

Refining diffuse large B-cell lymphoma subgroups using integrated analysis of molecular profiles

Sydney Dubois, Bruno Tesson, Sylvain Mareschal, Pierre-Julien Viailly, Elodie Bohers, Philippe Ruminy, Pascaline Etancelin, Pauline Peyrouze, Christiane Copie-Bergman, Bettina Fabiani, et al.

▶ **To cite this version:**

Sydney Dubois, Bruno Tesson, Sylvain Mareschal, Pierre-Julien Viailly, Elodie Bohers, et al.. Refining diffuse large B-cell lymphoma subgroups using integrated analysis of molecular profiles. EBioMedicine, 2019, 48, pp.58-69. 10.1016/j.ebiom.2019.09.034 . hal-02345428

HAL Id: hal-02345428

https://normandie-univ.hal.science/hal-02345428

Submitted on 18 Nov 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

EBioMedicine 48 (2019) 58–69

Contents lists available at ScienceDirect

EBioMedicine

journal homepage: www.elsevier.com/locate/ebiom

EBioMedicine
Published by THE LANCET

Research paper

Refining diffuse large B-cell lymphoma subgroups using integrated analysis of molecular profiles

Sydney Dubois[a], Bruno Tesson[b], Sylvain Mareschal[c], Pierre-Julien Viailly[a,d], Elodie Bohers[a], Philippe Ruminy[a], Pascaline Etancelin[a], Pauline Peyrouze[e], Christiane Copie-Bergman[f], Bettina Fabiani[g], Tony Petrella[h], Jean-Philippe Jais[i], Corinne Haioun[j], Gilles Salles[c], Thierry Jo Molina[k], Karen Leroy[l], Hervé Tilly[a], Fabrice Jardin[a,*], on behalf of Lymphoma Study Association (LYSA) investigators

[a] *Inserm U1245, Centre Henri Becquerel, Université de Rouen, IRIB, Rouen, France*
[b] *Institut Carnot CALYM, Paris, France*
[c] *Cancer Research Center of Lyon, INSERM U1052 UMR CNRS 5286, Lyon, France*
[d] *Normandie Univ, EdN BISE 497, Normandy, France*
[e] *Université de Lille 2, Lille, France*
[f] *Department of Pathology, Henri Mondor Hospital, APHP, INSERM U955, Université Paris-Est, Créteil, France*
[g] *Laboratoire de Pathologie, AP-HP Hôpital Saint Antoine, Paris, France*
[h] *Department of Pathology, Hôpital Maisonneuve-Rosemont, Montréal, Quebec, Canada*
[i] *Institut Imagine HGID, Inserm U1163, AP-HP Hôpital Necker, Université Paris Descartes, Paris, France*
[j] *Unité Hémopathies Lymphoïdes, AP-HP Hôpital Henri Mondor, Créteil, France*
[k] *Pathology, AP-HP, Hôpital Necker, Université Paris Descartes, Paris, France*
[l] *Inserm U1016 - CNRS UMR8104 - Université Paris Descartes Groupe Hospitalier Cochin, Paris, France*

ARTICLE INFO

Article history:
Received 5 July 2019
Revised 30 September 2019
Accepted 30 September 2019
Available online 21 October 2019

Keywords:
Diffuse large B-cell lymphoma
Independent component analysis
Transcriptomic variability
Gene signatures, prognosis

ABSTRACT

Background: Gene expression profiling (GEP), next-generation sequencing (NGS) and copy number variation (CNV) analysis have led to an increasingly detailed characterization of the genomic profiles of DLBCL. The aim of this study was to perform a fully integrated analysis of mutational, genomic, and expression profiles to refine DLBCL subtypes. A comparison of our model with two recently published integrative DLBCL classifiers was carried out, in order to best reflect the current state of genomic subtypes.

Methods: 223 patients with *de novo* DLBCL from the prospective, multicenter and randomized LNH-03B LYSA clinical trials were included. GEP data was obtained using Affymetrix GeneChip arrays, mutational profiles were established by Lymphopanel NGS targeting 34 key genes, CNV analysis was obtained by array CGH, and FISH and IHC were performed. Unsupervised independent component analysis (ICA) was applied to GEP data and integrated analysis of multi-level molecular data associated with each component (gene signature) was performed.

Findings: ICA identified 38 components reflecting transcriptomic variability across our DLBCL cohort. Many of the components were closely related to well-known DLBCL features such as cell-of-origin, stromal and MYC signatures. A component linked to gain of 19q13 locus, among other genomic alterations, was significantly correlated with poor OS and PFS. Through this integrated analysis, a high degree of heterogeneity was highlighted among previously described DLBCL subtypes.

Interpretation: The results of this integrated analysis enable a global and multi-level view of DLBCL, as well as improve our understanding of DLBCL subgroups.

© 2019 The Author(s). Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license.
(http://creativecommons.org/licenses/by-nc-nd/4.0/)

* Corresponding author at: 1 rue d'Amiens Rouen, Rouen 76000, France.
E-mail address: fabrice.jardin@chb.unicancer.fr (F. Jardin).

https://doi.org/10.1016/j.ebiom.2019.09.034
2352-3964/© 2019 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license.
(http://creativecommons.org/licenses/by-nc-nd/4.0/)

1. Research in context

1.1. Evidence before this study

The current gold standard of diffuse large B-cell lymphoma (DLBCL) subclassification is based on gene expression profiling (GEP) leading to three main molecular subtypes: Germinal Center B-cell like (GCB), Activated B-Cell like (ABC), and Primary Mediastinal B-cell Lymphoma (PMBL). However, the clinical relevance of these subtypes is currently under question, as multiple studies have noted their heterogeneity. Furthermore, several recent prospective clinical trials based on the GCB/non-GCB dichotomy for targeted therapy have published negative results, suggesting too reductive a classification.

1.2. Added value of this study

DLBCL as an entity is increasingly characterized on multiple molecular levels. This study is the first to use unsupervised independent component analysis applied to GEP data in order to define gene signatures which explain the variability among DLBCL patients. Furthermore, the thorough integrated analysis performed highlights the interplay between gene signatures and genomic alterations.

1.3. Implications of all the available evidence

We defined novel gene signatures which refine the understanding of DLBCL heterogeneity and highlight novel prognostic factors. Currently accepted subtypes were dissected using our unsupervised analysis in order to identify the molecular mechanisms behind their complexity. Comparative integrated analyses were performed in light of recent data in order to provide an up-to-date overview of DLBCL subtypes.

2. 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]. Since 2000, gene expression profiling (GEP) has subdivided the DLBCL entity into three main subtypes: Germinal Center B-cell like (GCB), Activated B-Cell like (ABC), and Primary Mediastinal B-cell Lymphoma (PMBL) [2–4]. The biological pathways behind each subtype and their clinical significance are now fairly well established, but approximately 20% of DLBCL patients remain unclassified according to this stratification, revealing further complexity that remains to be elucidated.

Standard DLBCL therapy remains rituximab, cyclophosphamide, doxorubicin, vincristine and prednisone (R-CHOP) chemotherapy [5–7]. However, recent studies have shown that the ABC/GCB dichotomy does not always translate to clinical relevance in the R-CHOP era, suggesting the need to more clearly refine these subtypes [8–10]. Furthermore, as targeted therapies become increasingly widespread, it is essential to thoroughly characterize each molecular subtype, irrespective of GCB/non-GCB subtype, in order to tailor each patient's optimal treatment [11].

In recent years, multiple studies have focused on single alteration type platforms, such as next-generation sequencing (NGS) to characterize DLBCL mutational profiles [12–18] or array comparative genomic hybridization (aCGH) to detail DLBCL copy number variations (CNVs) [19,20]. Even more recently, several studies have presented multi-platform results on DLBCL cohorts, in an attempt to deliver an integrated analysis of DLBCL genomic profiles [21–23].

In this study, we used independent component analysis (ICA) [24] applied to GEP and correlated NGS, aCGH, FISH and immunohistochemistry (IHC) data to perform a fully integrated analysis of mutational, genomic, and expression profiles of DLBCL. The resulting multi-level view of DLBCL enables a refined subclassification of this disease, novel prognostic characteristics, and should improve our understanding of DLBCL as a whole.

3. Materials and methods

3.1. Patients

The patient cohort has been previously described [15] and clinical trial results have been published. Briefly, 223 adult patients with *de novo* CD20+ DLBCL enrolled in the prospective and multicenter LNH-03B LYSA trials with available frozen tumor samples, centralized histopathologic review, and adequate DNA/RNA quality were selected (Supplementary Methods). 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.

3.2. Gene expression data

Gene expression data was available for all 223 patients (GEO GSE87371). Samples were analyzed with HGU133+2.0 Affymetrix GeneChip arrays (Affymetrix). The chips were scanned with an Affymetrix GeneChip Scanner 3000 and subsequent images were analyzed using Gene Chip® Operating Software (GCOS) 1.4. Raw feature normalization and quality check were handled using Bioconductor software (affy, affyQCReport, GCRMA). One probeset per gene was selected using JetSet annotations. COO signature was established as previously described [15]. Of note, PMBL in this study are GEP-defined PMBL, whose signatures were established using hierarchical clustering (complete distance, Ward agglomeration) of a previously published gene signature [4], excluding *TCL1A* and *E2F2* that could not be reliably measured on U133+2.0 arrays.

3.3. Independent component analysis

Independent Component Analysis (ICA) is a blind source separation method that has been shown to be particularly sensitive for identifying latent processes that underlie coordinated expression changes in transcriptome datasets [24–26]. While principal component analysis identifies components that best recapitulate the variance of the data under the constraint of orthogonality between the components, ICA performs a similar decomposition but constraining on statistical independence of the components instead. The result of the ICA decomposition of the gene expression matrix is a set of components which are characterized by the individual weights each component assigns to each gene and by the scores of the components in each of the samples. We applied the fastICA algorithm to GEP data (Supplementary methods) to extract a total of 38 independent components (**Suppl Fig 1**): 6 were eliminated due to probable batch effect. Each of the components was characterized by a score reflecting the activity of the latent process in the samples and by their "leading genes" (genes associated with the most significant weights for that component).

3.4. NGS data

The Lymphopanel was designed to identify mutations in 34 genes important for lymphomagenesis, as detailed previously [15,27,28]. NGS data was available for 213 patients.

Ion Torrent Personal Genome Machine (PGM) Sequencing and PGM data analysis were performed as described previously, using an in-house generated bioinformatics pipeline [15,27,28].

3.5. Copy number variations

Copy Number Variations (CNVs) were identified performing Comparative Genomic Hybridization (CGH) on 190 patients after whole-genome amplification, using Agilent SurePrint G3 4 × 180 K microarrays. Briefly, arrays were scanned with Agilent Feature Extraction and processed with cghRA [29]. Recurring CNVs were identified running GISTIC [30] version 2.0.22 (Supplementary Methods). Raw data is available via GEO (GSE136962); processed data with full annotations is available in Supplementary Methods and **Suppl Tables 1, 2 and 3**.

3.6. FISH data

FISH analysis was performed on 3 µm TMA tissue sections using break-apart FISH DNA probes for BCL2/18q21, and BCL6/3q27 (probes Y5407, and Y5408; Dako A/S) as previously described [31]. For cMYC/8q24, two different MYC break-apart FISH DNA probes were used: MYC FISH DNA probe Split Signal (Y5410, Dako A/S) and Vysis LSI MYC dual color, BA rearrangement probe (Abbott Laboratories, Chicago, IL), as previously described [32].

3.7. Immunohistochemistry

Immunohistochemistry was performed on 3 µm tissue sections for CD10, BCL6, MUM1, BCL2, MYC, FOXP1 and IgM as described previously [33,34]. The Hans algorithm was applied using standard cutoff levels of 30% for CD10, BCL6 and MUM1. BCL2 and MYC overexpression thresholds were respectively set at 50% and 40%, in accordance with previous publications [35]. IgM staining was considered positive when tumor cells significantly expressed this isotype, with a threshold set at 10% [34].

3.8. Integrative genetic classification

Both NMF consensus clustering, according to the method used by Chapuy and colleagues [21] and genClass clustering, according to the algorithm proposed by Schmitz and colleagues [22], were applied to our cohort with adjustments made to account for our data and cohort size (supplementary methods).

3.9. Survival analyses

All statistical analyses were performed using R software version 3.3.3. Progression-free survival (PFS) was evaluated from the date of random assignment to the date of disease progression, relapse, or death from any cause. Overall survival (OS) was evaluated from the date of enrollment to the date of death from any cause.

4. Results

4.1. Biological interpretation of components

The biological role of each component was interpreted by analyzing the overlap between components' leading genes and SignatureDB and MSigDB [36–39] database genesets (**Suppl Figure 2, Suppl Tables 4 and 5**).

32 batch-effect-free components were identified, of which 28 showed significant results in geneset over-representation analyses. Many of these were closely related to well-known DLBCL features such as cell-of-origin, stromal and MYC signatures (highlighted in **Table 1**).

Component 2 is highly affiliated with monocyte/macrophage activity: leading genes include *CD33* and *CD68*. Components 3 and 11 are both associated with the Stromal-1 geneset [40] but only have two leading genes in common (*CETP* and *PMEPA1*), suggesting differences between both components. Component 7 overlaps significantly with the Stromal-2 geneset [40]. Components 4, 5 and 9 are linked to proliferation, with components 4 and 9 being more specifically associated with *MYC* and *BCL6* signatures respectively. Component 6 is linked to T-cell development and activity: leading genes include *IL21* and *ZAP70*, which play an essential role in the regulation of the adaptive immune response by regulating T-cell activation. Components 8 and 15 are respectively enriched in GCB and ABC DLBCL signatures, while components 10 and 17 are highly linked to interferon and PMBL. Finally, certain components were enriched in specific cytogenetic locations, such as components 12 (18q21) and 23 (19q13).

4.2. Association of components with genomic alterations and immunohistochemistry

NGS, aCGH, FISH and IHC data were analyzed in order to establish correlations between components and mutations (**Fig. 1, Suppl Fig 3a/Table 6**), rearrangements (**Fig. 1, Suppl Fig 3b/Table 7**), protein expression levels (**Fig. 1, Suppl Fig 3c/Table 8**), and CNVs (**Fig. 2** and **Suppl Table 9**). Notable significant associations are highlighted in **Table 1**. Benjamini-Hochberg adjustment was applied to each of the analyses to account for multiple testing; for associations between components and CNVs, a custom permutation procedure was used to control False Discovery Rate (**Suppl Methods**).

High component 4 and 5 (proliferation) expression scores were associated with *CD79B* and *MYD88* mutations. Furthermore, component 4 was linked to 9p21 deletion, covering the CDKN2A/B/ANRIL locus of known tumor suppressor genes. MYC IHC expression was also positively correlated with component 4, consistent with geneset enrichment results.

Component 6 (T-cell lineage) was negatively correlated with *CD79B* and *MYD88* mutations. It was also linked to a small deletion on 14q11, hosting *TCR* loci, reflecting the TCR rearrangements of the T lymphocytes present in the tumor micro-environment.

As expected, components 8 and 15 (associated with GCB and ABC respectively) presented GCB and ABC mutational landscapes respectively [15]. Component 8 was correlated with REL amplification (2p16) and with *BCL2* rearrangement. Following the Hans algorithm [41], component 8 was positively correlated with BCL6 and CD10 expression and negatively correlated with MUM1 expression, whereas component 15 was negatively correlated with CD10 expression. Furthermore, FOXP1 was highly expressed in the ABC gene signature, and its IHC expression was negatively correlated with components 8 and 10 (GCB and PMBL). IgM was negatively correlated with component 8: this was expected, as GCB DLBCL typically switch to an IgG BCR, preferentially inducing plasma cell differentiation, while ABC DLBCL mostly express an IgM BCR [42].

Components 10 (interferon-related) and 17 (PMBL-related) share numerous positive correlations with mutations, including *XPO1*; furthermore, component 17 is linked to an even more complete PMBL mutational profile, including *STAT6* and *SOCS1* mutations [15,43]. Component 17 was linked to 9p24 amplification, including *PDCD1LG2* coding for PD-1-Ligand. Interestingly, GEP-defined PMBL with high interferon component expression scores, and therefore high PDL1 expression, also presented low T-cell component expression, in keeping with the negative regulation of T lymphocytes by PDL1 during the immune response (**Suppl Figure 4**).

Of note, high component 23 (19q13) expression score was positively associated with poor prognosis *CD79B* and *MYD88* mutations and negatively associated with favorable prognosis *B2M, CD58* and *TNFRSF14* mutations [44–46]. Other poor-prognosis factors associated with component 23 included 9p21 deletion as well as FOXP1

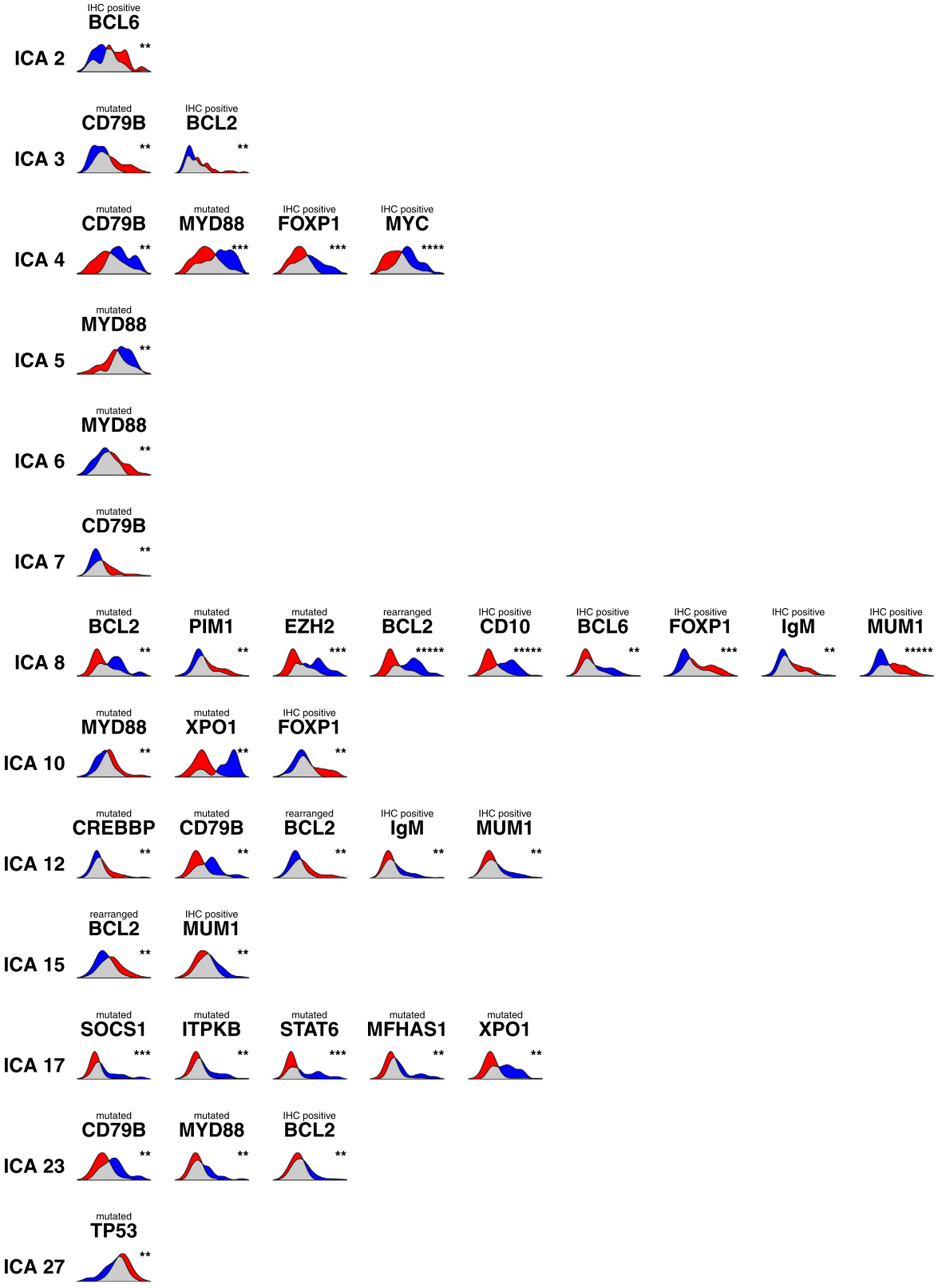


Fig. 1. Component expression associations with mutations, FISH and IHC.
Density plots are used in each cell. The X axis represents component expression level and the Y axis represents patient distribution (gaussian kernel density estimate). Red and blue respectively indicate the distribution of patients negative or positive for mutation, FISH or IHC as noted in cell title. Grey areas correspond to the overlap between the two distributions. Only plots where ICA expression is highly significantly different between positive and negative patients are shown here (Mann-Whitney, FDR<0.01). Significance of the difference is furthermore described by stars in the top right corner of each cell : ** for FDR < 0.01, *** for FDR < 0.001, **** for FDR < 0.0001, ***** for FDR < 0.00001. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Biological interpretation of components.

ICA defined components	Leading geneset associations	Biological Interpretation	Mutation associations	CNV associations	IHC associations	FISH associations
ICA_2	Mono-1/2/3, LN-1	Monocytes/Macrophages			FOXP1, BCL6	
ICA_3	LN-5, Stromal-1, Mesench-1	Stromal1	B2M, CD79B		FOXP1, BCL2	
ICA_4	MycUp-2/3/4, StarveDn-1	MYC/Proliferation	CARD11, CD79B, MYD88, PIM1	9p21 deletion (CDKN2A/B/ANRIL), 11q24 amplification	FOXP1, MYC, MUM1	
ICA_5	Prolif-8/10/3	Proliferation	CD79B, MYD88		FOXP1, MUM1	
ICA_6	Module-2.8, PanT-1/2, GCThUp-3, TNK-1/2	T-cell	CD79B, MYD88	14q11 deletion (TCR loci)	BCL6	
ICA_7	Stromal-2, CNS-1, Mesench-1	Stromal2	TNFAIP3, CD79B		BCL6	
ICA_8	GCBDLBCL-3, GCB-1/7/3, ABCDLBCL-1/2-3/4	GCB	BCL2, CREBBP, EZH2, GNA13,STAT6, PIM1	2p16 amplification (REL), 3q27 amplification (BCL6)	BCL6, CD10, FOXP1, IgM, MUM1	BCL2
ICA_9	Prolif-5, BCL6-4	Proliferation/BCL6				
ICA_10	Module-3.1, IFN-3/1/2, IRF3-1, PMBL-1	Interferon/Inflammation/PMBL	B2M, CIITA, STAT6, TNFAI3, XPO1, MYD88	6q21 deletion (ZAP70, PRDM1)	BCL6, FOXP1	
ICA_11	Stromal-1, LN-5/4/3	Stromal1-like	STAT6, TNFRSF14, CD79B	9p21 deletion (CDKN2A/B/ANRIL)	FOXP1, MUM1	MYC
ICA_12	chr18q21	Amplification hot spot 18q21	CD79B, PIM1, CREBBP	18q21 amplification (BCL2), 3q27 amplification (BCL6), 9p21 deletion (CDKN2A/B/ANRIL)	FOXP1, IgM, MUM1, CD10	BCL2
ICA_15	ABCDLBCL-4/3, GCBDLBCL-3	ABC	MYD88, PIM1		MUM1, CD10	BCL2
ICA_16	BCL6-4/2, Prolif-5	BCL6/Proliferation	CD79B, PIM1		MUM1	BCL2
ICA_17	PMBL-1, chr9q34/32, chr9p24	PMBL	B2M, CD58, ITPKB, MFHAS1, SOCS1, STAT6, TNFAI3, XPO1, KMT2D	9p24 amplification (PDL1)	CD10	
ICA_23	chr19q13	19q13	CD79B, MYD88, B2M, CD58, TNFRSF14	9p21 deletion (CDKN2A/B/ANRIL), 6q21 deletion (ZAP70, PRDM1)	FOXP1, IgM, BCL2	

ICA components of interest are highlighted in this table, along with a selection of most significant geneset associations, used to extrapolate a biological interpretation of each component. Significant mutation, CNV, IHC and FISH associations are shown for each component when applicable (blue text for positive association, red text for negative association).

and IgM expression, which might suggest preferential induction of NF-kB activation and other pro-survival pathways in this component.

4.3. Clinical data

A range of clinical traits were analyzed for correlations with the identified components (**Suppl Table 10**). IPI was negatively correlated with components 3 and 11 (Stromal1-like), as well as components 17 (PMBL) and 6 (T-cell). Age greater than 60 years was positively correlated with components 4 and 5 (proliferation and MYC/proliferation) as well as component 23 (19q13). Mediastinal localization was positively correlated with components 10 and 17 (PMBL-related) as well as component 3 (stromal1-like).

Association of component scores with survival was assessed first using a continuous Cox model adjusting the p-value for the number of components tested using Benjamini-Hochberg procedure and was then illustrated by grouping patients into component score tertiles. High component 11 (stromal1-like) scores were associated with better PFS and OS, both in patients treated by Rituximab-containing chemotherapy (R-chemo) and in patients treated specifically by R-CHOP, corroborating previous findings [40] (**Fig. 3a**). Component 23 (19q13) expression score was found to be correlated with poor OS and PFS in patients treated with either R-chemo or R-CHOP regimens (**Fig. 3b**). Component 8 (GCB) expression score was also correlated with significantly better PFS in R-chemo treated patients; however, this prognostic impact was not observed for OS or in other treatment subgroups (**Suppl Fig 5**).

We used Lenz and colleagues' cohort [40] as a validation cohort and were able to reproduce the positive survival impact of high component 11 (**Suppl Fig 6a**) and component 8 (**Suppl Fig 6b**) expression scores. We also used Affymetrix gene expression data available for part of Chapuy and colleagues' cohort [21] and were able to reproduce the negative prognostic impact of high component 23 (**Suppl Fig 7e/f**) and the positive prognostic impact of high component 11 (**Suppl Fig 7c/d**) scores, but not the survival impact of component 8 (**Suppl Fig 7a/b**) expression score. Finally, we also used RNAseq data available for part of Schmitz and colleagues' cohort [22] and were able to reproduce the positive prognostic impact of high component 8 (**Suppl Fig 8a/b**) and 11 (**Suppl Fig 8c/d**) expression scores, but not the survival impact of component 23 (**Suppl Fig 8e/f**).

4.4. Comparative analysis of integrated classifiers

Two recently published papers have also attempted integrated analyses of DLBCL cohorts in order to propose distinct genetic classifications of this disease [21,22]. We applied both a semi-

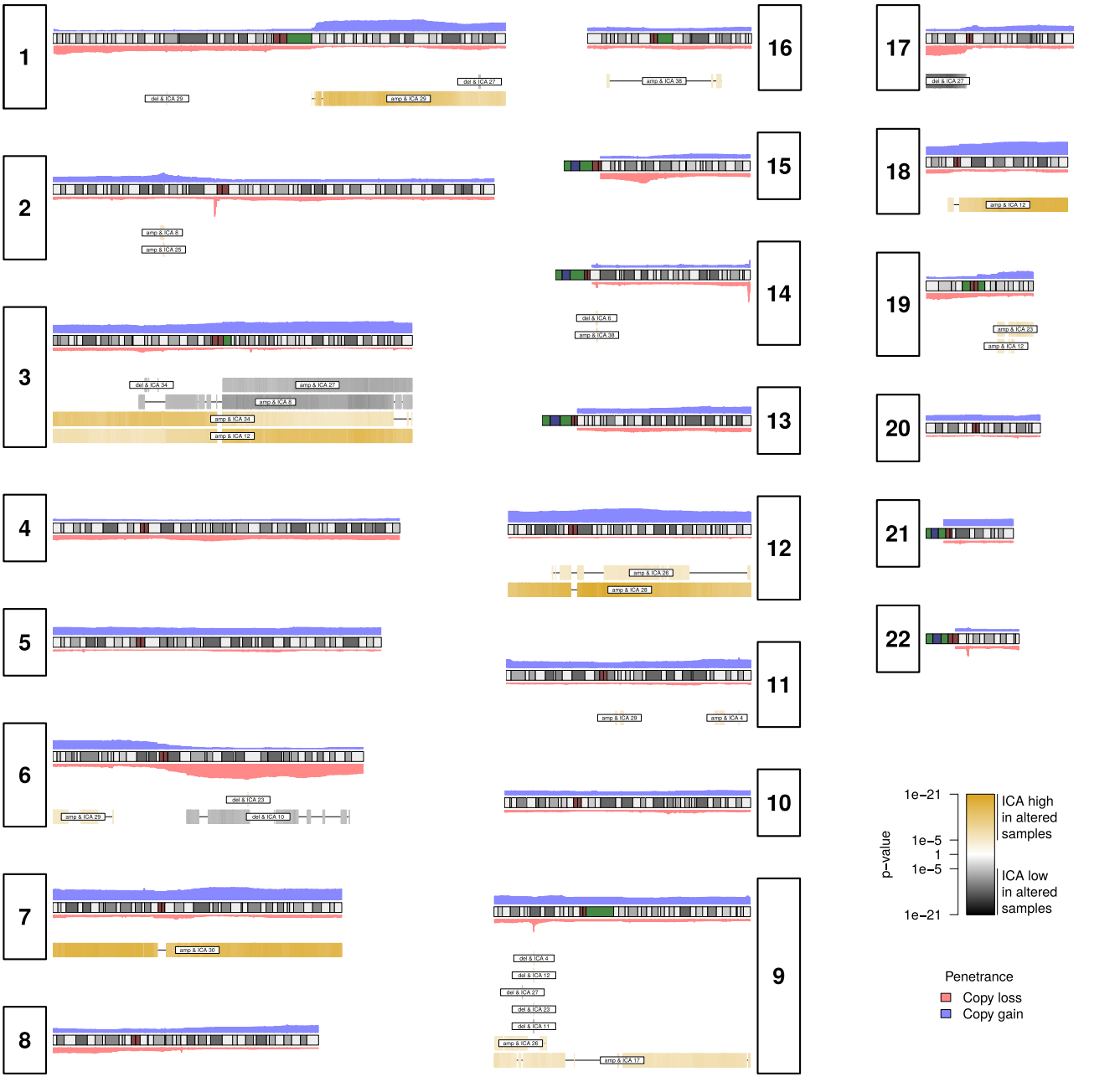


Fig. 2. Graphic representation of component and chromosomal alteration associations.
Chromosomes 1 to 22 are represented along with their cytogenetic bands. Proportions of the series presenting copy gain (blue) or loss (red) are presented as histograms around the chromosome ideogram. Significant associations of chromosomal regions with component expression levels are shown underneath each chromosome: yellow indicates positive associations and black indicates negative associations, with correlation intensity being depicted by variation in color intensity. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

supervised methodology using the GenClass algorithm [22] (**Fig. 4**) and unsupervised Nonnegative Matrix Factorization (NMF) consensus clustering to our data [21] (**Fig. 5**, **Supplementary Methods**). **Figure 6** offers an overview of the heterogeneity explained by both of these integrated analyses as well as their relationship to ICA-defined components in our cohort. To monitor the impact of missing values, we also reproduced these analyses on the 173 patients with fully available data on the three main molecular levels (GEP, NGS and CNV): neither genClass algorithm (**Suppl Figure 9a**) nor NMF clustering (**Suppl Figure 9b**) showed any major difference compared to the analysis on 223 patients.

Regarding genClass algorithm, the described pre-defined clusters (*NOTCH1*-mutated N1 subtype, *CD79B*- and MYD88 L265P-mutated MCD subtype, *NOTCH2*-mutated or *BCL6*-translocated BN2 subtype, and *EZH2*-mutated or *BCL2*-translocated EZB subtype) were assembled. Of note, MYC translocations, not available in the original paper, were associated with EZB subtype. 47.5% of our cohort was genetically classified into one of these four subtypes, similar to the original results (46.6%). A fifth subtype was added, termed STS for *STAT6* or *SOCS1* mutations, to account for the presence of GEP-defined PMBL in our cohort (**Fig. 4**). 11.7% of our cohort was classified into the STS subtype, leading to a total of 59% of our cohort classified into one of these five subtypes. CD274 (PDL1) gain was associated with STS subtype.

The expression scores of ICA-defined components were evaluated among the five defined subtypes (**Fig. 6**). Interestingly, com-

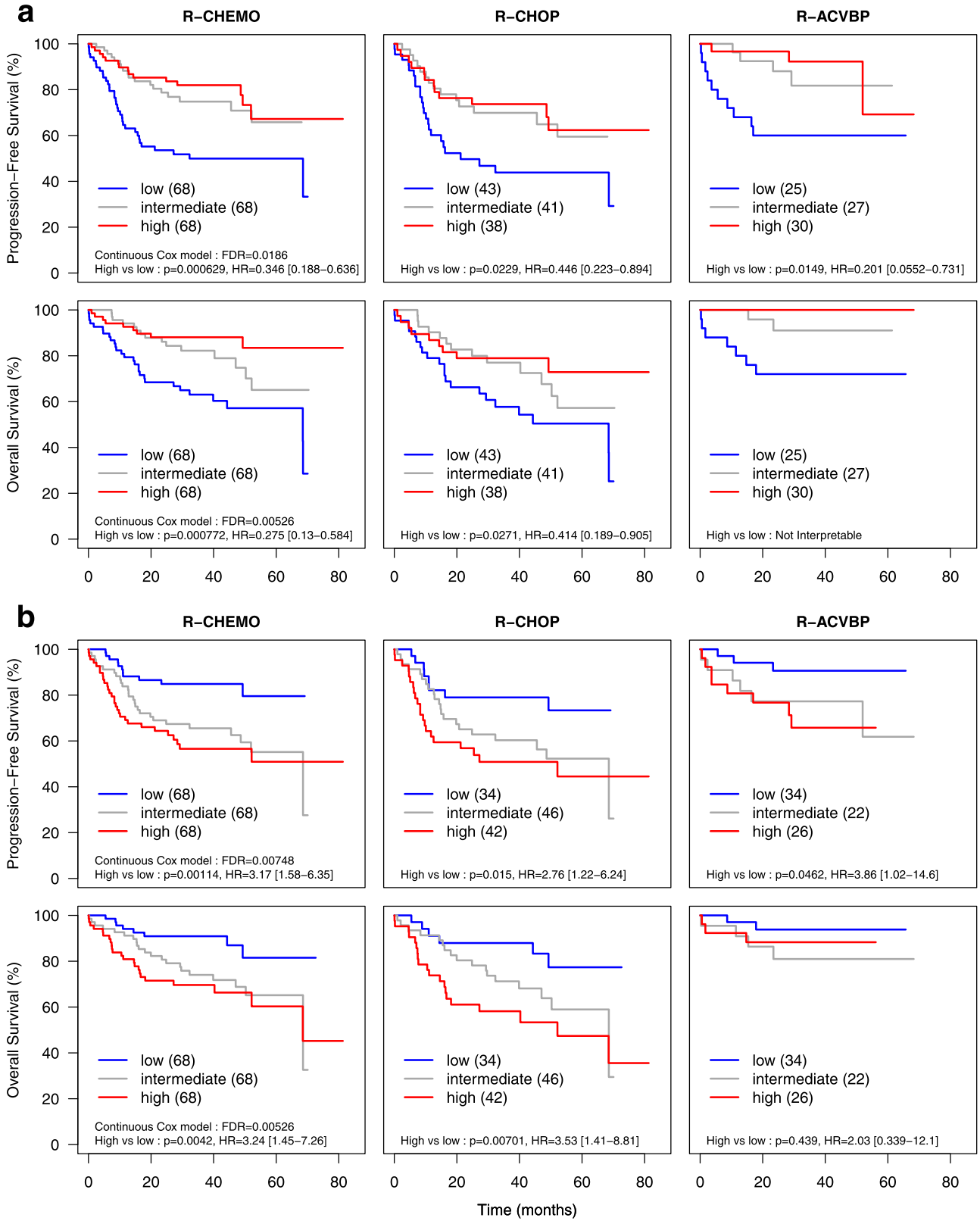


Fig. 3. Components with significant prognostic impact.
Fig. 3a.Positive prognostic impact of component 11. Fig. 3b.Poor prognostic impact of component 23. Survival analyses were performed according to treatment regimen: R-chemo includes all patients with Rituximab regardless of associated chemotherapy, R-CHOP includes patients with R-CHOP and R-CHOP like regimens (R-miniCHOP), R-ACVBP includes patients with R-ACVBP treatment. Patients were grouped into tertiles relative to component expression levels (low, intermediate and high). Progression-free survival analyses are shown in the upper panels; overall survival analyses are shown in the lower panels. FDR of continuous Cox model applied to R-CHEMO is shown. P-value is calculated for high tertile vs low tertile with hazard ratio (HR) indicated.

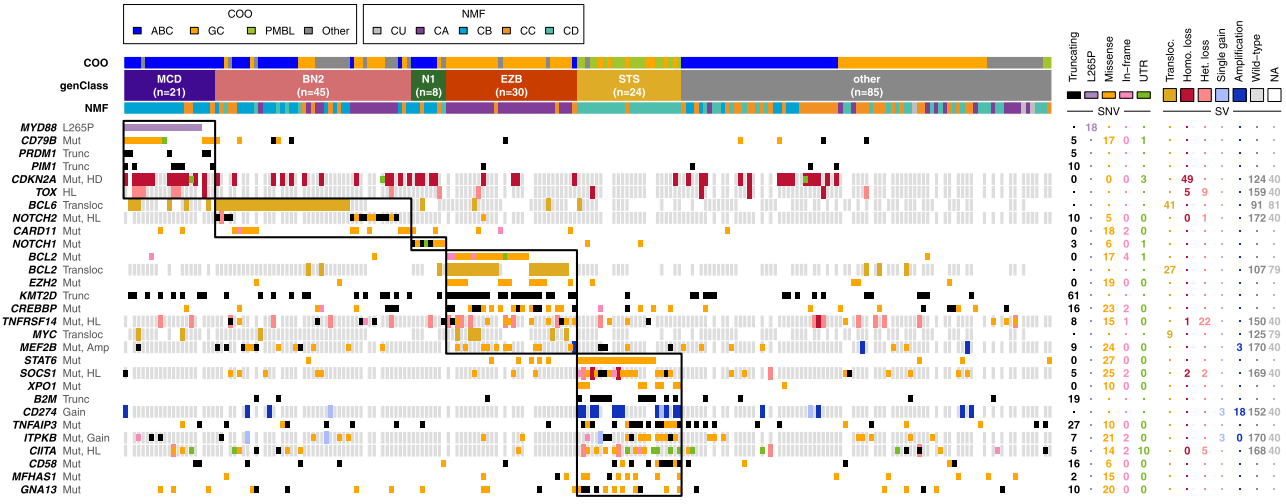


Fig. 4. Application of genClass Algorithm to Cohort.
Patients are distributed in columns and are grouped according to the genClass classification, cell of origin (COO) and NMF classification. Black squares highlight MCD, BN2, N1, EZB and STS clusters. Single Nucleotide Variants (SNV) and Structural Variants (SV) selected by the genClass algorithm are displayed in rows (one row per gene). The total number of each alteration is shown in the right-hand panel. "." indicates event types which are disregarded by the algorithm. Gray cell background highlights the availability of CGH or FISH data for the considered patient, only for genes (rows) where it is relevant. For SV, "Wild-type" and "NA" columns correspond respectively to patients wild-type for the chromosomal aberration in question or without SV data.

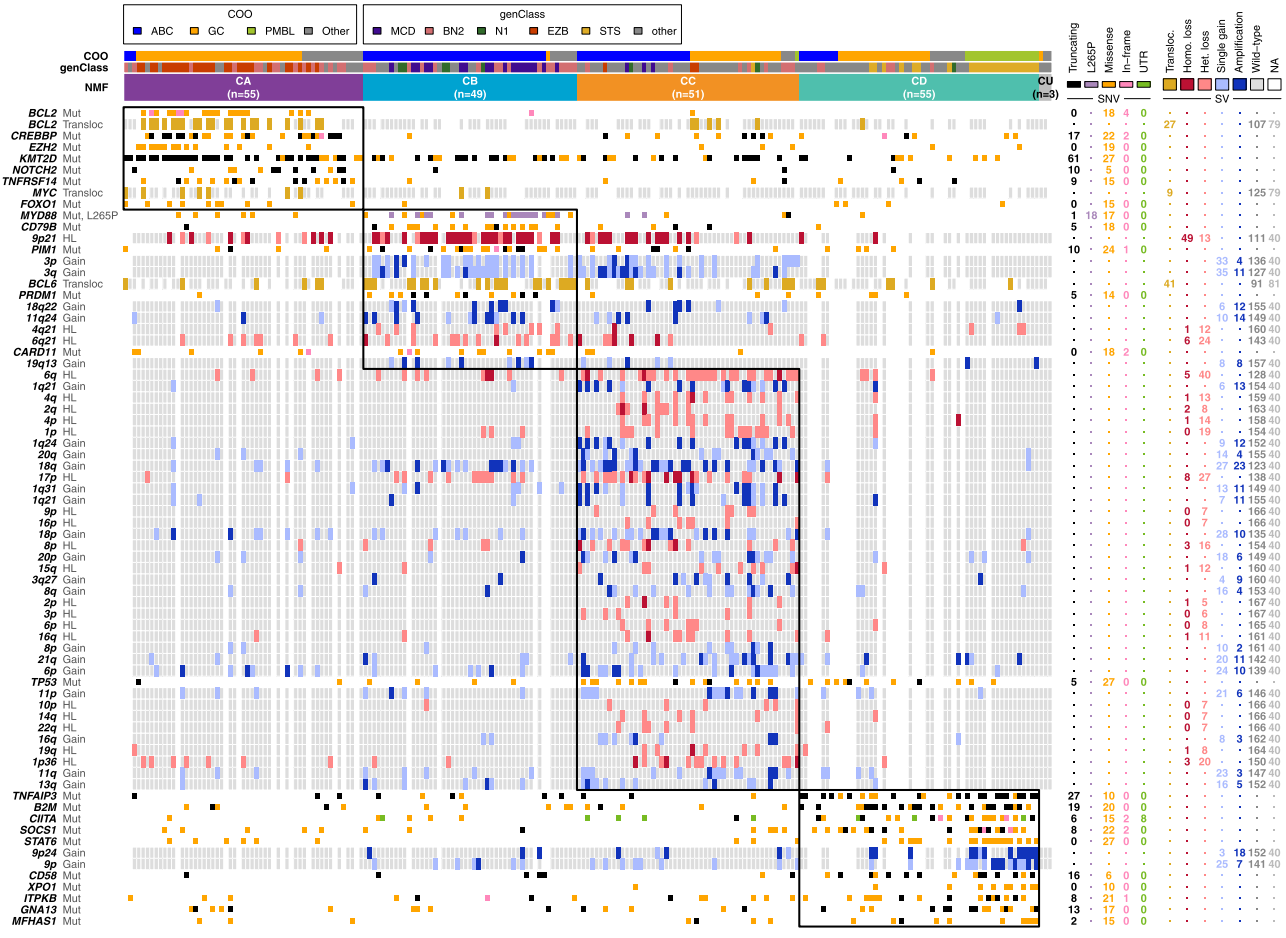


Fig. 5. Application of NMF algorithm to cohort.
Patients are distributed in columns and are grouped according to the NMF classification, cell of origin (COO) and genClass classification. Black squares highlight clusters A (CA), B (CB), C (CC) and D (CD). Single Nucleotide Variants (SNV) and Structural Variants (SV) selected by the NMF algorithm are displayed in rows (one row per event). The total number of each alteration is shown in the right-hand panel. "." indicates event types which are disregarded by the algorithm. Gray cell background highlights the availability of CGH or FISH data for the considered patient, only for genes (rows) where it is relevant. For SV, "Wild-type" and "NA" columns correspond respectively to patients wild-type for the chromosomal aberration in question or without SV data.

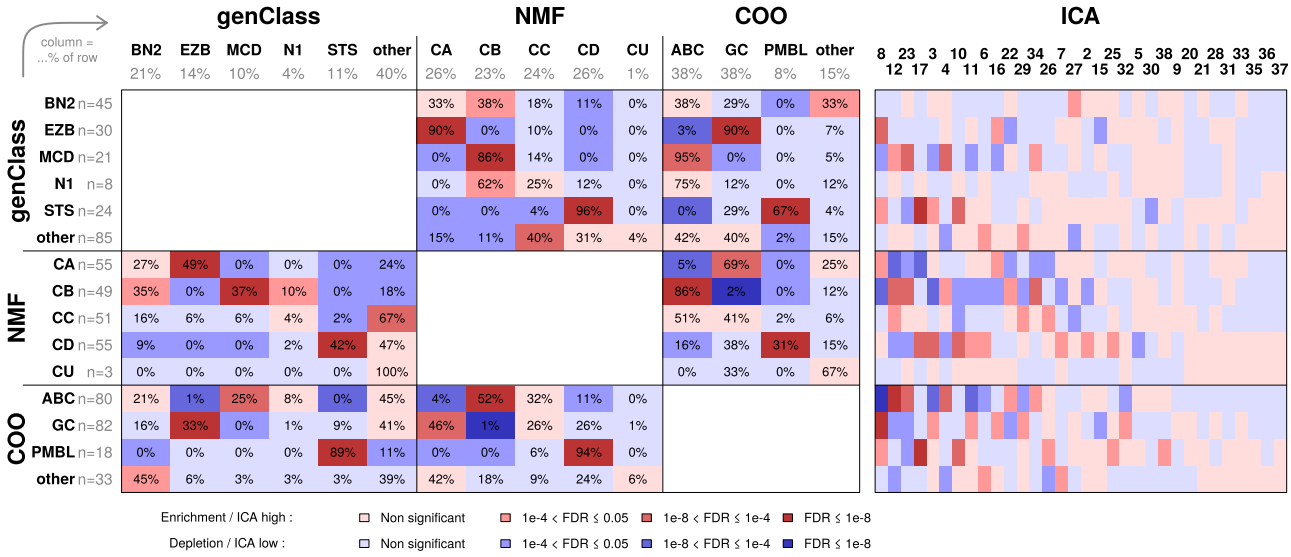


Fig. 6. Relationship between genClass, NMF, COO and ICA classifications.
Heatmap illustrating the interplay between genClass clusters (BN2, EZB, MCD, N1, STS and other), NMF clusters (CA, CB, CC, CD, CU), COO (ABC, GC, PMBL, other) and ICA-defined component expression level. Percentages within each cell of the heatmap indicate the proportion of patients within the cluster indicated in row that also belong to the cluster described in column. For example, 90% of the 30 EZB patients are clustered within NMF cluster CA while the remaining 10% belong to CC.
Percentages on top of each classification-related column indicate the proportion of patients within each group considering the whole-cohort. For example, COO defines 38% of the cohort as ABC, 38% as GC, 8% as PMBL and 15% as "other". Cell background highlights imbalances between these proportions, which would be expected in the absence of correlation, and the observed proportions (enrichments in red, depletions in blue). Color intensity varies according to the significance of a Fisher test (refer to the graphical legend at the bottom of the figure).
Cell colors in the ICA panel vary in a similar fashion: a red background indicates a higher ICA score in the considered group as compared to the rest of the patients, while a blue background indicates a lower ICA score. Color intensity varies according to the significance of a Mann-Whitney test, with the same legend described at the bottom of the figure. ICAs are ordered from the highest number of significant correlations (on the left) to the fewest (on the right), with their numeric ID printed at the top of the column.
Corresponding p-values and FDR are presented in Suppl Table 8. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

ponents 4 (MYC/proliferation) and 23 (19q13) were significantly overexpressed among MCD subtype patients, component 8 (GCB) was significantly overexpressed among EZB subtype patients, and components 10 and 17 (Interferon/PMBL) were significantly overexpressed among STS subtype patients.

In the unsupervised NMF approach (**Fig. 5**), gene mutations and copy number alterations were considered as distinct input features, with the copy number features defined as recurrently altered focal regions or chromosome arms. The optimal number of clusters in our cohort was determined to be 4 (termed CA, CB, CC and CD), in addition to a CU cluster (3 samples for whom no molecular alterations were detected for any of the considered features).

CA patients were overwhelmingly of GCB subtype and exhibited classic GCB alterations, such as BCL2 and MYC translocations (including several double-hit DLBCL patients), and *EZH2, CREBBP* and *KMT2D* mutations. Accordingly, most CA patients were of the EZB genClass subtype. Of note, in this unsupervised approach, *NOTCH2* mutations were not associated with this cluster.

CB DLBCLs were almost exclusively of ABC COO and either MCD or BN2 genClass subtype. Many characteristic ABC mutations were defining features of cluster CB. NFkB pathway alterations including MYD88L265P, *CD79B, PIM1, CARD11,* and *PRDM1* mutations were common among CB patients. BCL6 translocations were over-represented in CB, accounting for BN2 subtype representation among CB. Deletions of tumor suppressor locus 9p21 were also an identifying feature of CB DLBCL.

CC DLBCLs were split between GCB and ABC subtype patients. They seem mostly characterized by multiple CNVs, with *TP53* mutations being the only mutations significantly more prevalent among CC than other clusters, suggesting that this cluster is dependent on genetic instability. 17p deletion, including *TP53*, is also a defining feature of CC, further supporting this hypothesis. Given the pre-defined nature of genClass clusters, most CC DLBCLs did not belong to a specific GenClass subtype.

Cluster D included the vast majority of GEP-defined PMBL patients, who were therefore predominantly of the STS genClass subtype. *STAT6* and *SOCS1* mutations, but also *B2M, TNFAIP3, CIITA*, and *XPO1* mutations were significantly more common among CD DLBCLs. Both focal 9p24 amplification, including CD274, and full 9p chromosome arm gain were also predominant CD features.

The survival impacts shown in Chapuy and Schmitz's papers were not reproduced within our cohort [21,22].

4.5. Integrated analyses providing explanations for DLBCL subgroup heterogeneity

One of the goals of this integrated analysis was to better delineate DLBCL subtypes. GEP-defined PMBL seemed to be a relatively homogeneous group in our analyses. On the other hand, unclassified or "other" DLBCL seemed more difficult to characterize: GenClass and NMF analyses showed that "other" DLBCL are strongly represented among the BN2 genClass subtype and among NMF clusters linked to GCB (CA) and PMBL (CD). Interestingly, ICA was able to highlight that "other" DLBCL express component 6 (T-cell, $p = 2.8e\text{-}4$) and component 7 (stromal 2, $p = 1.17e\text{-}3$) significantly more strongly than GCB, ABC or PMBL patients (**Fig. 6, Suppl Table 11**). These data suggest a particularly strong involvement of T lymphocytes within "other" DLBCL, which might explain their inability to fit into the ABC/GCB GEP dichotomy and go against a specific B-cell oncogenic pathway linked to "other" DLBCL. Cibersort analysis [47] showed that component 6 was indeed more highly expressed among samples with higher total estimated T cell fraction (**Suppl Figure 10**; Pearson rho=0.75, $p = 1.4e\text{-}41$) and that "other" DLBCL

contained a higher total estimated T cell fraction as compared to ABC, GCB and PMBL patients (**Suppl Figure 10**).

We also identified variability among the GCB subtype, with heterogeneity observed with respect to their expression of components related to MYC and proliferation (components 4 and 5) and to T lymphocytes (component 6). Indeed, this revealed GCB patients with a proliferation-high/T lymphocyte-low profile and GCB patients with the opposite phenotype (**Suppl Figure 11**). The identification of GCB patients with a more aggressive phenotype might explain the presence of GCB COO patients among the poorer outcome BN2 subtype (**Fig. 6**).

5. Discussion

In this study, we provide a novel multi-platform classification of DLBCL among a large, prospective cohort with thorough clinical data. ICA proved to be an effective unsupervised approach for the analysis of DLBCL GEP and we were able to detect associations between the identified components and genomic, protein, and clinical data.

Using the current ABC/GCB dichotomy, attempts at subtype-directed treatment have not yet shown efficacy in terms of prognosis, suggesting that the current DLBCL subclassification is insufficient for treatment decision-making [8,48,49]. Recent examples include the Phoenix study, which did not show a significant impact of adding ibrutinib to R-CHOP in ABC subtype [9], the REMoDL-B trial, which found no survival impact of adding bortezomib to R-CHOP including in ABC subtype patients specifically [10], and the CAVALLI trial, which identified the addition of venetoclax to R-CHOP as beneficial in BCL2 IHC positive patients, but with no impact of COO [11].

Our integrated analysis has shown that there is a much higher heterogeneity than previously suspected among ABC, GCB and "other" subtypes, as highlighted in **Fig. 6**, perhaps explain in part the negative results of such recent clinical trials. A large proportion of GCB patients present the classical GCB/EZB/CA phenotype, but a non-negligible sub-population are of a more aggressive BN2 phenotype with higher proliferation/MYC component expression scores. ABC patients are mostly CB but are split between BN2, MCD and N1 genClass clusters, suggesting biological heterogeneity and therefore prognostic variability [22]. "Other" DLBCL strongly express T-cell component and are distributed in a fairly balanced way among NMF clusters; however, they are over-represented in BN2 genClass subtype, suggesting that their defining alterations have yet to be identified. GEP-defined PMBL patients were voluntarily kept in this study as they were part of initial DLBCL clinical trial cohorts and their presence helped to better define this particular subgroup of DLBCL patients according to their genomic characteristics. Finally, genetic instability seems to be an important factor as it defined an entire cluster both in our study (CC) and in Chapuy's (c2). Furthermore, CC patients were mostly unclassified using genClass-defined seed classes, suggesting that genetic instability is an often unrecognized factor in DLBCL subclassification.

Many components had biological roles defined by their overlap with geneset databases, without significant associations with either NGS, CNV or FISH/IHC data. Of note, our Lymphopanel was selected based on existing literature data. Although this method did not enable novel mutation detection, it showcased that our study was able to produce dependable results with only 34 genes analyzed for mutations, more easily reproducible in a routine clinical setting than whole-exome sequencing. Of particular interest, all recent integrative analyses of DLCBL highlighted the importance of CNV data: several ICA-identified components were defined exclusively by chromosomal hotspots and cluster 2 as defined by Chapuy depends almost entirely on chromosomal alterations. Monti previously identified the negative prognostic impact of complex CNV phenotypes, even increasing prognostic accuracy compared to IPI [19]. This indicates the importance of taking into account multiple molecular levels when defining DLBCL subgroups, not only GEP.

Our study also highlighted prognostic indicators in component 11 and 23 expression scores. Component 11 was determined to be strongly similar to the Stromal-1 gene signature defined by Lenz [40] and its positive prognostic impact was confirmed there, as well as in Schmitz et al's cohort [22]. Furthermore, our study suggests that component 23 expression highlights an all-around poor prognosis phenotype, whose negative impact on PFS and OS was confirmed in Chapuy et al's cohort [21]. In terms of CNV it is associated with gain of 19q13, which is represented in R/R phenotype [43] and has previously been shown to be linked to poor EFS at 24 months [50]; component 23 is also linked to both *CD79B* and *MYD88* mutations, associated with poor-prognosis ABC; in terms of IHC, it is associated with FOXP1, BCL2 and IgM expression. Finally, component 23 was shown to be over-expressed in the MCD subgroup. 19q13 gain/amplification has previously been shown to be specific to ABC DLBCL, notably encompassing *SPIB*, which has been described as an oncogene essential for ABC DLBCL cell line survival [19,20], but this is the first study showing that a gene expression signature associated with 19q13 gain, among other alterations, is linked to poor OS and PFS among a prospective cohort. Furthermore, *SPIB* was not among the leading genes of component 23, indicating that its prognostic impact is independent of this previously described oncogene.

On the other hand, we did not reproduce significant survival differences in the subtypes identified using GenClass and NMF clustering, unlike the results published by Schmitz and Chapuy respectively [21,22]. This might be due to variability among cohort subpopulation distribution (voluntary enrichment in ABC and unclassified patients in Schmitz's paper), overfitting or the lack of a uniform treatment regimen among our cohort, leading to smaller survival analysis subgroups.

One of the main goals of integrated analyses is to attempt to better pinpoint which groups of patients would benefit from targeted therapies. Of great interest today are immune checkpoint inhibitors, including PDL1/PD1 blockade. It is crucial to best define targetable patients, as despite the potential of PDL1/PD1 blockade in this disease, certain patients only experience a short-lived response before relapsing [51]. In this study, we identified a subpopulation of PMBL patients specifically showing high interferon component and low T-cell component expression scores that might be interesting candidates for PD1 blockade treatment, already approved by the FDA in PMBL.

In conclusion, we have provided a multi-level integrated analysis of DLBCL, using an innovative unsupervised approach applied to a prospective cohort. This yielded novel results as compared to the two other currently published integrated analyses of DLBCL. Further studies are needed to validate the components identified herein, as well as their multi-level molecular profiles. Such studies are crucial for tomorrow's precision medicine era as we advance toward treatment decisions based on combinations of alterations rather than single variants. Integrated analyses such as ours provide the framework for refining DLBCL subtypes and potentially improving patient outcome.

Author contributions

SD designed the study; analyzed and interpreted data; wrote the paper

BT designed ICA bio-informatics pipeline; collected, analyzed and interpreted data

SM designed aCGH bio-informatics pipeline; analyzed and interpreted data; designed figures

P-JV designed NGS bio-informatics pipeline; collected and analyzed data
EB collected, analyzed and interpreted data
PR collected, analyzed and interpreted data
PE collected and analyzed data
PP collected data
CC-B collected and analyzed data
BF collected and analyzed data
TP collected and analyzed data
J-PJ designed gene expression data analysis; collected and analyzed data
CH collected and analyzed data
GS designed the study
TJM collected and analyzed data
KL collected and analyzed data
HT designed and supervised the study
FJ designed and supervised the study

Declaration of Competing Interest

The authors declare no conflicts of interest relevant to this study. Dr. Haioun reports personal fees from Amgen, personal fees from Roche, personal fees from Celgene, personal fees from Janssen, personal fees from Gilead, personal fees from Takeda, outside the submitted work; Dr. Salles reports personal fees from Amgen, personal fees from BMS, personal fees from Abbvie, personal fees from Janssen, personal fees from Merck, personal fees from Novartis, personal fees from Gilead / Kite, personal fees from Epizyme, personal fees from Pfizer, personal fees from Celgene, personal fees from Roche, personal fees from Takeda, personal fees from Autolus, personal fees from MorphoSys, personal fees from ACERTA, personal fees from Servier, outside the submitted work; Dr. Molina reports personal fees from Merck, personal fees from Celgene, personal fees from Novartis, outside the submitted work; Dr. Leroy reports personal fees and non-financial support from Bristol Myers Squibb, personal fees and non-financial support from Roche, personal fees and non-financial support from Astra-Zeneca, personal fees and non-financial support from Nanostring, personal fees from MSD, outside the submitted work; Dr. Jardin reports personal fees from Gilead, grants and personal fees from Roche, personal fees from Janssen, personal fees from Servier, grants and personal fees from Celgene, outside the submitted work.

Acknowledgements

We warmly thank Céline Villenet (Université Lille 2) for her technical work in producing aCGH data.

Funding Sources

INCA (Institut National du Cancer) funded this study but had no role in study design; in the collection, analysis, and interpretation of data; in the writing of the report; or in the decision to submit the paper for publication.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ebiom.2019.09.034.

References

[1] Swerdlow SH, Campo E, Pileri SA, et al. The 2016 revision of the World Health Organization classification of lymphoid neoplasms. Blood 2016;127:2375–90.
[2] Alizadeh AAA, Eisen MBMB, Davis RE, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature 2000;403:503–11.
[3] Wright G, Tan B, Rosenwald A, Hurt EH, Wiestner A, Staudt LM. A gene expression-based method to diagnose clinically distinct subgroups of diffuse large B cell lymphoma. Proc Natl Acad Sci USA 2003;100:9991–6.
[4] Rosenwald A, Wright G, Leroy K, et al. Molecular diagnosis of primary mediastinal B cell lymphoma identifies a clinically favorable subgroup of diffuse large B cell lymphoma related to Hodgkin lymphoma. J Exp Med 2003;198:851–62.
[5] Coiffier B, Lepage E, Brière J, et al. CHOP chemotherapy plus rituximab compared with chop alone in elderly patients with diffuse large-B-cell lymphoma. N Engl J Med 2002;346:235–42.
[6] Czuczman MS, Grillo-López AJ, White CA, et al. Treatment of patients with low-grade B-cell lymphoma with the combination of chimeric anti-CD20 monoclonal antibody and Chop chemotherapy. J Clinic Oncol 1999;17:268–76.
[7] Vose JM, Link BK, Grossbard ML, et al. Phase II study of rituximab in combination with chop chemotherapy in patients with previously untreated, aggressive non-Hodgkin's lymphoma. J Clinic Oncol 2001;19:389–97.
[8] Staiger AM, Ziepert M, Horn H, et al. Clinical impact of the cell-of-origin classification and the MYC/ BCL2 dual expresser status in diffuse large B-cell lymphoma treated within prospective clinical trials of the German high-grade non-hodgkin's lymphoma study group. J Clinic Oncol 2017;35:2515–26.
[9] Younes A, Sehn LH, Johnson P, et al. Randomized phase III trial of Ibrutinib and Rituximab plus cyclophosphamide, doxorubicin, vincristine, and prednisone in non–germinal center B-cell diffuse large B-cell lymphoma. J Clin Oncol 2019;37:1285–95.
[10] Davies A, Cummin TE, Barrans S, et al. Gene-expression profiling of bortezomib added to standard chemoimmunotherapy for diffuse large B-cell lymphoma (REMoDL-B): an open-label, randomised, phase 3 trial. Lancet Oncol 2019;20:649–62.
[11] Morschhauser F, Feugier P, Flinn IW, et al. Venetoclax plus rituximab, cyclophosphamide, doxorubicin, vincristine and prednisolone (R-CHOP) improves outcomes in BCL2-positive first-line diffuse large B-cell lymphoma (DLBCL): first safety, efficacy and biomarker analyses from the phase II CAVALLI st. Blood 2018;132 782–782.
[12] 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. In: Proceedings of the National Academy of Sciences of the United States of America, 109; 2012. p. 3879–84.
[13] Morin RD, Mungall K, Pleasance E, et al. Mutational and structural analysis of diffuse large B-cell lymphoma using whole-genome sequencing. Blood 2013;122:1256–65.
[14] Pasqualucci L, Trifonov V, Fabbri G, et al. Analysis of the coding genome of diffuse large B-cell lymphoma. Nat Genet 2011;43:830–7.
[15] Dubois S, Viailly P-JP-J, Mareschal S, et al. Next generation sequencing in diffuse large B cell lymphoma highlights molecular divergence and therapeutic opportunities: a LYSA Study. Clin Cancer Res 2016;22:2919–28.
[16] Mullighan CG. Genome sequencing of lymphoid malignancies. Blood 2013;122:3899–907.
[17] Zhang J, Grubor V. Genetic heterogeneity of diffuse large B-cell lymphoma. Proc Natl Acad Sci 2013;110:1398–403.
[18] Steidl C, Gascoyne RD. The molecular pathogenesis of primary mediastinal large B-cell lymphoma. Blood 2011;118:2659–69.
[19] Monti S, Chapuy B, Takeyama K, et al. Integrative analysis reveals an outcome-associated and targetable pattern of p53 and cell cycle deregulation in diffuse large b cell lymphoma. Cancer Cell 2012;22:359–72.
[20] Lenz G, Wright GW, Emre NCT, et al. Molecular subtypes of diffuse large B-cell lymphoma arise by distinct genetic pathways. In: Proceedings of the National Academy of Sciences of the United States of America, 105; 2008. p. 13520–5.
[21] 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;24:679–90.
[22] Schmitz R, Wright GW, Huang DW, et al. Genetics and pathogenesis of diffuse large B-cell lymphoma. N Engl J Med 2018;378:1396–407.
[23] Reddy A, Zhang J, Davis NS, et al. Genetic and functional drivers of diffuse large b cell lymphoma. Cell 2017;171:481–94 e15.
[24] Rotival M, Zeller T, Wild PS, et al. Integrating genome-wide genetic variations and monocyte expression data reveals trans-regulated gene modules in humans. PLoS Genet 2011;7:e1002367.
[25] Liebermeister W. Linear modes of gene expression determined by independent component analysis. Bioinformatics 2002;18:51–60.
[26] Teschendorff AE, Journée M, Absil PA, Sepulchre R, Caldas C. Elucidating the altered transcriptional programs in breast cancer using independent component analysis. PLoS Comput Biol 2007;3:e161.
[27] Dubois S, Mareschal S, Picquenot J-M, et al. Immunohistochemical and genomic profiles of diffuse large B-cell lymphomas: implications for targeted EZH2 inhibitor therapy? Oncotarget 2015;6:16712–24.
[28] Dubois S, Viailly P-J, Bohers E, et al. Biological and clinical relevance of associated genomic alterations in MYD88 L265P and non-L265P–mutated diffuse large B-cell lymphoma: analysis of 361 cases. Clin Cancer Res 2017;23:2232–44.
[29] Mareschal S, Ruminy P, Alcantara M, et al. Application of the cghRA framework to the genomic characterization of diffuse large B-cell lymphoma. Bioinformatics 2017;33:2977–85.
[30] Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhim R, Getz G. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. Genome Biol 2011;12:R41.

[31] Copie-Bergman C, Gaulard P, Leroy K, et al. Immuno-fluorescence in situ hybridization index predicts survival in patients with diffuse large B-cell lymphoma treated with R-CHOP: a GELA study. J Clinic Oncol 2009;27:5573–9.
[32] Copie-Bergman C, Cuillière-Dartigues P, Baia M, et al. MYC-IG rearrangements are negative predictors of survival in DLBCL patients treated with immunochemotherapy: a GELA/LYSA study. Blood 2015;126:2466–74.
[33] Molina TJT, Canioni D, Copie-Bergman C, et al. Young patients with non-germinal center B-cell-like diffuse large B-cell lymphoma benefit from intensified chemotherapy with ACVBP plus rituximab compared with CHOP plus rituximab: analysis of data from the Groupe d'Etudes des Lymphomes de l'Adulte/lympho. J Clin Oncol 2014;32:3996–4003.
[34] Jardin F, Delfau-Larue MH, Molina TJ, et al. Immunoglobulin heavy chain/light chain pair measurement is associated with survival in diffuse large B-cell lymphoma. Leuk Lymph 2013;54:1898–907.
[35] Johnson NA, Slack GW, Savage KJ, et al. Concurrent expression of MYC and BCL2 in diffuse large B-cell lymphoma treated with rituximab plus cyclophosphamide, doxorubicin, vincristine, and prednisone. J Clinic Oncol 2012;30:3452–9.
[36] Shaffer AL, Wright G, Yang L, et al. A library of gene expression signatures to illuminate normal and pathological lymphoid biology. Immunol Rev 2006;210:67–85.
[37] Liberzon A, Subramanian A, Pinchback R, Thorvaldsdottir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. Bioinformatics 2011;27:1739–40.
[38] Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA 2005;102:15545–50.
[39] Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The molecular signatures database hallmark gene set collection. Cell Syst 2015;1:417–25.
[40] Lenz G, Wright G, Dave SS, et al. Stromal gene signatures in large-b-cell lymphomas. N Engl J Med 2008;359:2313–23.
[41] Hans CP, Weisenburger DD, Greiner TC, et al. Confirmation of the molecular classification of diffuse large B-cell lymphoma by immunohistochemistry using a tissue microarray. Blood 2004;103:275–82.
[42] Ruminy P, Etancelin P, Couronné L, et al. The isotype of the BCR as a surrogate for the GCB and ABC molecular subtypes in diffuse large B-cell lymphoma. Leukemia 2011;25:681–8.
[43] Mareschal S, Dubois S, Viailly P-J, et al. Whole exome sequencing of relapsed/refractory patients expands the repertoire of somatic mutations in diffuse large B-cell lymphoma. Genes Chromosomes Cancer 2015;55:251–67.
[44] Dubois S, Viailly P-J, Mareschal S, et al. Next-generation sequencing in diffuse large B-cell lymphoma highlights molecular divergence and therapeutic opportunities: a LYSA Study. Clin Cancer Res 2016;22:2919–28.
[45] Karube K, Enjuanes A, Dlouhy I, et al. Integrating genomic alterations in diffuse large B-cell lymphoma identifies new relevant pathways and potential therapeutic targets. Leukemia 2018;32:675–84.
[46] Pasqualucci L, Dalla-Favera R. Genetics of diffuse large B-cell lymphoma. Blood 2018;131:2307–19.
[47] Chen B., Khodadoust M.S., Liu C.L., Newman A.M., Alizadeh A.A. Profiling tumor infiltrating immune cells with CIBERSORT. In: Methods in molecular biology (Clifton, N.J.). 2018: 243–59.
[48] Offner F, Samoilova O, Osmanov E, et al. Frontline rituximab, cyclophosphamide, doxorubicin, and prednisone with bortezomib (VR-CAP) or vincristine (R-CHOP) for non-GCB DLBCL. Blood 2015;126:1893–901.
[49] Leonard JP, Kolibaba K, Reeves JA, et al. Randomized phase 2 open-label study of R-CHOP ± Bortezomib in patients (Pts) with untreated non-germinal center B-Cell-like (Non-GCB) subtype diffuse large cell lymphoma (DLBCL): results from the pyramid trial (NCT00931918). Blood 2015;126.
[50] Novak AJ, Asmann YW, Maurer MJ, et al. Whole-exome analysis reveals novel somatic genomic alterations associated with outcome in immunochemotherapy-treated diffuse large B-cell lymphoma. Blood Cancer J 2015;5:e346.
[51] Lesokhin AM, Ansell SM, Armand P, et al. Nivolumab in patients with relapsed or refractory hematologic malignancy: preliminary results of a phase IB study. J Clin Oncol 2016;34:2698–704.