Prognostic implications of 5-hydroxymethylcytosines from circulating cell-free DNA in diffuse large B-cell lymphoma

Brian C.-H. Chiu,[1,*] Zhou Zhang,[2] Qiancheng You,[3] Chang Zeng,[2] Elizabeth Stepniak,[4] Paige M. Bracci,[5] Kangkang Yu,[3] Girish Venkataraman,[6] Sonali M. Smith,[7] Chuan He,[3,8,*] and Wei Zhang[2,9,*]

[1]Department of Public Health Sciences, University of Chicago, Chicago, IL; [2]Department of Preventive Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL; [3]Department of Chemistry and [4]Comprehensive Cancer Center, University of Chicago, Chicago, IL; [5]Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, CA; [6]Department of Pathology and [7]Section of Hematology/Oncology, Department of Medicine, and [8]Department of Biochemistry and Molecular Biology, Institute for Biophysical Dynamics, and Howard Hughes Medical Institute, University of Chicago, Chicago, IL; and [9]The Robert H. Lurie Comprehensive Cancer Center, Northwestern University Feinberg School of Medicine, Chicago, IL

Key Points

- Genome-wide 5hmC loci can be profiled in 1 to 2 ng of cfDNA from blood plasma and correlate with clinical features of DLBCL.

- 5hmC in cfDNA collected at the time of DLBCL diagnosis is associated with EFS and OS, independent of established prognostic factors.

An elevated level of circulating cell-free DNA (cfDNA) has been associated with tumor bulk and poor prognosis in diffuse large B-cell lymphoma (DLBCL), but the tumor-specific molecular alterations in cfDNA with prognostic significance remain unclear. We investigated the association between 5-hydroxymethylcytosines (5hmC), a mark of active demethylation and gene activation, in cfDNA from blood plasma and prognosis in newly diagnosed DLBCL patients. We used 5hmC-Seal, a highly sensitive chemical labeling technique, to profile genome-wide 5hmC in plasma cfDNA from 48 DLBCL patients at the University of Chicago Medical Center between 2010 and 2013. Patients were followed through 31 December 2017. We found a distinct genomic distribution of 5hmC in cfDNA marking tissue-specific enhancers, consistent with their putative roles in gene regulation. The 5hmC profiles in cfDNA differed by cell of origin and were associated with clinical prognostic factors, including stage and the International Prognostic Index. We developed a 29 gene–based weighted prognostic score (wp-score) for predicting event-free survival (EFS) and overall survival (OS) by applying the elastic net regularization on the Cox proportional-hazards model. The wp-scores outperformed (eg, prognostic accuracy, sensitivity, specificity) established prognostic factors in predicting EFS and OS. In multivariate Cox models, patients with high wp-scores had worse EFS (hazard ratio, 9.17; 95% confidence interval, 2.01-41.89; $P = .004$) compared with those in the low-risk group. Our findings suggest that the 5hmC signatures in cfDNA at the time of diagnosis are associated with clinical outcomes and may provide a novel minimally invasive prognostic approach for DLBCL.

Introduction

Diffuse large B-cell lymphoma (DLBCL) is a heterogeneous group of malignancies with distinct genetic abnormalities, molecular alterations, clinical features, and prognosis.[1] Despite improved chemoimmuno-therapies, ~20% to 40% of patients will experience disease recurrence or mortality.[2,3] Emerging evidence suggests that elevated levels of tumor-derived circulating cell-free DNA (cfDNA) in DLBCL correlate with poor prognosis[4] and detect relapse months prior to clinically detectable disease by imaging.[5,6] However, the tumor-specific molecular targets in cfDNA with prognostic value remain largely unknown.

The pathogenesis of DLBCL is strongly linked to perturbation of epigenetic mechanisms. Greater epigenetic heterogeneity,[7,8] global hypomethylation,[9] and aberrant gene-specific promoter methylation[10-12] have been linked with poorer survival and relapse. However, previous studies have only investigated 5-methylcytosines (5mC) or interpreted all modified cytosines as 5mC. In the human genome, 5mC can be oxidized by the human TET enzymes to 5-hydroxymethylcytosines (5hmC) in an active DNA-demethylation process.[13,14] Although 5mC is typically associated with suppressed gene expression,[15] 5hmC is particularly enriched in gene bodies and enhancers that mark for specific gene/locus activation in the chromatin.[16,17] The 5hmC levels change in tumors, and sustained loss has been associated with prognosis.[18-20] Because 5mC represses protein-coding genes, as well as a vast amount of transposons in the human genome, targeting 5hmC for prognostication could better reflect gene-activation changes and a greater specificity. However, because of the low abundance of 5hmC loci in the genome (~0.5%-1% of CpG sites are hydroxymethylated vs 2%-8% for that of 5mC) and difficulties in distinguishing 5hmC from 5mC using conventional bisulfite conversion approaches,[21] no study has evaluated 5hmC in cfDNA for its prognostic value in DLBCL.

In this study, we applied the 5hmC-Seal, a highly sensitive chemical labeling–based sequencing technology, to profile genome-wide 5hmC in cfDNA from blood plasma of 48 patients with newly diagnosed DLBCL. The 5hmC-Seal technology has been shown to be a robust profiling approach for enriching and quantifying 5hmC-modified DNA fragments with as little as 1 to 2 ng of cfDNA from <5 mL of plasma.[22-24] We tested the hypothesis that 5hmC profiles in cfDNA at the time of diagnosis reflect the clinical characteristics of DLBCL and are associated with survival.

Materials and methods

Study subjects

The overall study design is shown in Figure 1. We prospectively enrolled patients aged 20 years and older who were newly diagnosed with non-Hodgkin lymphoma at the University of Chicago Medical Center from 2010 to 2013. All diagnoses were confirmed by hematopathologists according to the 2008 World Health Organization criteria.[25] Blood samples were drawn from consented patients and processed immediately to separate plasma. For this study, we included only DLBCL patients with blood plasma available. We excluded DLBCL patients with primary central nervous system lymphoma, posttransplantation lymphoproliferative disorder, transformation of a previously diagnosed indolent lymphoma, or with HIV infection. After exclusion, a total of 48 DLBCL patients was included in the cfDNA analysis. This study was approved by the Institutional Review Board at the University of Chicago.

Sample preparation and the 5hmC-Seal profiling

Approximately 2 to 3 mL of frozen plasma from each subject was processed by centrifuging at 1350g for 12 minutes twice and at 13 500g for 12 minutes once, followed by cfDNA extraction (1-2 ng per sample) using the QIAamp Circulating Nucleic Acid Kit (Qiagen). Genomic DNA from cfDNA-paired tumor blocks for 7 patients was isolated (30-50 ng per sample) using a DNeasy Blood & Tissue Kit (Qiagen) and fragmented by sonication. We constructed 5hmC-Seal libraries according to an established

protocol.[22] DNA samples were first repaired and ligated with adaptors. Next, the T4 bacteriophage enzyme β-glucosyltransferase was used to transfer an engineered glucose moiety containing an azide-group to 5hmC in duplex DNA. A biotin tag was added to the azide group using Huisgen cycloaddition ("Click") chemistry. Finally, the 5hmC-containing DNA fragments with biotin tags were captured by avidin beads. The 5hmC-Seal libraries were constructed through polymerase chain reaction amplification and sequenced using an Illumina NextSeq 500 platform (PE38) at the University of Chicago Genomics Core Facility. We randomly labeled the cfDNA samples for the 5hmC-Seal library constructions and sequencing. Technicians were blinded to clinical outcomes. Technical robustness, including reproducibility, of the 5hmC-Seal was demonstrated in our previous study.[22]

### Processing of the 5hmC-Seal data

Bioinformatics processing of the 5hmC-Seal data from cfDNA was described in detail in our previous report.[22] Briefly, raw sequencing reads were trimmed for adaptor sequences using Trimmomatic.[26] Low-quality bases were also trimmed to a minimum length of 30 bp, followed by alignment to the human genome reference (hg19) using Bowtie 2 with the end-to-end alignment mode.[27] Read pairs were concordantly aligned with fragment length ≤500 bp and with average ≤1 ambiguous base and up to 4 mismatched bases per 100-bp length. Alignments with Mapping Quality Score ≥10 were counted for gene bodies, according to the gene start and gene end annotations by the GENCODE Project (release 19),[28] using featureCounts[29] without strand information. The 5hmC-Seal libraries were sequenced to produce a median of ~25 million reads in each sample, and a median number of ~13.5 million unique reads (ie, >50%) mapped to ~22 000 gene bodies. The raw count data summarized for the gene bodies were then normalized using DESeq2[30] and corrected for library size for statistical analysis. To explore gene regulatory relevance of 5hmC in cfDNA, we also summarized the 5hmC-Seal data according to the genomic peaks of H3K4me1, a tissue-specific marker for enhancers,[31] as provided by the Roadmap Epigenomics Project[32] for the B cell and other tissues for comparison.

Linking 5hmC in cfDNA with cell of origin and clinical characteristics

We examined whether the 5hmC-Seal data reflected the cell of origin (ie, germinal center B-cell–like [GCB] and activated B-cell–like [ABC] DLBCLs), as determined by the Han's algorithm with immunohistochemistry staining,[33] or were associated with standard prognostic factors, such as Ann Arbor stage (3/4 vs 1/2), serum lactate dehydrogenase (LDH) levels (elevated vs normal), and the International Prognostic Index (IPI; high = 3/4/5 vs low = 0/1/2). For each comparison, the top differential 5hmC marker genes (P < .05) from logistic models adjusting for age and sex were retained as candidates for further feature selection based on the elastic net regularization, using the *glmnet* library for the R statistical package.[34] This feature-selection process was repeated 100 times, and a panel of 5hmC marker genes that were selected from ≥80% iterations was kept as final feature genes.

Developing a weighted prognostic model for DLBCL

We collected baseline clinical, laboratory, and treatment data, disease progression or relapse, and retreatment from electronic
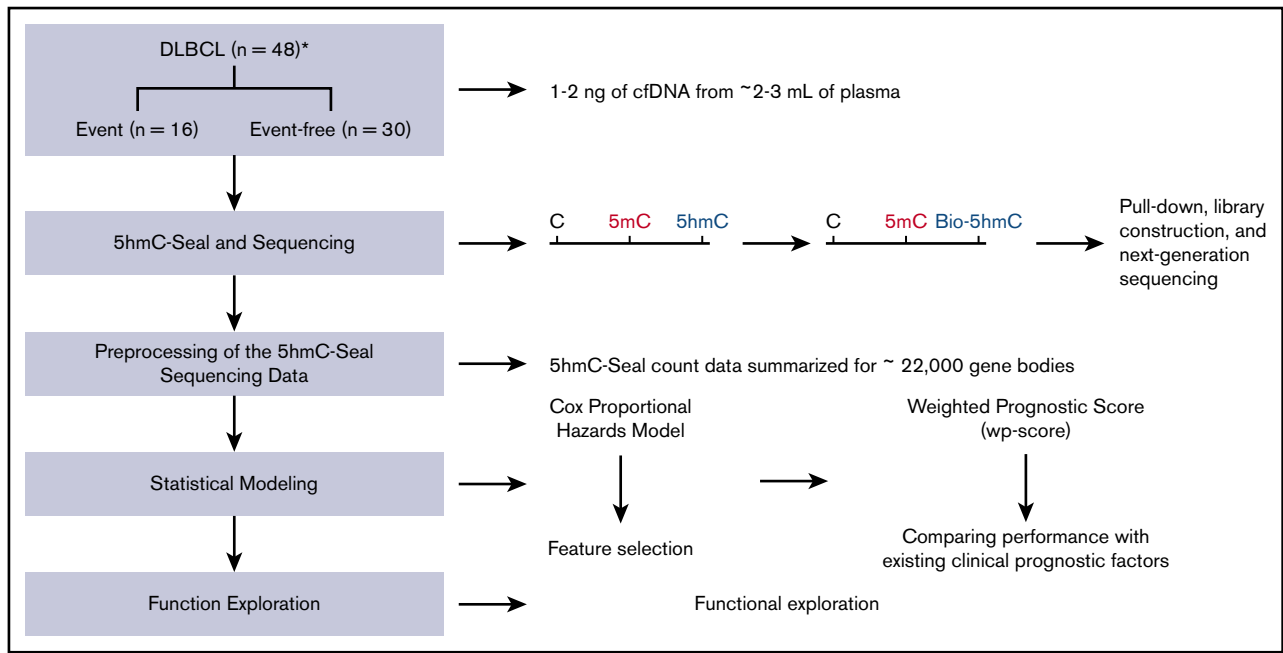
Figure 1. Study design and an overview of the 5hmC-Seal assay. A total of 48 cfDNA samples collected at the time of diagnosis from patients with DLBCL is included in this study. A weighted prognostic score based on the 5hmC marker genes is developed to evaluate prognosis after treatment. The 5hmC-Seal technique uses a chemical labeling strategy to sensitively profile 5hmC in cfDNA from nanogram-level DNA materials. *Clinical outcomes, including the development of clinical events (ie, relapse, death) are missing for 2 patients.

medical records. Deaths were ascertained using the National Death Index. We considered unplanned consolidative radiation therapy, but not radiation therapy as part of the initial treatment plan, as a retreatment. Event-free survival (EFS) was defined as time from diagnosis until relapse or progression, unplanned retreatment of lymphoma after initial immunochemotherapy, or death.[35] Overall survival (OS) was defined as time from diagnosis until death from any cause. Follow-up was through 31 December 2017.

Candidate marker genes associated with clinical events (ie, relapse, retreatment, death) were first detected with a less stringent cutoff ($P < .05$) under univariate Cox proportional hazards models, aiming to retain the most informative marker genes for further variable selection. Next, we applied the elastic net regularization on the multivariate Cox proportional hazards model, including age and sex as covariates, using *glmnet*[34] to select the final panel of marker genes for clinical events. The coefficients of final marker genes were used to compute a weighted prognostic score (wp-score) for each patient:

$$wp - score = \sum_{k=1}^{n}(\beta_k \times G_k)$$

β_k is the coefficient from the multivariate logistic model for gene k, and G_k is the normalized count of kth marker gene in the final panel. Kaplan-Meier curves were used to display survival curves based on the wp-scores (ie, risk scores). We then compared the prognostic accuracy, sensitivity, and specificity of the wp-score (high risk vs low risk) associated with clinical events with those using the established prognostic factors, including the serum LDH level (elevated vs normal), cell of origin (ABC vs GCB), Ann Arbor stages (1/2 vs 3/4), and the IPI (low = 0/1/2 vs high = 3/4/5). Multivariate Cox models were used to assess the association between the wp-scores and EFS or OS, controlling for age, sex, and standard prognostic

factors. Log-rank P values were used to evaluate statistical significance for the Cox models.

Pathway analysis and exploration of tissue relevance

The TiGER (Tissue-specific Gene Expression and Regulation) database[36] for tissue-specific expression was used to evaluate potential gene expression relevance of the 5hmC-Seal data derived from patient cfDNA. H3K4me1, a tissue-specific enhancer marker, peaks derived from various tissues from the Roadmap Epigenomics Project[32] (accessed on 15 December 2018) were used to explore the relationships between 5hmC-Seal profiles from DLBCL patients and *cis*-regulatory elements. To explore the underlying biological connections of the candidate marker genes, we conducted Kyoto Encyclopedia of Genes and Genomes[37] pathway enrichment analysis using the National Institutes of Health/DAVID tool.[38] We used the Reactome Functional Interaction (FI)[39] plug-in to explore FIs across the candidate marker genes associated with clinical events. Hubs of the Reactome FI networks were estimated based on the betweenness centrality, which detects the amount of influence that a node (ie, gene) has over the flow of information in a gene network.

Results

Patient characteristics

A total of 48 patients with newly diagnosed DLBCL was included in the study (Table 1). Median age at diagnosis was 59.5 years (range, 24-82 years), 63% (n = 30) were males, 50% were stage 1/2 based on the Ann Arbor staging system for lymphomas, 27% had an IPI score ≥ 3, and 68% had GCB-type DLBCL. In addition, most patients (67%) received R-CHOP (rituximab plus cyclophosphamide, doxorubicin, vincristine, and prednisone), followed by

Table 1. Characteristics of the study subjects (N = 48)

Characteristics	n (%)
Age, median (range), y	58 (24-82)
Sex	
Male	30 (62.5)
Female	18 (37.5)
Ann Arbor stage	
I	12 (26.1)
II	11 (23.9)
III	6 (13.0)
IV	17 (37.0)
Missing	2
IPI score	
0-1	15 (40.6)
2	14 (37.8)
3-5	8 (21.6)
Missing	11
Cell of origin	
GCB	23 (67.6)
ABC	11 (32.4)
Missing	14
LDH	
Elevated	25 (53.2)
Not elevated	22 (46.8)
Missing	1
Initial treatment	
R-CHOP	32 (66.6)
EPOCH-R	8 (16.7)
Other regimen	8 (16.7)
Vital status, alive	34 (70.8)

EPOCH-R (etoposide, prednisone, vincristine, cyclophospha-mide, and doxorubicin plus rituximab; 17%), as the front-line treatment. Outcomes for 2 subjects cannot be determined. At the end of the follow-up, 16 patients had a clinical event, and 30 did not (Figure 1).

Distinct distributions of 5hmC in patient-derived cfDNA

The 5hmC-Seal sequencing reads obtained from patient-derived cfDNA showed distinct genomic distributions (Figure 2A-B). The 5hmC-Seal sequencing reads in cfDNA were enriched in gene bodies, whereas they were depleted in the flanking regions relative to the transcription start sites and transcription end sites (Figure 2A). The distribution of 5hmC in cfDNA was consistent with their putative roles in gene activation and significantly overlapped with the B-cell–derived Roadmap Epigenomics Project H3K4me1 peaks (Figure 2B). We also found that DLBCL patient-derived cfDNA samples were more enriched with the H3K4me1 peaks derived from the B cell than from other tissue types (eg, lung, pancreas, liver, and brain) using the Roadmap Epigenomics Project data (Student t test, $P < .001$, Figure 2B), suggesting tissue relevance of the profiled 5hmC-Seal data in patients with DLBCL.

Next, in 7 patients with cfDNA-paired tissue samples, we compared 5hmC distributions between cfDNA samples and paired tumor tissue samples from the same patients (Figure 2C). We found that ~16 000 gene bodies contained ≥30 sequencing reads in cfDNA and paired tissue samples. The top-ranking most variable genes (ie, most informative) in cfDNA showed higher correlation in paired tissue samples from the same individuals (mean Pearson's $r = 0.91$) than from different patients (mean Pearson's $r = 0.88$) (Figure 2C), supporting the tumor origin of a patient's 5hmC profile in cfDNA. The most variable genes in cfDNA were also primarily enriched with genes specifically expressed in blood compared with other tissue types (hypergeometric $P < .001$), based on the TiGER database for tissue-specific gene expression (Figure 2D).

5hmC-Seal data reflect cell of origin and clinical characteristics

To evaluate the potential clinical utility and interpretation of cfDNA-based 5hmC prognostic markers for DLBCL, we compared the 5hmC profiles in DLBCL patient-derived cfDNA with the clinical characteristics of patients. We found that 5hmC marker genes detected in cfDNA differed by cell of origin and clinical characteristics of patients at diagnosis (supplemental Table 1). We found that 5hmC-Seal profiles in cfDNA distinguished GCB-type DLBCL from ABC-type DLBCL (Figure 3A), including genes involved in the glycosaminoglycan biosynthesis pathway (eg, EXTL1 encoding exostosin-like glycosyltransferase 1) that are related to the subtypes and aggressiveness of B-cell lymphoma.[40] We also found that 5hmC-Seal profiles in cfDNA differed by Ann Arbor stage (1/2 vs 3/4) (Figure 3B), LDH level (elevated vs normal) (Figure 3C), and the IPI (Figure 3D).

Prognostic value of 5hmC in cfDNA for DLBCL

Among the 46 DLBCL patients with available outcome data, 34 were alive at the end of the follow-up (ie, 31 December 2017). We identified 214 candidate marker genes potentially associated with clinical events (supplemental Table 2). We also explored functional annotations using these 214 candidate genes because the feature selection procedure that followed considered statistical significance, not biological relevance. Pathway analysis suggested that these 214 genes were involved in the Kyoto Encyclopedia of Genes and Genomes pathways (supplemental Table 3). Results from Reactome FI analysis suggested some functional interaction hubs important in the gene network among the candidate marker genes (supplemental Figure 1), such as HIST1H2BC (encoding histone cluster 1 H2B family member C) that was among the enriched pathways, as described above, TBP (encoding TATA-Box binding protein), and E2F1,[41] GATA-3,[42] and MLH1,[43] which have been associated with the prognosis of DLBCL.

These 214 candidate genes were trimmed to 29 final marker genes after feature selection for predicting patient outcomes (Figure 4A). A wp-score was then computed for each patient based on the 29 marker genes. Compared with patients in the low-risk group (ie, low wp-score), patients in the high-risk group had worse OS (Figure 4B, log-rank $P = .001$) or worse EFS (Figure 4C, log-rank $P = .002$). Specifically, in the multivariate analysis controlling for age and sex, high-risk scores (wp-scores) were associated with poorer EFS (hazard ratio, 9.17; 95% confidence interval [CI], 2.01-41.89; $P = .004$) compared with low-risk scores (Table 2). Moreover,
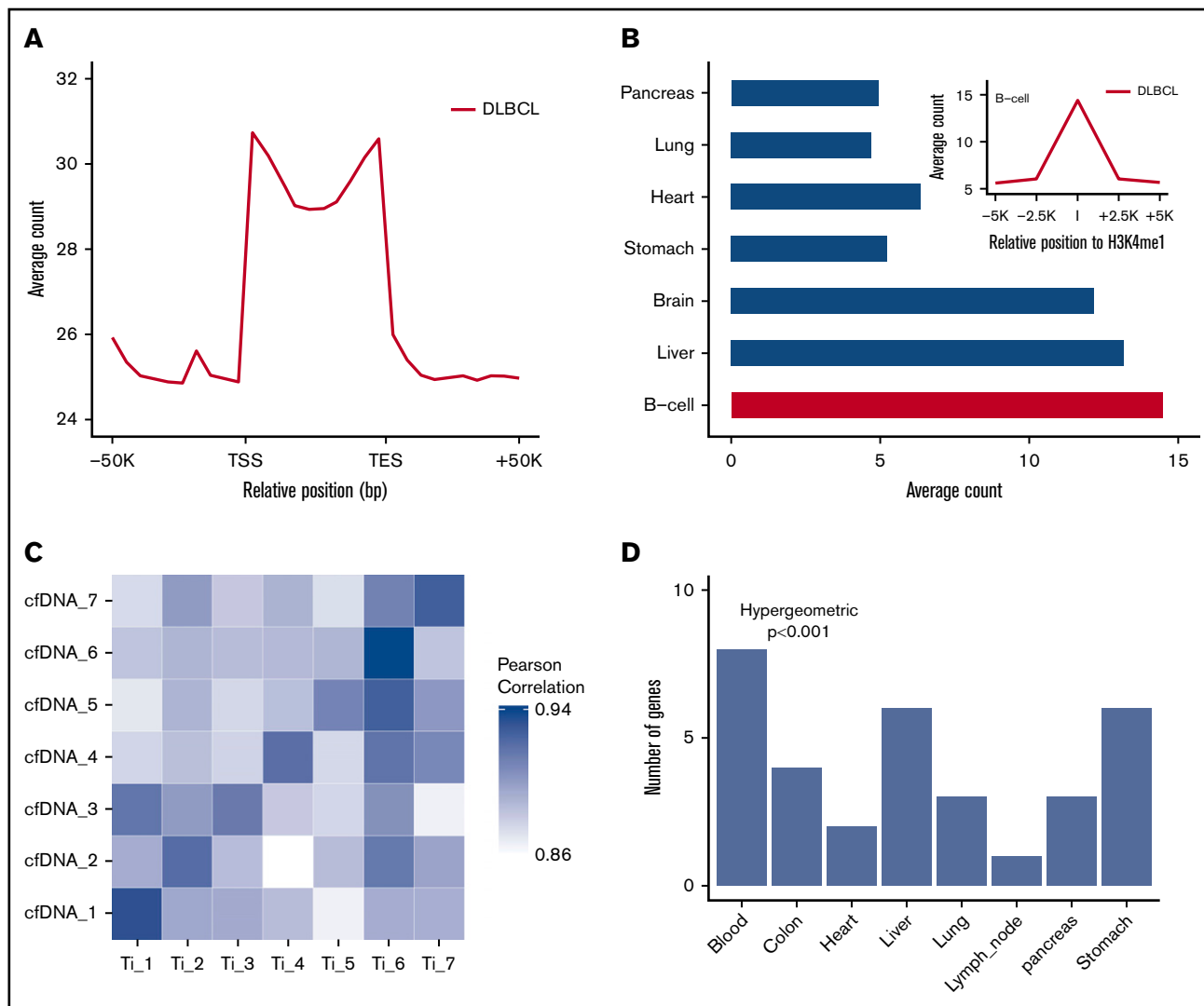
Figure 2. Distinct distributions of 5hmC in patient-derived cfDNA and tissue relevance. (A) The median and range of average counts are plotted against the relative genomic positions. The 5hmC-Seal data are enriched in gene bodies (split into 10 bins for all genes) relative to the flanking regions. (B) The 5hmC-Seal data are enriched in the B-cell–derived H3K4me1 loci compared with other tissue types (Student t test $P < .001$). H3K4me1 loci are obtained from the Roadmap Epigenomics Project data. (C) The top 100 highly variable (ie, most informative) 5hmC marker genes in cfDNA are highly correlated with those identified in the cfDNA-paired tissue samples from the same subject (n = 7). (D) The top 500 highly variable 5hmC marker genes in cfDNA are enriched within genes specific to blood compared with other tissues, based on the TiGER database for tissue-specific genes. TES, transcription end site; TSS, transcription start site.

the wp-scores remained significantly associated with EFS after additional adjustment for standard prognostic factors, suggesting that the wp-scores are an independent prognostic factor for DLBCL (Table 2). Importantly, results for overall accuracy, sensitivity, and specificity showed that the wp-scores had an overall superior performance for predicting, at diagnosis, patients at risk for having a clinical event during the follow-up compared with standard clinical prognostic factors, such as elevated LDH level, advanced stages (3/4), ABC-type DLBCL, and high IPI (≥3) (Figure 4D). Data on MYC, BCL2, and BCL6 expression determined by immunohistochemistry were available for 14 patients. In exploratory analyses, the 5hmC-based wp-scores also performed better than double or triple expression (ie, MYC and BCL2 and/or BCL6) in predicting prognosis (data not shown). However, these results should be interpreted with

caution, given the small sample size and exploratory nature of the analysis.

Discussion

In this prospective study of newly diagnosed patients with DLBCL, we profiled genome-wide 5hmC in cfDNA from blood plasma and investigated its association with prognosis and known prognostic markers. We found distinct genomic distributions of 5hmC in cfDNA and demonstrated the relevance of cfDNA-based 5hmC to tumor origin. In addition, 5hmC marker genes differed by cell of origin and clinical characteristics of patients at diagnosis. We identified a panel of 29 marker genes that were associated with the probability of having a clinical event. The wp-scores based on these 29 marker genes were associated with OS and EFS, independent of established
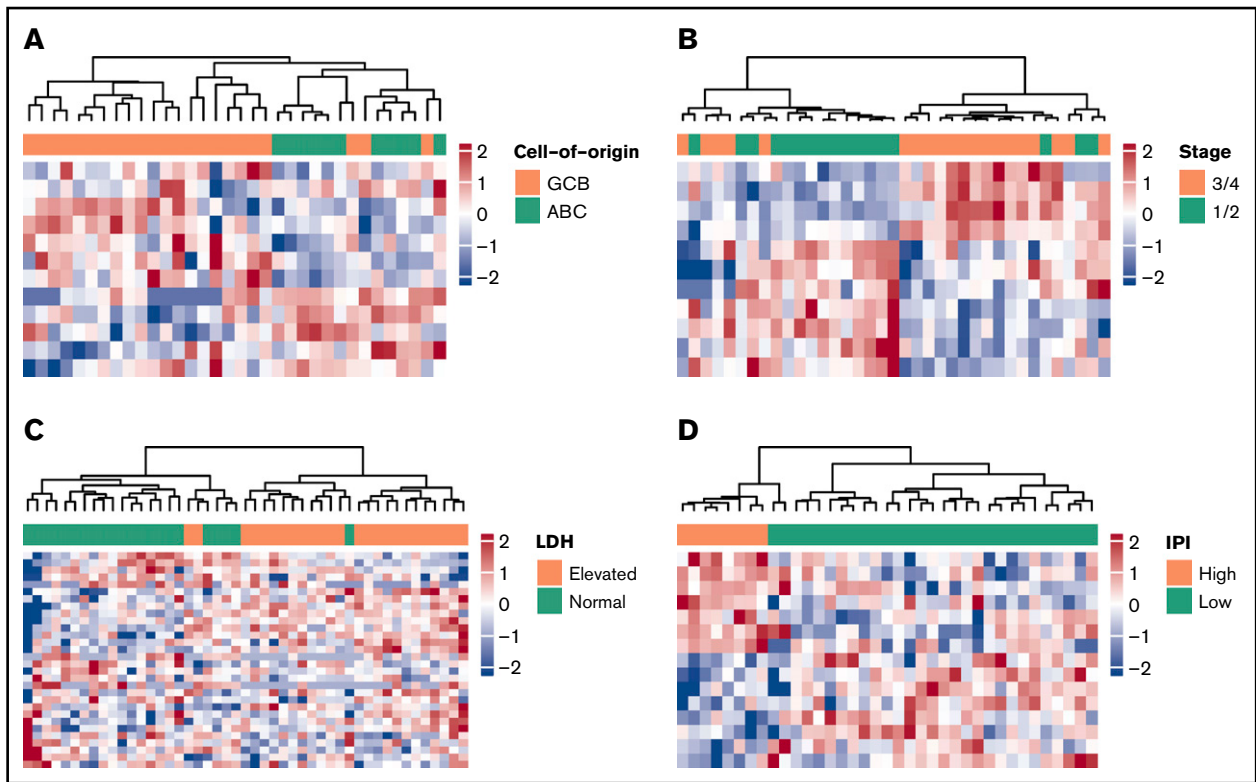
Figure 3. 5hmC in patient-derived cfDNA differs by cell of origin and clinical characteristics. The heat maps are plotted using the final selected genes in each comparison. The 5hmC profiles in cfDNA are shown to be associated with cell of origin (GCB vs ABC, 12 genes) (A), Ann Arbor stages (1/2 vs 3/4, 11 genes) (B), LDH level (elevated vs normal, 18 genes) (C), and the IPI (low = 0/1/2 vs high = 3/4/5, 15 genes) (D).

prognostic factors. To our knowledge, this is the first study to profile genome-wide 5hmC in cfDNA and provide suggestive evidence of the prognostic value of these epigenetic markers in DLBCL.

Despite convincing evidence that supports 5hmC as a novel class of epigenetic biomarkers for various solid tumors and hematological malignancies,[18] it remains technically challenging to profile 5hmC in cfDNA because of the scarcity of 5hmC. To address the gap, we applied the 5hmC-Seal, a highly sensitive and robust technique based on covalent chemical linkage[44] and requiring as little as ~1 to 2 ng of DNA from ~2 to 3 mL of plasma.[22] To our knowledge, the 5hmC-Seal is the only method that allows mapping genome-wide 5hmC, and it is highly sensitive for clinically feasible amounts of cfDNA samples. The assay has been validated and implemented for several cancers in our laboratories[22,45] and those of other investigators.[23,46]

Our findings that genome-wide 5hmC signatures in cfDNA correlated with established prognostic factors and were associated with the prognosis of DLBCL suggested that cfDNA-based 5hmC signatures could complement current biopsy-based clinical practice for DLBCL prognostication. Delineating cell of origin[33,47] or determining genetic alterations[48-50] is clinically important amid the rapid development of novel molecular targeted regimens for DLBCL.[2,51] The major limitation is that these approaches require tissue biopsies, which are invasive and are prone to sampling bias as a result of intratumoral and spatial heterogeneity.[52-54] Accumulating evidence suggests that circulating cfDNA from blood plasma contains epigenetic information released from the tumor/tumor microenvironment into the blood and reflects tumor pathobiology.[54-56] As such, cfDNA offers transformative opportunities to overcome some of the limitations of tissue-based approaches. Two recent studies reported that global hypomethylation[9] and aberrant *DAPK1* methylation[12] of cfDNA predicted poor outcomes in DLBCL. Our findings of the prognostic significance of 5hmC in cfDNA for DLBCL suggest that 5hmC may also play an important role in the progression of DLBCL, and it warrants further evaluation.

In our study, a weighted model consisting of 29 gene markers is associated with EFS and OS independent of standard prognostic factors, such as age, stage, LDH, and IPI. Some of these genes have been implicated in lymphoma, such as *PDSS1* (encoding prenyl [decaprenyl] diphosphate synthase, subunit 1), *NHP2* (encoding NHP2 ribonucleoprotein), and *ANGEL1* (encoding angel homolog). We also found that the wp-score based on 5hmC markers outperformed (eg, overall accuracy, sensitivity, and/or specificity) existing prognostic factors in predicting a clinical event. For example, cell of origin is a well-established prognostic factor in DLBCL and is a potential biomarker for future personalized therapies.[51,57] In this study, the sensitivity, specificity, and overall predictive accuracy of cell of origin for a clinical event is <50%: 0.56, 0.29, and 0.36, respectively. In contrast, the corresponding values for wp-score are 0.86, 1.00, and 0.96 (Figure 4D). LDH, one of the most commonly used biomarkers for DLBCL during scheduled clinical visits, also does not perform as well as the wp-score. These findings suggest that the 5hmC profiles in cfDNA hold
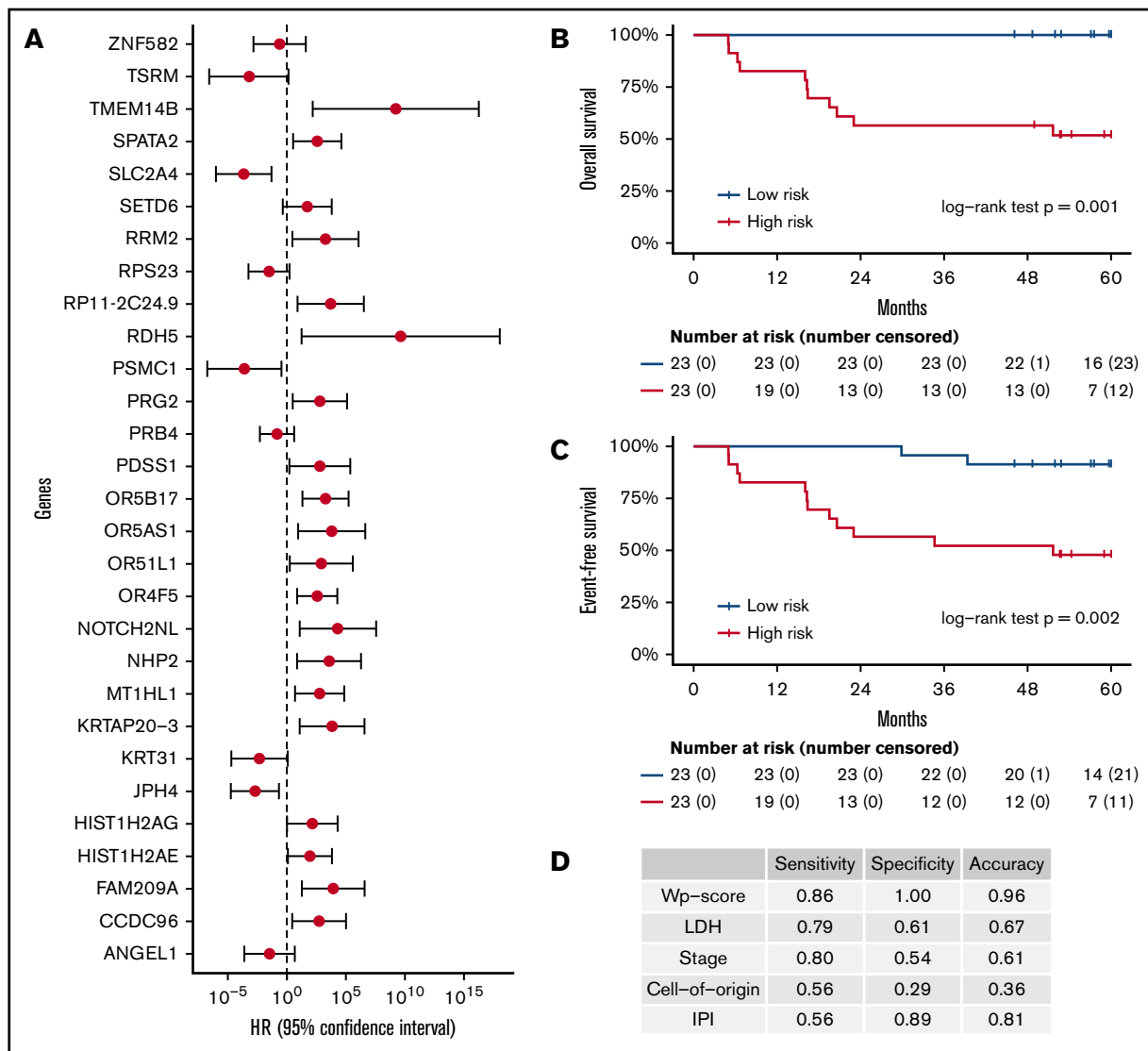
Figure 4. Prognostic implications of cfDNA-based 5hmC for DLBCL. The 5hmC marker genes are associated with the occurrence of clinical events (relapse, retreatment, or death) and can predict patient survival. (A) A weighted prognostic model consists of 29 marker genes that are associated with clinical events. The wp-scores are computed using these final marker genes for all patients with DLBCL. The wp-scores predict OS (B) and EFS (C) at the time of diagnosis. (D) The wp-scores show superior predictive performance (ie, sensitivity and specificity that maximized the Youden's index) for the risk of developing clinical events compared with standard clinical prognostic factors, including serum LDH level, Ann Arbor stages, cell of origin, and the IPI.

the promise to be a convenient alternative that could supplement the current clinical practice to provide relevant clinical information and risk stratification for DLBCL.

The current study has several strengths, including the confirmation of DLBCL diagnosis and outcomes, the prospective study design, and the use of 5hmC-Seal, a state-of-the-art technique. There are also limitations. First, the relatively small sample size does not allow us to control for treatment approaches or to validate the marker panel in independent samples. Although the majority of patients (67%) received R-CHOP as the standard front-line treatment, EPOCH-R and other regimens accounted for 33%. The wp-score was slightly higher for the EPOCH-R group than for the R-CHOP group, but the difference was not statistically significant. Second, we have limited data on MYC, BCL2, and BCL6 expression and tumor

burden. Comparing the prognostic significance of 5hmC-based wp-score with these prognostic factors is warranted in future work. Third, similar to other studies of DLBCL in European countries and North America, we were not able to evaluate the association between 5hmC in cfDNA and prognosis by race/ethnicity or population. Future studies with a large minority patient population are warranted to evaluate the generalizability of our results.

In conclusion, our findings suggest that 5hmC in patient-derived cfDNA profiled using the 5hmC-Seal, a highly robust and sensitive technique, has the potential to be a clinically convenient and minimally invasive prognostic approach for DLBCL. Future epigenetic studies of prognosis for DLBCL should include 5hmC as a stable and important epigenetic marker.

Table 2. Multivariate prognostic models for OS and EFS in patients with DLBCL

Outcome	Model	Variable	Hazard ratio	95% CI (lower)	95% CI (upper)	P	Log-rank P
OS	wp-score	wp-score (high)*	Infinite	0.000	Infinite	ND	.001
		Age	0.98	0.29	1.02	.29	
		Sex (female)	2.49	0.26	12.29	.26	
EFS	wp-score	wp-score (high)	9.17	2.01	41.89	.004	.002
		Age	0.98	0.94	1.02	.29	
		Sex (female)	2.23	0.59	8.41	.24	
	wp-score + LDH	wp-score (high)	10.85	2.24	52.64	.003	.001
		LDH (elevated)	5.99	1.47	24.49	.01	
		Age	0.96	0.92	1.00	.05	
		Sex (female)	1.23	0.29	5.12	.78	
	wp-score + IPI	wp-score (high)	14.28	1.54	132.84	.02	<.001
		IPI (high)	23.44	3.12	175.87	.002	
		Age	0.94	0.88	1.01	.09	
		Sex (female)	11.13	0.97	128.17	.05	
	wp-score + GCB vs ABC	wp-score (high)	20.08	1.37	293.60	.03	.02
		Cell of origin	4.47	0.78	25.73	.09	
		Age	0.93	0.86	1.00	.05	
		Sex (female)	4.27	0.50	36.92	.19	
	wp-score + stage	wp-score (high)	6.55	1.28	33.46	.02	.03
		Stage (advanced)	3.99	0.83	19.24	.08	
		Age	0.98	0.93	1.04	.52	
		Sex (female)	2.37	0.50	11.26	.28	

ND, not defined.

*All patients in the high wp-score group are alive at the end of the follow-up period. Thus, the β estimate is infinite for OS analysis.

Authorship

Contribution: B.C.-H.C., C.H., and W.Z. designed the research and provided oversight; Q.Y. and K.Y. performed the 5hmC-Seal experiment; Z.Z., C.Z., and W.Z. analyzed the 5hmC-Seal data and performed statistical analyses; E.S. coordinated sample collection; G.V. was the study hematopathologist; P.M.B. and S.M.S. provided clinical advice and helped to interpret the data; B.C.-H.C., Z.Z., and W.Z. drafted the manuscript with input from all authors; and all authors read and approved the final manuscript.

Conflict-of-interest disclosure: C.H. is a scientific founder of Accent Therapeutics, Inc. and a member of its scientific advisory board. C.H. and W.Z. are shareholders of Epican Genetech Co. Ltd. The remaining authors declare no competing financial interests.

ORCID profiles: B.C.-H.C., 0000-0001-8979-7244; C.Z., 0000-0003-3520-3724.

Correspondence: Brian C.-H. Chiu, University of Chicago, 5841 S Maryland Ave, MC200, Chicago, IL 60637; e-mail: bchiu@uchicago.edu; or Wei Zhang, Northwestern University, 680 N Lake Shore Dr, Suite 1400, Chicago, IL 60611; e-mail: wei.zhang1@northwestern.edu.

References

1. Menon MP, Pittaluga S, Jaffe ES. The histological and biological spectrum of diffuse large B-cell lymphoma in the World Health Organization classification. *Cancer J.* 2012;18(5):411-420.

2. Roschewski M, Staudt LM, Wilson WH. Diffuse large B-cell lymphoma-treatment approaches in the molecular era. *Nat Rev Clin Oncol.* 2014;11(1): 12-23.

3. Friedberg JW. Relapsed/refractory diffuse large B-cell lymphoma. *Hematology Am Soc Hematol Educ Program.* 2011;2011:498-505.

4. Hohaus S, Giachelia M, Massini G, et al. Cell-free circulating DNA in Hodgkin's and non-Hodgkin's lymphomas. *Ann Oncol*. 2009;20(8):1408-1413.

5. 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.

6. 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.

7. Chambwe N, Kormaksson M, Geng H, et al. Variability in DNA methylation defines novel epigenetic subgroups of DLBCL associated with different clinical outcomes. *Blood*. 2014;123(11):1699-1708.

8. De S, Shaknovich R, Riester M, et al. Aberration in DNA methylation in B-cell lymphomas has a complex origin and increases with disease severity. *PLoS Genet*. 2013;9(1):e1003137.

9. Wedge E, Hansen JW, Garde C, et al. Global hypomethylation is an independent prognostic factor in diffuse large B cell lymphoma. *Am J Hematol*. 2017; 92(7):689-694.

10. Asmar F, Punj V, Christensen J, et al. Genome-wide profiling identifies a DNA methylation signature that associates with TET2 mutations in diffuse large B-cell lymphoma. *Haematologica*. 2013;98(12):1912-1920.

11. Morin RD, Johnson NA, Severson TM, et al. Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin. *Nat Genet*. 2010;42(2):181-185.

12. Kristensen LS, Hansen JW, Kristensen SS, et al. Aberrant methylation of cell-free circulating DNA in plasma predicts poor outcome in diffuse large B cell lymphoma. *Clin Epigenetics*. 2016;8(1):95.

13. Ito S, D'Alessio AC, Taranova OV, Hong K, Sowers LC, Zhang Y. Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature*. 2010;466(7310):1129-1133.

14. Tahiliani M, Koh KP, Shen Y, et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science*. 2009; 324(5929):930-935.

15. Iyer LM, Abhiman S, Aravind L. Natural history of eukaryotic DNA methylation systems. *Prog Mol Biol Transl Sci*. 2011;101:25-104.

16. Yu M, Hon GC, Szulwach KE, et al. Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell*. 2012;149(6):1368-1380.

17. Branco MR, Ficz G, Reik W. Uncovering the role of 5-hydroxymethylcytosine in the epigenome. *Nat Rev Genet*. 2011;13(1):7-13.

18. Zeng C, Stroup EK, Zhang Z, Chiu BCH, Zhang W. Towards precision medicine: advances in 5-hydroxymethylcytosine cancer biomarker discovery in liquid biopsy. *Cancer Commun. (Lond)*. 2019;39(1):12.

19. Mariani CJ, Madzo J, Moen EL, Yesilkanal A, Godley LA. Alterations of 5-hydroxymethylcytosine in human cancers. *Cancers (Basel)*. 2013;5(3):786-814.

20. Gilat N, Tabachnik T, Shwartz A, et al. Single-molecule quantification of 5-hydroxymethylcytosine for diagnosis of blood and colon cancers. *Clin Epigenetics*. 2017;9(1):70.

21. Huang Y, Pastor WA, Shen Y, Tahiliani M, Liu DR, Rao A. The behaviour of 5-hydroxymethylcytosine in bisulfite sequencing. *PLoS One*. 2010;5(1):e8888.

22. Li W, Zhang X, Lu X, et al. 5-Hydroxymethylcytosine signatures in circulating cell-free DNA as diagnostic biomarkers for human cancers [published correction appears in Cell Res. 2019;29(7):599]. *Cell Res*. 2017;27(10):1243-1257.

23. Song CX, Yin S, Ma L, et al. 5-Hydroxymethylcytosine signatures in cell-free DNA provide information about tumor types and stages. *Cell Res*. 2017; 27(10):1231-1242.

24. Tian X, Sun B, Chen C, et al. Circulating tumor DNA 5-hydroxymethylcytosine as a novel diagnostic biomarker for esophageal cancer. *Cell Res*. 2018; 28(5):597-600.

25. Swerdlow SH, Campo E. WHO Classification: Pathology and Genetics of Tumors of Haematopoietic and Lymphoid Tissues. Lyon, France: IARC Press; 2008.

26. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114-2120.

27. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9(4):357-359.

28. Harrow J, Frankish A, Gonzalez JM, et al. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res*. 2012;22(9): 1760-1774.

29. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014; 30(7):923-930.

30. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550.

31. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489(7414):57-74.

32. Kundaje A, Meuleman W, Ernst J, et al; Roadmap Epigenomics Consortium. Integrative analysis of 111 reference human epigenomes. *Nature*. 2015; 518(7539):317-330.

33. Hans CP, Weisenburger DD, Greiner TC, et al. Confirmation of the molecular classification of diffuse large B-cell lymphoma by immunohistochemistry using a tissue microarray. *Blood*. 2004;103(1):275-282.

34. Friedman DL, Whitton J, Leisenring W, et al. Subsequent neoplasms in 5-year survivors of childhood cancer: the Childhood Cancer Survivor Study. *J Natl Cancer Inst*. 2010;102(14):1083-1095.

35. Maurer MJ, Ghesquières H, Jais JP, et al. Event-free survival at 24 months is a robust end point for disease-related outcome in diffuse large B-cell lymphoma treated with immunochemotherapy. *J Clin Oncol*. 2014;32(10):1066-1073.

36. Liu X, Yu X, Zack DJ, Zhu H, Qian J. TiGER: a database for tissue-specific gene expression and regulation. *BMC Bioinformatics*. 2008;9(1):271.

37. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res*. 2016; 44(D1):D457-D462.

38. Huang W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res*. 2009;37(1):1-13.

39. Fabregat A, Jupe S, Matthews L, et al. The Reactome Pathway Knowledgebase. *Nucleic Acids Res*. 2018;46(D1):D649-D655.

40. Yip GW, Smollich M, Götte M. Therapeutic value of glycosaminoglycans in cancer. *Mol Cancer Ther*. 2006;5(9):2139-2148.

41. Møller MB, Kania PW, Ino Y, et al. Frequent disruption of the RB1 pathway in diffuse large B cell lymphoma: prognostic significance of E2F-1 and p16INK4A. *Leukemia*. 2000;14(5):898-904.

42. Juskevicius D, Lorber T, Gsponer J, et al. Distinct genetic evolution patterns of relapsing diffuse large B-cell lymphoma revealed by genome-wide copy number aberration and targeted sequencing analysis. *Leukemia*. 2016;30(12):2385-2395.

43. Rossi D, Rasi S, Di Rocco A, et al. The host genetic background of DNA repair mechanisms is an independent predictor of survival in diffuse large B-cell lymphoma. *Blood*. 2011;117(8):2405-2413.

44. Song CX, Szulwach KE, Fu Y, et al. Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. *Nat Biotechnol*. 2011; 29(1):68-72.

45. Han D, Lu X, Shih AH, et al. A highly sensitive and robust method for genome-wide 5hmC profiling of rare cell populations. *Mol Cell*. 2016;63(4): 711-719.

46. Zhang J, Han X, Gao C, et al. 5-hydroxymethylome in circulating cell-free DNA as a potential biomarker for non-small-cell lung cancer. *Genomics Proteomics Bioinformatics*. 2018;16(3):187-199.

47. Rosenwald A, Wright G, Chan WC, et al; Lymphoma/Leukemia Molecular Profiling Project. 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.

48. Rosenthal A, Younes A. High grade B-cell lymphoma with rearrangements of MYC and BCL2 and/or BCL6: Double hit and triple hit lymphomas and double expressing lymphoma. *Blood Rev*. 2017;31(2):37-42.

49. Landsburg DJ, Petrich AM, Abramson JS, et al. Impact of oncogene rearrangement patterns on outcomes in patients with double-hit non-Hodgkin lymphoma. *Cancer*. 2016;122(4):559-564.

50. Petrich AM, Nabhan C, Smith SM. MYC-associated and double-hit lymphomas: a review of pathobiology, prognosis, and therapeutic approaches. *Cancer*. 2014;120(24):3884-3895.

51. 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.

52. Alizadeh AA, Aranda V, Bardelli A, et al. Toward understanding and exploiting tumor heterogeneity. *Nat Med*. 2015;21(8):846-853.

53. Drew L. Towards the better diagnosis of lymphoma. *Nature*. 2018;563(7731):S38-S40.

54. Corcoran RB, Chabner BA. Application of cell-free DNA analysis to cancer treatment. *N Engl J Med*. 2018;379(18):1754-1765.

55. Wan JCM, Massie C, Garcia-Corbacho J, et al. Liquid biopsies come of age: towards implementation of circulating tumour DNA. *Nat Rev Cancer*. 2017; 17(4):223-238.

56. Warton K, Samimi G. Methylation of cell-free circulating DNA in the diagnosis of cancer. *Front Mol Biosci*. 2015;2:13.

57. Nowakowski GS, LaPlant B, Macon WR, et al. Lenalidomide combined with R-CHOP overcomes negative prognostic impact of non-germinal center B-cell phenotype in newly diagnosed diffuse large B-cell lymphoma: a phase II study. *J Clin Oncol*. 2015;33(3):251-257.