide treatment in this subtype as well, calling for further investigation of these mutations’ effects.

Recently, small molecule inhibitors of PIM kinases have been developed and studied in DLBCL. In our cohort, PIM1 mutations were significantly skewed toward ABC (33.3%, FDR=3.3x10^{-5}) in our cohort and represented 20.2% of ABC variants (Figures 2A, 3 and S5A). Of note, truncating PIM1 variants were predominant in ABC patients (13 of 14 identified). Interestingly, DLBCL sensitivity to PIM kinase inhibitors is not correlated with the level of PIM kinase expression (33), suggesting that even expression-modifying PIM1 mutations might not necessarily hamper the efficacy of PIM1 inhibitors, although this remains to be verified. In any case, the high frequency of PIM1 mutations in our cohort seems to justify the pre-screening of PIM kinase inhibitor candidate patients by PIM1 sequencing, in order to identify those most likely to benefit from this targeted therapy.

The most highly mutated gene in GCB, BCL2(34), can notably be targeted by BH3-mimetics. Among our GCB cohort, BCL2 was the third most frequently mutated gene (24.1%,
FDR=2.5x10^{-5}) (Figures 3 and S5B), with mutations most frequently impacting the amino terminus and between the BH1 and BH3 domains. The BH3 domain was not impacted, indicating that BH3 mimetic activity would not be hampered in these patients (Figure S7).

Targeting epigenetic changes is also being increasingly explored in the realm of tailored therapy, notably among the GCB subtype, justifying the inclusion of genes such as *EZH2*, *CREBBP*, *KMT2D* and *EP300* within the Lymphopanel. In our cohort, 18.1% of GCB patients were *EZH2* mutated, mostly at the Y641 hotspot, and *EZH2* mutations were enriched in GCB (FDR=1.8x10^{-3}), as observed in previous studies, with no *EZH2* mutated ABC patients(13) (Figures 3 and S5B). Interestingly, one PMBL patient presented an *EZH2* Y641 mutation, adding to the debate of which DLBCL subtypes might benefit from *EZH2* inhibitor treatment(26). *CREBBP* mutations were identified in 31.3% of GCB patients and 6.2% of ABC patients, corroborating previous studies(17), but were also found in 24.2% of “other” and 11.1% of PMBL (Figures 1 and S5). *KMT2D* mutations, evenly distributed throughout the coding sequence, were identified in 40.9% of our cohort, with *KMT2D* being the most frequently mutated gene in all subtypes except PMBL (Figure 1). Of the 124 variants, 78 were truncating mutations, indicating bias toward loss of protein expression.

Mutations of master regulators *TP53* and *MYC* showed no significant difference in subtype distribution in our cohort (Figure 3). *TP53* variants were present in 14.9% of total patients (slightly less than previously reported(13,35)) (Figure 1) and several recurrent variants were identified, including six R248 variants, shown to be loss of function mutants in DLBCL(35). As for *MYC*, we discovered mutations in 6.5% of total patients (4.9% of ABC, 9.6% of GCB, 3% of “other” and 5.6% of PMBL). No mutations were identified at the known hotspots in Burkitt Lymphoma (BL), including S62 and T58.

Finally, although frequent mutations are most enticing for therapeutic potential, rare mutations can also be potentially very effectively actionable. For instance, *BRAF* and *NOTCH1* mutations have been identified as rare but functionally relevant in DLBCL(13). In our cohort, we identified one patient with an activating V600E *BRAF* variant. As for *NOTCH1*, ten patients in our cohort presented mutations (Figure 1). *TCF3* was mutated in three patients, all of whom harbored a pathogenic N551K variant, which has shown altered DNA-binding specificity(36). The negative regulator of *TCF3*, *ID3*, was mutated in 4.7% of patients (Figure 1), with most variants affecting the HLH domain (Figure S7), potentially impairing its ability to inactivate *TCF3*, as shown in BL(36). *ID3* and *TCF3* mutations have recently been shown to be of comparable frequency in BL and DLBCL, highlighting the overlap between these two entities(37).
The Lymphopanel highlights a unique PMBL mutational signature

STAT6 and SOCS1 mutations were both very significantly enriched in PMBL (FDR=6.8x10^{-14} and FDR=2.5x10^{-5} respectively), with STAT6 being the most frequently mutated gene in PMBL (72.2%, Figures 3 and S5C). All STAT6 mutations were non-synonymous SNVs and recurrent variants were frequent among PMBL, including seven N417 variants (Figure S7). SOCS1 mutations were present in 55.5% of PMBL patients, 30% of them harboring SOCS1 inactivating mutations (Figure S5C). TNFAIP3 mutations have also been frequently identified in PMBL (up to 36%)(38). We found TNFAIP3 mutations to be significantly enriched in PMBL (61.1%, FDR=3.1x10^{-5}), with all but one TNFAIP3 mutated PMBL patient harboring truncating mutations (Figures 3 and S5C). The frequent TNFAIP3 and SOCS1 mutations in PMBL highlight the similarities between PMBL and cHL(38,39). Furthermore, no MYD88, CD79B or CARD11 mutations were observed in PMBL patients (Figure 1), suggesting a PMBL dependence on TNFAIP3 mutations for the constitutive activation of NFkB. Interestingly, significant GNA13 mutation enrichment was shown in PMBL rather than GCB in our cohort (50%, FDR=2.8x10^{-4}) (Figures 3 and S5), and one third of GNA13 variants were potentially truncating mutations (Figure 4).

We discovered 46 B2M variants in 18.1% of total patients, significantly enriched in PMBL (50%, FDR=1.2x10^{-3}) (Figures 1, 3 and S5C). Six of nine B2M mutated PMBL patients harbored truncating mutations and the remaining three presented pathogenic non-synonymous SNVs, two of which were a highly recurrent p.M1R variant predicted to affect the start codon (Figure 4). Recently, frequent B2M mutations leading to lack of protein expression were identified in cHL(9). Our similar findings in PMBL further highlight the molecular similarities between these two entities. CD58 variants were also significantly enriched in PMBL (FDR=1.2x10^{-3}) (Figures 3 and S5), with five of seven CD58 mutated PMBL patients exhibiting truncating mutations, potentially leading to a lack of CD58 expression. Importantly, PD-1 blockade has recently shown promise as an effective treatment of relapsed or refractory cHL(40). Escape from T cell immunity via B2M and/or CD58 mutations could potentially affect the activity of PD-1 inhibitors in PMBL and cHL and warrants further investigation. Furthermore, we found 39 CIITA variants, enriched in PMBL as well (55.6% of PMBL versus 14.14% of total patients, FDR=3.1x10^{-5}) (Figures 1, 3 and S5C). To our knowledge, this is the first study to identify truncating immunity pathway mutations as major in PMBL, although CIITA/PDL1-2 translocations had been shown to impact PMBL survival(41).
ITPKB, MFHAS1 and XPO1 mutations were identified as significantly overrepresented in a relapsed/refractory DLBCL patient cohort (25), leading to their inclusion in the Lymphopanel. In our cohort, ITPKB, MFHAS1 and XPO1 mutations were all identified for the first time as being significantly enriched in PMBL patients (FDR=1.4x10^{-2}, FDR=3.9x10^{-3} and FDR=4.8x10^{-10} respectively, Figure 3). The roles of ITPKB and MFHAS1 in lymphomagenesis are not yet fully elucidated. ITPKB regulates B cell survival and function, and ITPKB deficiency leads to B cell antigen presentation unresponsiveness and decreased B cell survival, associated with specific overexpression of proapoptotic Bim protein (42). We found 48 ITPKB variants, present in 14% of total patients (Figures 1 and S5). ITPKB mutations were uniformly distributed along the protein sequence, excepting the seemingly conserved catalytic IPK domain, affected by only two non-synonymous SNVs (Figure 4). MFHAS1 is a potential oncogene with highly conserved leucine-rich tandem repeats, isolated from a minimal common amplified region in malignant fibrous histiocytomas at 8p23.1. MFHAS1 is also a target gene for chromosomal translocation t(8;14)(p23.1;q21) in B-cell lymphoma cell lines and has been shown to be tumorigenic in nude mice (43). In our cohort, MFHAS1 variants were identified in 7.9% of total patients, and were distributed homogeneously along the coding sequence, with no recurrent mutations identified (Figure 4). Although the role of these potential oncogenes in lymphomagenesis is still unclear, their high mutation frequency in PMBL especially warrants further investigation. As for XPO1, it encodes CRM1, an exporter of several tumor suppressor proteins (TSPs). Cytoplasmic export of TSPs renders them inactive, indicating that XPO1 acts as a proto-oncogene. XPO1 mutations were present in 4.7% of total patients (1.2% of ABC, 1.2% of GCB, 3% of “other” and 38.9% of PMBL) (Figures 1 and S5). The mutation identified is recurrent, with all variants affecting E571 (Figure 4), as previously found in rare CLL cases (<5%) (44). Importantly, Selective Inhibitors of Nuclear Export (SINEs), have been shown to effectively target CRM1 and retain TSPs in the nucleus (45). Given the frequency of XPO1 E571 mutations in PMBL, further investigation concerning their impact on response to SINEs is warranted (Jardin et al, under review).

Identification of gene mutations correlated with clinical characteristics and prognosis
We tested each gene in the Lymphopanel for correlation with clinical characteristics including age, stage, IPI, bone marrow involvement and presence of bulky disease. CD79B, KMT2D and MYD88 mutations were significantly correlated with age (Figure S8). For CD79B and MYD88 mutations, enriched in ABC, this might be linked to the significantly
higher age of ABC patients \( (p=1.01\times 10^{-7}) \) in our cohort(46). As for \textit{KMT2D} mutations, homogenous among subtypes, this result potentially suggests an accumulation of passenger-type mutations, particularly visible in this long gene \( (16.6 \text{ kb sequenced}) \). \textit{MYD88} mutations were also significantly correlated with higher IPI \( (\text{FDR}=0.04) \).

On the other hand, PMBL patients were significantly younger \( (p=1.21\times 10^{-8}) \) and the large majority of genes whose mutations were significantly enriched in PMBL were also correlated with younger age (Figure S8). In addition, \textit{B2M} and immunity pathway mutations were significantly inversely correlated with Ann Arbor stage \( (\text{FDR}=3.2\times 10^{-3} \text{ and FDR}=0.045 \text{ respectively}) \), and \textit{B2M} and \textit{STAT6} mutations were significantly inversely correlated with IPI \( (\text{FDR}=5.5\times 10^{-4} \text{ and FDR}=0.026 \text{ respectively}) \).

As expected, OS and PFS decreased with increasing IPI and ABC patients displayed significantly worse OS and PFS than GCB patients in the R-CHOP treatment group \( (p=1.9\times 10^{-2} \text{ and } p=6.6\times 10^{-3} \text{ respectively}) \) (Figure S9). Of note, PMBL patients in our cohort did not present the favorable prognosis typically associated with this subtype(47), perhaps due to the lack of dedicated treatment.

Prognostic impact was assessed for all Lymphopanel genes and subtypes, among patients treated with R-chemotherapy, separated into an R-CHOP group and an R-ACVBP group. Analyses over all subtypes were performed using multivariate Cox tests, while analyses of each subtype individually were performed using log-rank tests. In both cases, independent corrections of p-values were performed for OS and PFS. Gene mutations harboring prognostic impacts with uncorrected p-values \( < 0.05 \) are highlighted in Table 1 and Figure S10, and all survival analyses are presented in Table S7. We set an FDR threshold of 0.05 to detect gene mutations with significant prognostic impacts.

Among ABC DLBCL treated with R-CHOP, \textit{TNFAIP3} mutations were significantly associated with lower OS \( (\text{FDR}=7.86\times 10^{-4}) \) and PFS \( (\text{FDR}=3.04\times 10^{-2}) \), while \textit{GNA13} mutations were significantly associated with lower PFS \( (\text{FDR}=3.04 \times 10^{-2}) \), (Figure 5). Of note, the poor prognostic impact of these mutations was not observed in patients treated with R-ACVBP, which, for \textit{TNFAIP3} at least, might be linked to the improved survival observed in younger patients with DLBCL treated with R-ACVBP(48). These findings need to be confirmed in an independent cohort of patients, given the small sample size.

Unfortunately, the prognostic impact of \textit{TNFAIP3} mutations in PMBL patients could not be assessed due to the existence of only one WT \textit{TNFAIP3} PMBL patient treated with R-CHOP (Table S7). Moreover, previously described prognostic impacts of \textit{TP53}, \textit{FOXO1} or \textit{MYD88}...
mutations were not confirmed\cite{35,49,50}, although the impact of specific deleterious mutations was not analyzed on its own.

In conclusion, NGS is increasingly accessible in academic research settings, but its possibilities in routine clinical settings have yet to be thoroughly harnessed. This study has shown that, by using a restricted set of genes, not only can well-known mutations be tracked, but novel or rare mutations can also be identified and potentially targeted as well. Detailing subtype-enriched gene and pathway mutations is critical for optimal understanding and treatment of DLBCL: here, we have added to the field’s current knowledge of the genetic heterogeneity between DLBCL subtypes, and identified novel clinical and prognostic correlations, which might also have an impact on therapeutic decisions. NGS with a consensus targeted panel could contribute valuable information to multidisciplinary meetings and constitute a molecular breakthrough in the way treatment decisions are made, by providing personalized mutational profiles of actionable targets. Furthermore, although this study was performed on frozen tumor samples, which are of limited availability, the Lymphopanel was designed to be effective in FFPE samples as well, further cementing its capacity to play an essential role in clinical disease management. In addition, our team has recently demonstrated that Lymphopanel NGS can be successfully performed using plasma-extracted cell-free circulating DNA, highlighting its feasibility and potential role in the monitoring of DLBCL\cite{51}. Taken together, these results serve as proof-of-principle that NGS with a consensus gene panel can alter the manner in which DLBCL patients are subclassified and treated.

Acknowledgements

This study was funded by grants from the Institut National du Cancer (INCA). We thank Marie-Hélène Delfau, Camille Laurent and Loïc Ysebaert for their critical review.

Contributors

SD, P-JV, SM, EB, PB, PR, J-PJ, MF, GS, HT, KL and FJ contributed to the conception and design of the study. TJM, FD, TF, CH, TL, CC-B, JB, TP, DC, BF, BC, RD, FP, AB, MA and NK collected clinical data. EB and CM performed NGS experiments. PP performed GEP experiments. SD, P-JV and SM performed statistical analyses and contributed to data interpretation. P-JV and SM developed the bioinformatics pipeline necessary for data
analysis. SD wrote the paper and all authors provided critical revisions and gave final approval to submit for publication.

Figure and Table Legends

Figure 1: Mutation frequencies in the total cohort and by subtype. Mutation frequencies in the total cohort are represented as a stacked bar chart. The number at the top of each bar represents the mutation frequency of the indicated gene in the total cohort. Bars are subdivided by DLBCL subtype, with segments proportional to mutation frequency. Potentially actionable targets are highlighted by indicating the appropriate class of molecule, which could potentially be used.

Figure 2: Mutation pathway heterogeneity among DLBCL subtypes. Pie-chart representations of mutation frequencies represented by pathway are represented for ABC (A), GCB (B), PMBL (C) and Other (D) subtypes. Mutation frequencies per gene and per subtype are shown here as the percentage of the total number of variants. Genes were grouped into 8 specific pathways: Immunity (CIITA, B2M, TNFRSF14 and CD58), NOTCH (NOTCH1 and NOTCH2), Apoptosis/Cell Cycle (MFHAS1, XPO1, MYC, CDKN2A/B, FOXO1, TP53, GNA13 and BCL2), NFKB (TNFAIP3, MYD88, PIM1, CARD11, IRF4 and PRDM1), Epigenetic Regulation (EZH2, KMT2D, EP300, MEF2B and CREBBP), MAP Kinases (BRAF), JAK-STAT (SOCS1 and STAT6), and BCR (CD79A/B, ITPKB, TCF3 and ID3).

Figure 3: Heatmap of mutation frequencies by subtype. Mutation frequencies in each subtype are indicated for each gene of the Lymphopanel. Mutation frequencies per gene and per subtype are shown here as the percentage of the total number of patients. Statistical significance of gene mutation enrichment in a given subtype is indicated by the FDR column. The horizontal line separates genes with FDR<0.05, considered significant, from genes with FDR≥0.05.

Figure 4. Mutations at the protein level. Protein representations were rendered using domains imported from the Pfam database (green). Exons were located using the CCDS database and were numbered automatically, not counting non-coding and spliced exons. Substitutions are depicted as diamonds, insertions as triangles and deletions as rectangles whose length varies according to the deletion size. A color code is also applied, with blue indicating non-synonymous substitutions or non-frameshift insertions, yellow indicating insertions and frameshift deletions, and pink indicating stop-gain and stop-loss mutations. Blue shading indicates sequenced regions.

Figure 5: TNFAIP3 and GNA13 mutations are linked to worse prognosis in R-CHOP-treated ABC patients. Survival analysis was performed on ABC patients treated with R-CHOP according to the presence or absence of TNFAIP3 mutation (A, B) or GNA13 mutation (C, D). Both TNFAIP3 and GNA13 mutations are associated with significantly less favorable prognosis in R-CHOP treated ABC patients.

Table 1. Prognostic impact of gene mutations among the Lymphopanel. Gene mutations with a tendency toward prognostic impact (uncorrected p<0.05) are indicated, for OS and/or PFS. The subtype and treatment cohort in which a prognostic impact was identified is shown, as is the number of mutant and WT patients for each survival analysis. HR column indicates Hazard Ratio. Bold type highlights significant FDR values.


generation sequencing and key elements in the precision medicine era. Leukemia &

exome sequencing of relapsed/refractory patients expands the repertoire of somatic
mutations in diffuse large B-cell lymphoma. Genes, chromosomes & cancer. 2015;

Immunohistochemical and genomic profiles of diffuse large B-cell lymphomas:
Implications for targeted EZH2 inhibitor therapy? Oncotarget. 2015.

computing. Vienna, Austria;

cell receptor signaling with ibrutinib in diffuse large B cell lymphoma. Nature

cell-like diffuse large B cell lymphoma cells towards ibrutinib treatment. Cancer cell
international. 2014;14:32.

inhibitor sotrastaurin selectively inhibits the growth of CD79 mutant diffuse large B-

Targetable activating mutations are very frequent in GCB and ABC diffuse large B-cell

synthetic lethality for the therapy of ABC diffuse large B cell lymphoma. Cancer cell.

kinases are progression markers and emerging therapeutic targets in diffuse large B-cell

34. Schuetz JM, Johnson Na, Morin RD, Scott DW, Tan K, Ben-Nierah S, et al. BCL2
mutations in diffuse large B-cell lymphoma. Leukemia. Nature Publishing Group;
2012;26:1383–90.

profile and prognostic significance of TP53 in diffuse large B-cell lymphoma patients
treated with rituximab-CHOP: a report from an International DLBCL Rituximab-

lymphoma pathogenesis and therapeutic targets from structural and functional


<table>
<thead>
<tr>
<th>Gene</th>
<th>Subtype</th>
<th>Cohort</th>
<th>Type</th>
<th>Mutant (n)</th>
<th>WT (n)</th>
<th>HR</th>
<th>p</th>
<th>FDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>TNFAIP3</td>
<td>ABC</td>
<td>R-CHOP</td>
<td>OS</td>
<td>6</td>
<td>48</td>
<td>9.04</td>
<td>9.03E-06</td>
<td>7.86E-04</td>
</tr>
<tr>
<td>TNFAIP3</td>
<td>ABC</td>
<td>R-CHOP</td>
<td>PFS</td>
<td>6</td>
<td>48</td>
<td>5.59</td>
<td>4.29E-04</td>
<td>3.04E-02</td>
</tr>
<tr>
<td>GNA13</td>
<td>ABC</td>
<td>R-CHOP</td>
<td>PFS</td>
<td>5</td>
<td>49</td>
<td>5.89</td>
<td>6.47E-04</td>
<td>3.04E-02</td>
</tr>
<tr>
<td>B2M</td>
<td>all</td>
<td>R-CHOP</td>
<td>PFS</td>
<td>16</td>
<td>100</td>
<td>0.12</td>
<td>6.24E-03</td>
<td>3.25E-01</td>
</tr>
<tr>
<td>B2M</td>
<td>all</td>
<td>R-CHOP</td>
<td>OS</td>
<td>16</td>
<td>100</td>
<td>0.11</td>
<td>3.13E-02</td>
<td>5.43E-01</td>
</tr>
<tr>
<td>IRF4</td>
<td>all</td>
<td>R-CHOP</td>
<td>OS</td>
<td>10</td>
<td>106</td>
<td>2.53</td>
<td>2.98E-02</td>
<td>5.43E-01</td>
</tr>
<tr>
<td>CD79B</td>
<td>all</td>
<td>R-CHOP</td>
<td>OS</td>
<td>19</td>
<td>97</td>
<td>0.31</td>
<td>3.13E-02</td>
<td>5.43E-01</td>
</tr>
<tr>
<td>CD79B</td>
<td>all</td>
<td>R-CHOP</td>
<td>PFS</td>
<td>19</td>
<td>97</td>
<td>0.29</td>
<td>2.15E-02</td>
<td>5.58E-01</td>
</tr>
<tr>
<td>CD58</td>
<td>all</td>
<td>R-CHOP</td>
<td>PFS</td>
<td>12</td>
<td>104</td>
<td>2.36</td>
<td>3.27E-02</td>
<td>5.67E-01</td>
</tr>
<tr>
<td>GNA13</td>
<td>ABC</td>
<td>R-CHOP</td>
<td>OS</td>
<td>5</td>
<td>49</td>
<td>3.68</td>
<td>1.36E-02</td>
<td>5.94E-01</td>
</tr>
<tr>
<td>FOXO1</td>
<td>GC</td>
<td>R-CHOP</td>
<td>OS</td>
<td>5</td>
<td>35</td>
<td>3.87</td>
<td>3.53E-02</td>
<td>7.72E-01</td>
</tr>
<tr>
<td>CD79B</td>
<td>ABC</td>
<td>R-CHOP</td>
<td>OS</td>
<td>16</td>
<td>38</td>
<td>0.35</td>
<td>4.44E-02</td>
<td>7.72E-01</td>
</tr>
<tr>
<td>CD79B</td>
<td>ABC</td>
<td>R-CHOP</td>
<td>PFS</td>
<td>16</td>
<td>38</td>
<td>0.33</td>
<td>3.07E-02</td>
<td>8.30E-01</td>
</tr>
</tbody>
</table>

Table 1. Prognostic impacts of gene mutations among the Lymphopanel
<table>
<thead>
<tr>
<th>Gene</th>
<th>ABC n = 81</th>
<th>GCB n = 83</th>
<th>PMBL n = 18</th>
<th>other n = 33</th>
<th>FDR</th>
</tr>
</thead>
<tbody>
<tr>
<td>STAT6</td>
<td>0%</td>
<td>14%</td>
<td>72%</td>
<td>6%</td>
<td>6.8e-14</td>
</tr>
<tr>
<td>XPO1</td>
<td>1%</td>
<td>1%</td>
<td>39%</td>
<td>3%</td>
<td>4.8e-10</td>
</tr>
<tr>
<td>SOCS1</td>
<td>6%</td>
<td>16%</td>
<td>56%</td>
<td>12%</td>
<td>2.5e-05</td>
</tr>
<tr>
<td>BCL2</td>
<td>1%</td>
<td>24%</td>
<td>0%</td>
<td>3%</td>
<td>2.5e-05</td>
</tr>
<tr>
<td>CIITA</td>
<td>12%</td>
<td>10%</td>
<td>56%</td>
<td>9%</td>
<td>3.1e-05</td>
</tr>
<tr>
<td>TNFAIP3</td>
<td>15%</td>
<td>11%</td>
<td>61%</td>
<td>15%</td>
<td>3.1e-05</td>
</tr>
<tr>
<td>CD79B</td>
<td>25%</td>
<td>2%</td>
<td>0%</td>
<td>3%</td>
<td>3.2e-05</td>
</tr>
<tr>
<td>PIM1</td>
<td>33%</td>
<td>8%</td>
<td>0%</td>
<td>6%</td>
<td>3.3e-05</td>
</tr>
<tr>
<td>GNA13</td>
<td>9%</td>
<td>12%</td>
<td>50%</td>
<td>12%</td>
<td>2.8e-04</td>
</tr>
<tr>
<td>CD58</td>
<td>6%</td>
<td>10%</td>
<td>39%</td>
<td>6%</td>
<td>1.2e-03</td>
</tr>
<tr>
<td>CREBBP</td>
<td>6%</td>
<td>31%</td>
<td>11%</td>
<td>24%</td>
<td>1.2e-03</td>
</tr>
<tr>
<td>B2M</td>
<td>9%</td>
<td>18%</td>
<td>50%</td>
<td>24%</td>
<td>1.2e-03</td>
</tr>
<tr>
<td>EZH2</td>
<td>0%</td>
<td>18%</td>
<td>6%</td>
<td>9%</td>
<td>1.8e-03</td>
</tr>
<tr>
<td>TNFRSF14</td>
<td>2%</td>
<td>17%</td>
<td>0%</td>
<td>24%</td>
<td>1.8e-03</td>
</tr>
<tr>
<td>MFHAS1</td>
<td>1%</td>
<td>10%</td>
<td>28%</td>
<td>9%</td>
<td>3.9e-03</td>
</tr>
<tr>
<td>MYD88</td>
<td>28%</td>
<td>10%</td>
<td>0%</td>
<td>15%</td>
<td>4.7e-03</td>
</tr>
<tr>
<td>ITPKB</td>
<td>9%</td>
<td>16%</td>
<td>39%</td>
<td>9%</td>
<td>1.4e-02</td>
</tr>
<tr>
<td>PRDM1</td>
<td>16%</td>
<td>6%</td>
<td>0%</td>
<td>3%</td>
<td>5.1e-02</td>
</tr>
<tr>
<td>NOTCH2</td>
<td>2%</td>
<td>10%</td>
<td>0%</td>
<td>15%</td>
<td>7.6e-02</td>
</tr>
<tr>
<td>IRF4</td>
<td>14%</td>
<td>5%</td>
<td>11%</td>
<td>0%</td>
<td>8.7e-02</td>
</tr>
<tr>
<td>MEF2B</td>
<td>12%</td>
<td>23%</td>
<td>11%</td>
<td>6%</td>
<td>1.4e-01</td>
</tr>
<tr>
<td>BRAF</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>3%</td>
<td>2.1e-01</td>
</tr>
<tr>
<td>FOXL1</td>
<td>4%</td>
<td>12%</td>
<td>6%</td>
<td>3%</td>
<td>2.1e-01</td>
</tr>
<tr>
<td>KMT2D</td>
<td>41%</td>
<td>46%</td>
<td>17%</td>
<td>42%</td>
<td>2.2e-01</td>
</tr>
<tr>
<td>CARD11</td>
<td>14%</td>
<td>7%</td>
<td>0%</td>
<td>9%</td>
<td>3.5e-01</td>
</tr>
<tr>
<td>NOTCH1</td>
<td>7%</td>
<td>1%</td>
<td>6%</td>
<td>6%</td>
<td>3.7e-01</td>
</tr>
<tr>
<td>CD79A</td>
<td>0%</td>
<td>2%</td>
<td>0%</td>
<td>0%</td>
<td>4.5e-01</td>
</tr>
<tr>
<td>TP53</td>
<td>13%</td>
<td>16%</td>
<td>11%</td>
<td>6%</td>
<td>4.6e-01</td>
</tr>
<tr>
<td>CDKN2B</td>
<td>0%</td>
<td>1%</td>
<td>0%</td>
<td>3%</td>
<td>5.4e-01</td>
</tr>
<tr>
<td>ID3</td>
<td>5%</td>
<td>2%</td>
<td>6%</td>
<td>9%</td>
<td>5.5e-01</td>
</tr>
<tr>
<td>MYC</td>
<td>5%</td>
<td>10%</td>
<td>6%</td>
<td>3%</td>
<td>5.5e-01</td>
</tr>
<tr>
<td>CDKN2A</td>
<td>2%</td>
<td>1%</td>
<td>0%</td>
<td>0%</td>
<td>7.4e-01</td>
</tr>
<tr>
<td>TCF3</td>
<td>1%</td>
<td>2%</td>
<td>0%</td>
<td>0%</td>
<td>7.4e-01</td>
</tr>
<tr>
<td>EP300</td>
<td>15%</td>
<td>14%</td>
<td>17%</td>
<td>18%</td>
<td>9.6e-01</td>
</tr>
</tbody>
</table>