



## OPEN ACCESS

## EDITED BY

Gabor Mikala,  
Central Hospital of Southern Pest, Hungary

## REVIEWED BY

Prajish Iyer,  
City of Hope National Medical Center,  
United States  
Antonio Giovanni Solimando,  
University of Bari Aldo Moro, Italy

## \*CORRESPONDENCE

Mao Mao

✉ maomao@yuhs.ac

<sup>†</sup>These authors have contributed equally to this work

RECEIVED 21 November 2023

ACCEPTED 02 January 2024

PUBLISHED 19 January 2024

## CITATION

Chang Y, Li S, Li Z, Wang X, Chang F, Geng S, Zhu D, Zhong G, Wu W, Chang Y, Tu S and Mao M (2024) Non-invasive detection of lymphoma with circulating tumor DNA features and protein tumor markers. *Front. Oncol.* 14:1341997. doi: 10.3389/fonc.2024.1341997

## COPYRIGHT

© 2024 Chang, Li, Li, Wang, Chang, Geng, Zhu, Zhong, Wu, Chang, Tu and Mao. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

# Non-invasive detection of lymphoma with circulating tumor DNA features and protein tumor markers

Yu Chang<sup>1†</sup>, Shiyong Li<sup>2†</sup>, Zhiming Li<sup>3†</sup>, Xinhua Wang<sup>1†</sup>, Fangyuan Chang<sup>4</sup>, Shuaipeng Geng<sup>4</sup>, Dandan Zhu<sup>4</sup>, Guolin Zhong<sup>2</sup>, Wei Wu<sup>2</sup>, Yinyin Chang<sup>4</sup>, Shichun Tu<sup>4</sup> and Mao Mao<sup>5</sup> <sup>2,5\*</sup>

<sup>1</sup>Department of Oncology, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China,

<sup>2</sup>Research and Development, SeekIn Inc, Shenzhen, China, <sup>3</sup>Department of Internal Medicine, Sun

Yat-sen University Cancer Center, Guangzhou, China, <sup>4</sup>Clinical Laboratories, Shenyou Bio,

Zhengzhou, China, <sup>5</sup>Yonsei Song-Dang Institute for Cancer Research, Yonsei University,

Seoul, Republic of Korea

**Background:** According to GLOBOCAN 2020, lymphoma ranked as the 9<sup>th</sup> most common cancer and the 12<sup>th</sup> leading cause of cancer-related deaths worldwide. Traditional diagnostic methods rely on the invasive excisional lymph node biopsy, which is an invasive approach with some limitations. Most lymphoma patients are diagnosed at an advanced stage since they are asymptomatic at the beginning, which has significantly impacted treatment efficacy and prognosis of the disease.

**Method:** This study assessed the performance and utility of a newly developed blood-based assay (SeekInCare) for lymphoma early detection. SeekInCare utilized protein tumor markers and a comprehensive set of cancer-associated genomic features, including copy number aberration (CNA), fragment size (FS), end motif, and lymphoma-related virus, which were profiled by shallow WGS of cfDNA.

**Results:** Protein marker CA125 could be used for lymphoma detection independent of gender, and the sensitivity was 27.8% at specificity of 98.0%. After integrating these multi-dimensional features, 77.8% sensitivity was achieved at specificity of 98.0%, while its NPV and PPV were both more than 92% for lymphoma detection. The sensitivity of early-stage (I-II) lymphoma was up to 51.3% (47.4% and 55.0% for stage I and II respectively). After 2 cycles of treatment, the molecular response of SeekInCare was correlated with the clinical outcome.

**Conclusion:** In summary, a blood-based assay can be an alternative to detect lymphoma with adequate performance. This approach becomes particularly valuable in cases where obtaining tissue biopsy is difficult to obtain or inconclusive.

## KEYWORDS

lymphoma, early diagnosis, copy number aberration, fragment size, end motif, EBV

## 1 Introduction

Lymphoma is a type of common cancer worldwide, including a large group of lymphoid hematopoietic malignancies, which can be further classified as Non-Hodgkin lymphoma (NHL) and Hodgkin lymphoma (HL) two major types (1). Lymphoma was the ninth most common cancer and the 12<sup>th</sup> leading cause of cancer death according to GLOBOCAN 2020 (2). As counted in 2016, the incidence and mortality of lymphoma were 6.5 and 3.73 per 100,000, respectively, and lymphoma is considered as one of the top three cancer incidences in children for boys in China (3). Lymphoma is a diverse group of malignancies that originated from B, T, and NK cells, which includes more than 30 unique subtypes (4, 5). Currently, diagnosis and classification of lymphoma was mainly based on excisional or punctured tissues by immunohistochemistry and *in situ* hybridization (4, 6, 7). However, the invasive operation has significant limitations, including potential procedural risks and inter- and intra-tumor heterogeneity (8–10). Especially for patients with lesions that are difficult to reach (e.g., those in brain; deep lymph nodes in the thoracic cavity or abdomen).

In the past few years, an emerging technique named liquid biopsy could potentially improve these limitations. Analysis of circulating cell-free DNA (cfDNA) is the leading non-invasive liquid biopsy approach and has been extensively utilized for cancer early diagnosis (11–13), through capturing early cancer signals from circulating tumor DNA (ctDNA). Although the majority of cfDNA was often not of cancerous origin, technical advances in next-generation sequencing (NGS), together with advanced computational methods, allowing for identifying tumor specific alterations or features, such as single nucleotide variations (11, 14), copy number aberrations (CNA) (15–19) and fragment-omics (16, 20–22), and epigenetic changes (23–28). With regard to lymphoma, liquid biopsy has been used for classification of lymphoma subtype, assessment of tumor burden as a prognostic biomarker, and detection of molecular response to therapy (29–31). All of them were based on the PCR or NGS target region assays to detect lymphoma specific mutations. Most lymphoma patients did not have hotspot mutations, and most mutations detected from cfDNA came from hematopoietic stem cells and accumulated over time, which resulted in false positive prediction. However, CNAs were rarely detected in healthy individuals.

Here, we have developed a blood-based multi-omics assay (named SeekInCare) (18). It utilized traditional protein tumor markers (PTMs) and cancer genomic hallmarks including CNA, fragment size (FS), end motif, and cancer-related virus which were evaluated by shallow whole-genome sequencing with the aim to calculate the cancer risk score (CRS). SeekInCare making full use of the common cancer genomic features, CNA, FS, end motif, and a panel of seven protein markers, is not limited to a particular cancer type and has been reported in several cancers (16, 20–22). Therefore, it could be applied to a broad pan-cancer mode to detect multiple cancer types simultaneously. It showed good performance in hepatocellular carcinoma (HCC) (75.0% sensitivity was achieved at 98.0% specificity) (22). Here we presented its utility in another cancer type, lymphoma. This study aimed to assess the potential for broad clinical utility of SeekInCare for lymphoma detection across diverse lymphoma subtypes.

## 2 Materials and methods

### 2.1 Sample collection

A total of 144 newly diagnosed and treatment-naïve lymphoma patients from two hospitals, The First Affiliated Hospital of Zhengzhou University and Sun Yat-sen University Cancer Center, were enrolled in this study from March 2021 to February 2022. Diagnosis and classification were made according to the 2016 revision of the World Health Organization classification of lymphoid neoplasms (5). 396 healthy individuals without any history of cancer or other cancer related clinical symptoms were enrolled as a control group. 99 healthy individuals from the physical examination center of the Second Affiliated Hospital of Sun Yat-sen University, with the remaining samples collected from our employees and their families. The study protocol was approved by the ethics committee of leading site: The First Affiliated Hospital of Zhengzhou University (2022-KY-0719-001). All participants provided written informed consent upon enrollment. One tube of blood (Streck, La Vista, USA) was collected from each participant after enrollment.

cfDNA extraction from blood, library construction, and sequencing process were described in a previous study by Meng et al, 2021 (22). In brief, cfDNA was extracted from plasma through the QIAamp Circulating Nucleic Acid Kit (Qiagen, Hilden, Germany). cfDNA was subject to library construction using the Kapa Hyper Prep Kit (Kapa Biosystems, Wilmington, MA) according to the manufacturer's protocol. Prepared libraries were sequenced on NovaSeq system (Illumina, San Diego, CA), with pair-end 150 bp for WGS to generate approximately 10 Gb of raw sequencing data.

### 2.2 Protein tumor marker quantification

500  $\mu$ L of plasma was used to quantify the expression of seven common protein tumor markers (AFP, CA125, CA15-3, CA19-9, CA72-4, CEA, and CYFRA21-1) by using Roche cobas e411 analyzer (Roche Diagnostics GmbH, Mannheim, Germany) and commercially available reagent kits following manufacturer's instructions.

### 2.3 Sequence alignment and CNA analysis

The sequence alignment and CNA analysis were described in Meng et al, 2021 (22). In brief, human genome was divided into non-overlapping 1-Mb bins, using the reads number of each bin from healthy samples to calculate the mean value and standard deviation. Next, calculate the Z-score of each bin by subtracting the mean value and dividing by the standard deviation. Bins with an absolute Z-score greater than 3 were classified as CNA bins and samples with more than 10 such bins were deemed CNA positive.

In order to depict the significantly recurrent amplification/deletion regions in patients across different lymphoma subtypes, we divided the reads number of each bin by the mean value of corresponding bin, and performed log<sub>2</sub>-transformation (named log R ratio), which was used as input for QDNAseq (v1.22.0) package to detect CNA segments. Patients' CNA segments were used to detect recurrent CNA amplification/deletion regions through GISTIC2.0 software (32).

## 2.4 FS analysis

Read pairs with a MAPQ score below 30 for either read or PCR duplicates were removed. Then calculate the length of each cfDNA fragment. The short fragments ratio (defined as P150) was the proportion of DNA fragments within 50~150bp. The ratio difference (named FS.S-L) between short and long (180~220bp) fragments was calculated as follows:

$$FS.S-L = \frac{\# \text{ of short reads (50 ~ 150bp)}}{\# \text{ of total reads}} - \frac{\# \text{ of long reads (180 ~ 220bp)}}{\# \text{ of total reads}}$$

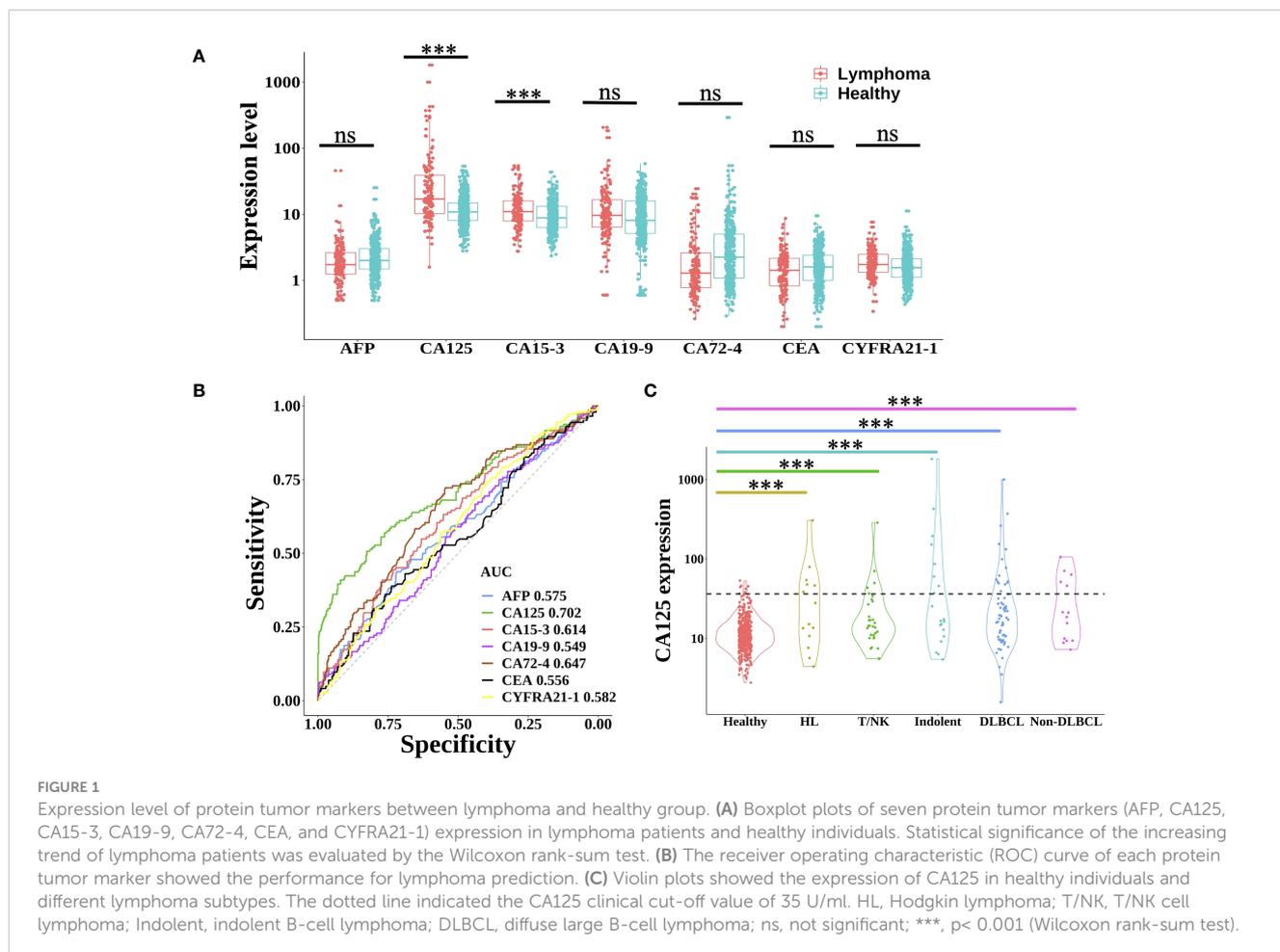
## 2.5 End motif analysis

The first 4-nucleotide sequence of each 5' fragment end (Watson and Crick strands) was extracted from the mapping results and defined as end motif (16, 33, 34). Then Wilcoxon rank-sum test was used to detect the significantly different frequency of each end motif between lymphoma patients and healthy individuals, and use of false discovery rate (FDR) method

to adjust for multiple testing. End motif with adjustment q-value less than 0.001 was selected to build a classifier to differentiate the lymphoma patients from the healthy subjects. To minimize the issue of overfitting and reduce the number of end motifs, least absolute shrinkage and selection operator regression (LASSO) method was applied to do the feature selection first. The selecting features were initially standardized, centering the data around a mean of zero and a standard deviation of one. This process was repeated 30 times and end motifs retained in all the repetitions were selected, which would be used to build the lymphoma predictive model by support vector machine (SVM). The average predicted value from multiple SVM models was used for further analysis.

## 2.6 EBV viral reads fraction calculation

Un-mapped reads were extracted from human mapping results. These reads were mapped to the Epstein-Barr virus (EBV) genome (AJ507799.2) with bowtie2 version 2.4.1 (<http://bowtie-bio.sourceforge.net/index.shtml>). After filtering the reads with mapping MAPQ score less than 30, the fraction of EBV reads per megabase was calculated as the following formula.



$$EBV \text{ viral read fraction} = \frac{\# \text{ of reads mapped to EBV}}{\# \text{ of total reads}} \times 10^6$$

## 2.7 Integrating multi-omics and multidimensional features to calculate cancer risk score

Through SeekInCare assay (Supplementary Figure 1), genomic features CNA, FS, end motif, and EBV value were analyzed by the sWGS data. Additionally, the protein tumor marker CA125 was incorporated as a feature. Feature selection was performed using LASSO followed by the construction of the lymphoma classifier model (35, 36). Before constructing the model, the five-dimensional features were initially normalized by subtracting the mean and dividing by the standard deviation. The sample set was randomly stratified in a 4:1 ratio into a training and validation set, then built the lymphoma classifier model with 10-fold cross-validation in the training set. This process was repeated 100 times. The average prediction value from these models was defined as CRS. If an individual's CRS surpassed the average prediction value corresponding to a specificity of 98.0%, they would be categorized as having lymphoma.

## 2.8 Statistical methods

The sample size for our study was determined using the Buderer's method (37), aiming for a sensitivity of 75% at a specificity of 98% with a 95% confidence level and a case-to-control ratio of 1:2. The calculated values for n1 and n2 were 219 and 28, respectively, and the larger value was selected, resulting in a sample size of 219. Considering a 5% dropout rate, the final sample size was set at 231. Notably, we were able to surpass this requirement by collecting a total of 540 samples for our study. This larger sample size not only meets the study's statistical power needs but also contributes to heightened result reliability.

All statistical analyses were performed using R statistical software (version 4.1.2), which was described in a previous study by Meng et al, 2021 (22).

## 3 Results

### 3.1 Characteristics of enrolled participants

We prospectively enrolled 144 treatment-naïve lymphoma patients and 396 healthy individuals in this study. The mean ages were 51.2 years in the lymphoma group and 46.9 years in the healthy group ( $p < 0.01$ , Table 1). Sub-classification of lymphoma patients was summarized in Table 1 (Table 1). There was no significant difference between the two groups in gender ( $p = 0.144$ ). As for lymphoma classification, except for 14 HL patients, 130 (90.3%) lymphoma patients were non-Hodgkin lymphoma (NHL), 76.2% of which originated from B cells and 23.8% from T/NK cells. Among B-cell NHLs, 79.8% were aggressive types,

including diffuse large B-cell lymphoma (DLBCL), accounting for 82.3% of the total subgroup. The detailed information of each sample was shown in Supplementary Table 1.

### 3.2 Protein tumor markers for lymphoma diagnosis

Protein tumor markers have been used for decades to aid in the diagnosis and management of a variety of cancers. Here, seven common plasma protein tumor markers (AFP, CA125, CA15-3, CA19-9, CA72-4, CEA, and CYFRA21-1) were selected to estimate the expression in each lymphoma patient and healthy individual (18). As shown in Figure 1A, only the expression of CA125 and CA153 in lymphoma was significantly higher than that in the healthy group (Wilcoxon rank-sum test,  $P < 0.001$ ). However, using receiver operating characteristic (ROC) analysis to evaluate these proteins' performance in differentiating lymphoma from healthy subjects (Figure 1B), only CA125 achieved a moderate AUC (area under the curve) of 0.702. Based on the clinical cut-off value: 35 U/ml, CA125 positive ratio was only 2.3% in healthy individuals and 27.8% in lymphoma patients (24.7% in male and 31.3% in female lymphoma respectively) (Supplementary Table 2), which indicated CA125 could be used as a biomarker for lymphoma detection, regardless of gender. As for the different subtypes of lymphoma, CA125 positive ratio was 42.9% in HL. In NHL, CA125 positive proportion of indolent B-cell lymphoma was 40.0%, which was higher than aggressive B-cell lymphoma (26.6%). The positive ratio for T/NK cell lymphoma was the least at 16.1% (Figure 1C).

### 3.3 Distinct CNA patterns across diverse lymphoma subtypes

Since somatic CNA has been widely detected in most cancer patients, sWGS was used to depict the CNA profile of lymphoma patients and healthy subjects. 88 (61.1%) patients were CNA positive, however only 5 (1.3%) healthy subjects were CNA positive, which suggested that CNA might be an efficient feature for lymphoma diagnosis with high specificity of 98.7% (Supplementary Figure 2A). Especially for HL, 85.7% of patients were CNA positive. Aggressive B-cell NHL patients had the second highest CNA mutational frequency. Among them, DLBCL and Non-DLBCL had similar CNA mutational frequency (64.6% vs 71.4%). The CNA positive ratio was lowest in T/NK-cell NHL (41.9%).

Segmented CNA profiles of each lymphoma patient was shown in Supplementary Figure 2B. The heatmap showed that aggressive B-cell lymphoma (including DLBCL and non-DLBCL) had a significantly higher absolute value of log R ratio of CNA segments than the other subtypes, which meant aggressive B-cell patients had a higher concentration of ctDNA in plasma. GISTIC analysis was used to find the significant recurrent CNA peaks in lymphoma and the genomic region mutational frequency across different lymphoma subtypes. In all the patients, significant mutational focal deletion peaks (1p36.32, 3p21.31, 4q21.3, 4q35.2,

**TABLE 1** Demographics and clinical information of lymphoma patients and healthy individuals.

	Lymphoma (n = 144)	Healthy (n = 396)	p value*
Age(mean ± SD)	51.2 ± 16.4	46.9 ± 13.1	0.005
<45	49 (34.0%)	160 (40.4%)	<0.001
45~55	26 (18.1%)	118 (29.8%)	
55~65	37 (25.7%)	78 (19.7%)	
>65	32 (22.2%)	40 (10.1%)	
Gender			
F	67 (46.5%)	214 (54.0%)	0.144
M	77 (53.5%)	182 (46.0%)	
Stage			
I	19 (13.2%)		
II	20 (13.9%)		
III	15 (10.4%)		
IV	65 (45.1%)		
NA	25 (17.4%)		
Major subtype			
HL	14 (9.7%)		
NHL	130 (90.3%)		
NHL Cell Origin			
T/NK-cell	31 (23.8%)		
B-cell	99 (76.2%)		
B-cell NHL aggressive or not			
No ( <i>Indolent</i> )	20 (20.2%)		
Yes ( <i>Aggressive</i> )	79 (79.8%)		
DLBCL or not in aggressive B-cell NHL group			
No	14 (17.7%)		
Yes	65 (82.3%)		

Values are n (%).

\*Student's t-test was used to compare mean age difference between healthy and lymphoma, and Fisher's exact test was used to estimate the difference of age and gender compositions between those two groups; SD, standard deviation; HL, Hodgkin lymphoma; NHL Non-Hodgkin lymphoma.

6q23.3, 6q26, 8p23.3, and 15q12) and broad amplification regions (1q, 3p, 3q, 5p, 5q, 7p, 7q, 9p, 12p, 12q, 18q, 19q, and 21q) were identified (Figure 2A, Supplementary Table 3). The CNA mutational pattern of each lymphoma subtype with depicted in Figure 2B. Among HL patients, chromosomal gains most frequently involved 2p, 9, 14, and 19, whereas losses primarily affected chromosomes 1p, 4q, and 13q. Amplification of 1q, 3, 7, 11q, 12, 18q and deletion of 1p, 6q, 8p, and 15p was frequently mutated in the DLBCL, which was similar to the Non-DLBCL CNA mutational

frequency. T/NK-cell NHL did not have frequent mutational regions, except for amplification of 1q and 5. CNA pattern differences among these major subtypes might contribute to identifying the lymphoma subtypes.

### 3.4 Characterizing aberrant fragment size distribution of cfDNA in lymphoma patients

It is well known that cfDNA from cancer patients had significantly more short fragments than that from healthy individuals, which could be used for cancer prediction (21). However, there are few published studies about the FS profile in lymphoma. In this study, the fragment size distribution in both lymphoma and healthy individuals was shown in Figure 3A. Similar to the FS distribution of solid tumors, lymphoma patients had a higher proportion of short fragments (50~150 bp, P150) than healthy subjects and the same profile of 10-periodic peaks (21, 22). Meanwhile, the proportion of long fragments (180~220 bp) in lymphoma patients was lower than that in healthy control. Based on these results, we developed a new index of FS (named FS.S-L, Materials and Methods), which had a high correlation with the previous FS method P150 (Pearson: 0.985) and had a better performance for cancer prediction. At the same specificity of 98.0%, FS.S-L predicted 6.9% more lymphoma patients than the P150 (Figure 3B), and FS.S-L index (AUC = 0.837) also has a significantly better performance ( $p < 0.0001$  by DeLong's test) to discriminate lymphoma and healthy individual than P150 (AUC = 0.800) (Figure 3C). The value of FS.S-L in each lymphoma subtype was also significantly higher than that in healthy individuals by the Wilcoxon rank-sum test ( $p < 0.001$ ) (Figure 3D), which suggested that FS.S-L was suitable for all the lymphoma detection, regardless of subtypes.

### 3.5 Differential end motif proportions between lymphoma patients and healthy individuals

The cfDNA end motif was identified using the first 4bp sequence on each 5' end of cfDNA fragment after alignment to the reference genome (Figure 4A). 189 out of 256 end motifs with a different frequency between lymphoma patients and healthy individuals were found by the Wilcoxon rank-sum test and p-value adjust method (FDR). For example, the frequency of motif ATAG showed a significant increase in lymphoma patients, while the frequency of motif CGAG was significantly lower in lymphoma patients (Figure 4B, Materials and Methods). To minimize the issue of overfitting, LASSO was used to select the end motifs and reduce the number of features. Finally, 24 end motifs were retained in the feature selection process by LASSO after repeating 30 times. Hierarchical clustering analysis was used to identify the different characteristics of selected end motifs between the cancer patients

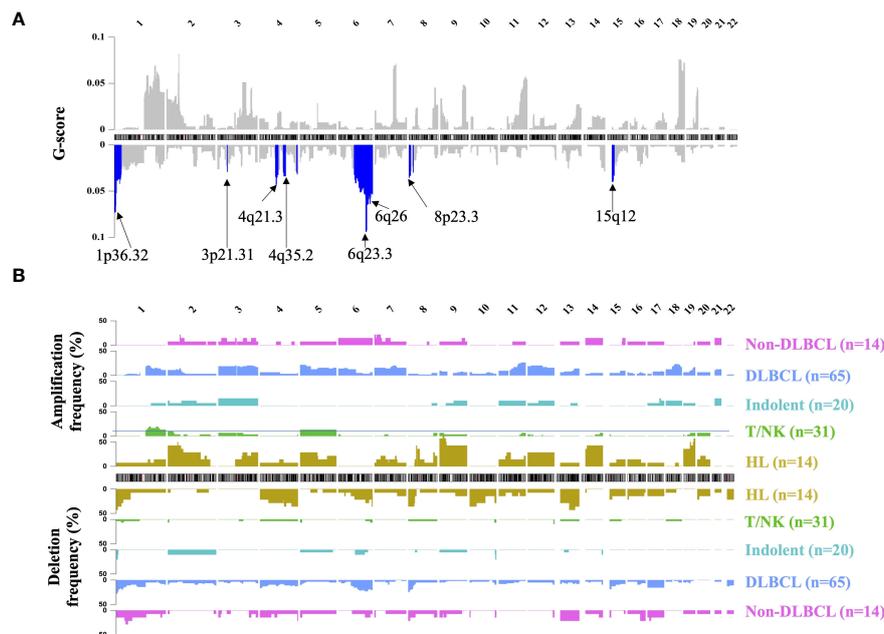


FIGURE 2

Landscape of genomic CNA detection in ctDNA from lymphoma patients. Chromosomal overview of focal recurrent amplification (upper) and deletion (lower) regions was shown in (A), and the G-score was calculated by the GISTIC analysis. The blue region represented the significant focal deletion region (q-value = 0.25) with the annotated position. Fractions of aberrant samples across genomic loci and different lymphoma subtypes were shown in panel (B) HL, Hodgkin lymphoma; T/NK, T/NK cell lymphoma; Indolent, indolent B-cell lymphoma; DLBCL, diffuse large B-cell lymphoma.

and healthy controls (Figure 4C). Heatmap formed different clusters between lymphoma patients and healthy subjects and showed that the frequency of selected end motifs in lymphoma patients was quite distinct from that in healthy individuals. Use the frequency of selected end motifs to build the lymphoma predictive model by SVM, and the predicted score of healthy subjects and lymphoma patients across different subtypes was shown in Figure 4D. All the lymphoma subtypes had a significantly higher predicted score than healthy controls.

### 3.6 Quantitative detection of EBV by the sWGS

In addition to the lymphoma specific ctDNA genomic feature, quantification of circulating EBV DNA was useful for diagnosis, monitoring, and prognostication of lymphoma, especially for HL and T/NK-cell NHL patients (38–40). Based on the sWGS data, the proportions of circulating EBV reads in lymphoma patients and healthy individuals were shown in Supplementary Figure 3. 42.9% HLs and 54.8% T/NK-cell NHLs were EBV positive, with more than 0.5 reads per million. However, the plasma EBV proportions of all healthy individuals were far less than 0.5 reads/M and the maximum was 0.16 reads/M. EBV status of other lymphoma subtypes (Indolent, DLBCL, and non-DLBCL) was similar to that of healthy subjects. 21 lymphoma patients also accepted EBV detection by real-time PCR in blood samples and *in situ*

hybridization in tissue samples. 5 patients with EBV negative in tissue were also negative in blood, regardless of method. 15 out of 16 patients (93.8%) with EBV positive in tissue were also EBV positive detected by sWGS in blood, but only 5 patients (31.3%) were detected by the real-time PCR method in the blood (Supplementary Table 4). These findings suggested that our assay was more sensitive for EBV detection in blood than the real-time PCR method, which could be used for lymphoma detection.

### 3.7 The blood-based multidimensional assay achieved optimal accuracy in lymphoma detection

Based on the sample source, samples were divided into training cohort and validation cohort. 45 patients and 99 healthy individuals were enrolled from different affiliated hospitals of Sun Yat-sen University were used as independent validation cohort, the remaining samples were used as training cohort to build the lymphoma classifier model, which achieved AUC of 0.937. SeekInCare performance in the validation cohort mirrored that of the training cohort, with an AUC of 0.964 (Delong's test  $p = 0.3$ , Figure 5A), affirming its generalizability. Therefore, we amalgamated all samples for further assessment of each feature's independent performance in distinguishing lymphoma patients from healthy individuals via ROC analysis and calculated the CRS value. CNA made the largest contribution to the model, which was

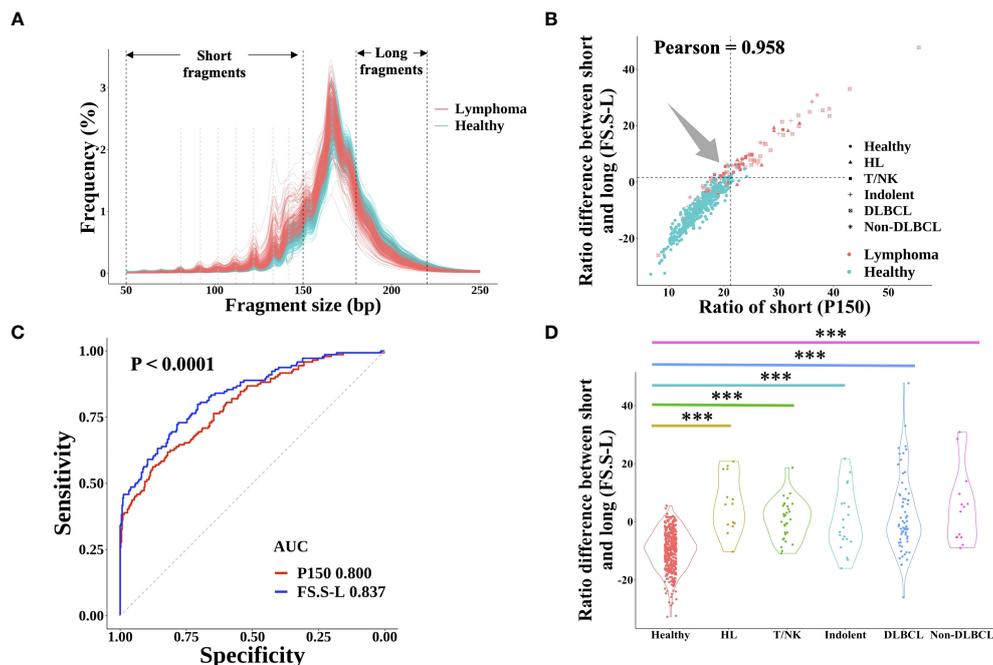


FIGURE 3

cfDNA fragment size distribution differentiated between lymphoma patients and healthy subjects. **(A)** The fragment size distribution of cfDNA was determined by sWGS of plasma samples from lymphoma patients and healthy subjects. The black dotted line showed the range of short fragments (50~150bp) and long fragments (180bp~220bp). The grey dotted line showed the location of 10-periodic peaks (81bp, 92bp, 102bp, 112bp, 122bp, 133bp, and 142bp). The correlation between the proportion of short fragments (defined P150) and the proportion difference between short and long fragments (defined FS,S-L) was shown in **(B)** The dotted line represented the cut-off value at 98% quantiles in the healthy group. **(C)** ROC curve of P150 and FS,S-L showed the performance for lymphoma prediction. DeLong's test was used to show that the AUC of FS,S-L (blue) was significantly better than the AUC of P150 (red) with  $p < 0.0001$ . **(D)** Compare the value of FS,S-L between healthy individuals and different lymphoma subtypes. Statistical significance of the increasing trend of P150 for each lymphoma subtype was determined by the Wilcoxon rank-sum test ( $***p < 0.001$ ). HL, Hodgkin lymphoma; T/NK, T/NK cell lymphoma; Indolent, indolent B-cell lymphoma; DLBCL, diffuse large B-cell lymphoma.

the highest sensitivity of all the five variables at specificity of 98.0% followed by the FS, end motif, EBV and CA125 (Figure 5B). Pearson correlation among these five features was shown in Supplementary Figure 4A. Apart from a correlation of 0.57 between CNA and FS,S-L, and a correlation of 0.55 between FS,S-L and the end motif, there was no correlation among the other features, CA125, end motif, and EBV. SeekInCare detected 112 out of 144 lymphoma patients [77.8% (95% CI, 70.1%~84.3%)] at 98.0% specificity (95% CI, 96.1%~99.1%) (Supplementary Table 5), and the overall accuracy is 92.6% (95% CI, 90.0%~94.7%). Both the positive predictive value (PPV) and negative predictive value (NPV) were more than 92%. Meanwhile, the AUC of CRS was 0.947 (Figure 5B) and the AUC of precision-recall curve was 0.920 (Supplementary Figure 4B). The sensitivity of lymphoma detection was increased with the advancement of stages. While the sensitivity of early stage was 51.3% (47.4% (95% CI, 24.4%~71.1%) and 55.0% (95% CI, 31.5%~76.9%) for stage I and II respectively), and the sensitivity of advanced stage was 90.0% (stage III: 86.7% (95% CI, 59.5%~98.3%); Stage IV: 92.3% (95% CI, 83.0%~97.5%)), with high specificity of 98.0% (Figure 5C). Among lymphoma subtypes, all the HL patients were successfully detected, including 33.3% of early stage HLs. For the aggressive B cell lymphoma, including both DLBCL and non-DLBCL with a sensitivity of 72.3% (95% CI, 59.8%~82.7%) and 85.7% (95% CI, 57.2%~98.2%), respectively (Figure 5D). These results suggested that SeekInCare was an

efficient non-invasive assay for lymphoma detection with high accuracy.

### 3.8 Dynamic changes of blood-based assay correlate with treatment outcome

We explored the CRS dynamic changes after lymphoma related treatment. Blood samples were collected from three patients with stage IV lymphoma prior to treatment and after completing two treatment cycles. Notably, the amplitude of CNAs in patient ZL0019, exhibited a significant increase in regions such as 1q, 9p, and X (Figure 6A). In contrast, CNAs that were initially detected in the samples from patients ZL0020 and ZL0060 prior to treatment had disappeared, with no further detection of CNAs in their plasma after treatment. Concurrently, the other dimensions, including FS,S-L, CA125, end motif, and EBV, exhibited the changes consistent with the observed changes of CNAs. Specifically, The value of these dimensions increased in ZL0019 post-treatment, while they decreased to normal levels in ZL0020 and ZL0060, except EBV in ZL0020 and FS,S-L in ZL0019 remained negative (Supplementary Figure 5). Furthermore, our analysis revealed that CRS of ZL0019, as predicted by GLM, approached more closely to 1 after treatment (from its pre-treatment value of 0.9979 to 0.9995). Although several individual dimensions exhibited a noteworthy

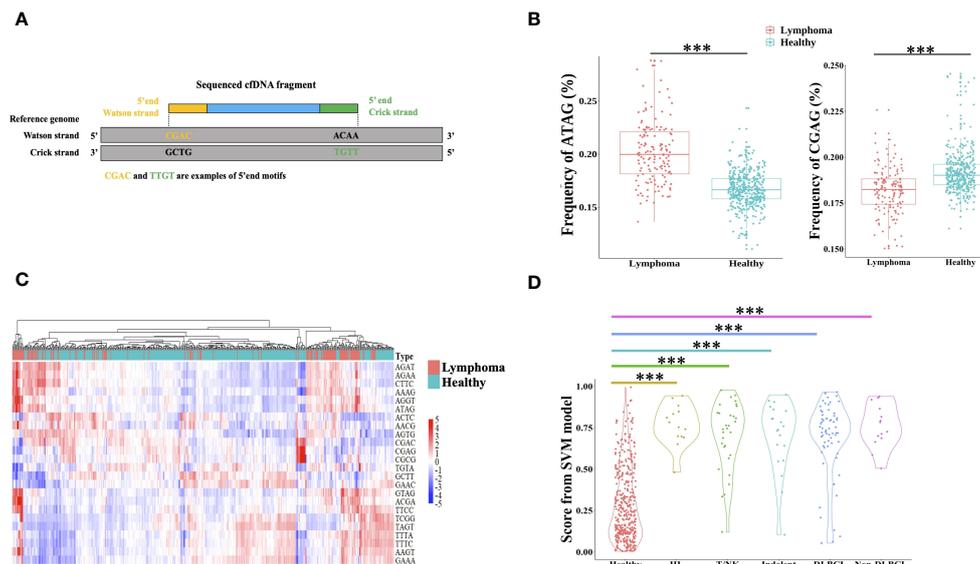


FIGURE 4

Differential end motif profiles between lymphoma patients and healthy individuals. (A) illustrated the determination of plasma DNA end motifs. After aligning the pair-end reads to the reference genome, the first 4-nucleotide sequence on each 5' fragment end (Watson and Crick strands) of one fragment DNA was defined as end motif. As shown in the example, CGAC and TTGT was the example of end motifs. One sample included a total of 256 4-mer motifs. (B) Boxplot showed two examples of the frequency difference of selected end motifs between lymphoma subjects and healthy individuals. (C) Heatmap analyzed the frequencies of 24 selected end motifs between healthy and lymphoma subjects, and samples were clustered based on the frequency of selected end motifs. The data are row normalized. (D) Violin plot showed the lymphoma predicted score by support vector machine. HL, Hodgkin lymphoma; T/NK, T/NK cell lymphoma; Indolent, indolent B-cell lymphoma; DLBCL, diffuse large B-cell lymphoma. \*\*\*, Wilcoxon rank-sum test  $P$ -value < 0.001.

increase in ZL0019, the nature of the sigmoid curve within the GLM model posed a challenge in precisely quantifying this change (Figure 6B). Additionally, the CRS values for ZL0020 and ZL0060 decreased to the normal level.

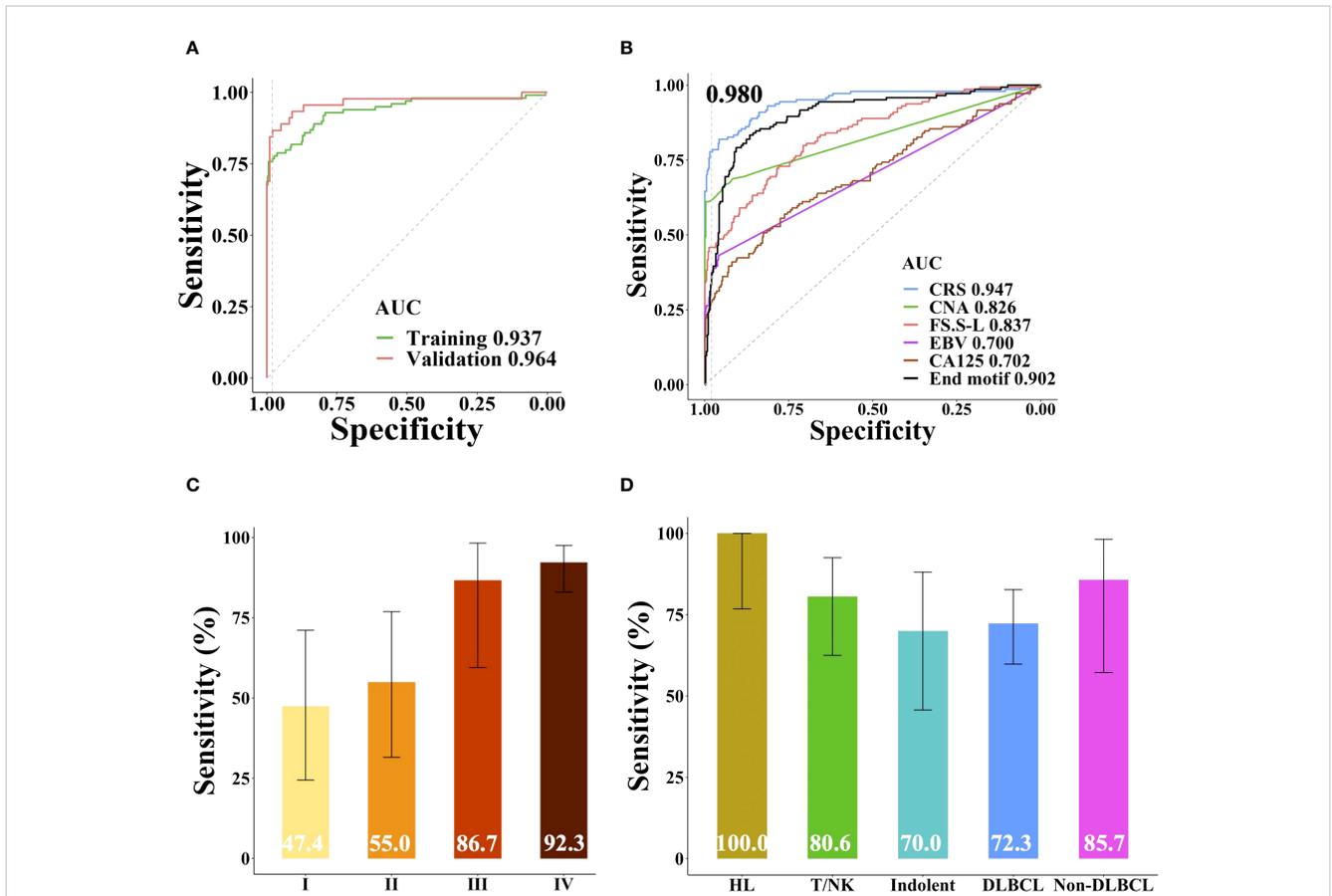
Remarkably, after two treatment cycles, all three patients exhibited a Partial Response (PR) according to image evaluations. However, after four cycles, ZL0019 progressed to Progressive Disease (PD), while ZL0060 achieved a Complete Response (CR), and ZL0020 also reached CR after six treatment cycles (Figure 6C). These findings illustrated the potential of liquid biopsy in promptly assessing the effectiveness of cancer treatment, as they align with clinical evaluations following two treatment cycles.

## 4 Discussion

Lymphomas, encompassing diverse neoplasms arising from B lymphocytes, T lymphocytes, and NK cells, are characteristic tumors surrounded by an inflammatory microenvironment. Among them, DLBCL constitutes a major portion of NHLs, displaying high clinical and biological heterogeneity because it arises from germinal center B cells at different stages of differentiation (41). Its classification is intricate, constantly evolving due to heterogenic variations in morphology, phenotype, genetic anomalies, prognosis, and clinical features (42). The tumor microenvironment (TME) is a complex composed environment enveloping cancer cells, comprising both cellular and extracellular elements, as well as a vascular network (43, 44). TME components

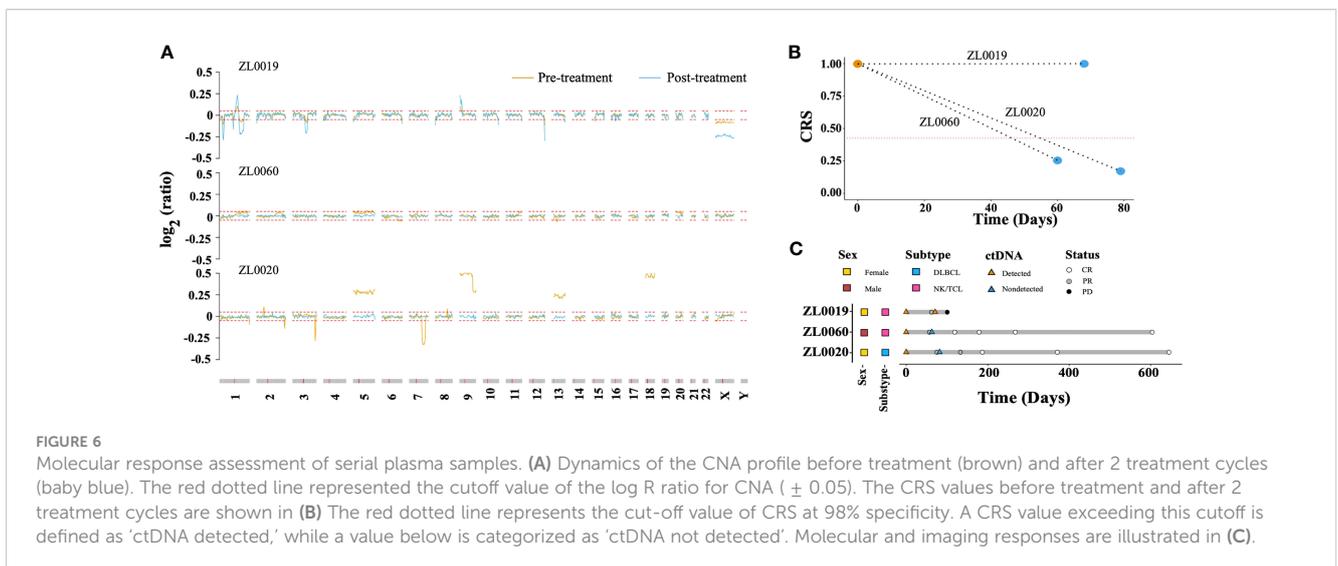
play pivotal roles in both initiating and sustaining carcinogenesis (45). The composition of TME cells holds relevance for the prognosis of B-cell lymphomas, including cHL, DLBCL and FL (46). It plays a pivotal role in various biological processes such as pathogenesis, progression, metastasis and drug resistance, facilitating sustained proliferation and immune escape (47, 48). Tumor progression is profoundly shaped by the TME. Via various pathways, tumor cells adeptly recruit stromal cells, providing tumor cell growth signals, intermediate metabolites, and a conducive environment for tumor progression and metastasis (49). Angiogenic factors, like vascular endothelial growth factor and its receptors, along with other TME components, are crucial in the progression and maintenance of lymphoproliferative disorders (50). Increased interfollicular microvascular density in FL patients predicts inferior overall survival and an increased transformation to DLBCL (51, 52). Antiangiogenic therapy has emerged as a vital tool for lymphoma treatment (53). In the bone marrow microenvironment, interactions with stromal cells and the extracellular matrix contribute to cell adhesion-mediated drug resistance, impacting chemotherapy efficiency and prognosis in B-NHL (54). Knockdown of small glutamine-rich TPR-containing protein A (SGTA) in Non-Hodgkin's Lymphomas induced CAM-DR, offering the potential for a novel therapeutic approach for CAM-DR in NHL (55).

Lymphoma is often diagnosed at an advanced stage when it has caused symptoms, and the diagnosis and classification are mainly based on the excisional or punctured tissues. It has been challenging to develop a more sensitive and specific method of detecting



lymphoma when it was still asymptomatic. With the advance of 'liquid biopsy' and next-generation sequencing technology, we developed a blood-based non-invasive assay, SeekInCare, which integrated the protein marker and multidimensional cancer

genomic features, such as CNA, FS, end motif, and oncogenic viral DNA. In this study, the SeekInCare assay has achieved an impressive 77.8% sensitivity with the high specificity of 98.0%, and 92.6% accuracy, while its NPV and PPV were both more than 92%



for lymphoma detection. The assay also performed well in detection of early stage lymphoma as the sensitivity of early stage was 51.3% (47.4% and 55.0% for stage I and II respectively).

CA125 is used to monitor certain cancers during and after treatment, especially ovarian cancer in symptomatic women presenting to primary care (56). Interestingly, among 7 commonly used plasma protein tumor markers for cancer, we found that only CA125 could be effectively used for lymphoma detection. It is especially effective in HL with a sensitivity of 42.9% and high specificity of 98.0%, compared to a moderate sensitivity of 27.8% at the same specificity for all the lymphoma patients in this study. Moreover, CA125 for lymphoma detection had the same performance between female and male patients, suggesting that CA125 is not female-specific as we may think in ovarian cancer. Therefore, our study suggests that CA125 can be a potential biomarker in lymphoma, especially in HL, for diagnosis.

The CNA analysis demonstrated that aggressive B-cell patients had a higher concentration of ctDNA in plasma which might be the reason for aggressive B-cell lymphoma patients who usually had a poor prognosis (57). We also found CNA from our assay could also reveal some lymphoma critical genes, such as *TNFRSF14* in 1p36.32, *TNFAIP3* in 6q23.3, and *SETD2* in 3p21.31. These cancer genes play an important role in immune escape and apoptotic response, which might benefit the strategy for lymphoma therapy and predict the prognosis. Specifically, mutations and deletions in the *TNFRSF14* gene are common in follicular lymphoma (FL), increasing the ability of lymphoma cells to stimulate allogeneic T-cell responses. FL patients with *TNFRSF14* aberrations may benefit from more aggressive immunosuppression to reduce harmful graft-versus-host disease after transplantation (58). Loss of *TNFAIP3* enhances MYD88<sub>L265P</sub>-driven signaling in non-Hodgkin lymphoma, presenting a potential opportunity for therapeutic targeting (59). Each subtype had a different CNA positive ratio and some unique frequently amplified/deleted regions, such as HL are frequently amplified at 2p, 9, 14, 19 and deleted at 1p, 4q, 13q. These genomic regions were also reported to be frequently affected by CNA in previous HL studies (60, 61). In addition to lymphoma diagnosis, the CNA profiles could be used to identify lymphoma subtypes and predict prognosis to some extent.

Similar fragmentation patterns were found in lymphoma as in solid tumors, such as HCC, showing the 10-bp periodic peaks located between 80 and 150 bp size ranges as previously reported (22). We have tried several methods to estimate the difference in FS between lymphoma patients and healthy individuals. We found a method to calculate the FS.S-L, which had a better performance than the previously published method P150.

As cleavage and fragmentation of cfDNA are nonrandom processes, cleavage site preference can be associated with tissue sources, disease status, chromatin accessibility, and nuclease activities (16, 33, 62). Thus, we were the first to depict the fragment-omics character of lymphoma based on end motifs and found a specific cluster of end motif frequency. Just like the HCC study (63), we also demonstrated that the end motif could be used to differentiate lymphoma from healthy subjects, which achieved an AUC of 0.902.

Quantitative detection of EBV by sWGS in blood achieved high sensitivity (93.8%) at 100% specificity, which was much more

sensitive than real-time PCR (31.3% sensitivity). One fact that people with EBV positive might be caused by other diseases (e.g., ongoing infectious mononucleosis) irrespectively of lymphoma. Nonetheless, the combination of EBV quantification and typical lymphoma alterations of other dimensions (CNA/end motif) strongly indicates an EBV-driven lymphoma. Meanwhile, EBV was also used as a molecular biomarker for disease monitoring and prognosis prediction in EBV positive patients, since pathogenesis and genetics were different between EBV-positive patients and EBV-negative patients, especially in HL, Burkitt lymphoma and some subtypes of T/NK-cell NHL (38, 39, 64). Even more, the SeekInCare assay was not only more sensitive to monitoring EBV expression in the blood than the real-time PCR method, but also reflected the dynamic change of lymphoma-associated mutations (CNA).

While the blood samples were collected from only three patients after two treatment cycles, this nevertheless underscores the potential of plasma-based molecular dynamics in promptly and accurately assessing treatment effectiveness. This finding is consistent with previous studies that demonstrated ctDNA levels after two treatment cycles serving as prognostic factors, with ctDNA dynamics significantly associated with clinical outcomes in DLBCLs (57, 65). This capability is particularly crucial in distinguishing genuine treatment responses from pseudo-responses (66).

This study had some limitations. More prospectively collected samples from multiple centers or real-world scenarios are needed to validate the generalizability of SeekInCare. Even though the assay demonstrated with high specificity of 98.0%, there are still 2.0% positive that require those individuals to go through routine clinical workup for confirmation including CT scan and/or biopsy. SeekInCare for lymphoma detection utilized common cancer hallmarks, not specific characteristics of lymphoma (such as AFP for HCC). We found the blood markers associated with lymphoma tumor burden including lactate dehydrogenase (LDH) and beta 2 microglobulin ( $\beta$ 2M) were also informative with 38.6% positivity for LDH and 30.8% for  $\beta$ 2M based on the clinical cut-offs in the current study. We could build a baseline in a healthy cohort to optimize the specificity of LDH and  $\beta$ 2M combining with the features of SeekInCare for lymphoma early detection in a future study.

Recent studies have demonstrated that fragment-omics features from sWGS could be used to predict tissue origin of cancer (21, 67), and it would be optimized in future studies. Therefore, lymphoma would be differentiated from other cancers through the TOO algorithm and lymphoma patients could benefit from blood-based non-invasive early detection. Indeed we were able to identify nine lymphoma patients out of total 41 cancer cases in our proof-of-concept study utilizing CNAs and the seven protein markers (18). SeekInCare could be used as a complementary test to medical imaging to distinguish the benign and malignant lymph nodes, especially for patients with that affected nodes were difficult to reach, or tissue-based subtyping was inconclusive. In summary, a non-invasive assay (SeekInCare) was validated in our case-control study, achieving an AUC of 0.947 for lymphoma detection, which also provided insights into lymphoma monitoring and multi-cancer early detection.

## Data availability statement

The raw data for both healthy individuals and lymphoma patients are accessible with controlled access through the Chinese National Genomics Data Center under PRJCA022396 (<https://ngdc.cncb.ac.cn/bioproject/browse/PRJCA022396>).

## Ethics statement

The studies involving humans were approved by The First Affiliated Hospital of Zhengzhou University. 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

YC: Data curation, Resources, Writing – review & editing. SL: Formal Analysis, Writing – original draft. ZL: Data curation, Resources, Writing – review & editing. XW: Resources, Supervision, Writing – review & editing. FC: Project administration, Writing – review & editing. SG: Formal Analysis, Writing – review & editing. DZ: Formal Analysis, Methodology, Writing – review & editing. GZ: Writing – review & editing. WW: Formal Analysis, Writing – review & editing. YYC: Investigation, Project administration, Writing – review & editing. ST: Writing – review & editing. MM: Funding acquisition, Methodology, Supervision, Writing – review & editing.

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This study was supported by National Natural Science Foundation of China (82000204), and Science and Technology Project of Henan Province (232102311114).

## Acknowledgments

The authors thank all the participating patients, and investigators for their contributions to the study. We also thank Dr. Jun Zhou for helpful comments on the manuscript.

## Conflict of interest

Authors MM, SL, GZ, and WW are full-time employees and stock shareholders of the company SeekIn. DZ and YYC are also the stock shareholders of SeekIn. DZ, SG, FC, YYC, and ST are full-time employees of Clinical Laboratories, Shenyou Bio, a wholly-owned subsidiary of SeekIn Inc.

The remaining 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.2024.1341997/full#supplementary-material>.

### SUPPLEMENTARY FIGURE 1

Schematic of multi-omics and multidimensional SeekInCare approach. One tube (8ml) of peripheral blood was collected from healthy individuals and lymphoma patients. cfDNA was extracted and then analyzed the cancer genomics, epigenetics, and virus infection through shallow whole genome sequencing. Meanwhile, the expression of seven protein tumor markers was also measured in the plasma. Cancer risk score (CRS) was calculated by machine learning method, which integrated CNA, FS, end motif, virus, and protein marker. CNA, copy number aberration, WGS, whole genome sequencing.

### SUPPLEMENTARY FIGURE 2

(A) Violin plot showed the number of CNA bins in healthy individuals and patients across different lymphoma subtypes. Since the y-axis was a logarithmic scale, sample without CNA bins was changed to 0.1. The dotted line indicated the cut-off value of CNA positive. (B) The Heatmap showed the segmented CNA profiles of all lymphoma patients. The chromosomes were arranged vertically from top to bottom, and lymphoma samples were sorted by the genomic region size of CNA among each lymphoma subtype from left to right. Red and blue represent amplification and deletion, respectively. HL, Hodgkin lymphoma; T/NK, T/NK cell lymphoma; Indolent, indolent B-cell lymphoma; DLBCL, diffuse large B-cell lymphoma.

### SUPPLEMENTARY FIGURE 3

The proportion of plasma EBV DNA reads among the total number of sequenced plasma DNA reads in healthy individuals and lymphoma patients. The cut-off value was 0.5 reads/M (dotted line). HL, Hodgkin lymphoma; T/NK, T/NK cell lymphoma; Indolent, indolent B-cell lymphoma; DLBCL, diffuse large B-cell lymphoma.

### SUPPLEMENTARY FIGURE 4

(A) The Pearson correlation between different features among all samples. (B) Precision-recall curve for the CRS values obtained from the SeekInCare assay in the classification of lymphoma patients. CNA, copy-number aberration. FS.S-L, the ratio difference between short and long fragments. EBV, Epstein-Barr virus.

### SUPPLEMENTARY FIGURE 5

The dynamic changes in each feature before and after treatment were illustrated. The dotted line represented the cut-off value set at the 98th percentile within the healthy group.

## References

1. Campo E, Swerdlow SH, Harris NL, Pileri S, Stein H, Jaffe ES. The 2008 WHO classification of lymphoid neoplasms and beyond: evolving concepts and practical applications. *Blood* (2011) 117(19):5019–32. doi: 10.1182/blood-2011-01-293050
2. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *Ca-cancer J Clin* (2021) 71(3):209–49. doi: 10.3322/caac.21660
3. Zheng R, Zhang S, Zeng H, Wang S, Sun K, Chen R, et al. Cancer incidence and mortality in China, 2016. *J Natl Cancer Center* (2022) 2(1):1–9. doi: 10.1016/j.jncc.2022.02.002. S2667005422000047.
4. Matasar MJ, Zelenetz AD. Overview of lymphoma diagnosis and management. *Radiol Clin N Am* (2008) 46(2):175–198, vii. doi: 10.1016/j.rcl.2008.03.005
5. 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
6. Ferry JA. Scientific advances and the evolution of diagnosis, subclassification and treatment of lymphoma. *Arch Med Res* (2020) 51(8):749–64. doi: 10.1016/j.jarmed.2020.05.022
7. Shi Y. Current status and progress of lymphoma management in China. *Int J Hematol* (2018) 107(4):405–12. doi: 10.1007/s12185-018-2404-8
8. McGranahan N, Swanton C. Biological and therapeutic impact of intratumor heterogeneity in cancer evolution. *Cancer Cell* (2015) 27(1):15–26. doi: 10.1016/j.ccell.2014.12.001
9. Gerlinger M, Endesfelder D, Stewart A, Tarpey P, McDonald NQ, Santos CR, et al. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *n Engl J Med* (2012) 366(10):883–92. doi: 10.1056/nejmoa1113205
10. Burrell RA, McGranahan N, Bartek J, Swanton C. The causes and consequences of genetic heterogeneity in cancer evolution. *Nature* (2013) 501(7467):338–45. doi: 10.1038/nature12625
11. Cohen JD, Li L, Wang Y, Thoburn C, Afari B, Danilova L, et al. Detection and localization of surgically resectable cancers with a multi-analyte blood test. *Science* (2018) 359(6378):926–30. doi: 10.1126/science.aar3247
12. Phallen J, Sausen M, Adleff V, Leal A, Hruban C, White J, et al. Direct detection of early-stage cancers using circulating tumor DNA. *Sci Transl Med* (2017) 9(403):eaan2415. doi: 10.1126/scitranslmed.aan2415
13. Crowley E, Di Nicolantonio F, Loupakis F, Bardelli A. Liquid biopsy: monitoring cancer-genetics in the blood. *Nat Rev Clin Oncol* (2013) 10(8):472–84. doi: 10.1038/nrclinonc.2013.110
14. Chan KA, Jiang P, Zheng YW, Liao GJ, Sun H, Wong J, et al. Cancer genome scanning in plasma: detection of tumor-associated copy number aberrations, single-nucleotide variants, and tumoral heterogeneity by massively parallel sequencing. *Clin Chem* (2013) 59(1):211–24. doi: 10.1373/clinchem.2012.196014
15. Molparia B, Nichani E, Torkamani A. Assessment of circulating copy number variant detection for cancer screening. *PLoS One* (2017) 12(7):e0180647. doi: 10.1371/journal.pone.0180647. Galli A, ed.
16. Jin C, Liu X, Zheng W, Su L, Liu Y, Guo X, et al. Characterization of fragment sizes, copy number aberrations and 4-mer end motifs in cell-free DNA of hepatocellular carcinoma for enhanced liquid biopsy-based cancer detection. *Mol Oncol* (2021) 15(9):2377–89. doi: 10.1002/1878-0261.13041
17. Lenaerts L, Vandenberghe P, Brison N, Che H, Neofytou M, Verheecke M, et al. Genomewide copy number alteration screening of circulating plasma DNA: potential for the detection of incipient tumors. *Ann Oncol* (2019) 30(1):85–95. doi: 10.1093/annonc/mdy476
18. Ji X, Li J, Huang Y, Sung PL, Yuan Y, Liu Q, et al. Identifying occult maternal malignancies from 1.93 million pregnant women undergoing noninvasive prenatal screening tests. *Genet Med* (2019) 21(10):2293–302. doi: 10.1038/s41436-019-0510-5
19. Dharajiya NG, Grosu DS, Farkas DH, McCullough RM, Almasri E, Sun Y, et al. Incidental detection of maternal neoplasia in noninvasive prenatal testing. *Clin Chem* (2018) 64(2):329–35. doi: 10.1373/clinchem.2017.277517
20. Moulriere F, Chandrananda D, Piskorz AM, Moore EK, Morris J, Ahlborn LB, et al. Enhanced detection of circulating tumor DNA by fragment size analysis. *Sci Transl Med* (2018) 10(466):eaat4921. doi: 10.1126/scitranslmed.aat4921
21. Cristiano S, Leal A, Phallen J, Fiksels J, Adleff V, Bruhm DC, et al. Genome-wide cell-free DNA fragmentation in patients with cancer. *Nature* (2019) 570(7761):385–9. doi: 10.1038/s41586-019-1272-6
22. Meng Z, Ren Q, Zhong G, Li S, Chen Y, Wu W, et al. Noninvasive detection of hepatocellular carcinoma with circulating tumor DNA features and alpha-Fetoprotein. *J Mol Diagn* (2021) 23(9):1174–84. doi: 10.1016/j.jmoldx.2021.06.003
23. Song CX, Yin S, Ma L, Wheeler A, Chen Y, Zhang Y, et al. 5-Hydroxymethylcytosine signatures in cell-free DNA provide information about tumor types and stages. *Cell Res* (2017) 27(10):1231–42. doi: 10.1038/cr.2017.106
24. Li W, Zhang X, Lu X, You L, Song Y, Luo Z, et al. 5-Hydroxymethylcytosine signatures in circulating cell-free DNA as diagnostic biomarkers for human cancers. *Cell Res* (2017) 27(10):1243–57. doi: 10.1038/cr.2017.121
25. Cai J, Chen L, Zhang Z, Zhang X, Lu X, Liu W, et al. Genome-wide mapping of 5-hydroxymethylcytosines in circulating cell-free DNA as a non-invasive approach for early detection of hepatocellular carcinoma. *Gut* (2019) 68(12):2195–205. doi: 10.1136/gutjnl-2019-318882
26. Chan KCA, Jiang P, Chan CWM, Sun K, Wong J, Hui EP, et al. Noninvasive detection of cancer-associated genome-wide hypomethylation and copy number aberrations by plasma DNA bisulfite sequencing. *Proc Natl Acad Sci* (2013) 110(47):18761–8. doi: 10.1073/pnas.1313995110
27. Liu MC, Oxnard GR, Klein EA, Swanton C, Seiden MV, Liu MC, et al. Sensitive and specific multi-cancer detection and localization using methylation signatures in cell-free DNA. *Ann Oncol* (2020) 31(6):745–59. doi: 10.1016/j.annonc.2020.02.011
28. Chen X, Gole J, Gore A, He Q, Lu M, Min J, et al. Non-invasive early detection of cancer four years before conventional diagnosis using a blood test. *Nat Commun* (2020) 11(1):3475. doi: 10.1038/s41467-020-17316-z
29. Rossi D, Spina V, Brusca G, Gaidano G. Liquid biopsy in lymphoma. *Haematologica* (2019) 104(4):648–52. doi: 10.3324/haematol.2018.206177
30. Desch AK, Hartung K, Botzen A, Brobeil A, Rummel M, Kurch L, et al. Genotyping circulating tumor DNA of pediatric Hodgkin lymphoma. *Leukemia* (2020) 34(1):151–66. doi: 10.1038/s41375-019-0541-6
31. Spina V, Brusca G, Cuccaro A, Martini M, Di Trani M, Forestieri G, et al. Circulating tumor DNA reveals genetics, clonal evolution, and residual disease in classical Hodgkin lymphoma. *Blood* (2018) 131(22):2413–25. doi: 10.1182/blood-2017-11-812073
32. Mermel CH, Schumacher SE, Hill B, Meyerson ML, Beroukhi R, Getz G. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol* (2011) 12(4):R41. doi: 10.1186/gb-2011-12-4-r41
33. Jiang P, Sun K, Peng W, Cheng SH, Ni M, Yeung PC, et al. Plasma DNA end motif profiling as a fragmentomic marker in cancer, pregnancy and transplantation. *Cancer Discovery* (2020) 10(5):664–73. CD-19-0622. doi: 10.1158/2159-8290.cd-19-0622
34. Yu SCY, Jiang P, Peng W, Cheng SH, Cheung YTT, Tse OYO, et al. Single-molecule sequencing reveals a large population of long cell-free DNA molecules in maternal plasma. *PNAS* (2021) 118(50):e2114937118. doi: 10.1073/pnas.2114937118
35. Teschendorff AE. Avoiding common pitfalls in machine learning omic data science. *Nat Mater* (2019) 18(5):422–7. doi: 10.1038/s41563-018-0241-z
36. Ma X, Chen Y, Tang W, Bao H, Mo S, Liu R, et al. Multi-dimensional fragmentomic assay for ultrasensitive early detection of colorectal advanced adenoma and adenocarcinoma. *J Hematol Oncol* (2021) 14(1):175. doi: 10.1186/s13045-021-01189-w
37. Buderer NMF. Statistical methodology: I. Incorporating the prevalence of disease into the sample size calculation for sensitivity and specificity. *Acad Emergency Med* (1996) 3(9):895–900. doi: 10.1111/j.1553-2712.1996.tb03538.x
38. Vrzalikova K, Pugh M, Mundo L, Murray P. The contribution of ebv to the pathogenesis of classical hodgkin lymphoma. *Ann Lymphoma*. (2021) 5:30–0. doi: 10.21037/aol-21-8
39. Wang H, Fu B, Gale RP, Liang Y. NK-/T-cell lymphomas. *Leukemia* (2021) 35(9):2460–8. doi: 10.1038/s41375-021-01313-2
40. Kwong YL, Pang AWK, Leung AYH, Chim CS, Tse E. Quantification of circulating Epstein-Barr virus DNA in NK/T-cell lymphoma treated with the SMILE protocol: diagnostic and prognostic significance. *Leukemia* (2014) 28(4):865–70. doi: 10.1038/leu.2013.212
41. Schneider C, Pasqualucci L, Dalla-Favera R. Molecular pathogenesis of diffuse large B-cell lymphoma. *Semin Diagn Pathol* (2011) 28(2):167–77. doi: 10.1053/j.semdp.2011.04.001
42. Morin RD, Scott DW. DLBCL subclassification: divide and conquer? *Blood* (2020) 135(20):1722–4. doi: 10.1182/blood.2020005335
43. Tamma R, Ranieri G, Ingravalle G, Annesse T, Oranger A, Gaudio F, et al. Inflammatory cells in diffuse large B cell lymphoma. *J Clin Med* (2020) 9(8):2418. doi: 10.3390/jcm9082418
44. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell* (2011) 144(5):646–74. doi: 10.1016/j.cell.2011.02.013
45. Wang L, Ding K, Zheng C, Xiao H, Liu X, Sun L, et al. Detachable nanoparticle-enhanced chemimmunotherapy based on precise killing of tumor seeds and normalizing the growing soil strategy. *Nano Lett* (2020) 20(9):6272–80. doi: 10.1021/acs.nanolett.0c01415
46. Scott DW, Gascoyne RD. The tumour microenvironment in B cell lymphomas. *Nat Rev Cancer*. (2014) 14(8):517–34. doi: 10.1038/nrc3774

47. Ingravallo G, Tamma R, Opinto G, Annese T, Gaudio F, Specchia G, et al. The effect of the tumor microenvironment on lymphoid neoplasms derived from B cells. *Diagnostics (Basel)*. (2022) 12(3):573. doi: 10.3390/diagnostics12030573
48. Sharma P, Hu-Lieskovan S, Wargo JA, Ribas A. Primary, adaptive, and acquired resistance to cancer immunotherapy. *Cell* (2017) 168(4):707–23. doi: 10.1016/j.cell.2017.01.017
49. Yuan Y, Jiang YC, Sun CK, Chen QM. Role of the tumor microenvironment in tumor progression and the clinical applications (Review). *Oncol Rep* (2016) 35(5):2499–515. doi: 10.3892/or.2016.4660
50. Ribatti D, Nico B, Ranieri G, Specchia G, Vacca A. The role of angiogenesis in human non-hodgkin lymphomas. *Neoplasia* (2013) 15(3):231–8. doi: 10.1593/neo.121962
51. Farinha P, Kyle AH, Minchinton AI, Connors JM, Karsan A, Gascoyne RD. Vascularization predicts overall survival and risk of transformation in follicular lymphoma. *Haematologica* (2010) 95(12):2157–60. doi: 10.3324/haematol.2009.021766
52. Jørgensen JM, Sørensen FB, Bendix K, Nielsen JL, Olsen ML, Funder AMD, et al. Angiogenesis in non-Hodgkin's lymphoma: Clinico-pathological correlations and prognostic significance in specific subtypes. *Leukemia Lymphoma*. (2007) 48(3):584–95. doi: 10.1080/10428190601083241
53. Ruan J, Hajjar K, Rafii S, Leonard JP. Angiogenesis and antiangiogenic therapy in non-Hodgkin's lymphoma. *Ann Oncol* (2009) 20(3):413–24. doi: 10.1093/annonc/mdn666
54. Zhang PP, Wang YC, Cheng C, Zhang F, Ding DZ, Chen DK. Runt-related transcription factor 2 influences cell adhesion-mediated drug resistance and cell proliferation in B-cell non-Hodgkin's lymphoma and multiple myeloma. *Leuk Res* (2020) 92:106340. doi: 10.1016/j.leukres.2020.106340
55. Wang Y, Huang Y, Xu X, Tang J, Huang X, Zhu J, et al. Expression of small glutamine-rich TPR-containing protein A (SGTA) in Non-Hodgkin's Lymphomas promotes tumor proliferation and reverses cell adhesion-mediated drug resistance (CAM-DR). *Leukemia Res* (2014) 38(8):955–63. doi: 10.1016/j.leukres.2014.05.013
56. Funston G, Hamilton W, Abel G, Crosbie EJ, Rous B, Walter FM. The diagnostic performance of CA125 for the detection of ovarian and non-ovarian cancer in primary care: A population-based cohort study. *PLoS Med* (2020) 17(10):e1003295. doi: 10.1371/journal.pmed.1003295. Shapiro SD, ed.
57. 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
58. Kotsiou E, Okosun J, Besley C, Iqbal S, Matthews J, Fitzgibbon J, et al. TNFRSF14 aberrations in follicular lymphoma increase clinically significant allogeneic T-cell responses. *Blood* (2016) 128(1):72–81. doi: 10.1182/blood-2015-10-679191
59. Wenzl K, Manske MK, Sarangi V, Asmann YW, Greipp PT, Schoon HR, et al. Loss of TNFAIP3 enhances MYD88L265P-driven signaling in non-Hodgkin lymphoma. *Blood Cancer J* (2018) 8(10):97. doi: 10.1038/s41408-018-0130-3
60. Joos S, Menz CK, Wrobel G, Siebert R, Gesk S, Ohl S, et al. Classical Hodgkin lymphoma is characterized by recurrent copy number gains of the short arm of chromosome 2. *Blood* (2002) 99(4):1381–7. doi: 10.1182/blood.V99.4.1381
61. Steidl C, Telenius A, Shah SP, Farinha P, Barclay L, Boyle M, et al. Genome-wide copy number analysis of Hodgkin Reed-Sternberg cells identifies recurrent imbalances associated with hepatocellular carcinoma. *Proc Natl Acad Sci USA* (2010) 116(3):418–27. doi: 10.1182/blood-2009-12-257345
62. Jiang P, Sun K, Tong YK, Cheng SH, Cheng THT, Heung MMS, et al. Preferred end coordinates and somatic variants as signatures of circulating tumor DNA associated with hepatocellular carcinoma. *Proc Natl Acad Sci USA* (2018) 115(46):E10925–33. doi: 10.1073/pnas.1814616115
63. Chen L, Abou-Alfa GK, Zheng B, Liu JF, Bai J, Du LT, et al. Genome-scale profiling of circulating cell-free DNA signatures for early detection of hepatocellular carcinoma in cirrhotic patients. *Cell Res* (2021) 31(5):589–92. doi: 10.1038/s41422-020-00457-7
64. Vockerodt M, Yap LF, Shannon-Lowe C, Curley H, Wei W, Vrzalikova K, et al. The Epstein-Barr virus and the pathogenesis of lymphoma: EBV and lymphomas. *J Pathol* (2015) 235(2):312–22. doi: 10.1002/path.4459
65. Li M, Mi L, Wang C, Wang X, Zhu J, Qi F, et al. Clinical implications of circulating tumor DNA in predicting the outcome of diffuse large B cell lymphoma patients receiving first-line therapy. *BMC Med* (2022) 20(1):369. doi: 10.1186/s12916-022-02562-3
66. Thompson JC, Carpenter EL, Silva BA, Rosenstein J, Chien AL, Quinn K, et al. Serial monitoring of circulating tumor DNA by next-generation gene sequencing as a biomarker of response and survival in patients with advanced NSCLC receiving pembrolizumab-based therapy. *JCO Precis Oncol* (2021) 5(5):510–24. doi: 10.1200/PO.20.00321
67. Ulz P, Perakis S, Zhou Q, Moser T, Belic J, Lazzeri I, et al. Inference of transcription factor binding from cell-free DNA enables tumor subtype prediction and early detection. *Nat Commun* (2019) 10(1):4666. doi: 10.1038/s41467-019-12714-4