

NGS-determined molecular markers and disease burden metrics from ctDNA correlate with PFS in previously untreated DLBCL

Ehsan Tabari, Alexander F. Lovejoy, Hai Lin, Christopher R. Bolen, Seng Lor Saelee, Joshua P. Lefkowitz, David M. Kurtz, Alessia Bottos, Tina G. Nielsen, Joana M. Parreira & Khai T. Luong

To cite this article: Ehsan Tabari, Alexander F. Lovejoy, Hai Lin, Christopher R. Bolen, Seng Lor Saelee, Joshua P. Lefkowitz, David M. Kurtz, Alessia Bottos, Tina G. Nielsen, Joana M. Parreira & Khai T. Luong (2024) NGS-determined molecular markers and disease burden metrics from ctDNA correlate with PFS in previously untreated DLBCL, *Leukemia & Lymphoma*, 65:5, 618-628, DOI: [10.1080/10428194.2024.2301924](https://doi.org/10.1080/10428194.2024.2301924)

To link to this article: <https://doi.org/10.1080/10428194.2024.2301924>

 View supplementary material 

 Published online: 09 Feb 2024.

 Submit your article to this journal 

 Article views: 350

 View related articles 

 View Crossmark data 

 Citing articles: 2 View citing articles 

NGS-determined molecular markers and disease burden metrics from ctDNA correlate with PFS in previously untreated DLBCL

Ehsan Tabari^a, Alexander F. Lovejoy^{at}, Hai Lin^a, Christopher R. Bolen^b, Seng Lor Saelee^a, Joshua P. Lefkowitz^a, David M. Kurtz^c, Alessia Bottos^d, Tina G. Nielsen^d, Joana M. Parreira^d and Khai T. Luong^{at}

^aRoche Sequencing Solutions, Pleasanton, CA, USA; ^bGenentech, Inc., South San Francisco, CA, USA; ^cDivision of Oncology, Department of Medicine, Stanford University, Stanford, CA, USA; ^dF. Hoffmann-La Roche Ltd., Basel, Switzerland

ABSTRACT

Personalized risk stratification and treatment may help improve outcomes among patients with diffuse large B-cell lymphoma (DLBCL). We developed a next-generation sequencing (NGS)-based method to assess a range of potential prognostic indicators, and evaluated it using pretreatment plasma samples from 310 patients with previously untreated DLBCL from the GOYA trial (NCT01287741). Variant calls and DLBCL subtyping with the plasma-based method were concordant with corresponding tissue-based methods. Patients with a tumor burden greater than the median ($p = .003$) and non-germinal center B-cell-like (non-GCB) DLBCL ($p = .049$) had worse progression-free survival than patients with a tumor burden less than the median or GCB DLBCL. Multi-factor assessment combining orthogonal features from a single pretreatment plasma sample has promise as a prognostic indicator in this setting ($p = .085$). This minimally invasive plasma-based NGS assay could enable comprehensive prognostic assessment of patients in a clinical setting, with greater accessibility than current methods.

ARTICLE HISTORY

Received 13 June 2023
Received 14 December 2023
Accepted 2 January 2024

KEYWORDS

DLBCL; circulating tumor DNA; next-generation sequencing; biomarkers; GOYA

Introduction

Diffuse large B-cell lymphoma (DLBCL) is the most prevalent type of non-Hodgkin lymphoma in adults, accounting for approximately 40% of all lymphomas, and exhibits a wide range of biological and molecular heterogeneity [1, 2]. Although standard of care therapies are curative in most cases, as many as 40% of patients ultimately relapse or become refractory to treatment [3].

Personalized risk stratification and treatment are promising avenues to improving outcomes for patients with DLBCL. However, conventional methods (such as the International Prognostic Index [IPI], which uses clinical parameters to classify patients into risk groups) have been shown to offer suboptimal risk stratification [3–6]. Furthermore, treatment strategies determined by metabolic imaging with positron emission tomography/computed tomography have been unsuccessful in improving survival [7, 8]. Previous studies have shown

biomarkers based on tumor molecular features to be encouraging for risk stratification and therapeutic targeting [3, 9–11], but these are not widely available in clinical settings.

The most common molecular classification scheme in DLBCL relies on determining the tumor's cell of origin (COO) [9, 12, 13]. Most DLBCL tumors can be classified into two transcriptionally distinct subtypes, germinal center B-cell like (GCB) and activated B-cell like (ABC) [9]. These subtypes are prognostic, with the ABC subtype showing inferior progression-free survival (PFS) compared with the GCB subtype [14]. As reported in many studies, patients with ABC DLBCL experience treatment failure with rituximab plus cyclophosphamide, doxorubicin, vincristine and prednisone (R-CHOP) therapy more frequently than patients with GCB DLBCL [15]. Patients with ABC DLBCL may show favorable responses to immunomodulatory regimens or Bruton's tyrosine kinase inhibitors [16–18]; however, data are limited and further investigation is warranted.

CONTACT Ehsan Tabari  ehsan@tabari.info  Roche Sequencing Solutions, Freenome, 279 E Grand Ave, 5th Floor, South San Francisco, CA 94080, USA

 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/10428194.2024.2301924>.

Prior presentation: Previously presented at the 61st ASH Annual Meeting & Exposition, December 7–10, 2019. Title: 'Molecular characteristics and disease burden metrics determined by NGS on ctDNA correlate with PFS in previously untreated DLBCL' (oral presentation).

Present address: Alexander F. Lovejoy, Freenome, South San Francisco, CA, USA; Khai T. Luong Illumina Inc., San Diego, CA, USA.

The phase 3 POLARIX trial recently demonstrated the superiority of polatuzumab vedotin plus R-CHP over R-CHOP in patients with untreated DLBCL [19]. An exploratory analysis of centrally confirmed biomarker subgroups showed a favorable trend toward polatuzumab vedotin plus R-CHP compared with R-CHOP in patients with ABC DLBCL [19]. Despite the exploratory nature of these analyses and the fact that COO only partially describes the heterogeneity of DLBCL disease [20], rapid and accurate subtype identification could still be very attractive.

While molecular methods show promise in dissecting DLBCL heterogeneity, reliance on assays not available at most institutions can limit their clinical applicability. Blood plasma samples are typically easier to collect than tissue samples, and the nucleic acid derived from plasma samples is of high quality. Circulating tumor DNA (ctDNA)-based approaches enable minimally invasive tumor genotyping using the detection of variants [3, 21], and assessment of tumor burden by the prevalence of these variants [3, 22–25]. These approaches can be used for molecular subtyping analyses [3], making plasma-based assays a promising route for patient risk stratification.

In this study, we report sequencing results and analyses of plasma samples from patients with previously untreated DLBCL from GOYA, a randomized phase 3 trial comparing rituximab with obinutuzumab as part of an immunochemotherapy regimen (R-CHOP versus G-CHOP; NCT01287741) [26–28]. Based on these targeted next-generation sequencing (NGS) ctDNA data, we explored multiple methods to use the plasma-specific data to risk stratify patients with DLBCL, including correlating ctDNA levels with prognosis, and developing a new COO subtyping method.

Table 1. Baseline and disease characteristics in the GOYA ITT and biomarker-evaluable populations.

ITT population <i>N</i> =1418	Biomarker-evaluable population <i>n</i> =310	
Age, median (range), years	62 (18-86)	64 (18-86)
Male, n (%)	752 (53)	154 (53)
Treated with G-CHOP, n (%)	706 (50)	153 (53)
ECOG PS 2/3, n (%)	186 (13)	28 (10)
IPI score 4/5 (high risk), n (%)	220 (16)	46 (16)
COO subtypes		
GCB	540 (58)	152 (49)
ABC	243 (26)	120 (39)
Unclassified	150 (16)	37 (12)
Region, n (%)		
Asia	518 (37)	50 (16)
Eastern Europe	196 (14)	64 (21)
North America	216 (15)	65 (21)
Western Europe	426 (30)	114 (37)
Other	62 (4)	17 (5)

ECOG PS: Eastern Cooperative Oncology Group performance status; G-CHOP: obinutuzumab plus cyclophosphamide, doxorubicin, vincristine and prednisone; IPI: International Prognostic Index; ITT: intent to treat.

Methods

Patient cohort

For this investigational exploratory analysis, a subset of patients who had pretreatment plasma samples sufficient for analysis was chosen from the GOYA trial (*n*=310). The sample size was considered to have sufficient statistical power to demonstrate the prognostic potential of ctDNA. The study design for GOYA has been previously described [26, 27]. Patients had previously untreated DLBCL, and baseline characteristics are described in Table 1. Similar baseline characteristics were observed between the intention-to-treat and biomarker-evaluable populations (Table S1).

The GOYA trial was conducted in accordance with the updated Declaration of Helsinki, the International Conference on Harmonization guidelines for Good Clinical Practice, and all applicable local laws and regulations. The protocol was approved by the ethics committees of participating centers and written informed consent was provided by all patients.

NGS

The NGS assay used in this analysis is research-use, and has not been validated by other data sets. The assay is described in the [Supplementary Methods](#)

Variant calling, filtering and tumor burden estimation

Single nucleotide variants (SNVs), small insertions and deletions (indels), and translocations were called from sequencing data using updated versions of the AVENIO ctDNA analysis variant callers. Full details can be found in the [Supplementary Methods](#).

Sample tumor burden was estimated by calculating the number of tumor genome copies per mL of plasma. This calculation, labeled mutant molecule per mL (MMPM), is described in the [Supplementary Methods](#).

Variant call concordance between tissue and plasma

The variant calls from plasma were compared with those from tissue reported by FoundationOne™ Heme (FMI) [29], where tissue-based variant calls were available. Details of the concordance evaluation can be found in the [Supplementary Methods](#) and Table S2. All correlations were performed using the Pearson correlation.

COO classifier

We developed a machine learning model to predict DLBCL COO based on the variants detected in plasma. The cohort was divided into two groups: 80 samples were set aside for blind validation (the validation set), while the remaining 230 samples were used to train and test the model (the training set). For each sample, a list of genes with any non-silent somatic SNVs, indels or translocations was compiled, regardless of the loci of the variants within the gene; for example, SNV in EZH2, indel in PIM1 or translocation in BCL2. In one case, a variant within the MYD88 gene (L265P/L273P) was singled out as it is a known common driver for ABC DLBCL [30]. The model was an ensemble of 100 XGBoost predictors that were parameterized similarly (subsample = 1, colsample_bytree = .2, min_child_weight = 1). Each predictor was trained independently against the linear predictive scores reported by the NanoString Lymph2Cx assay [13]. The median of scores from the 100 predictors was used as the final score for each sample (Figure S1A). While the range of scores for the ABC and GCB subtypes were differentiable, the score ranges for the unclassified overlapped with the scores of the ABC subtype (Figure S1B). Hence, the subtypes reported in this study were subsequently labeled GCB and non-GCB. The top 22 features were chosen based on the improvement of accuracy (gain) for the final model, comprising 18 SNVs, 3 indels and 1 translocation (Table 2).

Table 2. Features of the final COO classifier model. Chosen features consisted of 18 SNVs, 3 indels and 1 translocation.

Feature	Gain
EZH2 SNV	39.02
BCL2 translocation	35.98
PIM1 SNV	27.41
TNFRSF14 SNV	25.43
MYD88p Leu273Pro SNV	22.89
PIM1 indel	17.24
CD58 SNV	16.04
PTEN SNV	15.91
EGR1 SNV	15.77
PRDM1 indel	15.73
GNA13 SNV	15.29
SGK1 SNV	14.93
P2RY8 SNV	14.61
CNTNAP2 SNV	13.87
IRF4 SNV	12.84
SOCS1 SNV	12.09
KLHL14 indel	12.08
CD79B SNV	12.01
BCL10 SNV	9.38
ACTB SNV	9.12
B2M SNV	8.98
BCL2 SNV	8.49

COO: cell of origin; indel: insertions and deletions; SNV: single nucleotide variant.

Prognostic associations

The analysis of the prognostic value of known molecular markers is described in the [Supplementary Methods](#).

Results

Variant landscape and concordance with tissue

SNVs and short indels from a 314 kilobase (kb) region of the genome, as well as three gene translocations (BCL2, BCL6, and MYC) with unspecified partners, were detected in the plasma samples from 310 patients. No plasma samples were excluded from analysis for technical reasons. After filtering to exclude likely non-tumor-specific variants, between 4 and 877 tumor-specific SNVs (average 190 SNVs; 20 non-silent SNVs) were detected per sample, with a median allele frequency (AF) ranging from 0.07% to 46.2%, and an average of 10 indels detected per patient (range, 0-42). In addition, 170 translocations of BCL2, BCL6 and MYC genes were detected (Table S2).

Of the 230 patients with plasma samples in the training set, 163 had corresponding tissue samples that were previously tested using the FMI Heme assay [31]. Overall, 1170/3298 SNVs reported from tissue fell within the boundaries of our panel; of these, 854 variants were detected in plasma (73.0% positive percentage agreement [PPA]; Figure 1A). Also, 419 short indels were reported from tissue, 158 of which fell within the boundaries of our panel; of these, 112 indels were detected in the corresponding plasma samples (70.9% PPA; Figure 1A). Finally, 49/67 gene translocations reported from tissue for BCL2, BCL6 and MYC, were detected in the plasma samples (73.1% PPA), and for samples without translocation calls in tissue, no translocation was found at a PPA of 96.7% (409/423; Figure 1B). Additionally, as expected, higher concordance was seen with higher input masses of cell-free DNA (cfDNA; $p < .0001$), allele frequency ($p < 0.0001$; Figure S2) and advanced stages of disease ($p < .0001$; Figure 1C,D; Figure S3). In particular, when the maximum recommended input of 50 ng of cfDNA was used in the assay, 88.2% PPA was achieved (compared with 60.8% for samples with < 30 ng input).

The sensitivity of translocation detection was assessed by comparing NGS calls with the results obtained through fluorescence *in situ* hybridization (FISH) from tissue samples. BCL2 FISH results were

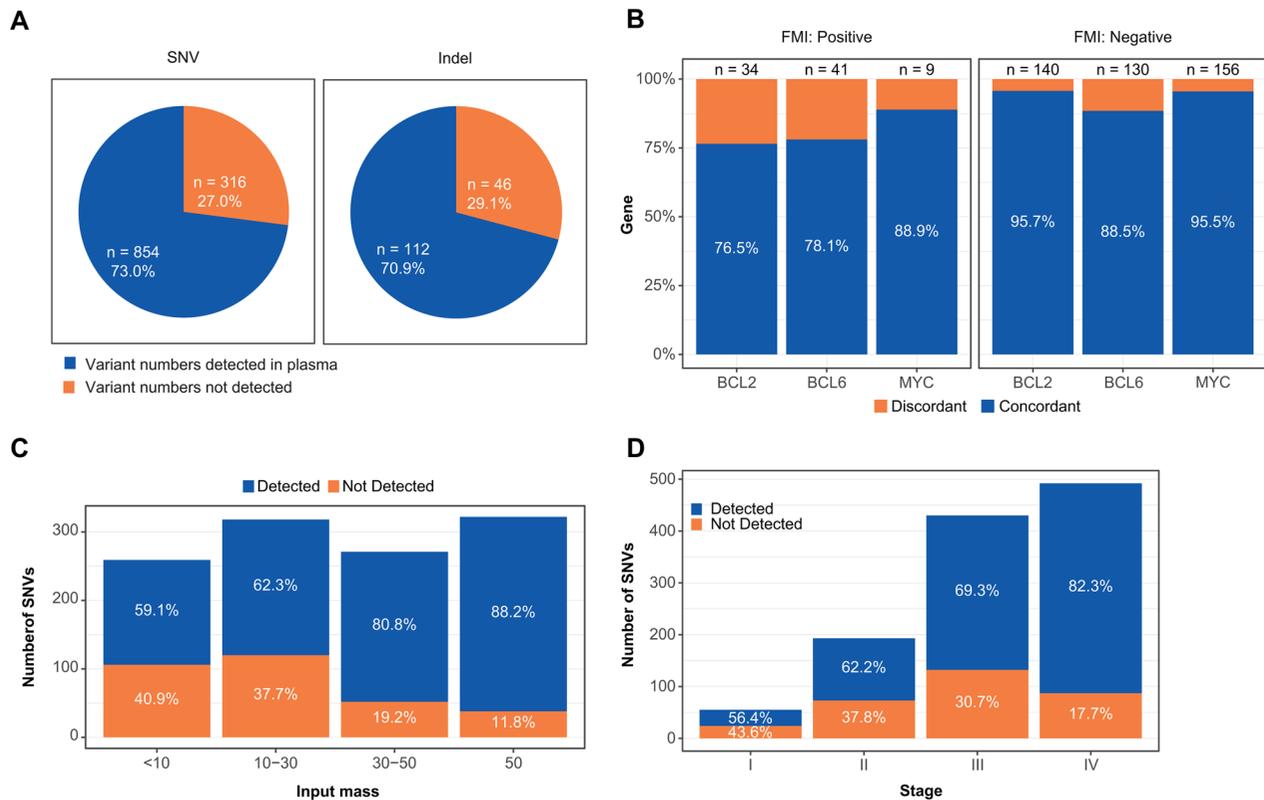


Figure 1. Concordance between variant calls detected in plasma and those detected in a corresponding tumor tissue assay. (A) Comparison between number and percentage of SNV (left) and indel (right) calls detected in tumor tissue in regions overlapping with the plasma panel. (B) PPA (left) and NPA (right) for BCL2, BCL6, and MYC translocation calls from plasma samples compared with paired tumor tissue samples. For samples with translocation calls in tissue, a corresponding translocation was detected in plasma 73.1% of the time (49/67). For samples without translocation calls in tissue, no translocation was found 96.7% of the time (409/423). (C and D) PPA of SNV calls between tissue and plasma by (C) input mass of cfDNA into library preparation and (D) stage of disease.

cfDNA: cell-free DNA; FMI: FoundationOne; indel: insertions and deletions; NPA: negative percent agreement; PPA: positive percent agreement; SNV: single nucleotide variant.

available for 224 samples, and MYC FISH results for 191 samples. Out of 39 samples that had positive FISH BCL2 calls, our assay only detected 9 (sensitivity of 23.01%). Similarly, out of 18 samples that had positive MYC FISH results, only 7 (38.89%) were detected from plasma.

Comparison of translocation detection by FMI versus FISH showed that results were generally consistent, with 17/39 samples (44%) with positive FISH BCL2 calls detected by FMI. Out of 18 samples with positive MYC FISH results, only 3 (17%) were detected by FMI.

ctDNA variant and tumor burden association with prognosis

Next, the association of known molecular markers detected at baseline in plasma and PFS was examined. Variants in BCL2, BCL6, CARD11, CD79B, MYC, MYD88, and TP53 were assessed (Figure 2A) [32–34]. No individual variants were found to be significantly

associated with PFS at a level of $p < .05$. The closest associations with molecular markers and PFS were found in samples with short variants (indels and SNVs) in TP53 (HR, 1.99; $p = .0027$; Figure 2B), and translocations in MYC (HR 1.34, $p = .41$; Figure 2C).

Across the full dataset, the range of MMPM was 3.45-8055. Higher MMPM values corresponded with higher IPI scores (analysis of variance $p < .001$; Figure 3A) and a higher likelihood of bulky disease ($p = .019$; Figure 3B).

The prognostic value of MMPM was then analyzed. The samples were divided into two sets, high MMPM and low MMPM, with high MMPM samples having MMPM values > median MMPM. Higher MMPM values corresponded with worse PFS (HR, 1.97; $p = .003$). This trend was also observed on different MMPM split points of 25%, 50%, and 75%, and regardless of treatment received (Figure 3C; Figure S4). When controlled for COO, IPI score, antibody treatment and number of planned chemotherapy cycles in a multivariate analysis, a weak trend for MMPM and PFS was observed

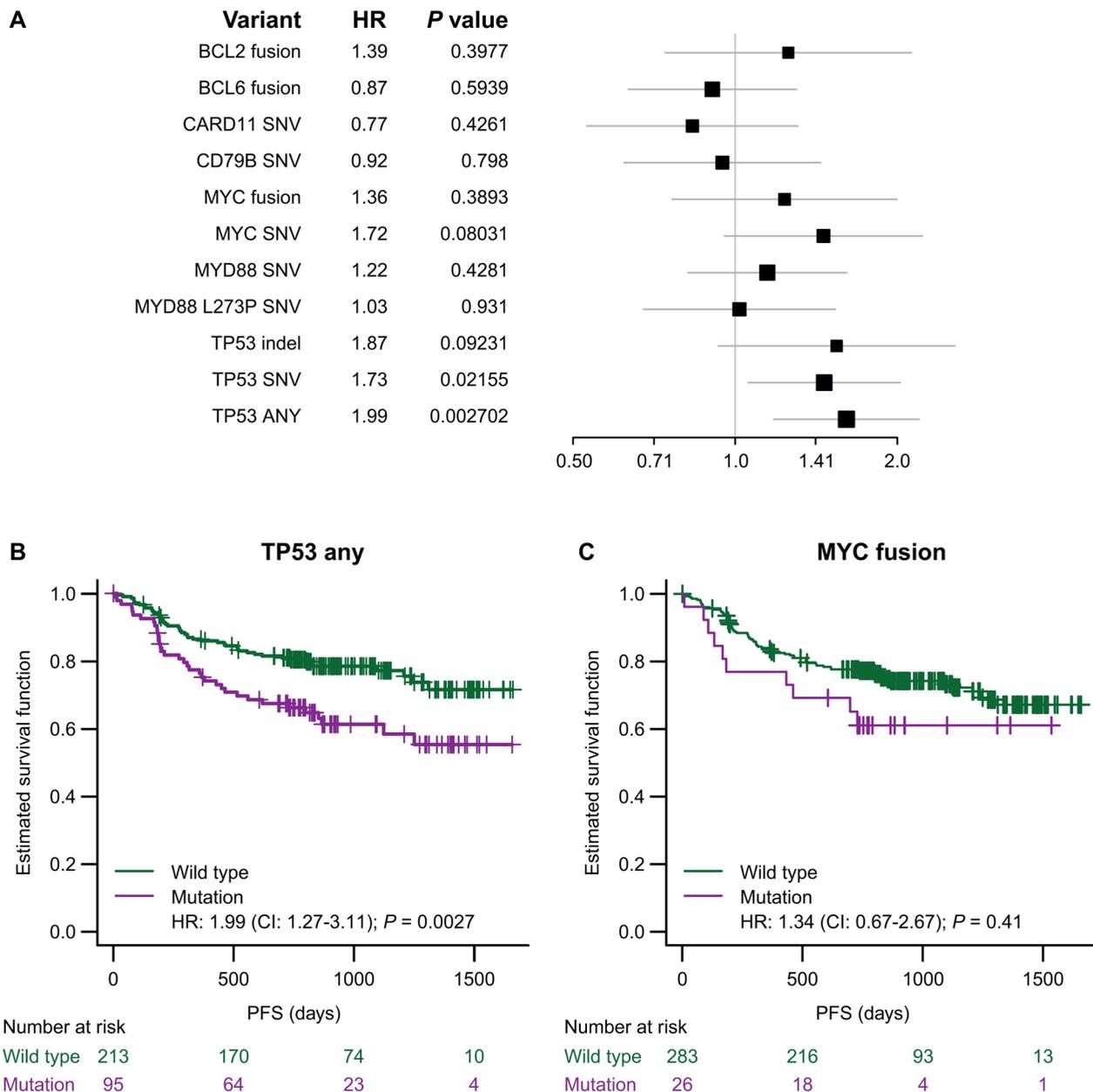


Figure 2. Association between gene variants detected in plasma and PFS. (A) HR, 95% CIs, and P values for the presence of specific variants of BCL2, BCL6, CARD11, CD79B, MYC, MYD88, and TP53 in plasma samples. (B and C) KM curves of PFS for presence versus absence of (B) TP53 SNVs or indels or (C) MYC translocations.

CI: confidence interval; HR: hazard ratio; indel: insertions and deletions; KM: Kaplan–Meier; PFS: progression-free survival; SNV: single nucleotide variant.

(HR, 1.48; $p = .11$). Additionally, MMPM showed a weak association with pretreatment total metabolic tumor volume with $r=0.41$ (Figure 3D), a low association with sum of product of the greatest diameters (SPD; $r=0.13$; Figure 3E) and a significant association with PFS after accounting for SPD (HR, 1.97; $p = .0031$).

Molecular subtyping from plasma

The validation set ($n=80$) contained 44 GCB, 21 ABC, and 15 unclassified samples, as determined by NanoString. Using the machine learning-based method

developed for this analysis, we classified these samples into GCB and non-GCB. No COO-relevant variants were detected for six samples and were labeled as indeterminate and excluded from the concordance analysis; 41 samples were classified as GCB, and 33 samples were classified as non-GCB (Figure 4A). The plasma-based method was concordant to the NanoString results in 33/40 (82.5%) GCB samples and 17/20 (85.0%) ABC samples. The overall agreement between the plasma-based method and NanoString was 75.8% when all samples (including those not classified with either method) were considered (Figure

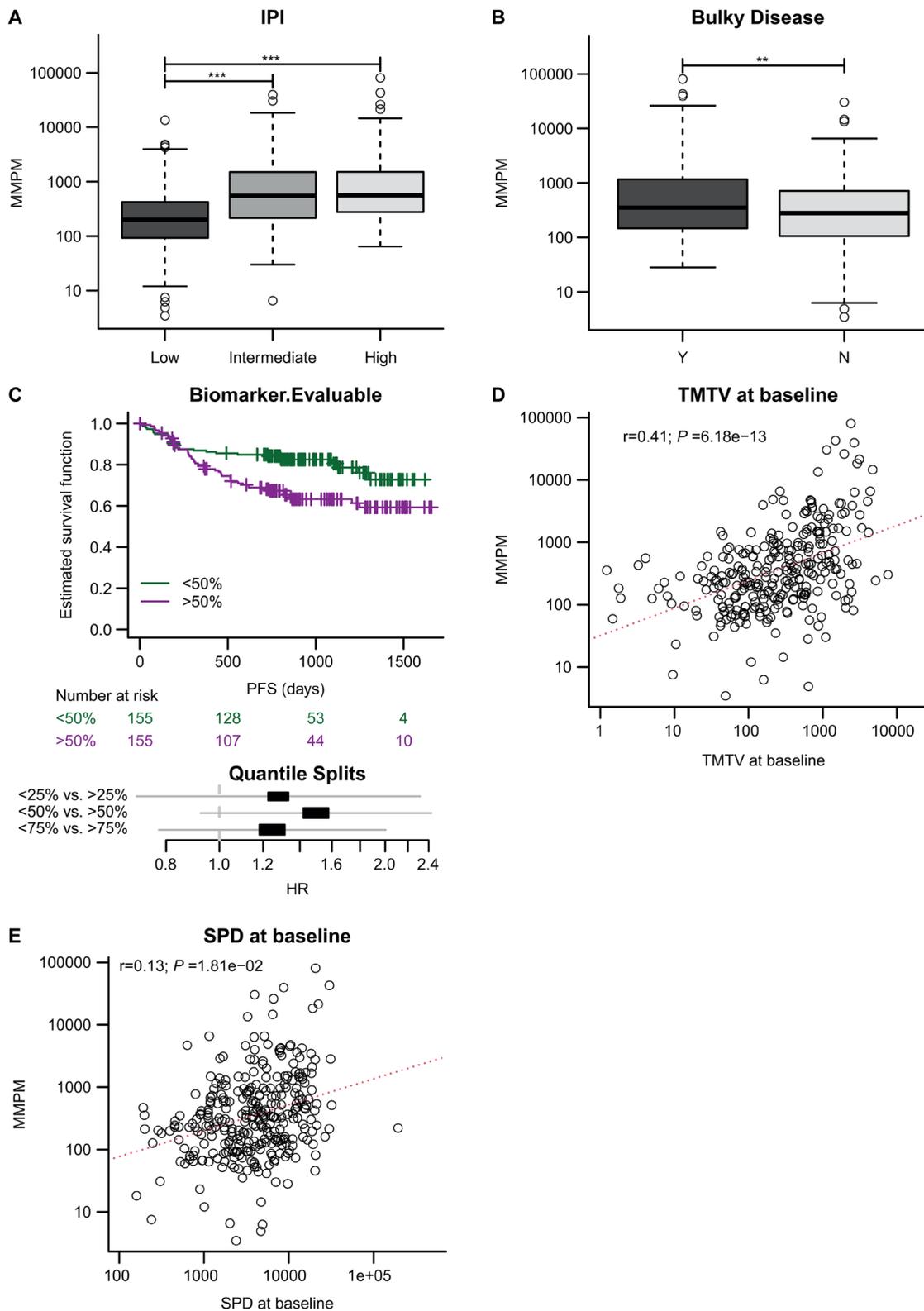


Figure 3. Association between tumor burden assessment from plasma (MMPM), other prognostic indicators, and prognosis. Higher MMPM values correlated with (A) higher IPI scores ($p < .001$), and (B) bulky disease ($p = .019$). (C) KM curves showing PFS split by above (purple) or below (green) the median MMPM value at baseline for all patients in the cohort. HRs and 95% CIs for quantile splits are also shown for MMPM split points of 25%, 50%, and 75%. (D and E) Comparison of MMPM with scan-based measures of (D) TMTV and (E) SPD.

** $P \leq .01$; *** $P \leq .001$.

CI: confidence interval; HR: hazard ratio; IPI: International Prognostic Index; KM: Kaplan–Meier; MMPM: mutant molecule per mL; N: no; PFS: progression-free survival; SPD: sum of product of the greatest diameters; TMTV: total metabolic tumor volume; vs: versus; Y: yes.

4B,C). Additionally, associations between patient outcomes and COO subtypes were similar between the two approaches (Figure 4D,E). Using the plasma-based method, samples unclassified by NanoString were classified into 5 GCB and 9 non-GCB subtypes. Improved PFS for GCB versus non-GCB was observed (HR, 2.57; $p = .4$; Figure S5).

Combined prognostic modeling from ctDNA features

Given the diversity of biomarkers available from plasma samples, we explored the possibility of a combined biomarker using baseline ctDNA features. A prognostic model was created by combining known prognostic features from the ctDNA assay (MMPM, TP53 SNV/indels, BCL2 translocations, MYC translocations, and COO subtypes with IPI scores and was developed on the 230 samples from the training set (Table 3). Patients with three or more high-risk biomarkers (or two, including high IPI score) were considered as ctDNA high risk, and had significantly worse PFS (Figure 5A; HR, 3.39; 95% confidence interval [CI]: 2.18-5.29; $p < .001$) than patients with low-risk biomarkers. This trend remained consistent in the test set, although the association was no longer significant (Figure 5B; HR, 2.15; 95% CI: 0.89-5.19). Notably, the prognostic model showed greater differentiation between patients at high risk versus patients at low risk when the plasma-based COO was used in the model, as opposed to the tissue-based COO (Figure S6A,B).

Discussion

Current methods for pretreatment prognostic assessment and risk stratification for patients with DLBCL involve multiple tests, and can often be limited by tumor tissue sample quality or availability, making early identification of patients at high risk difficult [3, 9–11]. A plasma-based assay could overcome these limitations, and potentially enable assessment of multiple prognostic modalities with a single test. Here, we provide proof-of-concept that a plasma-based DLBCL assay can enable such multiple modalities, including variant calling (SNVs, indels, and translocations), COO subtyping, and tumor burden measurements, with a single assay.

As with any ctDNA assay, the sensitivity of variant calls and molecular subtyping used in the assay described here was limited by the tumor burden in plasma. While ctDNA levels in DLBCL are typically

higher than those in solid tumor samples, mean AFs are still usually $< 10\%$, and can be $< 1\%$ in some cases [22–25], so it is expected that not all variants found in tissue will be detected in plasma. The tissue-plasma concordance in this study was 73.0% for SNVs, which is in line with previous publications [3] and, as expected, improved with advanced-stage patients (who typically have higher tumor burden). Particularly high concordance was seen in samples with higher input masses (88.2% PPA with 50ng input, as opposed to 60.8% with < 30 ng input), showing that increasing input cfDNA can improve imperfect plasma sensitivity. A greater volume of plasma (i.e. > 2 -2.5 mL) would be desirable to maximize the proportion of samples that have optimal input, enabling improved sensitivity of detection in plasma.

It is notable that no significant associations were seen between individual variants and patient outcomes in this dataset, since Bolen et al. found a clear link between BCL2 alterations and prognosis in the same dataset from the GOYA study [28]. There are a few possible explanations for this. Firstly, the size of this substudy – it may be expected that a larger sample set would result in more significant associations, particularly for variants in TP53 or MYC translocations, although not many significant associations were observed in a previous larger substudy of patients from the GOYA trial [28]. Secondly, the reduced sensitivity for variant calling in plasma versus tissue, given lower tumor burden levels, as false negatives will occur in any ctDNA assay. Finally, it seems apparent that other metrics such as tumor burden or molecular subtype have a greater association with patient outcomes than individual variants. These factors suggest that we may not expect to see a strong association with outcomes for individual variants; considering ctDNA data more holistically could provide better outcome predictions.

Previous studies of ctDNA in DLBCL have also identified an association between MMPM and PFS, with higher tumor burden being associated with poorer patient outcomes [23, 25]. Although this association was not found to be statistically significant in the current study, this could be attributed to a variety of reasons such as the relatively small sample size or differences in the statistical models and covariates used. It is important to note that these results do not contradict those observed in previous studies, with a strong trend between MMPM and PFS observed even after correcting for COO and IPI ($p = 0.11$), consistent with the interpretation from prior ctDNA studies that MMPM is independently prognostic for survival [23, 25].

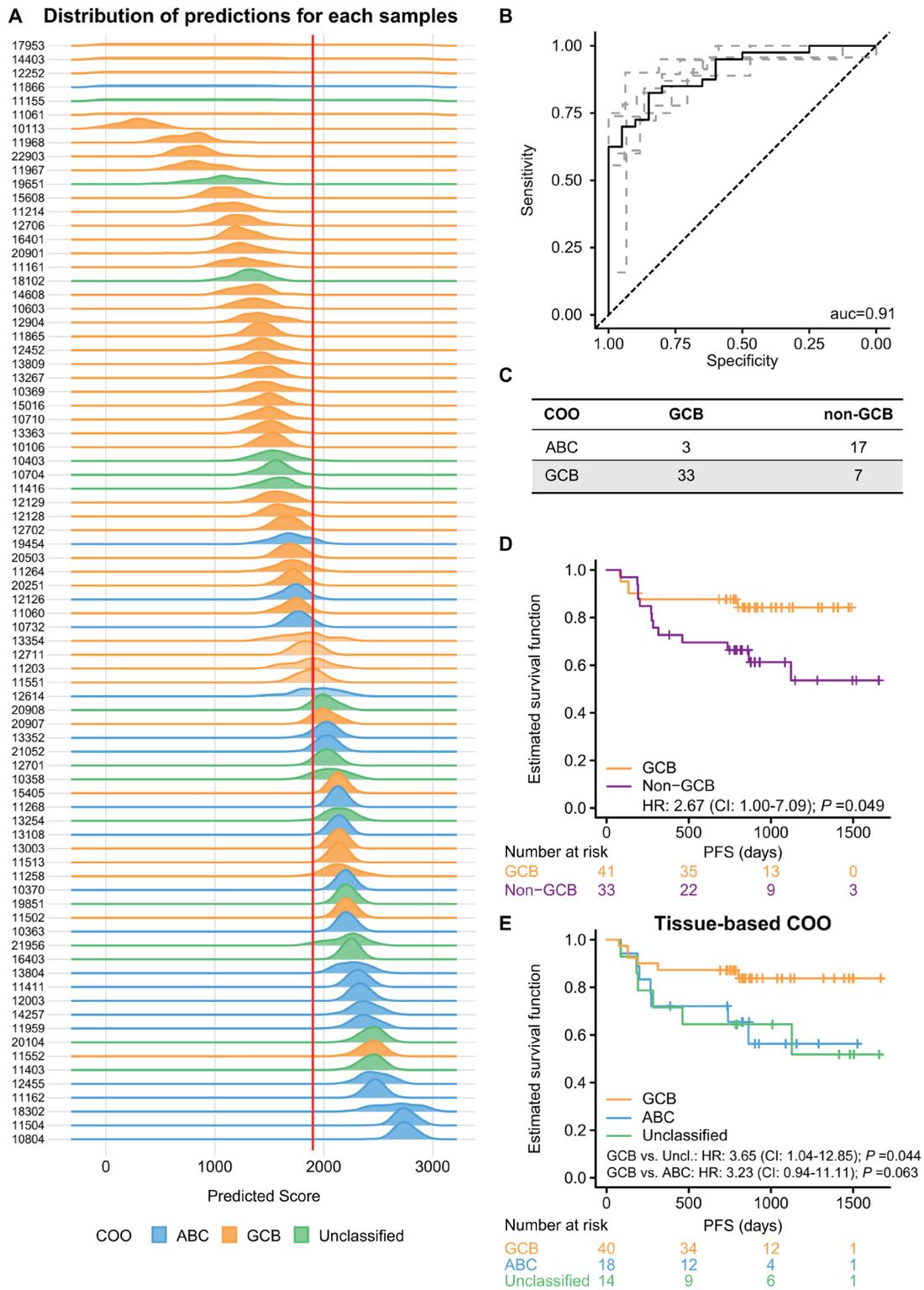


Figure 4. Molecular subtyping from plasma. (A) Prediction results of the validation set ($n=80$). Each row corresponds to the distribution of scores from the machine learning model for one sample. Distributions are colored by the corresponding NanoString Lymph2Cx COO call. The median value of the distribution gives the COO prediction, with medians to the left of the decision boundary (red vertical line) representing non-GCB calls, and those to the right representing GCB calls. (B) ROC curve for plasma machine learning-based variant calls compared with corresponding NanoString Lymph2Cx calls for the 5-fold cross validation (dashed lines) and for the validation set (solid line, AUC = 0.91). (C) Plasma machine learning-based COO calls compared with tissue NanoString Lymph2Cx-based calls. KM curves showing PFS for the validation set split by COO classification as determined by (D) the plasma-based machine learning method or (E) the tissue-based NanoString Lymph2Cx. ABC: activated B-cell like; AUC: area under curve; COO: cell of origin; CI: confidence interval; GCB: germinal center B-cell like; HR: hazard ratio; KM: Kaplan-Meier; PFS: progression-free survival; ROC: receiver operating characteristic; Uncl: unclassified.

Table 3. Combined prognostic model using ctDNA features.

Feature	Model weight
IPI high	2
IPI intermediate	1
MMPM > median	1
TP53 SNV/indel	1
BCL2 translocation	1
MYC translocation	1
COO non-GCB	1

COO: cell of origin; ctDNA: circulating tumor DNA; GCB: germinal center B-cell like; indel: insertions and deletions; IPI: International Prognostic Index; MMPM: mutant molecule per mL; SNV: single nucleotide variant.

Despite the limited sensitivity for variant calling in plasma, the concordance of molecular subtypes between tissue and plasma remained high; COO subtype concordance was 83.3% for patients classified by both NanoString and our plasma-based method. A small number of samples unclassified by NanoString were classified into GCB and non-GCB subtypes using the plasma-based method. The improved PFS with GCB versus non-GCB observed in this study showcases the potential benefit of defining otherwise unclassified samples. It is notable that both the bi-level and multi-level subtyping methods could be recapitulated in plasma, highlighting the utility of this noninvasive technique for identifying transcriptional subtypes. Many subtyping methods require fresh frozen tissue, which is often unavailable in a clinical setting, or isolation of RNA from formalin-fixed paraffin-embedded samples that can vary greatly in quality. Being able to determine subtypes like these from plasma samples could enable more widespread sample availability.

Another key benefit of cfDNA-based methods is that tumor variant AF can be used as a proxy for tumor burden. Furthermore, the simultaneous reporting of multiple biomarkers in this plasma NGS assay can be combined for a more comprehensive prognostic assessment accounting for both the size and the molecular characteristics of the malignancy. This prognostic assessment is straightforward but limited, as it is trained on a limited number of samples ($n=230$) from patients split between two treatments. A more comprehensive algorithm based on molecular and tumor burden information could be established by using this assay with a larger and more diverse patient population. Finally, while this study explores the benefits of a plasma NGS assay using pretreatment samples, one significant benefit of a plasma-based assay is the ability to take multiple measurements over time. The advantages of tumor-burden tracking and molecular assessment in longitudinal samples have been shown previously [3, 22–25], and the assay described here could improve prognostic prediction through repeated sampling. Likely the optimal assessment schedule for patients with DLBCL would involve multi-modal evaluations

prior to and throughout treatment, with prognosis updated after each assessment [34].

The limitations of this study include its exploratory nature and lack of external validation. Further research is required to validate these models in an independent dataset.

In conclusion, the plasma-based NGS assay described here could enable a more rapid, comprehensive and accessible prognostic assessment of patients with

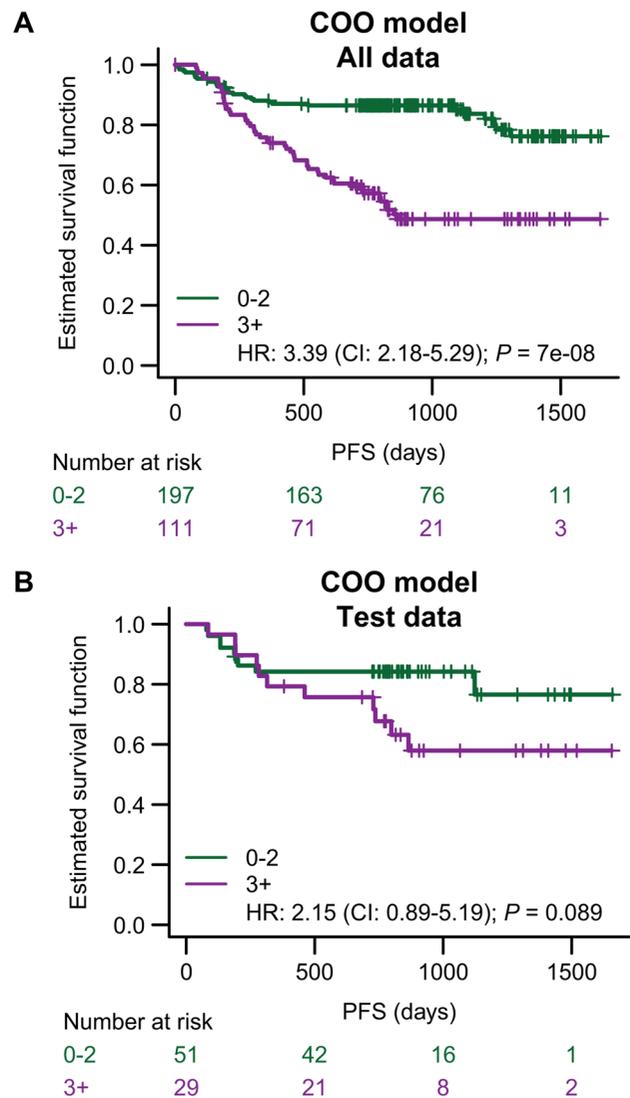


Figure 5. A multi-modal model for prognostic prediction using IPI and a ctDNA-based NGS assay. Methods that assigned points for high-risk features were developed based on the training set of 230 patients, then applied to all 310 patients (full dataset), or the 80 patients in the test set. KM curves showing PFS for high versus low risk are shown for each model. The model incorporated MMPM, TP53 status, BCL2/MYC translocation status, COO calling from plasma, and IPI. (A) PFS in the full dataset and (B) PFS in the test set.

CI: confidence interval; COO: cell of origin; ctDNA: circulating tumor DNA; HR: hazard ratio; IPI: International Prognostic Index; KM: Kaplan–Meier; MMPM: mutant molecule per mL; NGS: next-generation sequencing; PFS: progression-free survival.

previously untreated DLBCL in a clinical setting compared with currently available methods.

Acknowledgments

The authors thank the patients and their families, all GOYA trial team members, and investigators. GOYA was supported by F. Hoffmann-La Roche Ltd. Third-party medical writing assistance under the direction of Ehsan Tabari was provided by Helen Cathro, PhD, of Ashfield MedComms, an Inizio company, and was funded by F. Hoffmann-La Roche Ltd.

Authors' contributions

Study design: ET, AL, KTL. Study conduct: ET, AL, CB, KTL. Data collection: SLS, JPL, TN. Data analysis: ET, CB, HL. Data interpretation: ET, AL, CB, DMK, AB, TN, JP.

Disclosure statement

E.T is employed by Freenome, formerly employed by Roche Sequencing Solutions and has equity ownership interests in Freenome. A.F.L is employed by Freenome, formerly employed by Roche Sequencing Solutions, and has equity ownership interests in Freenome. H.L is employed by Roche Sequencing Solutions and has other financial relationships with Veractye (spouse). C.R.B is employed by Genentech, Inc. and has equity ownership interests in F. Hoffmann-La Roche Ltd. S.L.S is employed by Roche Sequencing Solutions. J.P.L is employed by Roche Sequencing Solutions. D.M.K reports consultancy with Roche Molecular, Genentech, Inc., and equity ownership in Foresight Diagnostics. A.B, T.G.N and J.M.P are employed by and have equity ownership interests in F. Hoffmann-La Roche Ltd. K.T.L is employed by Illumina Inc., formerly employed by F. Hoffmann-La Roche Ltd., and has equity ownership interests in Natera and Illumina Inc.

Funding

The author(s) reported there is no funding associated with the work featured in this article.

Data availability statement

Data are not shared in an unrestricted way, but individual requests for data sharing will be assessed on a case-by-case basis.

References

- [1] Camus V, Sarafan-Vasseur N, Bohers E, et al. Digital PCR for quantification of recurrent and potentially actionable somatic mutations in circulating free DNA from patients with diffuse large B-cell lymphoma. *Leuk Lymphoma*. 2016;57(9):2171–2179. doi:10.3109/10428194.2016.1139703
- [2] Coccaro N, Anelli L, Zagaria A, et al. Molecular complexity of diffuse large B-cell lymphoma: can it be a roadmap for precision medicine? *Cancers (Basel)*. 2020;12(1):185. doi:10.3390/cancers12010185
- [3] Scherer F, Kurtz DM, Newman AM, 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
- [4] Sehn LH, Berry B, Chhanabhai M, et al. The revised international prognostic index (R-IPI) is a better predictor of outcome than the standard IPI for patients with diffuse large B-cell lymphoma treated with R-CHOP. *Blood*. 2007;109(5):1857–1861. doi:10.1182/blood-2006-08-038257
- [5] Ziepert M, Hasenclever D, Kuhnt E, et al. Standard international prognostic index remains a valid predictor of outcome for patients with aggressive CD20+ B-cell lymphoma in the rituximab era. *J Clin Oncol*. 2010;28(14):2373–2380. doi:10.1200/jco.2009.26.2493
- [6] Younes A. Prognostic significance of diffuse large B-cell lymphoma cell of origin: seeing the Forest and the trees. *J Clin Oncol*. 2015;33(26):2835–2836. doi:10.1200/jco.2015.61.9288
- [7] Casasnovas RO, Ysebaert L, Thieblemont C, et al. FDG-PET-driven consolidation strategy in diffuse large B-cell lymphoma: final results of a randomized phase 2 study. *Blood*. 2017;130(11):1315–1326. doi:10.1182/blood-2017-02-766691
- [8] Dührsen U, Müller S, Hertenstein B, 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):2024–2034. doi:10.1200/jco.2017.76.8093
- [9] Alizadeh AA, Eisen MB, Davis RE, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature*. 2000;403(6769):503–511. doi:10.1038/35000501
- [10] Rosenwald A, Wright G, Chan WC, et al. The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. *N Engl J Med*. 2002;346(25):1937–1947. doi:10.1056/NEJMoa012914
- [11] Moffitt AB, Dave SS. Clinical applications of the genomic landscape of aggressive non-hodgkin lymphoma. *J Clin Oncol*. 2017;35(9):955–962. doi:10.1200/jco.2016.71.7603
- [12] Lenz G, Wright GW, Emre NC, et al. Molecular subtypes of diffuse large B-cell lymphoma arise by distinct genetic pathways. *Proc Natl Acad Sci U S A*. 2008;105(36):13520–13525. doi:10.1073/pnas.0804295105
- [13] Scott DW, Wright GW, Williams PM, et al. Determining cell-of-origin subtypes of diffuse large B-cell lymphoma using gene expression in formalin-fixed paraffin-embedded tissue. *Blood*. 2014;123(8):1214–1217. doi:10.1182/blood-2013-11-536433
- [14] Kim M, Suh C, Kim J, et al. Difference of clinical parameters between GCB and non-GCB subtype DLBCL. *Blood*. 2017;130(Supplement 1):5231–5231. doi:10.1182/blood.V130.Suppl_1.5231.5231%J Blood.
- [15] Gutiérrez-García G, Cardesa-Salzmann T, Climent F, et al. Gene-expression profiling and not immunophenotypic algorithms predicts prognosis in patients with diffuse large B-cell lymphoma treated with immunochemotherapy.

- Blood. 2011;117(18):4836–4843. doi:[10.1182/blood-2010-12-322362](https://doi.org/10.1182/blood-2010-12-322362)
- [16] Wilson WH, Young RM, Schmitz R, et al. Targeting B cell receptor signaling with ibrutinib in diffuse large B cell lymphoma. *Nat Med*. 2015;21(8):922–926. doi:[10.1038/nm.3884](https://doi.org/10.1038/nm.3884)
- [17] Gupta E, Accurso J, Sluzevich J, et al. Excellent outcome of immunomodulation or bruton's tyrosine kinase inhibition in highly refractory primary cutaneous diffuse large B-cell lymphoma, leg type. *Rare Tumors*. 2015;7(4):6067–6166. doi:[10.4081/rt.2015.6067](https://doi.org/10.4081/rt.2015.6067)
- [18] 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(15):1285–1295. doi:[10.1200/jco.18.02403](https://doi.org/10.1200/jco.18.02403)
- [19] Tilly H, Morschhauser F, Sehn LH, et al. Polatuzumab vedotin in previously untreated diffuse large B-cell lymphoma. *N Engl J Med*. 2022;386(4):351–363. doi:[10.1056/NEJMoa2115304](https://doi.org/10.1056/NEJMoa2115304)
- [20] Chapuy B, Stewart C, Dunford AJ, et al. Molecular subtypes of diffuse large B cell lymphoma are associated with distinct pathogenic mechanisms and outcomes. *Nat Med*. 2018;24(5):679–690. doi:[10.1038/s41591-018-0016-8](https://doi.org/10.1038/s41591-018-0016-8)
- [21] Rossi D, Diop F, Spaccarotella E, et al. Diffuse large B-cell lymphoma genotyping on the liquid biopsy. *Blood*. 2017;129(14):1947–1957. doi:[10.1182/blood-2016-05-719641](https://doi.org/10.1182/blood-2016-05-719641)
- [22] Kurtz DM, Green MR, Bratman SV, et al. Noninvasive monitoring of diffuse large B-cell lymphoma by immunoglobulin high-throughput sequencing. *Blood*. 2015;125(24):3679–3687. doi:[10.1182/blood-2015-03-635169](https://doi.org/10.1182/blood-2015-03-635169)
- [23] Roschewski M, Dunleavy K, Pittaluga S, 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):541–549. doi:[10.1016/s1470-2045\(15\)70106-3](https://doi.org/10.1016/s1470-2045(15)70106-3)
- [24] Roschewski M, Staudt LM, Wilson WH. Dynamic monitoring of circulating tumor DNA in non-Hodgkin lymphoma. *Blood*. 2016;127(25):3127–3132. doi:[10.1182/blood-2016-03-635219](https://doi.org/10.1182/blood-2016-03-635219) *J Blood*.
- [25] Kurtz DM, Scherer F, Jin MC, et al. Circulating tumor DNA measurements as early outcome predictors in diffuse large B-cell lymphoma. *J Clin Oncol*. 2018; 36(28):2845–2853. doi:[10.1200/JCO.2018.78.5246](https://doi.org/10.1200/JCO.2018.78.5246)
- [26] Vitolo U, Trněný M, Belada D, et al. Obinutuzumab or rituximab plus cyclophosphamide, doxorubicin, vincristine, and prednisone in previously untreated diffuse large B-cell lymphoma. *J Clin Oncol*. 2017;35(31):3529–3537. doi:[10.1200/jco.2017.73.3402](https://doi.org/10.1200/jco.2017.73.3402)
- [27] Sehn LH, Martelli M, Trněný M, et al. A randomized, open-label, phase III study of obinutuzumab or rituximab plus CHOP in patients with previously untreated diffuse large B-cell lymphoma: final analysis of GOYA. *J Hematol Oncol*. 2020;13(1):71. doi:[10.1186/s13045-020-00900-7](https://doi.org/10.1186/s13045-020-00900-7)
- [28] Bolen CR, Klanova M, Trneny M, et al. Prognostic impact of somatic mutations in diffuse large B-cell lymphoma and relationship to cell-of-origin: data from the phase III GOYA study. *Haematologica*. 2020;105(9):2298–2307. doi:[10.3324/haematol.2019.227892](https://doi.org/10.3324/haematol.2019.227892)
- [29] Newman AM, Bratman SV, Stehr H, et al. FACTERA: a practical method for the discovery of genomic rearrangements at breakpoint resolution. *Bioinformatics*. 2014;30(23):3390–3393. doi:[10.1093/bioinformatics/btu549](https://doi.org/10.1093/bioinformatics/btu549)
- [30] Ngo VN, Young RM, Schmitz R, et al. Oncogenically active MYD88 mutations in human lymphoma. *Nature*. 2011;470(7332):115–119. doi:[10.1038/nature09671](https://doi.org/10.1038/nature09671)
- [31] He J, Abdel-Wahab O, Nahas MK, et al. Integrated genomic DNA/RNA profiling of hematologic malignancies in the clinical setting. *Blood*. 2016;127(24):3004–3014. doi:[10.1182/blood-2015-08-664649](https://doi.org/10.1182/blood-2015-08-664649)
- [32] 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(3):359–372. doi:[10.1016/j.ccr.2012.07.014](https://doi.org/10.1016/j.ccr.2012.07.014)
- [33] Xu-Monette ZY, Wu L, Visco C, et al. Mutational profile and prognostic significance of TP53 in diffuse large B-cell lymphoma patients treated with R-CHOP: report from an international DLBCL rituximab-CHOP consortium program study. *Blood*. 2012;120(19):3986–3996. doi:[10.1182/blood-2012-05-433334](https://doi.org/10.1182/blood-2012-05-433334)
- [34] Schuetz JM, Johnson NA, Morin RD, et al. BCL2 mutations in diffuse large B-cell lymphoma. *Leukemia*. 2012;26(6):1383–1390. doi:[10.1038/leu.2011.378](https://doi.org/10.1038/leu.2011.378)