Exploring the cell-free total RNA transcriptome in diffuse large B-cell lymphoma and primary mediastinal B-cell lymphoma patients as biomarker source in blood plasma liquid biopsies

Philippe Decruyenaere[1,2,3]*, Edoardo Giuili[4,5], Kimberly Verniers[2,3], Jasper Anckaert[2,3], Katrien De Grove[1], Malaïka Van der Linden[6], Dries Deeren[7], Jo Van Dorpe[6], Fritz Offner[1] and Jo Vandesompele[2,3]

[1]Department of Hematology, Ghent University Hospital, Ghent, Belgium, [2]OncoRNALab, Cancer Research Institute Ghent (CRIG), Ghent University, Ghent, Belgium, [3]Department of Biomolecular Medicine, Ghent University, Ghent, Belgium, [4]Interuniversity Institute of Bioinformatics in Brussels (IB²), Free University of Brussels, Brussels, Belgium, [5]Department of Biotechnology and Pharmacy, University of Bologna, Bologna, Italy, [6]Department of Pathology, Ghent University Hospital, Ghent, Belgium, [7]Department of Hematology, Algemeen Ziekenhuis (AZ) Delta Roeselare-Menen, Roeselare, Belgium

Introduction: Diffuse large B-cell lymphoma (DLBCL) and primary mediastinal B-cell lymphoma (PMBCL) are aggressive histological subtypes of non-Hodgkin's lymphoma. Improved understanding of the underlying molecular pathogenesis has led to new classification and risk stratification tools, including the development of cell-free biomarkers through liquid biopsies. The goal of this study was to investigate cell-free RNA (cfRNA) biomarkers in DLBCL and PMBCL patients.

Materials and methods: Blood plasma samples (n=168) and matched diagnostic formalin-fixed paraffin-embedded (FFPE) tissue samples (n=69) of DLBCL patients, PMBCL patients and healthy controls were collected between 2016-2021. Plasma samples were collected at diagnosis, at interim evaluation, after treatment, and in case of refractory or relapsed disease. RNA was extracted from 200 µl plasma using the miRNeasy serum/plasma kit and from FFPE tissue using the miRNeasy FFPE kit. RNA was subsequently sequenced on a NovaSeq 6000 instrument using the SMARTer Stranded Total RNA-seq pico v3 library preparation kit.

Results: Higher cfRNA concentrations were demonstrated in lymphoma patients compared to healthy controls. A large number of differentially abundant genes were identified between the cell-free transcriptomes of DLBCL patients, PMBCL patients, and healthy controls. Overlap analyses with matched FFPE samples showed that blood plasma has a unique transcriptomic profile that significantly differs from that of the tumor tissue. As a good concordance between tissue-

derived gene expression and the immunohistochemistry Hans algorithm for cell-of-origin (COO) classification was demonstrated in the FFPE samples, but not in the plasma samples, a 64-gene cfRNA classifier was developed that can accurately determine COO in plasma. High plasma levels of a 9-gene signature (*BECN1, PRKCB, COPA, TSC22D3, MAP2K3, UQCRHL, PTMAP4, EHD1, NAP1L1* pseudogene) and a 5-gene signature (*FTH1P7, PTMAP4, ATF4, FTH1P8, ARMC7*) were significantly associated with inferior progression-free and overall survival in DLBCL patients, respectively, independent of the NCCN-IPI score.

Conclusion: Total RNA sequencing of blood plasma samples allows the analysis of the cell-free transcriptome in DLBCL and PMBCL patients and demonstrates its unexplored potential in identifying diagnostic, cell-of-origin, and prognostic cfRNA biomarkers.

1 Introduction

Diffuse large B-cell lymphoma (DLBCL) is the most common histological subtype of non-Hodgkin's lymphomas (NHL), representing approximately 25% of new diagnoses. DLBCL can occur *de novo* or because of transformation from low-grade B-cell lymphomas (1). Although a sustained complete response (CR) can be obtained in most patients with first-line R-CHOP immunochemotherapy (rituximab, cyclophosphamide, vincristine, doxorubicin, and prednisone), patients with refractory or relapsed (R/R) disease have a poor prognosis, despite second-line treatments (2).

Two decades ago, gene expression profiling (GEP) was used to discriminate between different cell-of-origin (COO) DLBCL subgroups, germinal center B-cell–like (GCB) and activated B-cell–like (ABC), which was clinically implemented through the use of surrogate immunohistochemistry (IHC) algorithms (3, 4). Since then, DLBCL has increasingly been recognized as a highly heterogeneous disease with respect to morphology, genetics, and biological behavior. This evolution was exemplified by recognizing primary mediastinal B-cell lymphoma (PMBCL) as a separate entity by the World Health Organization (WHO) in 2008 based on clinicopathologic features and a distinct molecular signature that overlapped with nodular sclerosis classical Hodgkin lymphoma (5–7). PMBCL accounts for 2-4% of NHL, and although outcomes have improved in the modern rituximab era, standard treatment is still a topic of debate and refractory disease correlates with poor outcome. Moreover, approximately 5% of DLBCL patients show a molecular PMBCL phenotype in the absence of mediastinal involvement, demonstrating the overlap, complexity, and challenges of current classifications (8). More recent evolutions include the identification of the double expressor (DEL) DLBCL subtype, characterized by an overexpression of MYC and BCL2 proteins not related to

underlying chromosomal rearrangement, and double-hit lymphomas (DHL), a separate entity defined as a dual rearrangement of MYC and BCL2 and/or BCL6, both associated with poor prognosis (1, 9). The increasing number of subtypes highlights that, although major progress has been made, further in-depth molecular analyses are needed to identify and characterize high-risk patients, to elucidate the biological pathways involved and reveal therapeutic targets, and to ultimately improve outcomes. In this setting, profiling through massively parallel sequencing has proven to be a powerful tool.

A liquid biopsy is the process of investigating tumor-derived cells or biomaterials like cell-free nucleic acids, metabolites, proteins, or extracellular vesicles through biofluid sampling, without the need of a tissue biopsy. Potential advantages include its non-invasive nature, its ability to reflect inter- and intra-tumor heterogeneity, and the possibility of repeated measurements through longitudinal profiling during disease course or treatment (10). Although most studies have focused on the use of cell-free or circulating tumor DNA fragments (cfDNA/ctDNA), there has been increased interest in different forms of coding and non-coding cfRNA, both circulating and extracellular vesicle/platelet-encapsulated, including messenger RNA (mRNA), microRNA (miRNA), long non-coding RNA (lncRNA), and circular RNA (circRNA) (11). Higher circulating transcript levels have been found in patients with solid and hematological malignancies, in which these are considered to play crucial roles in intercellular communication and to contribute to proliferation, malignant transformation, angiogenesis, and immune response escape (12, 13). The goal of this study was to explore the cell-free transcriptome of DLBCL and PMBCL patients using longitudinally collected plasma and matched formalin-fixed paraffin-embedded (FFPE) tissue samples, and to identify diagnostic, COO, and prognostic biomarkers, complementary to current risk stratification tools.

2 Materials and methods

2.1 Sample collection and processing

2.1.1 Sample collection

A total of 168 longitudinally collected plasma samples derived from 41 DLBCL NOS (hereinafter referred to as DLBCL) and 14 PMBCL patients were collected during the period 06/2016 and 09/2021 at Ghent University Hospital in Ghent (Belgium) and AZ Delta hospital in Roeselare (Belgium). Plasma samples were drawn at the time of diagnosis, at interim evaluation (after 4 cycles of immunochemotherapy), at final evaluation, and in the case of R/R disease. An overview of the collected plasma samples at each timepoint can be found in Supplementary Figure 1. Diagnostic FFPE samples were obtained from all patients that had a diagnostic plasma sample, except for 2 DLBCL and 1 PMBCL patient, of which insufficient FFPE material was available for RNA extraction. The study also included plasma samples from 22 healthy controls, as well as 29 FFPE samples derived from non-malignant lymph node tissue. Blood samples were drawn using PAXgene blood ccfDNA tubes (BD Biosciences, 768165) and processed to plasma within 4 hours after blood draw. Immunohistochemical staining was routinely performed on each DLBCL FFPE tissue sample for the markers CD10, BCL6, MUM1, BCL2 and MYC to determine COO classification and double expressor status. Fluorescence *in situ* hybridization (FISH) was performed for MYC and BCL2 rearrangements to identify DHL. Therapy response was assessed by F-fluorodeoxyglucose Positron Emission Tomography/computerized tomography (FDG-PET/CT) according to the Lugano guidelines (14). Response to first-line therapy was defined as obtaining a sustained complete remission (CR) without evidence of relapsed disease within the follow-up period (mean follow-up of 1018 and 1049 days for the DLBCL and PMBCL patients, respectively).

2.1.2 Preparation of plasma and FFPE scrolls

Plasma was obtained by using a one-step centrifugation protocol (1900g x 15min without brake at room temperature) and subsequently frozen and stored at -80°C. FFPE samples were obtained from the Pathology Department of Ghent University Hospital and were re-examined by an experienced pathologist to confirm the correct diagnosis, to determine the percentage of tumor invasion, and to select adequate tumor blocks. Five scrolls of 10 μm were cut from FFPE blocks on a Leica RM2125 RTS manual rotary microtome (Leica Microsystems, Germany) and put into RNase-free Eppendorf tubes (Eppendorf, Germany). If the FFPE also included non-malignant tissue, macrodissection was performed to enrich lymphoma invaded tissue. The FFPE scrolls were processed within 4 hours after preparation.

2.1.3 RNA extraction

RNA extraction was performed on 200 μl of plasma and on 5 scrolls of 10 μm using the miRNeasy serum/plasma kit and miRNeasy FFPE kit, respectively. In each plasma sample, 2 μL of Sequin spike-in controls were added to the sample lysates (1/5,000 of stock solution mix A; Garvan Institute of Medical Research), as well as 2 μL of External RNA Control Consortium (ERCC) spike-in

controls (1/12,500; ThermoFisher Scientific, 4456740) were added to the RNA eluate. Genomic DNA was removed by adding 1 μL HL-dsDNase (ArcticZymes, 70800-202) and 1.6 μL reaction buffer (ArcticZymes, 66001) to 12 μL RNA eluate, followed by 10 min incubation at 37°C, and 10 min incubation at 55°C to inactivate the DNase. RNA was subsequently stored at -80°C and thawed on ice immediately before library preparation.

2.1.4 Library preparation and sequencing

Total RNA sequencing libraries were prepared starting from 8 μL of RNA eluate (plasma) or 40 ng total RNA (FFPE) using the SMARTer Stranded Total RNA-Seq Kit v3 – Pico Input Mammalian (Takara, 634487) according to the manufacturer's protocol. Equimolar library pools were prepared based on qPCR quantification with KAPA Library Quantification Kit (Roche Diagnostics, Belgium, KK4854). The libraries were paired-end sequenced (2x100 nucleotides) on a NovaSeq 6000 instrument using a NovaSeq S2 kit (Illumina, 20028315) with standard workflow loading of 0.65 nM (2% PhiX). BCL files generated by the Illumina sequencing system were processed using the Illumina Bcl2fastq (v. 2.20) software to generate and demultiplex fastq files. Raw reads were assessed for quality using FastQC (v. 0.11.9) (15). The unique molecular identifier (UMI) sequences in the Pico v3 SMART UMI adapters were first extracted and added to the read name with UMI-tools (v. 1.0.0) (16). Next, the sequencing data were processed with an in-house developed and validated pipeline using cutadapt (v. 1.16) (17), SAMtools (v. 1.9) (18), and STAR (v. 2.7.3a) (19) to obtain deduplicated, high-quality and aligned RNA counts against the human genome (hg38). Ribosomal contamination was assessed using the BBMap tool (v. 38.87) (20). Duplicate reads were identified using the Picard tool (v. 2.21.6) (21). Additional QC statistics were generated by using multiQC (v. 1.9) software package (22). Gene counts were determined with HTSeq (v. 0.11.0) in reverse stranded mode, only considering uniquely mapping reads (23). Ensembl release 91 was used to annotate reads within human genes (24). Circtools (v. 1.2.0) was used to identify, annotate, and quantify back-spliced junction (BSJ) reads from circRNAs (25).

2.2 Data analysis

2.2.1 RNA concentration

Blood plasma RNA concentration was determined as previously described (26). Briefly, the mass of the top abundant spike-in control (ERCC00130) was calculated based on the input concentration and volume of spike-in mix added to the sample. The corresponding RNA concentration was then estimated by multiplying the ERCC00130 spike mass by the ratio of total reads mapped to the endogenous human genome and the number of reads mapped to the specific ERCC00130 spike, and finally dividing the obtained mass by the plasma volume of the sample.

2.2.2 Differential abundance analysis

Differential abundance analysis on normalized counts was performed using DESeq2 (v. 1.36.0) (27). Counts were pre-filtered by requiring a minimum of 10 counts in at least half of the samples

in one of the compared groups. CircRNA counts were pre-filtered by requiring a minimum of 4 back-splice junction counts in at least half of the samples in one of the compared groups. In the DESeq2 result table, genes with a Benjamini-Hochberg corrected p-value (q-value) below 0.05 were considered differentially abundant. Volcano plots were visualized using EnhancedVolcano (v. 1.14.0) and Venn diagrams were made using VennDiagramm (v. 1.7.3) (28, 29). Concerning the longitudinal differential abundance analysis, the ImpulseDE2 algorithm (v. 1.6.1) was applied on all patients for which all plasma samples during first-line therapy (diagnosis, interim evaluation, and final evaluation) were available using a q-value below 0.05. ImpulseDE2 models the gene-wise abundance trajectories over time with a descriptive single-pulse function, which is based on a negative binomial noise model with dispersion trend smoothing by DESeq2 (30).

2.2.3 COO classification using GEP and IHC

To investigate the performance of a tissue-derived GEP to correctly classify FFPE and plasma samples based on COO as predicted by IHC, both a normalized rank and a standardized abundance procedure were applied (for more details, cfr. Supplementary Table 1). The tissue-derived GEP was constructed starting from the 100 best COO class-predicting genes as initially identified in the Lymphochip microarray data by Alizadeh et al., of which 59 genes were retained following the removal of clone duplicates and uncertain clones (3). For the diagnostic plasma samples, besides the tissue-derived GEP, the same procedures were performed using a plasma-derived GEP as input, i.e. the differential abundant genes (DAG) identified between non-GCB and GCB samples in our cohort.

2.2.4 Survival analysis

To determine which genes best correlated with the different prognostic outcomes studied, a univariate Cox model was run for each gene, and only those genes with a LogRank p-value <=0.01 and absolute beta coefficient >=2 were retained. The significant genes were then categorized as favorable and unfavorable genes according to the beta coefficients (a negative beta coefficient represented a hazard ratio <1, meaning that a gene correlated with a favorable outcome; and vice versa for the positive beta coefficients). Subsequently, a hierarchical clustering was performed on the selected genes using a Euclidean distance and average agglomeration method ("stats" R library v. 3.6.2). Next, the clusters were detected using the Dynamic Tree Cut (v. 1.63-1) R library to avoid the arbitrary choice of the dissimilarity cut-off between clusters when performing a fixed height tree cut (31). For each of the clusters, a gene signature was calculated, which was computed as the average expression value of the genes present in the corresponding cluster. The gene signatures together with NCCN-IPI score were used as covariates of a multivariable Cox regression model. Finally, the samples of each prognostic subgroup were divided into a high-risk group and low-risk group according to the gene signatures. High abundance of the unfavorable gene signature defined a high-risk group, while a low abundance defined a low-risk subgroup and vice versa for favorable

signatures. Optimal cutoff point analyses were performed with maximally selected rank statistics using the *maxstat* package (v. 0.7-25) (32). Progression-free survival (PFS) was evaluated from the date of enrollment to the date of disease progression, relapse, death from any cause, or date of last follow-up in case of no event. Overall survival (OS) was evaluated from the date of enrollment to the date of death from any cause or date of last follow-up in case of no event. LogRank tests were used to assess differences in the OS and PFS rates calculated by Kaplan–Meier estimates.

2.2.5 Gene set enrichment analysis

A pre-ranked gene set enrichment analysis using GSEA (v. 4.2.3) was performed to explore the functionally enriched pathways and hallmark gene sets related to subgroups, based on the $\log_2$ transformed fold changes between the different groups obtained from DESeq2 differential abundance analysis (33). Significant enrichment was defined by a false discovery rate ≤ 0.05. Hallmark and Canonical Pathways gene sets were obtained via the Molecular Signatures Database MSigDB (v. 7.5.1) (34). Pathways were up- or downregulated according to the enrichment score (ES) which represents the degree to which a set was overrepresented at the top or bottom of the ranked list.

2.2.6 Statistical analysis

All analyses were conducted using the R statistical software package (v. 4.0.5) (35). Kruskal-Wallis tests were used to compare multiple groups, followed by Wilcoxon rank-sum tests for pairwise comparisons with Bonferroni-Holm multiple testing correction. Significance was defined as q-values smaller than 0.05. Using GeneOverlap (v. 1.32.0), significance of overlap between gene sets was determined by Fisher's exact test with the odds ratio representing the strength of the association. The universe consisted of all genes obtained in the corresponding differential abundance analysis. A Jaccard index was calculated to assess similarity between the gene sets, with 0 indicating no similarity and 1 indicating identical sets (36). Correlation analysis was carried out using the Pearson correlation coefficient (R) between different metrics.

3 Results

3.1 Clinical characteristics

A total of 41 DLBCL and 14 PMBCL patients were included. The mean age at inclusion was 65.3 and 37.0 years, respectively (Table 1). Most patients had stage IV disease at time of diagnosis, as defined by the Ann Arbor criteria (61% and 57.1%, respectively). Bulky disease was more common in PMBCL compared to DLBCL patients (50% and 22%, respectively). The National Comprehensive Cancer Network International Prognostic Index score (NCCN-IPI) score was low, low-intermediate, high-intermediate, and high in 2.4%, 39.0%, 39.0%, and 19.5% and in 21.4%, 71.4%, 7.1% and 0% of DLBCL and PMBCL patients, respectively (37). Based on the Hans algorithm, 20/41 and 21/41 DLBCL patients were classified as GCB

TABLE 1 Demographics, clinical characteristics, first-line treatments, and outcomes of DLBCL patients, PMBCL patients, and healthy controls included in this study.

	DLBCL NOS	PMBCL	healthy controls
clinical characteristics			
sex (m/v)			
male	17/41 (41.5%)	8/14 (57.1%)	13/22 (59%)
female	24/41 (58.5%)	6/14 (42.9%)	9/22 (41%)
age at diagnosis (yr)	65.41 (+/- 14.94)	37.00 (+/- 14.35)	58.2 (+/- 20.61)
Ann Arbor stage			
I	4/41 (9.8%)	1/14 (7.14%)	
II	7/41 (17.1%)	4/14 (28.57%)	
III	5/41 (12.2%)	1/14 (7.14%)	
IV	25/41 (61.0%)	8/14 (57.14%)	
bulky disease	9/41 (22.0%)	7/14 (50.0%)	
NCCN-IPI score			
low (0–1)	1/41 (2.4%)	3/14 (21.4%)	
low-intermediate (2–3)	16/41 (39.0%)	10/14 (71.4%)	
high-intermediate (4–5)	16/41 (39.0%)	1/14 (7.1%)	
high (>=6)	8/41 (19.5%)	0/14 (0%)	
tumor characteristics			
COO: GCB	20/41 (48.8%)	NA	
COO: non-GCB	21/41 (51.2%)	NA	
DEL	17/41 (41.5%)	NA	
MYC expression	23/41 (56.1%)	3/14 (21.4%)	
BCL2 expression	28/41 (68.3%)	6/14 (42.9%)	
MYC rearrangement	3/41 (7.3%)	NA	
BCL2 rearrangement	8/41 (19.5%)	NA	
first-line treatment			
R-(mini)CHOP	41/41 (100%)	0/14 (0%)	
DA-EPOCH-R	NA	3/14 (21.4%)	
R-ACVBP	NA	11/14 (78.6%)	
intrathecal chemotherapy	23/41 (56.1%)	14/14 (100%)	
laboratory values			
mean LDH (IU/L)	428.15 (+/- 406.25)	424.57 (+/- 258.50)	
mean ß2 microglobulin (mg/l)	2.44 (+/- 0.95)	1.92 (+/- 0.42)	
mean igG (g/l)	9.66 (+/- 5.43)	11.53 (+/- 2.68)	
outcome after first-line therapy			
CR	21/41 (51.2%)	12/14 (85.7%)	
R/R	19/41 (46.3%)	2/14 (14.3%)	
death	1/41 (2.4%)	0/14 (0%)	

(Continued)

TABLE 1 Continued

	DLBCL NOS	PMBCL	healthy controls
outcome of R/R patients after second- and third-line treatments			
CR	8/19 (42.1%)	2/2 (100%)	
death	11/19 (57.9%)	0/2 (0%)	
outcome after any therapy			
CR	29/41 (70.73%)	14/14 (100%)	
death	12/41 (29.27%)	0/14 (0%)	

Percentages and standard deviations are written in parentheses. COO, cell-of-origin; CR, complete remission; DEL, double-expressor lymphoma; DLBCL NOS, diffuse large B-cell lymphoma not otherwise specified; GCB, germinal center B-cell type; LDH, serum lactate dehydrogenase; NCCN-IPI, National Comprehensive Cancer Network International Prognostic Index score; PMBCL, primary mediastinal B-cell lymphoma; R/R, refractory/relapsed disease.

and non-GCB, respectively (4). Double expression for MYC and BCL2 was demonstrated in 17/41 patients. MYC and BCL2 rearrangements were present in 3/41 and 8/41 of DLBCL cases, respectively, without the presence of a DHL. The mean serum lactate dehydrogenase (LDH) value (reference interval 105-250 IU/L) was 428.2 IU/L and 424.6 IU/L and the mean ß2-microglobulin (reference interval 1.09-2.53 mg/l) was 2.4 mg/l and 1.9 mg/l for the DLBCL and PMBCL patients, respectively.

The DLBCL patients were treated with first-line R-(mini) CHOP and the PMBCL patients with dose-intensive rituximab, doxorubicin, cyclophosphamide, vindesine, bleomycin, and prednisone (R-ACVBP) with subsequent consolidation or dose-adjusted EPOCH-rituximab therapy (DA-EPOCH-R) (38, 39). After first-line therapy, 21/41 of DLBCL patients obtained a sustained complete remission (CR), 19/41 had R/R disease, and 1/41 died during first-line treatment. With second- and third-line treatments, CR could be obtained in 8/19 of R/R DLBCL patients. All treatments combined, 29/41 DLBCL patients obtained a CR and 12/41 died. In the PMBCL patients, CR could be obtained in 12/14 of patients after first-line treatment and in 2/2 of R/R patients with second-line treatments. Lastly, a total of 22 healthy age and gender matched controls were included.

3.2 cfRNA concentration in blood plasma

The blood plasma cfRNA concentration was higher in the lymphoma patients at the time of diagnosis compared to healthy controls (p = 0.0098; Figure 1A). A LDH value above normal reference level at diagnosis was associated with increased cfRNA levels (p=0.025; Figure 1B). The correlation between cfRNA concentration and LDH levels was, however, weak in the DLBCL samples (r = 0.4; p = 0.029) with low predictive value (AUC of 0.65) and not significant in the PMBCL samples (r = -0.28; p = 0.36) (Supplementary Figure 2). Within the DLBCL samples, a significant decrease in cfRNA concentration was demonstrated between the diagnostic and the post-treatment timepoint in first-line therapy responders (p=0.02), as opposed to non-responders (Figures 1C, D). The individual cfRNA evolution of several patients during therapy is visualized in Supplementary Figure 3.

3.3 cfRNA biomarkers in DLBCL and PMBCL patients

3.3.1 Diagnostic markers

A large number of plasma-derived DAG were identified between DLBCL patients, PMBCL patients and healthy controls in the diagnostic samples. A total of 604, 4076, and 2925 genes were differentially abundant between DLBCL and healthy control samples, between PMBCL and healthy control samples, and between PMBCL and DLBCL samples, respectively. The majority of DAG were protein-coding, but other RNA biotypes included lncRNAs, pseudogenes, snRNAs, snoRNAs, mt-tRNAs, rRNAs, and circRNAs (Supplementary Figures 4, 5 and Supplementary Table 2). Of the differentially abundant circRNAs, 40.2% of their linear counterparts were also significantly differentially abundant. When both the circRNA and the linear counterpart were significantly differentially abundant, the direction corresponded in 93.9% of the cases.

GSEA between the lymphoma samples and healthy controls showed enrichment of several hallmark gene sets, including interferon type I and II response, interleukin-6/Janus Kinase 2/Signal Transducer and Activator of Transcription 3 (IL-6/JAK/STAT3) signaling, Tumor Necrosis Factor (TNF) alpha signaling via Nuclear factor kappa-light-chain-enhancer of activated B-cells (NF-κB), epithelial mesenchymal transition, inflammatory response, and organization/degradation of the extracellular matrix (ECM). Moreover, compared to PMBCL samples, DLBCL samples showed enrichment for oxidative phosphorylation, KRAS signaling, G2M checkpoint signaling, NOTCH signaling, stabilization of p53, and regulation of Phosphatase and Tensin homolog (PTEN) stability and activity. Compared to the DLBCL samples, the PMBCL samples demonstrated enrichment for coagulation, epithelial mesenchymal transition, MAPK activation, as well as ECM organization and regulation. (Supplementary Figure 6).

To investigate the differences between the plasma and tissue compartment, the overlap was assessed between the DAG obtained in the diagnostic plasma samples and the matched FFPE samples. (Figure 2 and Supplementary Tables 2, 3) A total of 335 DAG were shared for DLBCL versus controls (p<0.001; Jaccard Index (JI) 0.051; odds ratio (OR) 3.4), 1579 DAG for PMBCL versus controls (p<0.001; JI 0.21; OR 2.7), and 174 DAG for PMBCL versus DLBCL
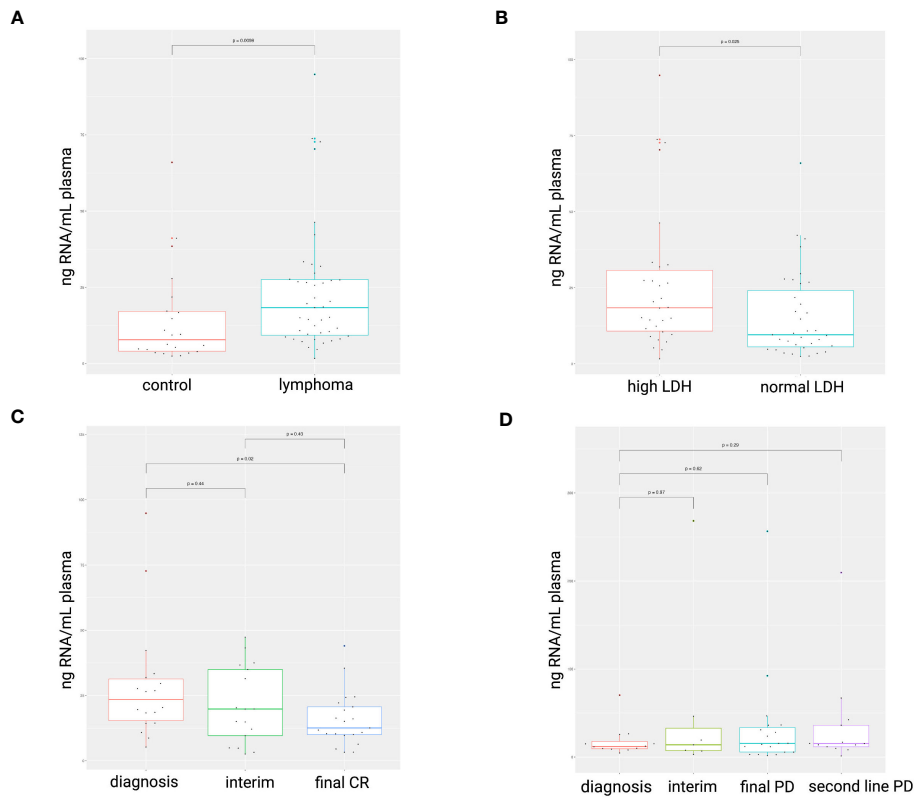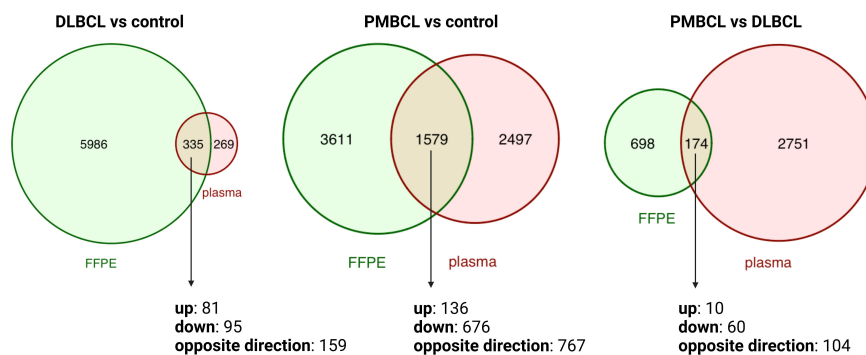
FIGURE 1
cfRNA concentration (ng RNA/mL plasma) in diagnostic plasma samples. Higher cfRNA concentrations were found in lymphoma patients compared to healthy controls **(A)**. Total cfRNA concentrations were increased in plasma samples with elevated LDH levels compared to normal levels (105-250 IU/L) **(B)**. A decrease in cfRNA concentration was observed in DLBCL patients successfully treated with R-CHOP **(C)**, whereas concentrations remained unchanged in non-responders **(D)**. LDH, serum lactate dehydrogenase; CR, complete response; PD, progressive disease.

(p<0.001; JI 0.048; OR 1.8), respectively. This corresponds with 5.30%, 30.42%, and 19.95% of the FFPE DAG also present in the blood plasma, respectively. Of these shared DAG, 52.5%, 51.4%, and 40.2% had the same direction of dysregulation, respectively. The top 10 genes that had the highest Log2FC in FFPE tissue that were also significantly dysregulated in the same direction in the plasma were identified for DLBCL versus controls (upregulated: *LAT2*, *SEMA4A*, *LOX*, *ADAM8*, *PTAFR*; downregulated: *FABP4*, *FAM107A*, *LIFR*,

LPL, *FMO2*), for PMBCL versus controls (upregulated: *PTGIR*, *CLIP2*, *ANKRD33B*, *ZNF185*, *TREM1*; downregulated: *IGLC2*, *LIFR*, *PDK4*, *PLAC8*, *IGKC*), and for PMBCL versus DLBCL (upregulated: *ANK1*, *PTGIR*, *ANKRD33B*, *SPINT2*, *UNC13B*; downregulated: *PLEKHG1*, *PLAC8*, *FOXP1*, *CCND2*, *PRDX2*).

Secondly, normalized mRNA counts of lymphoma-related genes were compared between DLBCL patients, PMBCL patients, and controls in both the tissue and blood plasma samples, including

FIGURE 2
Venn diagrams depicting the overlap between differentially abundant transcripts from matched FFPE and diagnostic plasma samples for DLBCL versus controls, PMBCL versus controls, and PMBCL versus DLBCL. DLBCL, diffuse large B-cell lymphoma; FFPE, formalin-fixed paraffin-embedded tissue; PMBCL, primary mediastinal B-cell lymphoma.

markers of B-cell differentiation (*CD19, CD20, PAX5, IRF4, BCL6*), T-cell differentiation (*CD3, CD4, CD5, CD8*), thymic B-cell differentiation (*CD23, MAL*), NF-κB (*REL, TRAF1*), cell proliferation (MKI67), activation (*CD30*), and apoptosis (*BCL2*). As expected, many of these genes were differentially abundant between the lymphoma subtypes and healthy controls in tissue samples. These differences were, however, not significant for most markers in the matched plasma samples. Moreover, when comparing DLBCL to control samples, discordant results were seen for *PAX5* (GCB p=0.98, non-GCB p=0.043) and *BCL2* (GCB p=0.61, non-GCB p=0.021) compared to the findings in the FFPE samples. When comparing PMBCL to control samples, concordant results were seen for *BCL2* (p=0.0025) and *CD3* (p=0.0018) and discordant results were noted for *CD20* (p=0.0095), *IRF4* (p=0.0011), and *PAX5* (p=0.041). (Supplementary Figures 7, 8).

Lastly, a differential abundance analysis was performed for cfRNA biomarkers that were previously reported in liquid biopsy studies involving DLBCL and/or PMBCL patients (40–45). For each of these genes, results are shown for both the plasma and FFPE samples (Table 2).

3.3.2 COO and DEL status

In clinical practice, IHC markers are used to discern the COO subtypes of DLBCL (e.g., CD10, BCL6, and MUM1 in the Hans algorithm), as well as DEL status (MYC and BCL2) (4, 9). To compare IHC protein expression with the gene abundance profiles in the tissue or blood plasma, normalized mRNA counts for these markers were compared between IHC positive and negative DLBCL patients. Our results showed increased gene counts in the FFPE samples for which there was also positive IHC staining for the

TABLE 2 Differential abundance analysis for cfRNA markers that have previously been reported for DLBCL or PMBCL patients.

RNA marker	biotype	sample	DLBCL vs control	non-GCB vs GCB	PMBCL vs control	PMBCL vs DLBCL
CCND2 (40)	mRNA	FFPE	↓ (q=0.026)	↑ (q=0.0061)	↓ (q<0.001)	NS
		plasma	↓ (q=0.003)	NS	↓ (q<0.001)	↓ (q=0.0018)
BCL2 (40)	mRNA	FFPE	NS	NS	↓ (q<0.001)	NS
		plasma	↓ (q=0.038)	NS	↓ (q=0.0012)	NS
MYC (40)	mRNA	FFPE	NS	NS	↓ (q=0.0042)	↓ (q=0.0093)
		plasma	NS	NS	↓ (q=0.002)	NS
BCL6 (40, 42)	mRNA	FFPE	↑ (q<0.001)	↓ (q=0.033)	↑ (q<0.001)	NS
		plasma	NS	NS	NS	NS
FN1 (40)	mRNA	FFPE	↑ (q<0.001)	↓ (q=0.02)	↑ (q<0.001)	↑ (q=0.012)
		plasma	NS	↓ (q=0.037)	NS	NS
PTEN (42)	mRNA	FFPE	NS	NS	↓ (q=0.0011)	NS
		plasma	NS	NS	NS	NS
CREBBP (41)	mRNA	FFPE	↓ (q<0.001)	NS	↓ (q<0.001)	NS
		plasma	NS	NS	NS	NS
LMO2 (40)	mRNA	FFPE	NS	↓ (q=0.033)	↑ (q<0.001)	↑ (q<0.001)
		plasma	NS	NS	NS	NS
TUG1 (43)	lncRNA	FFPE	NS	NS	NS	NS
		plasma	NS	↑ (q=0.013)	NS	NS
GAS5 (44)	lncRNA	FFPE	↓ (q<0.001)	NS	↓ (q=0.0055)	NS
		plasma	NS	NS	↓ (q=0.0053)	↓ (q=0.033)
XIST (44)	lncRNA	FFPE	NS	NS	NS	NS
		plasma	↑ (q=0.011) NS when sex matched	NS	NS	NS
circAPC (45)	circRNA	plasma	NS	NS	NS	NS

Genes with q-values >0.05 are labeled as NS. CircRNA, circular RNA; DLBCL, diffuse large B-cell lymphoma; lncRNA, long non-coding RNA; mRNA, messenger RNA; NS, not significant; PMBCL, primary mediastinal B-cell lymphoma.
arrow up means upregulated.
arrow down means downregulated.

corresponding protein. For the plasma samples, however, no significant difference in mRNA counts was observed. (Figure 3).

Next, we checked if a tissue-derived GEP, containing the best COO class-predicting genes as initially identified in the Lymphochip microarray data by Alizadeh et al., could correctly classify the DLBCL samples (using the Hans algorithm as ground truth) (3). When applied on the FFPE samples, a good concordance with the Hans algorithm could be demonstrated with an accuracy

FIGURE 3
Normalized mRNA counts of *MYC, BCL2, CD10, BCL6,* and *MUM1* within IHC positive and negative DLBCL patients. DLBCL, diffuse large B-cell lymphoma; FFPE, formalin-fixed paraffin-embedded; IHC, immunohistochemistry; mRNA, messenger RNA.

over 85% and an AUC of 0.93 using both a normalized rank and a standardized abundance approach (Figure 4A). When applied on the diagnostic plasma samples, however, this tissue-derived classifier failed (accuracy of 0.6 and 0.53 with an AUC 0.56 and 0.48 using a normalized rank and standardized abundance approach, respectively) (Figure 4B). Differential abundance analysis on the plasma samples identified a total of 64 DAG between non-GCB and GCB DLBCL cases, as determined by the Hans algorithm (Supplementary Table 2). A plasma-derived GEP based on these DAG was able to differentiate GCB from non-GCB patients with high accuracy (accuracy of 92% and 86% with an AUC of 0.97 and 0.92 using a normalized rank and standardized abundance approach, respectively) (Figure 4C). Notably, there was no overlap between the tissue- and plasma-derived GEP.

Lastly, no DAG could be identified between DEL and non-DEL samples, illustrating that the transcriptional differences of these subgroups are not clearly reflected in the plasma compartment. This is in line with the observation of only 3 DAG between the matched FFPE samples. (Supplementary Table 3).

3.3.3 Prognostic markers in DLBCL
3.3.3.1 Survival analysis

When comparing the diagnostic plasma samples of DLBCL responders to non-responders after first line R-CHOP immunochemotherapy, a total of 24 genes were significantly associated with PFS after R-CHOP in univariate cox regression analysis. (Supplementary Table 4) The significant genes were clustered according to the beta coefficients (a negative coefficient represented a hazard ratio <1, meaning that a gene correlated with a favorable outcome; and vice versa for a positive beta coefficient). For both the favorable and unfavorable cluster, a gene signature was computed as the average expression value of its genes. A higher abundance of the unfavorable 9-gene signature (*BECN1, PRKCB, COPA, TSC22D3, MAP2K3, UQCRHL, PTMAP4, EHD1, NAP1L1* pseudogene) was significantly associated with decreased PFS, both in univariate (p<0.001) and in multivariable regression analysis (p<0.001), the latter independent of the NCCN-IPI score. (Figure 5 and Supplementary Figure 9) A similar analysis was performed for the CR versus death groups after any treatment for OS analysis. A total of 10 genes were significantly associated with OS after any therapy in univariate cox regression analysis, of which five were categorized as unfavorable and five as favorable. (Supplementary Table 4) A higher abundance of the unfavorable 5-gene signature (*FTH1P7, PTMAP4, ATF4, FTH1P8, ARMC7*) was significantly associated with worse OS in both univariate (p<0.001) and multivariable regression analysis (p=0.004). (Figure 5 and Supplementary Figure 9).

3.3.3.2 Longitudinal gene abundance analysis during first line therapy

When comparing matched pre- and post-therapy samples in DLBCL responder patients after successful R-CHOP therapy, a total of 184 DAG were identified. (Supplementary Table 4) Among the downregulated genes with the highest log2FC were B-cell related markers, including *MS4A1 (CD20), PAX5, IGHD, FCRL1, CD79A, TNFRSF13C, IGHM, FCER2, CD72, CD22, CARD11, IRF4, EBF1, BLK, BLNK, STAP1, FAM129C, BACH2, TCL1A,* POU2AF1 and
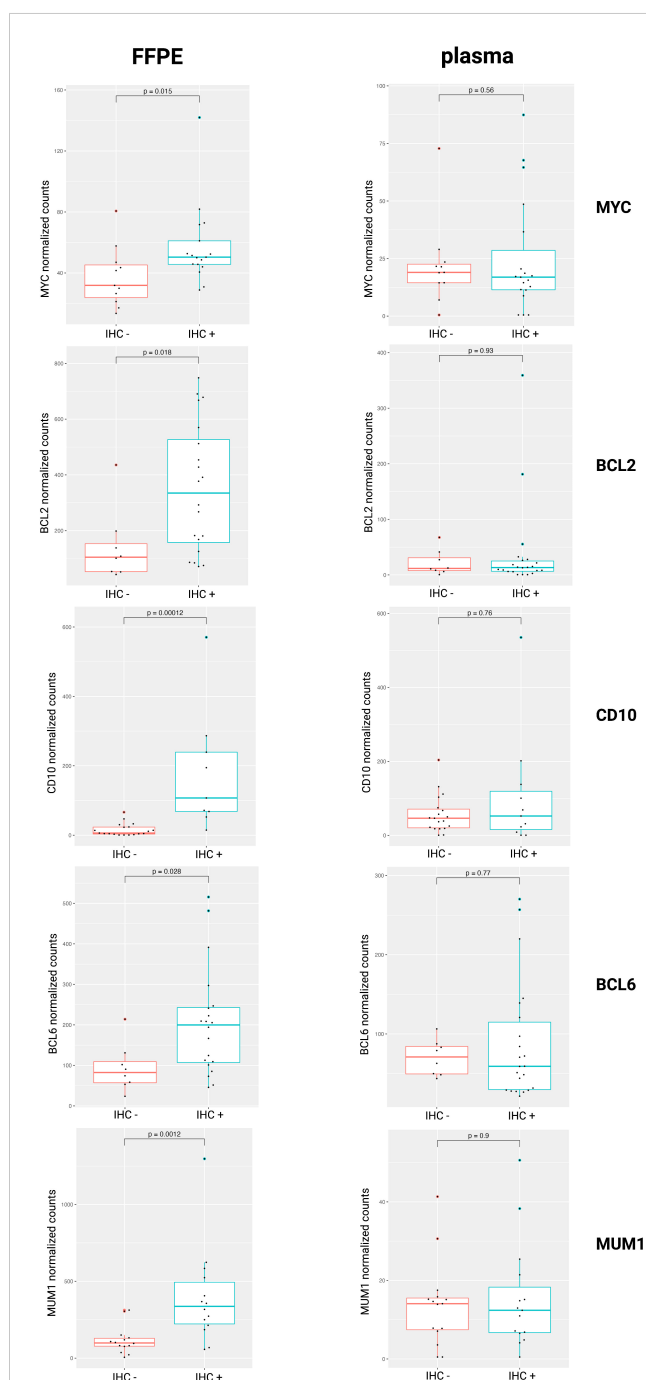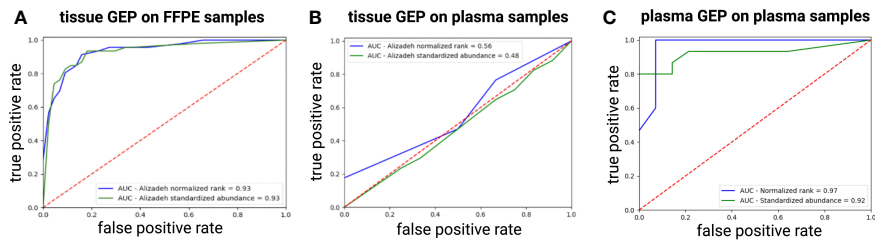
FIGURE 4

ROC curve showing the performance of the tissue-derived GEP, containing the best COO class-predicting genes as initially identified in the Lymphochip microarray data, to classify DLBCL samples by COO as predicted by the Hans algorithm for FFPE **(A)** and plasma samples **(B)** using both a normalized rank and a standardized abundance approach. Panel **(C)** shows the ROC curve depicting the performance of our plasma-derived GEP to classify DLBCL samples by COO as predicted by the Hans Algorithm using both a normalized rank and a standardized abundance approach **(C)**. AUC, area under the ROC curve; COO, cell-of-origin; DLBCL, diffuse large B-cell lymphoma; FFPE, formalin-fixed paraffin-embedded; GEP, gene expression profiling; ROC, Receiver Operating Characteristic.

FOXP1. Four of these DAG (*JAK3*, *LRP1*, *PHYKPL*, and *PLXNB2)* were also upregulated at pre-therapy timepoint compared to control samples, but decreased significantly during therapy below healthy control levels, with the opposite being true for *IGF2*. When analyzing the interim time point during therapy, a clear decrease in the aforementioned B-cell related markers was already noted to similar levels as the post-therapy time point. Additionally, a transient increase in neutrophil and inflammation associated markers, including *CD177*, *MMP8*, *CEACAM8*, *OLFM4*, *BPI*,

MS4A3, and *MPO*, was demonstrated at the interim timepoint. In contrast, within the DLBCL non-responders, only a transient decrease of several B-cell related markers (*IGHM*, *PAX5*, *CD22*, *POU2AF1*) was noted when comparing the diagnostic samples to the interim samples, no DAG could be identified between the diagnostic and the R/R samples, and no transient increase in inflammation associated markers was noted.

Finally, a longitudinal analysis using impulseDE2 was performed to identify gene abundance trajectories that
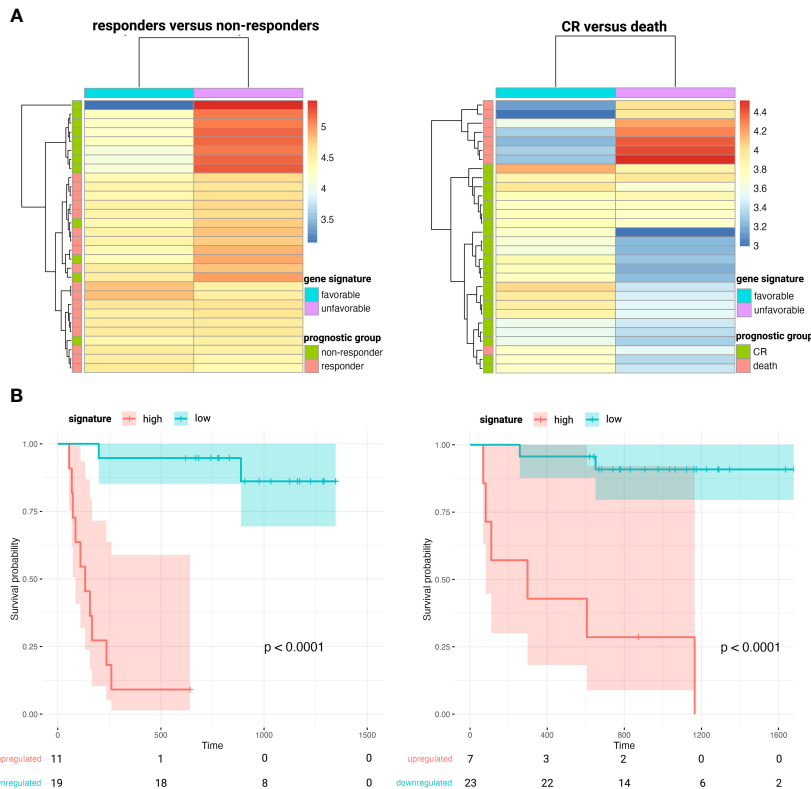

FIGURE 5

Heatmaps clustering DLBCL responders versus non-responders after first-line R-CHOP therapy (**A**, left) and patients that obtain CR versus death after any therapy (**A**, right) based on the abundance of the corresponding favorable and unfavorable gene signatures. Kaplan-Meier survival curves depicting poorer PFS (**B**, left) and OS (**B**, right) in patients with higher plasma levels of the unfavorable 9- and 5-gene signature, respectively. CR, complete remission; DLBCL, diffuse large B-cell lymphoma; OS, overall survival; PFS, progression-free survival.

significantly differ between DLBCL responders and non-responders during therapy course (30). A total of 504 cfRNA transcripts were found to be differentially longitudinally abundant between both groups, including PTMAP4 and NAP1L1 pseudogene, which were also significantly associated with PFS at diagnostic timepoint (Supplementary Table 4 and Supplementary Figure 10). As their abundance trajectory patterns differ between responders and non-responders over the course of R-CHOP, this could indicate that these transcripts could potentially serve as longitudinal prognostic markers beyond the diagnostic timepoint.

4 Discussion

Cell-free nucleic acids have increasingly been recognized as valuable precision medicine biomarkers in cancer research, including lymphoproliferative malignancies (10, 46–51). Although the origin of cfRNA remains largely unknown, both human and xenograft studies have illustrated its potential to reflect intra- and intertumoral heterogeneity, as well as functional changes during disease course and treatment, both of the tumor and the non-malignant compartment (11, 52, 53). We present the first study to longitudinally explore the total cell-free transcriptome in a cohort of DLBCL and PMBCL patients.

Higher ctDNA levels have been demonstrated in DLBCL patients compared to healthy controls (54, 55). Concordantly, our results showed a higher cfRNA concentration in the DLBCL and PMBCL samples. A significant decrease was observed after successful R-CHOP treatment, potentially indicating that a decline in cfRNA concentration levels by itself may reflect therapy response. CtDNA levels at diagnosis have been shown to correlate with LDH levels, a surrogate marker of tumor burden that is incorporated in the NCCN-IPI score due to its correlation with prognosis in NHL (37, 47, 50, 56–58). No firm correlation could, however, be demonstrated between cfRNA concentration and LDH levels in this study.

A large number of DAG were identified between DLBCL patients, PMBCL patients, and healthy controls. Considering mRNA, multiple widely used lymphoma markers were well reflected in the FFPE tissue, but poorly in the cell-free RNA compartment. Similarly, the majority of FFPE-derived DAG were not differential in the matched blood plasma, and only approximately half of the shared DAG showed the same direction of dysregulation. This finding is in concordance with a previous study in which extracellular mRNA of CCND2, BCL2, MYC, LMO2, and BCL6 was detected in only 14%, 10%, 10%, 10%, and 5% of DLBCL plasma samples, respectively, but in all of the matched tissue samples (40). Similar discrepancies have been reported for non-coding RNAs, in which no correlation between the EV-derived and the tissue derived miRNA repertoire, or even an opposite miRNA expression profile in serum versus matched tumor tissues was demonstrated (59, 60). Potential reasons include variability in (vesicle-mediated) secretion or passive release of RNA molecules in the bloodstream, differences in the rate and manner of degradation, and the shedding of these markers from other tissues besides the tumor compartment. Our overlap analysis pinpointed multiple

differentially abundant genes in the same direction in both compartments, likely representing tumor-specific cfRNA markers.

Several diagnostic cfRNA markers have been reported in DLBCL. In contrast to a previous study, decreased G1/S-specific cyclin-D2 (CCND2) and B-cell lymphoma 2 (BCL2) plasma levels were shown in DLBCL patients compared to healthy controls (40). More specifically CCND2 levels were decreased in both GCB and non-GCB samples compared to controls, and BCL2 levels were only significantly lower in the non-GCB samples compared to controls. In the tissue samples, higher CCND2 levels were found in the non-GCB compared to the GCB tissue samples, which has been reported and associated with an inferior prognosis within this COO subtype (3, 61). Similar to previous findings, B-cell lymphoma 6 (BCL6) and Fibronectin 1 (FN1) levels were not differentially abundant in the DLBCL plasma samples as opposed to control samples, although our results revealed higher levels in the FFPE samples (40). Concerning lncRNAs, decreased Growth Arrest Specific 5 (GAS5), as well as increased Taurine Up-Regulated 1 (TUG1) and increased X-inactive specific transcript (XIST) plasma levels have been demonstrated in DLBCL compared to control plasma samples (43, 44). In our cohort, GAS5 levels were found significantly decreased in DLBCL FPPE tissue, but not in blood plasma. TUG1 levels were found to be upregulated in the plasma of non-GCB DLBCL, compared to GCB DLBCL and control samples. Our results showed higher plasma levels of XIST, located on the long arm of the X chromosome, in the DLBCL patients. When comparing sex-matched samples, however, no significant result could be retained, indicating that the difference is sex-based and not disease-driven in our cohort. Lastly, circAPC, derived from the host gene adenomatous polyposis coli (APC), has been reported as downregulated in DLBCL patients (45). Although a similar trend was visible, statistical significance was not reached.

The COO classification provided a framework on which our current understanding of DLBCL biology is built, classifying tumors based on distinct patterns of GEP into GCB and ABC subtypes that have significant differences in survival (3, 62). Our results showed a good concordance between IHC staining and corresponding mRNA abundance of CD10, BCL6, and MUM1 for the DLBCL FFPE tissue, but not the plasma samples. Moreover, a good performance of a tissue-derived GEP, containing the best class-predicting genes as initially identified in the Lymphochip microarray data, to classify FFPE samples according to Hans algorithm COO was demonstrated, comparable with what was previously reported (4, 63). This tissue classifier, however failed for the diagnostic plasma samples. Besides miR-21 and miR-155, no cfRNA markers have been described for COO classification (11, 64, 65). Therefore, a plasma-derived COO classifier was developed, based on the DAG identified in our cohort, which demonstrated a high concordance with the Hans algorithm. Notably, there was no overlap between our plasma-derived and the tissue-derived GEP. Potential reasons include that the latter is mainly based on mRNA and the former also incorporated the non-coding transcriptome, including lncRNA, mt-RNA and pseudogenes, as well as the use of distinct RNA characterization methods. The lack of overlap supports the observation that cfRNA abundance does not necessarily reflect the transcriptomic differences found in the tissue but has a unique

abundance profile that besides tumor-derived transcripts also contains healthy tissue-derived signals and exhibits its own dynamics.

Pretreatment ctDNA levels have been shown as independent prognostic marker for event-free survival (EFS), PFS, and OS in DLBCL patients (46, 51, 54, 58, 66). Here, we have established a 9-gene and a 5-gene cfRNA signature, whose increased abundance was negatively associated with PFS and OS, respectively, independent of the NCCN-IPI score. The majority of the protein-coding genes in these signatures have been attributed an oncogenic role and have been associated with adverse prognostic outcomes (67–76). *FTH1P7/8*, *PTMAP4*, and *NAP1L1* pseudogene are pseudogenes of Ferritin Heavy Chain 1 (*FTH1*), Prothymosin Alpha (*PTMA*), and Nucleosome Assembly Protein 1 Like 1 (*NAP1L1*), respectively, which have been found dysregulated in various malignancies (77–79). Pseudogenes are being increasingly recognized as regulators of essential biological processes as they can compete with their parental gene for binding to miRNAs, while others generate small interference RNAs to dampen gene expression or encode functional mutated proteins (78). Protein Kinase C Beta (*PRKCB*) is an integral component of signaling via the B-cell receptor in DLBCL, activating NF-κB and VEGF–mediated angiogenesis, and its tissue expression was shown to be an independent predictor of poor OS (68, 69). Decreased Activating Transcription Factor 4 (*ATF4)* levels have been associated with improved survival probability in DLBCL patients (70). A recent study identified a Sirtuin 3 (SIRT3)–ATF4 axis required to maintain survival of DLBCL cells, regardless of subtype, by enabling them to optimize amino acid uptake and utilization. Targeting *ATF4* translation may, therefore, potentiate the cytotoxic effect of SIRT3 inhibition and serve as a potential therapeutic target (80).

PET-guided treatment of DLBCL remains debatable as direct evidence of improved patient outcomes is still lacking. Outside of a clinical trial, biopsy confirmation of an abnormal interim PET-CT scan is still recommended before switching therapy (81–83). In recent years, serial cell-free nucleic acids analyses have demonstrated potential to complement the predictive value of imaging results, as ctDNA monitoring has shown to facilitate (minimal) residual disease assessment and early relapse detection (46, 47, 56, 58, 66, 84–86). Only few studies have, however, investigated longitudinal abundance of (EV-derived) cfRNA markers (87, 88). When comparing pre- to interim-, and to post-therapy samples in DLBCL responders, a clear decrease in cell-free mRNA of B-cell related markers was noted, many of which have been proposed as potential diagnostic and/or prognostic biomarkers in DLBCL tissue (89–98). Moreover, the interim timepoint showed a clear, transient increase in neutrophil- and inflammation-related markers, reflecting an acute-phase reaction during therapy. In the R/R patients, only a transient decrease at interim timepoint of several B-cell markers was noted, and no acute phase response was discerned, potentially reflecting the poorer response to R-CHOP, and illustrating that repetitive sampling may allow to discern dynamic transcriptomic changes in relation to therapy response.

Our study has several limitations. Our findings have been compared with previous studies that have mainly used RT-qPCR or microarray-based abundance analyses and have collected and processed samples using other blood collection tubes and protocols, which may hamper direct comparison of results. Moreover, specific genes can be differentially enriched in EVs compared to the circulating-free compartment, underlining that caution should be applied when comparing the results of studies investigating different biofluid fractions (11, 88). The impact of these pre-analytical variables on cfRNA abundance results has been increasingly reported (99, 100). Secondly, because of the total RNA sequencing library preparation method, the miRNA fraction is underrepresented in our analyses as compared to the protein-coding and other non-coding RNA classes. Third, although the control samples were relatively equally distributed by sex and age, other variables may influence the cfRNA profile, including comorbidities, lifestyle conditions and medication intake, which may affect differential abundance analysis. Lastly, the gene signatures identified in this study need validation in an independent cohort. Currently, no other dataset with plasma-derived (long) RNA-seq data from a cohort of DLBCL and/or PMBCL patients is publicly available. A cohort is being prospectively collected for validation purposes of the results obtained in this study.

In conclusion, we present the first study to longitudinally explore the total cell-free transcriptome in DLBCL and PMBCL patients using blood plasma samples. A large number of differentially abundant genes were identified and compared against previous cfRNA biomarker studies. Overall, overlap analyses with matched FFPE samples showed that blood plasma has a unique transcriptomic profile, that significantly differs from that of the tissue. Our results demonstrated a good concordance between tissue-derived gene expression and the Hans algorithm for COO classification in FFPE samples, but not in the plasma samples. Therefore, a cfRNA classifier was developed that can accurately determine COO in plasma. Lastly, high plasma levels of a 9-gene signature and a 5-gene signature were associated with inferior PFS and OS in DLBCL patients, respectively, independent of the NCCN-IPI score. Our results may serve as a reference point for future cfRNA studies, and the biomarkers identified may represent potential targets for in-depth functional analysis.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://ega-archive.org, ID EGAS0000100758 (dataset EGAD00001011679).

Ethics statement

The studies involving humans were approved by medical Ethical Committee of Ghent University Hospital, Ghent, Belgium (EC/2016/0307). The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

Funding

Acknowledgments

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2023.1221471/full#supplementary-material

SUPPLEMENTARY FIGURE 1
Overview of the blood plasma samples included per time point in the study (total of 168 samples). Response at each timepoint was assessed by PET-CT. CR: complete remission; PD: progressive disease; PET-CT: positron emission tomography/computerized tomography.

SUPPLEMENTARY FIGURE 2
Correlation between LDH (IU/L) and cfRNA concentration (ng/ml blood plasma) in diagnostic plasma samples of DLBCL (red) and PMBCL (green) samples (A). ROC curve of cfRNA concentration for predicting abnormal LDH levels in diagnostic plasma samples of DLBCL patients (AUC=0.65) (B). AUC:

area under the ROC curve; DLBCL: diffuse large B-cell lymphoma; LDH: lactate dehydrogenase; PMBCL: primary mediastinal B-cell lymphoma; ROC: Receiver Operating Characteristic.

SUPPLEMENTARY FIGURE 3
Evolution of cfRNA concentrations in two responders to first-line treatment (A) and two non-responders to first-line treatment (B). CR: complete remission; DLBCL: diffuse large B-cell lymphoma; PD: progressive disease; PMBCL: primary mediastinal B-cell lymphoma; PR: partial response.

SUPPLEMENTARY FIGURE 4
Differentially abundant genes in the diagnostic plasma samples between different groups for each RNA subclass. CircRNA: circular RNA; DLBCL: diffuse large B-cell lymphoma; GCB: germinal center B-cell; lncRNA: long non-coding RNA; misc_RNA: miscellaneous RNA; mt-tRNA: mitochondrial tRNA; PMBCL: primary mediastinal B-cell lymphoma; rRNA: ribosomal RNA; snRNA: small nuclear RNA; snoRNA: small nucleolar RNA; TEC: To be Experimentally Confirmed.

SUPPLEMENTARY FIGURE 5
Volcano plots of the differentially abundant genes for the DLBCL versus control, PMBCL versus control, and PMBCL versus DLBCL comparisons in the diagnostic plasma samples (cutoff q-value =<0.05 and |log2FC|>=1) (A). Venn diagram illustrating the overlap of significantly differentially abundant transcripts between the different comparisons. Separate diagrams are shown for all genes, for mRNAs, lncRNAs, and circRNAs subtypes. (B) circRNA: circular RNA; DLBCL: diffuse large B-cell lymphoma; lncRNA: long non-coding RNA; log2FC: log2 fold change; mRNA: messenger RNA; PMBCL: primary mediastinal B-cell lymphoma.

SUPPLEMENTARY FIGURE 6
Gene set enrichment analysis (GSEA) on the diagnostic plasma samples for DLBCL patients versus healthy controls (A), PMBCL patients versus healthy controls (B), and PBMCL versus DLBCL patients (C). DLBCL: diffuse large B-cell lymphoma; PMBCL: primary mediastinal B-cell lymphoma.

SUPPLEMENTARY FIGURE 7
Normalized mRNA counts for CD3, CD4, CD5, CD8, CD19, CD20, CD23, CD30, PAX5, BCL2, BCL6, MKI67, IRF4, MAL, REL, and TRAF1 within the FFPE samples of healthy controls, GCB-DLBCL patients, non-GCB-DLBCL patients, and PMBCL patients. Multiple testing corrected p-values are shown for pairwise comparisons. DLBCL: diffuse large B-cell lymphoma; GCB: germinal center derived B-cell lymphoma; PMBCL: primary mediastinal B-cell lymphoma.

SUPPLEMENTARY FIGURE 8
Normalized mRNA counts of CD3, CD4, CD5, CD8, CD19, CD20, CD23, CD30, PAX5, BCL2, BCL6, MKI67, IRF4, MAL, REL, and TRAF1 within the diagnostic plasma samples of healthy controls, GCB-DLBCL patients, non-GCB-DLBCL patients, and PMBCL patients. Multiple testing corrected p-values are shown for pairwise comparisons. DLBCL: diffuse large B-cell lymphoma; GCB: germinal center derived B-cell lymphoma; PMBCL: primary mediastinal B-cell lymphoma.

SUPPLEMENTARY FIGURE 9
The genes in the unfavorable gene signatures (A) that are significantly associated with PFS (left) and OS (right), respectively, in univariate (B) and multivariable analysis (C), the latter also including the NCCN-IPI score. CI: confidence interval; HR: hazard ratio; NCCN-IPI: National Comprehensive Cancer Network International Prognostic Index score; OS: overall survival; PFS: progression-free survival.

SUPPLEMENTARY FIGURE 10
Abundance trajectories of PTMAP4 and NAP1L1 pseudogene between DLBCL responder and non-responder patients (n=27 per timepoint) in matched samples at diagnosis, at interim evaluation and at final evaluation after R-CHOP therapy. Multiple testing corrected p-values are shown for pairwise comparisons. DLBCL: diffuse large B-cell lymphoma; R-CHOP: rituximab, cyclophosphamide, vincristine, doxorubicin, and prednisone.

SUPPLEMENTARY TABLE 1
Procedure to evaluate the performance of a tissue- and plasma-derived GEP for IHC COO classification.

References

1. Swerdlow SH, Campo E, Pileri SA, Harris NL, Stein H, Siebert R, et al. The 2016 revision of the World Health Organization classification of lymphoid neoplasms. *Blood* (2016) 127(20):2375–90. doi: 10.1182/blood-2016-01-643569

2. Crump M, Neelapu SS, Farooq U, Van Den Neste E, Kuruvilla J, Westin J, et al. Outcomes in refractory diffuse large B-cell lymphoma: results from the international SCHOLAR-1 study. *Blood* (2017) 130(16):1800–8. doi: 10.1182/blood-2017-03-769620

3. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* (2000) 403(6769):503–11. doi: 10.1038/35000501

4. Hans CP, Weisenburger DD, Greiner TC, Gascoyne RD, Delabie J, Ott G, et al. Confirmation of the molecular classification of diffuse large B-cell lymphoma by immunohistochemistry using a tissue microarray. *Blood* (2004) 103(1):275–82. doi: 10.1182/blood-2003-05-1545

5. Savage KJ, Monti S, Kutok JL, Cattoretti G, Neuberg D, De Leval L, et al. The molecular signature of mediastinal large B-cell lymphoma differs from that of other diffuse large B-cell lymphomas and shares features with classical Hodgkin lymphoma. *Blood* (2003) 102(12):3871–9. doi: 10.1182/blood-2003-06-1841

6. Rosenwald A, Wright G, Leroy K, Yu X, Gaulard P, Gascoyne RD, 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(6):851–62. doi: 10.1084/jem.20031074

7. Gaulard P, Harris NL, Pileri SA, Kutok JL, Stein H, Kovrigina AM, et al. *Primary mediastinal (thymic) large B-cell lymphoma. WHO classification of tumours of haematopoietic and lymphoid tissues Lyon*. IARC, Lyon (2008) p. 250–3.

8. Savage KJ. Primary mediastinal large B-cell lymphoma. *Blood* (2022) 140(9):955–70. doi: 10.1182/blood.2020008376

9. Riedell PA, Smith SM. Double hit and double expressors in lymphoma: Definition and treatment. *Cancer* (2018) 124(24):4622–32. doi: 10.1002/cncr.31646

10. Rossi D, Diop F, Spaccarotella E, Monti S, Zanni M, Rasi S, et al. Diffuse large Bcell lymphoma genotyping on the liquid biopsy. *Blood* (2017) 129(14):1947–57. doi: 10.1182/blood-2016-05-719641

11. Decruyenaere P, Offner F, Vandesompele J. Circulating RNA biomarkers in diffuse large B-cell lymphoma: a systematic review. *Exp Hematol Oncol* (2021) 10(1):13. doi: 10.1186/s40164-021-00208-3

12. Sole C, Arnaiz E, Manterola L, Otaegui D, Lawrie CH. The circulating transcriptome as a source of cancer liquid biopsy biomarkers. *Semin Cancer Biol* (2019) 58:100–8. doi: 10.1016/j.semcancer.2019.01.003

13. Schwarzenbach H, Hoon DSB, Pantel K. Cell-free nucleic acids as biomarkers in cancer patients. *Nat Rev Cancer* (2011) 11(6):426–37. doi: 10.1038/nrc3066

14. Cheson BD, Fisher RI, Barrington SF, Cavalli F, Schwartz LH, Zucca E, et al. Recommendations for initial evaluation, staging, and response assessment of Hodgkin and non-Hodgkin lymphoma: the Lugano classification. *J Clin Oncol* (2014) 32 (27):3059–68. doi: 10.1200/JCO.2013.54.8800

15. Andrews S. *FastQC: A Quality Control Tool for High Throughput Sequence Data*. Babraham Institute, Cambridge, UK. (2010) Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

16. Smith T, Heger A, Sudbery I. UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res* (2017) 27 (3):491–9. doi: 10.1101/gr.209601.116

17. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* (2011) 17(1):10. doi: 10.14806/ej.17.1.200

18. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. *Gigascience* (2021) 10(2):giab008. doi: 10.1093/gigascience/giab008

19. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* (2013) 29(1):15–21. doi: 10.1093/bioinformatics/bts635

20. Bushnell B. *BBMap: A Fast, Accurate, Splice-Aware Aligner. No. LBNL-7065E.* Ernest Orlando Lawrence Berkeley National Laboratory, Berkeley, CA. (2014).

21. Broad Institute. *Picard tools* . Broad Institute, GitHub repository. Available at: http://broadinstitute.github.io/picard/ (Accessed 2022/06/22).

22. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* (2016) 32 (19):3047–8. doi: 10.1093/bioinformatics/btw354

23. Anders S, Pyl PT, Huber W. HTSeq–a Python framework to work with high-throughput sequencing data. *Bioinformatics* (2015) 31(2):166–9. doi: 10.1093/bioinformatics/btu638

24. Cunningham F, Allen JE, Allen J, Alvarez-Jarreta J, Amode MR, Armean IM, et al. Ensembl 2022. *Nucleic Acids Res* (2022) 50(D1):D988–95. doi: 10.1093/nar/gkab1049

25. Jakobi T, Uvarovskii A, Dieterich C. circtools-a one-stop software solution for circular RNA research. *Bioinformatics* (2019) 35(13):2326–8. doi: 10.1093/bioinformatics/bty948

26. Hulstaert E, Decock A, Morlion A, Everaert C, Verniers K, Nuytens J, et al. Messenger RNA capture sequencing of extracellular RNA from human biofluids using a comprehensive set of spike-in controls. *STAR Protoc* (2021) 2(2):100475. doi: 10.1016/j.xpro.2021.100475

27. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* (2014) 15(12):550. doi: 10.1186/s13059-014-0550-8

28. Blighe K, Rana S, Lewis M. *EnhancedVolcano: Publication-ready volcano plots with enhanced colouring and labeling* (2018). Available at: https://github.com/kevinblighe/EnhancedVolcano.

29. Chen H, Boutros PC. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinf* (2011) 12:35. doi: 10.1186/1471-2105-12-35

30. Fischer DS, Theis FJ, Yosef N. Impulse model-based differential expression analysis of time course sequencing data. *Nucleic Acids Res* (2018) 46(20):e119. doi: 10.1093/nar/gky675

31. Langfelder P, Zhang B, Horvath S. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics* (2008) 24(5):719–20. doi: 10.1093/bioinformatics/btm563

32. Hothorn T. Maximally selected rank statistics. *R package version 0.7-25.* (2017) Available at: https://CRAN.R-project.org/package=maxstat.

33. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* (2005) 102(43):15545–50. doi: 10.1073/pnas.0506580102

34. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* (2015) 1(6):417–25. doi: 10.1016/j.cels.2015.12.004

35. R Core Team. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing (2021). Available at: https://www.R-project.org/.

36. Shen L. *GeneOverlap: Test and visualize gene overlaps* (2022). Available at: http://shenlab-sinai.github.io/shenlab-sinai/.

37. Zhou Z, Sehn LH, Rademaker AW, Gordon LI, Lacasce AS, Crosby-Thompson A, et al. An enhanced International Prognostic Index (NCCN-IPI) for patients with diffuse large B-cell lymphoma treated in the rituximab era. *Blood* (2014) 123(6):837–42. doi: 10.1182/blood-2013-09-524108

38. Récher C, Coiffier B, Haioun C, Molina TJ, Fermé C, Casasnovas O, et al. Intensified chemotherapy with ACVBP plus rituximab versus standard CHOP plus rituximab for the treatment of diffuse large B-cell lymphoma (LNH03-2B): an open-label randomised phase 3 trial. *Lancet* (2011) 378(9806):1858–67. doi: 10.1016/S0140-6736(11)61040-4

39. Dunleavy K, Pittaluga S, Maeda LS, Advani R, Chen CC, Hessler J, et al. Dose-adjusted EPOCH-rituximab therapy in primary mediastinal B-cell lymphoma. *N Engl J Med* (2013) 368(15):1408–16. doi: 10.1056/NEJMoa1214561

40. Garcia V, Garcia JM, Silva J, Martin P, Peña C, Dominguez G, et al. Extracellular tumor-related mRNA in plasma of lymphoma patients and survival implications. *PloS One* (2009) 4(12):e8173. doi: 10.1371/journal.pone.0008173

41. Zhao H, Kan Y, Wang X, Chen L, Ge P, Qian Z. Genetic polymorphism and transcriptional regulation of CREBBP gene in patient with diffuse large B-cell lymphoma. *Biosci Rep* (2019) 39(8):BSR20191162. doi: 10.1042/BSR20191162

42. Provencio M, Rodríguez M, Cantos B, Sabín P, Quero C, García-Arroyo FR, et al. mRNA in exosomas as a liquid biopsy in non-Hodgkin Lymphoma: a multicentric study by the Spanish Lymphoma Oncology Group. *Oncotarget* (2017) 8(31):50949–57. doi: 10.18632/oncotarget.16435

43. Wang Y, Zhang M, Xu H, Wang Y, Li Z, Chang Y, et al. Discovery and validation of the tumor-suppressive function of long noncoding RNA PANDA in human diffuse large B-cell lymphoma through the inactivation of MAPK/ERK signaling pathway. *Oncotarget* (2017) 8(42):72182–96. doi: 10.18632/oncotarget.20053

44. Senousy MA, El-Abd AM, Abdel-Malek RR, Rizk SM. Circulating long non-coding RNAs HOTAIR, Linc-p21, GAS5 and XIST expression profiles in diffuse large B-cell lymphoma: association with R-CHOP responsiveness. *Sci Rep* (2021) 11(1):2095. doi: 10.1038/s41598-021-81715-5

45. Hu Y, Zhao Y, Shi C, Ren P, Wei B, Guo Y, et al. A circular RNA from APC inhibits the proliferation of diffuse large B-cell lymphoma by inactivating Wnt/β-catenin signaling *via* interacting with TET1 and miR-888. *Aging* (2019) 11(19):8068–84. doi: 10.18632/aging.102122

46. Kurtz DM, Scherer F, Jin MC, Soo J, Craig AFM, Esfahani MS, et al. Circulating tumor DNA measurements as early outcome predictors in diffuse large B-cell lymphoma. *J Clin Oncol* (2018) 36(28):2845–53. doi: 10.1200/JCO.2018.78.5246

47. Scherer F, Kurtz DM, Newman AM, Stehr H, Craig AFM, Esfahani MS, et al. Distinct biological subtypes and patterns of genome evolution in lymphoma revealed by circulating tumor DNA. *Sci Transl Med* (2016) 8(364):364ra155. doi: 10.1126/scitranslmed.aai8545

48. Camus V, Viennot M, Lévêque E, Viailly PJ, Tonnelet D, Veresezan EL, et al. Circulating tumor DNA in primary mediastinal large B-cell lymphoma versus classical Hodgkin lymphoma: a retrospective study. *Leuk Lymphoma* (2022) 63(4):834–44. doi: 10.1080/10428194.2021.2010060

49. Rivas-Delgado A, Nadeu F, Andrade-Campos M, López C, Enjuanes A, Mozas P, et al. Cell-free DNA for genomic analysis in primary mediastinal large B-cell lymphoma. *Diagn (Basel)* (2022) 12(7):1575. doi: 10.3390/diagnostics12071575

50. Rivas-Delgado A, Nadeu F, Enjuanes A, Casanueva-Eliceiry S, Mozas P, Magnano L, et al. Mutational landscape and tumor burden assessed by cell-free DNA in diffuse large B-cell lymphoma in a population-based study. *Clin Cancer Res* (2021) 27(2):513–21. doi: 10.1158/1078-0432.CCR-20-2558

51. Yao L, Xu H, Wo J, Zhao M, Liu Z, Dong T, et al. Prognostic value of circulating tumor DNA in lymphoma: a meta-analysis. *Clin Exp Med* (2022) 22(1):1–7. doi: 10.1007/s10238-021-00718-8

52. Roskams-Hieter B, Kim HJ, Anur P, Wagner JT, Callahan R, Spiliotopoulos E, et al. Plasma cell-free RNA profiling distinguishes cancers from pre-malignant conditions in solid and hematologic Malignancies. *NPJ Precis Oncol* (2022) 6(1):28. doi: 10.1038/s41698-022-00270-y

53. Zhuang J, Ibarra A, Acosta A, Karns AP, Aballi J, Nerenberg M, et al. Survey of extracellular communication of systemic and organ-specific inflammatory responses through cell free messenger RNA profiling in mice. *EBioMedicine* (2022) 83:104242. doi: 10.1016/j.ebiom.2022.104242

54. Eskandari M, Manoochehrabadi S, Pashaiefar H, Zaimy MA, Ahmadvand M. Clinical significance of cell-free DNA as a prognostic biomarker in patients with diffuse large B-cell lymphoma. *Blood Res* (2019) 54(2):114–9. doi: 10.5045/br.2019.54.2.114

55. Shirouchi Y, Mishima Y, Takayama T, Minowa S, Ishihara Y, Tamba M, et al. Serum cell-free DNA concentration as a possible prognostic marker in newly diagnosed diffuse large B-cell lymphoma. *BioMed Res* (2022) 43(4):99–106. doi: 10.2220/biomedres.43.99

56. Roschewski M, Dunleavy K, Pittaluga S, Moorhead M, Pepin F, Kong K, et al. Circulating tumour DNA and CT monitoring in patients with untreated diffuse large B-cell lymphoma: a correlative biomarker study. *Lancet Oncol* (2015) 16(5):541–9. doi: 10.1016/S1470-2045(15)70106-3

57. Bohers E, Viailly PJ, Becker S, Marchand V, Ruminy P, Maingonnat C, et al. Non-invasive monitoring of diffuse large B-cell lymphoma by cell-free DNA high-throughput targeted sequencing: analysis of a prospective cohort. *Blood Cancer J* (2018) 8(8):74. doi: 10.1038/s41408-018-0111-6

58. Herrera AF, Tracy S, Croft B, Opat S, Ray J, Lovejoy AF, et al. Risk profiling of patients with relapsed/refractory diffuse large B-cell lymphoma by measuring circulating tumor DNA. *Blood Adv* (2022) 6(6):1651–60. doi: 10.1182/bloodadvances.2021006415

59. Rinaldi F, Marchesi F, Palombi F, Pelosi A, Di Pace AL, Sacconi A, et al. MiR-22, a serum predictor of poor outcome and therapy response in diffuse large B-cell lymphoma patients. *Br J Haematol* (2021) 195(3):399–404. doi: 10.1111/bjh.17734

60. Baldasici O, Balacescu L, Cruceriu D, Roman A, Lisencu C, Fetica B, et al. Circulating small EVs miRNAs as predictors of pathological response to neo-adjuvant therapy in breast cancer patients. *Int J Mol Sci* (2022) 23(20):12625. doi: 10.3390/ijms232012625

61. Wang D, Zhang Y, he YQ. CCND2 mRNA expression is correlated with R-CHOP treatment efficacy and prognosis in patients with ABC-DLBCL. *Front Oncol* (2020) 10:1180. doi: 10.3389/fonc.2020.01180

62. Scott DW, Mottok A, Ennishi D, Wright GW, Farinha P, Ben-Neriah S, et al. Prognostic significance of diffuse large B-cell lymphoma cell of origin determined by digital gene expression in formalin-fixed paraffin-embedded tissue biopsies. *J Clin Oncol* (2015) 33(26):2848–56. doi: 10.1200/JCO.2014.60.2383

63. Meyer PN, Fu K, Greiner TC, Smith LM, Delabie J, Gascoyne RD, et al. Immunohistochemical methods for predicting cell of origin and survival in patients with diffuse large B-cell lymphoma treated with rituximab. *J Clin Oncol* (2011) 29(2):200–7. doi: 10.1200/JCO.2010.30.0368

64. Chen W, Wang H, Chen H, Liu S, Lu H, Kong D, et al. Clinical significance and detection of microRNA-21 in serum of patients with diffuse large B-cell lymphoma in Chinese population. *Eur J Haematol* (2014) 92(5):407–12. doi: 10.1111/ejh.12263

65. Bedewy AML, Elmaghraby SM, Shehata AA, Kandil NS. Prognostic value of miRNA-155 expression in B-cell non-hodgkin lymphoma. *Turk J Haematol* (2017) 34(3):207–12. doi: 10.4274/tjh.2016.0286

66. Meriranta L, Alkodsi A, Pasanen A, Lepistö M, Mapar P, Blaker YN, et al. Molecular features encoded in the ctDNA reveal heterogeneity and predict outcome in high-risk aggressive B-cell lymphoma. *Blood* (2022) 139(12):1863–77. doi: 10.1182/blood.2021012852

67. Vega-Rubín-de-Celis S. The role of beclin 1-dependent autophagy in cancer. *Biol (Basel)* (2019) 9(1):4. doi: 10.3390/biology9010004

68. Hans CP, Weisenburger DD, Greiner TC, Chan WC, Aoun P, Cochran GT, et al. Expression of PKC-beta or cyclin D2 predicts for inferior survival in diffuse large B-cell lymphoma. *Mod Pathol* (2005) 18(10):1377–84. doi: 10.1038/modpathol.3800434

69. Shipp MA, Ross KN, Tamayo P, Weng AP, Kutok JL, Aguiar RCT, et al. Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. *Nat Med* (2002) 8(1):68–74. doi: 10.1038/nm0102-68

70. Xiong D, Li M, Zeng C. Construction and validation of a risk scoring model for diffuse large B-cell lymphoma based on ferroptosis-related genes and its association with immune infiltration. *Transl Oncol* (2022) 16:101314. doi: 10.1016/j.tranon.2021.101314

71. Hart T, Chandrashekhar M, Aregger M, Steinhart Z, Brown KR, MacLeod G, et al. High-resolution CRISPR screens reveal fitness genes and genotype-specific cancer liabilities. *Cell* (2015) 163(6):1515–26. doi: 10.1016/j.cell.2015.11.015

72. Bao H, Li X, Cao Z, Huang Z, Chen L, Wang M, et al. Identification of COPA as a potential prognostic biomarker and pharmacological intervention target of cervical cancer by quantitative proteomics and experimental verification. *J Transl Med* (2022) 20(1):18. doi: 10.1186/s12967-021-03218-1

73. André F, Trinh A, Balayssac S, Maboudou P, Dekiouk S, Malet-Martino M, et al. Metabolic rewiring in cancer cells overexpressing the glucocorticoid-induced leucine zipper protein (GILZ): Activation of mitochondrial oxidative phosphorylation and sensitization to oxidative cell death induced by mitochondrial targeted drugs. *Int J Biochem Cell Biol* (2017) 85:166–74. doi: 10.1016/j.biocel.2017.02.011

74. Meng Q, Sun W, Li M, Zhao Y, Chen X, Sun L, et al. Increased expression of eps15 homology domain 1 is associated with poor prognosis in resected small cell lung cancer. *J Cancer* (2015) 6(10):990–5. doi: 10.7150/jca.11650

75. Baldari S, Ubertini V, Garufi A, D'Orazi G, Bossi G. Targeting MKK3 as a novel anticancer strategy: molecular mechanisms and therapeutical implications. *Cell Death Dis* (2015) 6(1):e1621. doi: 10.1038/cddis.2014.591

76. Park ER, Kim SB, Lee JS, Kim YH, Lee DH, Cho EH, et al. The mitochondrial hinge protein, UQCRH, is a novel prognostic factor for hepatocellular carcinoma. *Cancer Med* (2017) 6(4):749–60. doi: 10.1002/cam4.1042

77. Xiaohua Z, Xie Y, Huang W, Chen Z, Guo S. NAP1L1 promotes tumor proliferation through HDGF/C-JUN signaling in ovarian cancer. *BMC Cancer* (2022) 22(1):339. doi: 10.1186/s12885-022-09356-z

78. Di Sanzo M, Quaresima B, Biamonte F, Palmieri C, Faniello MC. FTH1 pseudogenes in cancer and cell metabolism. *Cells* (2020) 9(12):2554. doi: 10.3390/cells9122554

79. Zhu Y, Qi X, Yu C, Yu S, Zhang C, Zhang Y, et al. Identification of prothymosin alpha (PTMA) as a biomarker for esophageal squamous cell carcinoma (ESCC) by label-free quantitative proteomics and Quantitative Dot Blot (QDB). *Clin Proteomics* (2019) 16:12. doi: 10.1186/s12014-019-9232-6

80. Li M, Teater MR, Hong JY, Park NR, Duy C, Shen H, et al. Translational activation of ATF4 through mitochondrial anaplerotic metabolic pathways is required for DLBCL growth and survival. *Blood Cancer Discov* (2022) 3(1):50–65. doi: 10.1158/2643-3230.BCD-20-0183

81. Moskowitz CH, Schöder H, Teruya-Feldstein J, Sima C, Iasonos A, Portlock CS, et al. Risk-adapted dose-dense immunochemotherapy determined by interim FDG-PET in Advanced-stage diffuse large B-Cell lymphoma. *J Clin Oncol* (2010) 28(11):1896–903. doi: 10.1200/JCO.2009.26.5942

82. Le Gouill S, Casasnovas RO. Interim PET-driven strategy in de novo diffuse large B-cell lymphoma: do we trust the driver? *Blood* (2017) 129(23):3059–70. doi: 10.1182/blood-2016-05-672196

83. Dührsen U, Müller S, Hertenstein B, Thomssen H, Kotzerke J, Mesters R, et al. Positron emission tomography-guided therapy of aggressive non-hodgkin lymphomas (PETAL): A multicenter, randomized phase III trial. *J Clin Oncol* (2018) 36(20):2024–34. doi: 10.1200/JCO.2017.76.8093

84. Macaulay C, Alig S, Kurtz DM, Jin MC, Opat S, Soo J, et al. Interim circulating tumor DNA as a prognostic biomarker in the setting of interim PET-based adaptive therapy for DLBCL. *Blood* (2019) 134(Supplement_1):1600–0. doi: 10.1182/blood-2019-131278

85. Armand P, Oki Y, Neuberg DS, Faham M, Cummings C, Klinger M, et al. Detection of circulating tumour DNA in patients with aggressive B-cell non-Hodgkin lymphoma. *Br J Haematol* (2013) 163(1):123–6. doi: 10.1111/bjh.12439

86. Kurtz DM, Esfahani MS, Scherer F, Soo J, Jin MC, Liu CL, et al. Dynamic risk profiling using serial tumor biomarkers for personalized outcome prediction. *Cell* (2019) 178(3):699–713.e19. doi: 10.1016/j.cell.2019.06.011

87. Cui Q, Vari F, Cristino AS, Salomon C, Rice GE, Sabdia MB, et al. Circulating cell-free miR-494 and miR-21 are disease response biomarkers associated with interim-positron emission tomography response in patients with diffuse large B-cell lymphoma. *Oncotarget* (2018) 9(78):34644–57. doi: 10.18632/oncotarget.26141

88. Bang YH, Shim JH, Ryu KJ, Kim YJ, Choi ME, Yoon SE, et al. Clinical relevance of serum-derived exosomal messenger RNA sequencing in patients with non-Hodgkin lymphoma. *J Cancer* (2022) 13(5):1388–97. doi: 10.7150/jca.69639

89. Johnson NA, Boyle M, Bashashati A, Leach S, Brooks-Wilson A, Sehn LH, et al. Diffuse large B-cell lymphoma: reduced CD20 expression is associated with an inferior survival. *Blood* (2009) 113(16):3773–80. doi: 10.1182/blood-2008-09-177469

90. Desouki MM, Post GR, Cherry D, Lazarchick J. PAX-5: a valuable immunohistochemical marker in the differential diagnosis of lymphoid neoplasms. *Clin Med Res* (2010) 8(2):84–8. doi: 10.3121/cmr.2010.891

91. Reddy A, Zhang J, Davis NS, Moffitt AB, Love CL, Waldrop A, et al. Genetic and functional drivers of diffuse large B cell lymphoma. *Cell* (2017) 171(2):481–494.e15. doi: 10.1016/j.cell.2017.09.027

92. Yousefi Z, Sharifzadeh S, Zare F, Eskandari N. Fc receptor-like 1 (FCRL1) is a novel biomarker for prognosis and a possible therapeutic target in diffuse large B-cell lymphoma. *Mol Biol Rep* (2023) 50(2):1133–45. doi: 10.1007/s11033-022-08104-7

93. Bobée V, Ruminy P, Marchand V, Viailly PJ, Abdel Sater A, Veresezan L, et al. Determination of molecular subtypes of diffuse large B-cell lymphoma using a reverse transcriptase multiplex ligation-dependent probe amplification classifier: A CALYM study. *J Mol Diagn* (2017) 19(6):892–904. doi: 10.1016/j.jmoldx.2017.07.007

94. Choi S, Lee YJ, Choi Y, Kim M, Kim HJ, Kim JE, et al. Prognostic significance of BLK expression in R-CHOP treated diffuse large B-cell lymphoma. *J Pathol Transl Med* (2022) 56(5):281–8. doi: 10.4132/jptm.2022.07.26

95. Thomsen EA, Rovsing AB, Anderson MV, Due H, Huang J, Luo Y, et al. Identification of BLNK and BTK as mediators of rituximab-induced programmed cell death by CRISPR screens in GCB-subtype diffuse large B-cell lymphoma. *Mol Oncol* (2020) 14(9):1978–97. doi: 10.1002/1878-0261.12753

96. Ichikawa S, Fukuhara N, Katsushima H, Takahashi T, Yamamoto J, Yokoyama H, et al. Association between BACH2 expression and clinical prognosis in diffuse large B-cell lymphoma. *Cancer Sci* (2014) 105(4):437–44. doi: 10.1111/cas.12361

97. Ramuz O, Bouabdallah R, Devilard E, Borie N, Groulet-Martinec A, Bardou VJ, et al. Identification of TCL1A as an immunohistochemical marker of adverse outcome in diffuse large B-cell lymphomas. *Int J Oncol* (2005) 26(1):151–7. doi: 10.3892/ijo.26.1.151

98. Gascoyne DM, Banham AH. The significance of FOXP1 in diffuse large B-cell lymphoma. *Leuk Lymphoma* (2017) 58(5):1037–51. doi: 10.1080/10428194.2016.1228932

99. European Committee for Standardization Technical Committee (CEN/TC) 140. *In vitro diagnostic medical devices" published standard CEN/TS 17742:2022. Molecular in vitro diagnostic examinations - Specifications for pre-examination processes for venous whole blood - Isolated circulating cell free RNA from plasma.* Brussels (2022).

100. Anckaert J, Cobos FA, Decock A, Decruyenaere P, Deleu J, De Preter K, et al. Performance evaluation of RNA purification kits and blood collection tubes in the Extracellular RNA Quality Control (exRNAQC) study. *BioRxiv [Preprint].* (2023) doi: 10.1101/2021.05.11.442610

Glossary

ABC	activated B-cell lymphoma
APC	adenomatous polyposis coli
ATF4	Decreased Activating Transcription Factor 4
BCL2	B-cell lymphoma 2
BCL6	B-cell lymphoma 6
CCND2	G1/S-specific cyclin-D2
CD10	Membrane Metalloendopeptidase
CREBBP	CREB Binding Protein
cfDNA	cell-free DNA
cfRNA	cell-free RNA
circRNA	circular RNA
COO	cell-of-origin
CR	complete remission
ctDNA	circulating tumor DNA
DA-EPOCH-R	dose-adjusted EPOCH-rituximab therapy
DAG	differentially abundant gene
DEL	double-expressor lymphoma
DFS	disease-free survival
DHL	double-hit lymphoma
DLBCL	diffuse large B-cell lymphoma
ECM	extracellular matrix
EFS	event-free survival
ES	enrichment score
EV	extracellular vesicle
FDG-PET/CT	F-fluorodeoxyglucose Positron Emission Tomography/Computerized Tomography
FFPE	formalin-fixed paraffin-embedded tissue
FISH	Fluorescence in situ hybridization
FN1	Fibronectin 1
GAS5	Growth Arrest Specific 5
GEP	gene expression profiling
GSEA	gene set enrichment analysis
IGF	Insulin-like Growth Factor
IHC	immunohistochemistry
IL2	interleukin-2
IL6	interleukin-6
IRF4	Interferon Regulatory Factor 4
JAK	Janus kinase

(Continued)

Continued

JI	Jaccard Index
LDH	lactate dehydrogenase
LMO2	LIM domain only 2
lncRNA	long non-coding RNA
Log2FC	log2 fold change
misc_RNA	miscellaneous RNA
mRNA	messenger RNA
mt-tRNA	mitochondrial transfer RNA
MYC	MYC proto-oncogene protein
NCCN-IPI	National Comprehensive Cancer Network International Prognostic Index score
NF-κB	Nuclear factor kappa-light-chain-enhancer of activated B-cells
NHL	non-Hodgkin's lymphoma
OS	overall survival
PD	progressive disease
PFS	progression-free survival
PMBCL	primary mediastinal B-cell lymphoma
PR	partial response
PTEN	Phosphatase and Tensin homolog
R-ACVBP	dose-intensive rituximab, doxorubicin, cyclophosphamide, vindesine, bleomycin, and prednisone
R-CHOP	rituximab, cyclophosphamide, vincristine, doxorubicin, and prednisone
ROC	Receiver Operating Characteristic
R/R	relapse/refractory disease
rRNA	ribosomal RNA
RT-qPCR	reverse transcription quantitative real-time PCR
SD	stable disease
SIRT3	Sirtuin 3
snRNA	small nuclear RNA
snoRNA	small nucleolar RNA
STAT3	signal transducer and activator of transcription 3
TGF-β	Transforming Growth Factor Beta
TNF	Tumor Necrosis Factor
TUG1	Taurine Up-Regulated 1
UMI	Unique Molecular Identifier
VEGF	vascular endothelial growth factor
WHO	World Health Organization
XIST	X-inactive specific transcript.