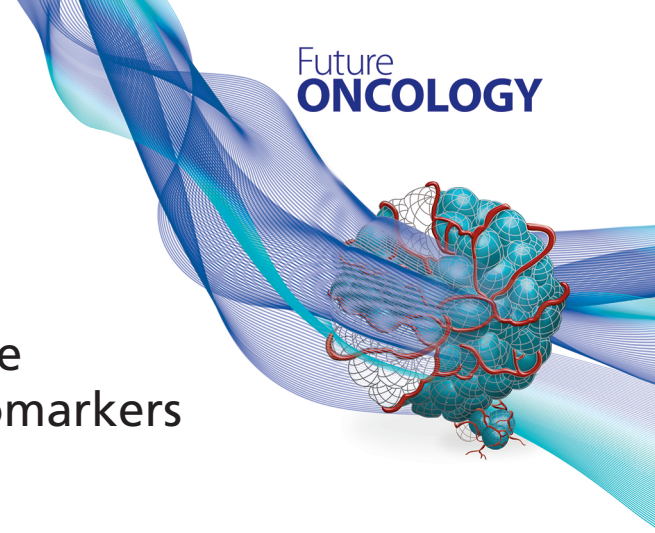
Research Article

For reprint orders, please contact: reprints@futuremedicine.com



Circulating DNA methylation profile improves the accuracy of serum biomarkers for the detection of nonmetastatic hepatocellular carcinoma

Thanh Hai Phan[‡,1], Van Thien Chi Nguyen[‡,2,3], Thu Thuy Thi Pham[1], Van-Chu Nguyen[4,5], Tan Dat Ho[1], Thi Mong Quynh Pham[2,3], Thanh-Huong Tran[4,5], Thanh Dat Nguyen[2,3], Nguyen Duy Khang Le[2,3], Trong-Hieu Nguyen[2,3], Minh-Long Duong[4], Hoai-Phuong Thi Bach[4], Van-Vu Kim[4,5], The-Anh Pham[4], Bao Toan Nguyen[1], Thanh Nhan Vo Nguyen[1], Thanh Dang Nguyen[1], Dung Thai Bieu Phu[1], Boi Hoan Huu Phan[1], Duy-Sinh Nguyen[6], Dinh-Kiet Truong[2], Thanh-Thuy Thi Do[2], Hoa Giang[2,3], Hoai-Nghia Nguyen[2,3], Minh-Duy Phan**[,2,3] & Le Son Tran*[,2,3]

[1]MEDIC Medical Center, Ho Chi Minh City, Vietnam
[2]Medical Genetics Institute, Ho Chi Minh City, Vietnam
[3]Gene Solutions, Ho Chi Minh City, Vietnam
[4]National Cancer Hospital, Hanoi, Vietnam
[5]Hanoi Medical University, Hanoi, Vietnam
[6]Department of Oncology, Faculty of Medicine, Nguyen Tat Thanh University, Ho Chi Minh City, Vietnam
*Author for correspondence: Tel.: +84 705 196 257; leson1808@gmail.com
**Author for correspondence: pmduy@yahoo.com
‡Authors contributed equally

Aim: This study exploited hepatocellular carcinoma (HCC)-specific circulating DNA methylation profiles to improve the accuracy of a current screening assay for HCC patients in high-risk populations. **Methods:** Differentially methylated regions in cell-free DNA between 58 nonmetastatic HCC and 121 high-risk patients with liver cirrhosis or chronic hepatitis were identified and used to train machine learning classifiers. **Results:** The model could distinguish HCC from high-risk non-HCC patients in a validation cohort, with an area under the curve of 0.84. Combining these markers with the three serum biomarkers (AFP, lectin-reactive AFP, des-γ-carboxy prothrombin) in a commercial test, μTASWako®, achieved an area under the curve of 0.87 and sensitivity of 68.8% at 95.8% specificity. **Conclusion:** HCC-specific circulating DNA methylation markers may be added to the available assay to improve the early detection of HCC.

Plain language summary: The early detection of liver cancer in high-risk populations can help people with the disease have a higher chance of survival and better quality of life. However, this is still a healthcare challenge. Current commercial blood tests measuring protein signatures in the blood have low accuracy due to increased levels of these proteins being detected in both liver cancer patients and patients with chronic liver diseases. In this study, we identified a set of signatures in DNA released by cancer cells into the bloodstream and used them as biomarkers to distinguish liver cancer patients from high-risk patients. We also demonstrated that adding those signatures to a commercial blood test currently used in clinics could improve the accuracy in detecting liver cancer patients.

Tweetable abstract: Cancer-specific methylation markers combined with the three serum biomarkers (AFP, AFP-L3 and DCP) in a commercial diagnostic test, μTASWako, improve accuracy of liver cancer screening in high-risk populations.

First draft submitted: 5 December 2022; Accepted for publication: 23 January 2023; Published online: 14 February 2023

Keywords: cell-free DNA • circulating tumor DNA • DNA methylation • early detection • GALAD • hepatocellular carcinoma • machine learning • μTASWako

Future Medicine

Liver cancer was the sixth most common cancer and the third most lethal malignancy in 2020, with an estimated 905,677 new cases and 830,180 deaths worldwide [1]. The most common liver cancer, hepatocellular carcinoma (HCC), is three-fold more common in males than in females, with its highest incidence (17.9 per 100,000) in east Asia and lowest incidence (3 per 100,000) in south Central America [2]. As with most solid tumors, early detection is essential for successful treatment of HCC. The prognosis for HCC patients is driven by the tumor stages [3]. HCC patients detected at a very early stage (Barcelona Clinic Liver Cancer stage 0/A, or American Joint Committee on Cancer tumor node metastasis [TNM] stage I) can achieve 5-year survival rates of 70% [4]. However, improved earlier detection of nonmetastatic HCC patients who are diagnosed with more developed stages, including stage II and IIIA, when the cancer cells have not spread to other tissues or organs also substantially increases survival rates. According to the Surveillance, Epidemiology and End Results database, HCC patients with localized tumors (nonmetastatic stages) have a median 5-year survival rate of 32%, while that number drops to 3% for those diagnosed in advanced stages with metastatic tumors [5]. Populations at high risk for HCC include those with cirrhosis of all etiologies and hepatitis B virus (HBV) infection, which is endemic in Asia. On the other hand, HCC incidence is rising in western nations due to the increased prevalence of hepatitis C virus (HCV) infection, nonalcoholic fatty liver disease and other cirrhotic disorders [6]. Recommended surveillance of at-risk persons in the USA and Europe is by biannual liver ultrasonography, but a recent meta-analysis showed that the performance of liver ultrasonography is suboptimal for early cancer detection, with a sensitivity of only 45% for early HCC [7]. While imaging co-assessment with levels of AFP and its fucosylated glycoform, lectin-reactive AFP (AFP-L3), increases sensitivity, this also increases false-positive rates [8,9]. The elevation of AFP, AFP-L3 and des-γ-carboxy prothrombin (DCP) in blood samples of HCC patients has been widely reported in previous studies, suggesting their utility as serum biomarkers for HCC diagnosis [10,11]. Most recently, a commercially available diagnostic test assessing serum levels of AFP, AFP-L3 and DCP on a microfluidics-based immunoanalyzer, namely μTASWako® i30, demonstrated up to 78% sensitivity and 62% specificity for early HCC detection [12]. Moreover, incorporation of those three markers into an HCC risk-scoring system, GALAD (AFP-L3, AFP, DCP, gender and age), further increased both sensitivity and specificity to 93.3 and 85.6%, respectively [13]. The model has recently been granted a breakthrough device designation by the US FDA [14]. However, the majority (67.8%) of HCC patients enrolled in this study were diagnosed at an intermediate or advanced stage (Barcelona Clinic Liver Cancer B or C) [13]. Thus the ability of the GALAD model to detect early-stage HCC patients has not been evaluated in this study [15]. Furthermore, diagnostic assays based on those serum markers have been shown to provide high false-positive rates due to their elevation being encountered in certain benign lesions such as active chronic hepatitis B [10,11]. It is thought that the combination of multiple biomarkers covering different molecular pathways of HCC carcinogenesis could provide incremental prediction power.

The use of circulating tumor DNA as a noninvasive and highly sensitive marker for HCC detection has now been evaluated by several studies [16–18]. Indeed, plasma circulating tumor DNA can be distinguished from other nontumor cell-free DNA (cfDNA) via several features, including genetic mutations, aberrant fragment length profiles and epigenetic modifications [19]. Of those features, 'signature' changes in DNA methylation patterns, an epigenetic mark, occur early in carcinogenesis and thus represent promising markers for early cancer screening [20]. There have been many studies reporting the utility of methylation markers in early cancer detection. As a result, various DNA methylation tests are now commercially available for assessing specific cancers, including Epi proColon® (Epigenomics AG, Berlin, Germany) for colon cancer screening and Cervi-M assay™ (EpiGene, New Taipei City, Taiwan) for the detection of cervical cancer [21,22]. In HCC, several studies have reported promoter hypermethylation leading to silencing of various tumor suppressor genes, including *CDKN2A, RASSF1A* and *GSTP1* [23–25]. Xu *et al.* showed a high accuracy of circulating methylation markers in differentiating HCC patients from healthy subjects by using a panel of ten CpG sites [26]. To date, most of these studies have only examined a single gene or a small number of genes, and the performance of such methylated markers, both individually and in panels, has not achieved the necessary sensitivity and specificity to be feasible for detection of patients with early-stage HCC in at-risk populations. A recent study by Liu *et al.* exploited genome-wide methylation profiles to detect more than 50 cancer types, with an overall specificity of 99.3% and sensitivity of 18, 43, 81 and 93% for stage I, II, III and IV, respectively [27]. However, the evaluation of predictive performance of these assays was mostly based on the discrimination of HCC patients from healthy individuals. Thus, it is essential to develop a

blood-based assay capable of distinguishing early-stage HCC patients from high-risk patients, for whom HCC screening is highly recommended by current guidelines [28].

In the present study we examined differences in circulating methylated DNA profiles between 58 HCC patients with very early (TNM stage I and II) and nonmetastatic stage disease (TNM stage IIIA) and 121 high-risk individuals with liver cirrhosis or chronic hepatitis, in a discovery cohort, using a panel of 450 target regions consisting of 18,000 CpG sites. The resulting HCC-specific methylation regions were employed to build machine learning models for differentiation of HCC patients from high-risk individuals. We subsequently validated the predictive performance of our methylation-based assay either alone or in combination with the µTASWako markers in an independent cohort of 48 HCC and 72 high-risk patients.

Patients & methods

Clinical sample collection

A total of 110 HCC patients and 196 high-risk individuals with liver cirrhosis or hepatitis were enrolled from May 2019 to January 2022. All patients were treatment naive at time of sample collection. HCC stage was determined according to the TNM scoring system recommended by The American Joint Committee on Cancer (8th edition) [29]. Of 110 HCC patients, four were excluded because they were diagnosed with metastatic HCC (stage IV) (Figure 1). High-risk patients with cirrhosis and chronic hepatitis were assessed by imaging test (magnetic resonance elastography, Fibroscan) and serological tests. High-risk patients were confirmed to have no history of cancer at the time of enrollment and were followed up for 6 months after enrollment to ensure that they did not develop HCC. Of 196 recruited high-risk patients, three had blood samples with heavy hemolysis and were excluded from the study. As a result, 106 HCC and 193 high-risk patients met all eligibility criteria and were divided into discovery and validation cohorts (Figure 1). The discovery cohort (n = 179) comprised 58 nonmetastatic HCC patients and 121 high-risk patients, including 18 cirrhosis and 103 hepatitis patients. The validation cohort had a total of 120 participants: 48 nonmetastatic HCC patients and 72 high-risk patients (18 with liver cirrhosis and 54 with hepatitis). All patients in the validation cohort were also subjected to a serological test (µTASWako i30, FujiFilm Healthcare, MA, USA). The µTASWako i30 is a commercially available microfluidic-based clinical assay that was employed to quantitatively measure AFP, AFP-L3 and DCP in human serum to assess the risk of developing primary HCC (Figure 1).

All eligible patients provided written informed consent prior to participation in the study. Patient characteristics are listed in Supplementary Table 1 and summarized in Table 1.

cfDNA isolation

For each patient, 10ml of peripheral blood was collected in a Streck tube (Cell-Free DNA BCT, Streck, NE, USA), and plasma separated by two rounds of centrifugation (2000 × *g* for 10 min and then 16,000 × *g* for 10 min). The plasma fraction was collected, aliquoted (1 ml) and stored at -80°C. cfDNA extraction from the aliquoted plasma was carried out using the MagMAX™ Cell-Free DNA Isolation Kit (Thermo Fisher), following the manufacturer's instructions, and then quantified using the QuantiFluor® dsDNA system (Promega, WI, USA).

Library preparation for targeted methylation

A minimum of 2 ng cfDNA was denatured and bisulfite-converted using an EZ DNA Methylation-Gold™ Kit (Zymo Research, CA, USA). Bisulfite-converted DNA PCR and sequencing were performed by dual-indexed library preparation using Accel-NGS™ Methyl-Seq DNA Library Kit (Swift Biosciences, MI, USA), following the manufacturer's instructions. DNA library concentrations were then quantified using the QuantiFluor dsDNA system (Promega). The Unique Dual Indexing Primers kit of Swift Biosciences (Cat. no. 39096) containing primer mix was used for multiplexing samples for next-generation sequencing.

After library construction, equal amounts of DNA libraries were pooled together, hybridized and captured using a customized panel of xGen® Lockdown® probes, targeting 450 regions and covering 18,000 CpG sites (Integrated DNA Technologies, Inc., IA, USA). The panel was constructed based on previous publications [23,26,30–33] and targeted regions from gene-regulatory regions (including promoters, exons, introns and intergenic regions) involved in cancer progression (Supplementary Table 2). Hybridization was performed using the xGen hybridization and wash kit (Integrated DNA Technologies).

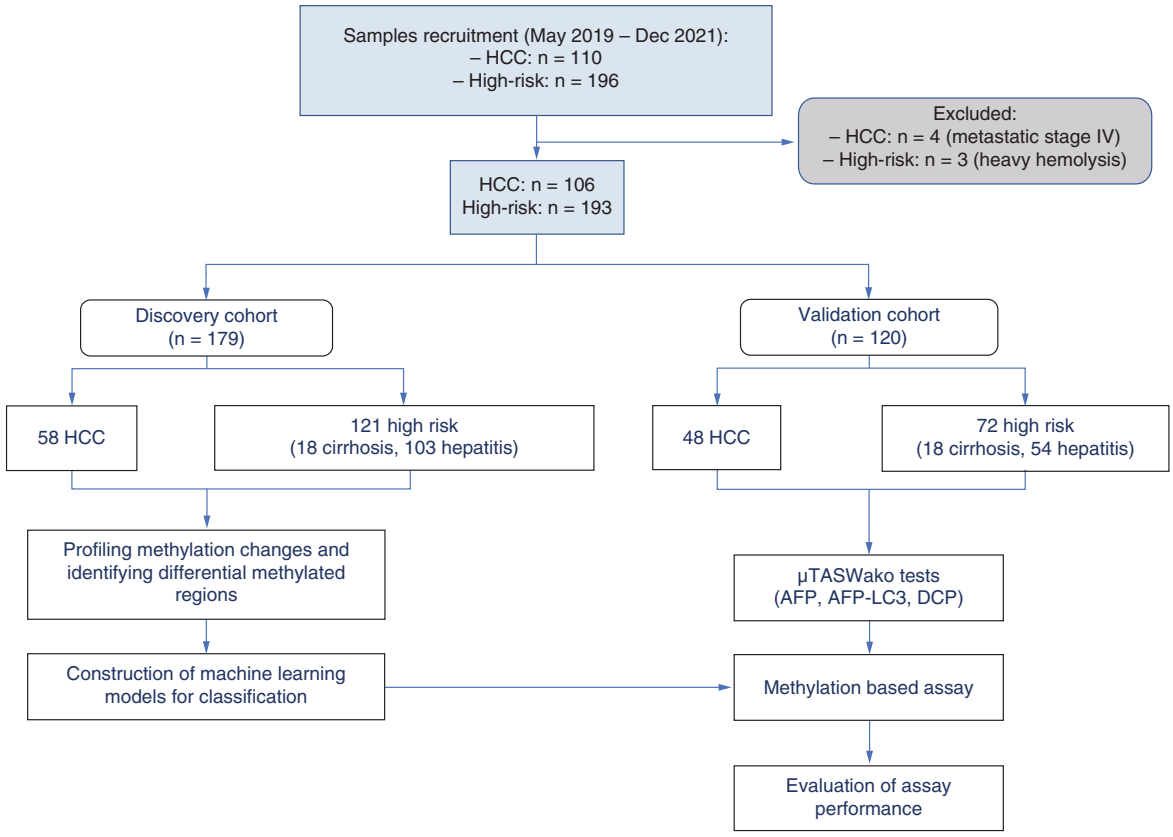


Figure 1. Patient flow diagram and study design. A total of 110 HCC patients and 196 high-risk patients with liver cirrhosis or chronic hepatitis were recruited from May 2019 to December 2021. Of those, four HCC patients diagnosed with metastatic HCC (stage IV, n = 4) and three high-risk patients who had hemolyzed blood samples (n = 3) were excluded. As a result, 106 nonmetastatic HCC and 193 high-risk participants who met all the eligibility criteria were grouped into the discovery and validation cohorts. The discovery cohort, consisting of 58 patients with nonmetastatic HCC and 121 high-risk patients (18 liver cirrhosis and 103 hepatitis), was employed to profile cancer-specific methylation changes and develop classifiers for discrimination of HCC and high-risk patients. The validation cohort, consisting of 48 HCC and 72 high-risk patients (18 cirrhosis and 54 hepatitis) with μTASWako test results, was employed to evaluate the performance of our classification model or its combination with μTASWako test biomarkers in segregating HCC from high-risk patients with liver cirrhosis and hepatitis.
HCC: Hepatocellular carcinoma.

Target methylation analysis

Each sample was run on an DNBSEQ-G400 DNA sequencing system (MGI Tech, Shenzhen, China) to generate sequencing data with 100-bp paired-end reads, at 20 million reads. For each sample, paired-end-read FASTQ files were adapter-trimmed sequencing adapters using trimmomatic [34]. Bisulfite-sequenced reads were then aligned to a reference genome, and methylation calling was performed using the Bismark aligner [35]. Methylation percentages were calculated individually for each region following the equation: %methylation = methylated C / (methylated C + unmethylated C). The methylated cytosines were those cytosines recovered among the CpG dinucleotides on the sequenced reads mapped to the target regions, whereas unmethylated cytosines were converted to thymines within CpG dinucleotides. Based on the methylation percentages of each region, methylation patterns were analyzed to identify differences between the sample groups.

For 450 targeted regions, the Wilcoxon rank sum test was performed for identifying the differentially methylated regions (DMRs) in the pairwise comparisons between HCC and high-risk patients (Figure 1). The p-values were corrected by the Benjamini–Hochberg method for multiple comparisons (with a corrected p cut-off $\alpha \leq 0.05$). Based on the set of genes associated with the DMRs, biologically significant pathways were determined using g:Profiler [36].

Table 1. Summary of patients' clinical features in the discovery and validation cohorts.

Criteria	Discovery cohort (n = 179)					Validation cohort (n = 120)				
	HCC (n = 58)		High risk (n = 121)		HCC vs high risk (p-value)	HCC (n = 48)		High risk (n = 72)		HCC vs high risk (p-value)
	n	%	n	%		n	%	n	%	
Gender:					0.07					0.43
– Female	8	13.8	7	5.8		14	29.2	26	36.1	
– Male	50	86.2	114	94.2		34	70.8	46	63.9	
Age (years):					<0.0001					0.1518
– Median	58		49			61		58		
– Min	23		40			32		21		
– Max	86		86			85		81		
Stage:										
– I	10	17.2								
– II	30	51.7								
– III	3	5.2								
– Unknown	15	25.9								
Risk factor:										
– HBV	43	74.1	77	63.6	0.16	32	66.7	36	50	0.07
– HCV	3	5.2	11	9.1	0.36	13	27.1	16	22.2	0.54
– HBV/HCV	1	1.7	5	4.1	0.4	0	0.0	1	1.4	NA
– Unknown viral infection status	6	10.3	10	8.3	0.64	0	0.0	1	1.4	NA
– Cirrhosis	5	8.6	18	14.9	0.24	3	6.3	18	25	0.018

$\chi 2$ test (sample size >5) was performed to estimate p-value to compare the gender ratios across different patient groups. The one-tailed Mann–Whitney U test was used to compare the median age between HCC and high-risk patients.
%: Percentage of particular cases among the total number of cases.
HBV: Hepatitis B virus; HCC: Hepatocellular carcinoma; HCV: Hepatitis C virus.

Construction of machine learning models

The DMRs selected from the pairwise comparisons of HCC versus high-risk patients were subsequently used as input features to construct machine learning models for distinguishing HCC from non-HCC patients. We implemented three popular algorithms for binary classification: logistic regression (LR), support vector machine (SVM) and extreme gradient boosting (XGB). Hyperparameter tuning was performed for each learning algorithm by using the grid-search function in the scikit-learn library (v. 1.1.1) [37]. After feature selection and hyperparameter tuning, the leave-one-out crossvalidation was applied to the discovery dataset to evaluate our models. Specifically, each algorithm was trained by all samples in the discovery cohort minus one, while the remaining sample was used for blind classification. This process was repeated for all samples until each sample was predicted once. The receiver operating characteristic (ROC) curve analysis and area under the ROC Curve (AUC) values were used as the main metrics to evaluate the predictive performance of each model. The model cut-off was set based on the threshold specificity of >90% or >95% to meet the requirement for early cancer detection assays [38,39]. During the independent validation phase, optimal hyperparameters found from the grid-search were fixed and applied to predict samples in the validation cohort to evaluate the model's ability to discriminate HCC patients from at-risk patients.

Statistical analysis

The Wilcoxon rank sum test was performed for comparing the methylation score values across different groups of patients. The p-values were corrected by the Benjamini–Hochberg method for multiple comparisons (with a corrected p cut-off $\alpha \leq 0.05$). The $\chi 2$ test was performed to compare the gender ratios between HCC and high-risk patients. The one-tailed Mann–Whitney U test was used to compare the median age across different patient groups (Table 1).

DeLong's test was used to compare the difference between AUCs (Table 2) [40]. All statistical analyses were carried

Table 2. Classification performance of methylation assay, GALAD model and their combination on the validation cohort.

Feature	AUC	Sensitivity (%)	Specificity (%)	Cut-off
Methylation markers	0.82 (0.80–0.91)	60.4 (56.3–73.0)	95.8 (88.9–98.9)	>0.67
GALAD	0.80 (0.78–0.86)	56.3 (52.1–69.3)	95.8 (88.9–98.9)	>4.619
Methylation markers + GALAD	0.87 (0.85–0.94)	68.8 (64.5–80.1)	95.8 (88.9–98.9)	>0.503

AUC: Area under the receiver operating characteristic curve; GALAD: Lectin-reactive AFP, AFP, des-γ-carboxy prothrombin, gender and age.

out using Python (v. 3.7) with some common data analysis packages (numpy, scipy, pandas). The Scikit-learn library (v. 1.1.1) was used to implement machine learning algorithms [37].

Results

Clinical characteristics of HCC patients & high-risk participants in discovery & validation cohorts

For the discovery cohort a total of 179 eligible patients, including 58 nonmetastatic HCC and 121 high-risk patients (18 liver cirrhosis and 103 chronic hepatitis), were enrolled from May 2019 to December 2021 (Figure 1). Both HCC and high-risk groups had higher percentages of male patients than female patients (Table 1). HCC patients had a significantly higher median age than high-risk patients (58 vs 49 years; $p < 0.0001$; Table 1). The majority of HCC patients (68.9%) were diagnosed with early-stage HCC (17.2% for stage I and 51.7% for stage II), while 5.2% had nonmetastatic stage IIIA HCC. The remaining HCC patients (25.9%) with unknown stages were confirmed by clinicians to have nonmetastatic tumors by imaging diagnosis. Most HCC patients (74.1%) were positive for HBV infection. The high-risk group comprised patients diagnosed with liver cirrhosis (14.9%), chronic hepatitis B (63.6%), hepatitis C (9.1%) or both (4.1%). Of those high-risk patients, 8.3% had chronic hepatitis with unknown viral infection status.

To independently validate the performance of our assay, we recruited an additional 120 subjects – 48 HCC patients and 72 high-risk patients (18 with liver cirrhosis and 54 with hepatitis) – from another hospital, from July 2019 to January 2022 (Figure 1). HCC patients in this cohort visited the hospital for diagnostic imaging examinations and were confirmed to have nonmetastatic HCC. Subsequently, those patients went to other hospitals for treatment, and therefore tumor tissue samples were not available for histopathological analysis. Despite the missing histological information, those patients were confirmed by trained clinicians to have nonmetastatic HCC by imaging diagnosis. There was no significant difference in the distribution of age and gender between the discovery and validation cohorts for HCC patients, but high-risk patients in the validation cohort had a significantly higher proportion of female patients (36.1 vs 5.8%; $p < 0.00001$; Table 1) and median age (58 vs 49 years; $p < 0.0001$; Table 1) than high-risk patients in the discovery cohort.

Patients with nonmetastatic HCC display aberrant methylation changes in circulating cfDNA

To identify HCC-specific methylation signatures of circulating cfDNA, as candidate biomarkers for HCC early screening in high-risk populations, we performed pairwise comparisons of 58 HCC versus 121 high-risk patients in the discovery cohort. We profiled methylation levels of 450 target regions which were selected based on their potential association with carcinogenesis, as reported in previous publications (Supplementary Table 2) [23,26,30–33]. Methylation levels of these regions were determined by bisulfite next-generation sequencing, with comparable sequencing quality control metrics including mapped read percentages of >80%, median-mapped depth of coverage of >100× and on-target ratios of >94% for the two groups ($p > 0.05$; Supplementary Figure 1A & Supplementary Table 3). To demonstrate the ability of our assay to accurately quantify methylation percentages of these regions, we created a series of standard samples by spiking different percentages of synthetic fully methylated DNA into unmethylated DNA (D5014, Zymo Research). After fragmenting 2 ng of each standard sample to create fragments of 100–200 bp (corresponding to the general length of plasma cfDNA), we tested them in our assay. We observed a good linear relationship ($R^2 = 0.99$; Supplementary Figure 1B) between the assay and spike-in mean methylation percentages across standard samples, indicating the high accuracy of our assay for quantification of methylation percentages. The lowest level that could be reliably detected by our assay was 1%.

HCC patients showed higher DNA methylation levels, in multiple regions, than high-risk patients, with 295/450 DMRs ($p < 0.05$, Wilcoxon rank-sum test, Benjamini–Hochberg false discovery rate; Figure 2A). The majority of these regions (255/295; 86.4%) were significantly hypermethylated, with 68 DMRs having twofold or greater

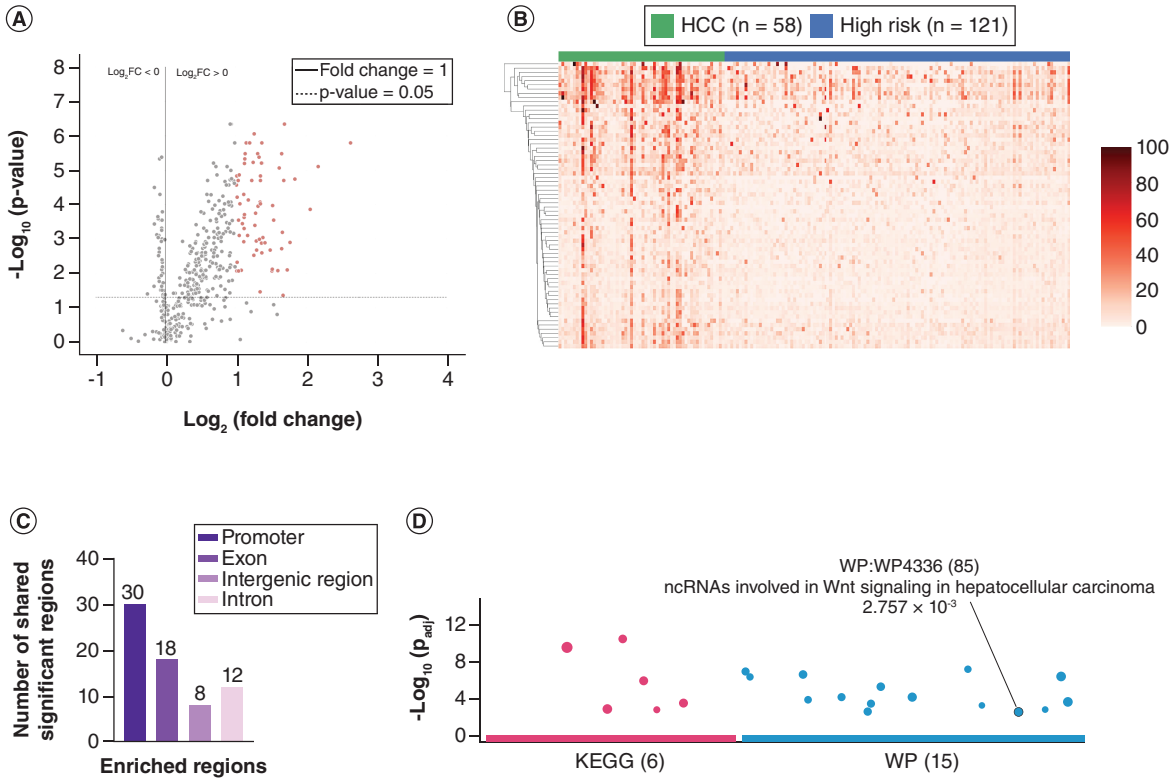


Figure 2. Aberrant circulating DNA methylation changes in liquid biopsy samples from patients with nonmetastatic hepatocellular carcinoma, compared with high-risk patients. (A) Volcano plot showing log2 fold change (x-axis) and significance (-log10 * adjusted p-value; y-axis) of differentially methylated regions (DMRs). Wilcoxon's rank sum test was used to identify DMRs between different patient groups ($p < 0.05$, corrected by Benjamini–Hochberg method). Dashed lines represent regions with log2 fold change >0 and p-value < 0.05. DMRs with log2 fold change >1 and p-value < 0.05 are shown in red. **(B)** Heat map showing 68 DMRs with log2 fold change >1 in plasma cell-free DNA samples of 58 HCC and 121 high-risk patients. Heat map visualization was analyzed with ComplexHeatmap package v. 2.8.0. Regions with different levels of methylation changes are clustered. **(C)** Distribution of the genomic regions enriched by the 68 DMRs differentiating HCC patients from high-risk patients. **(D)** KEGG and WP enrichment analysis using g:Profiler (https://biit.cs.ut.ee/gprofiler) for genes associated with the DMRs differentiating HCC patients from high-risk patients.
HCC: Hepatocellular carcinoma; KEGG: Kyoto Encyclopedia of Genes and Genomes; WP: WikiPathways.

hypermethylation which could differentiate HCC from high-risk patients (Figure 2B & Supplementary Table 4). Of those DMRs, 30 were in promoter regions, while 18, eight and 12 mapped to exons, intergenic regions and introns, respectively (Figure 2C). We performed pathway enrichment analysis for the set of genes mapped to these DMRs using gProfiler and identified 21 significantly enriched pathways, six from the Kyoto Encyclopedia of Genes and Genomes and 15 from the WikiPathway databases (Figure 2D & Supplementary Table 5). One such pathway was found to regulate the expression of noncoding RNAs involved in Wnt signaling in HCC (Supplementary Table 5), suggesting that DMRs in cfDNA may be reflective of methylation profiles in liver tumorigenesis. These findings indicated that cfDNA shed by early and nonmetastatic tumors are hypermethylated at multiple regulatory regions in cancer-associated genes and thus could serve as specific markers for distinguishing HCC patients from non-HCC high-risk patients.

Construction of machine learning models for distinguishing HCC patients from high-risk patients

The identification of cancer-specific methylation signatures in plasma cfDNA of HCC patients prompted us to assess whether a machine learning classifier with methylation markers could be constructed to detect HCC. After performing Wilcoxon's rank sum test to compare the levels of methylation percentages (methylation density) across 450 targeted regions between HCC versus high-risk patients to select the significant DMRs, these significant DMRs (n = 295) were used to construct three different machine learning algorithms (LR, SVM and XGB) which

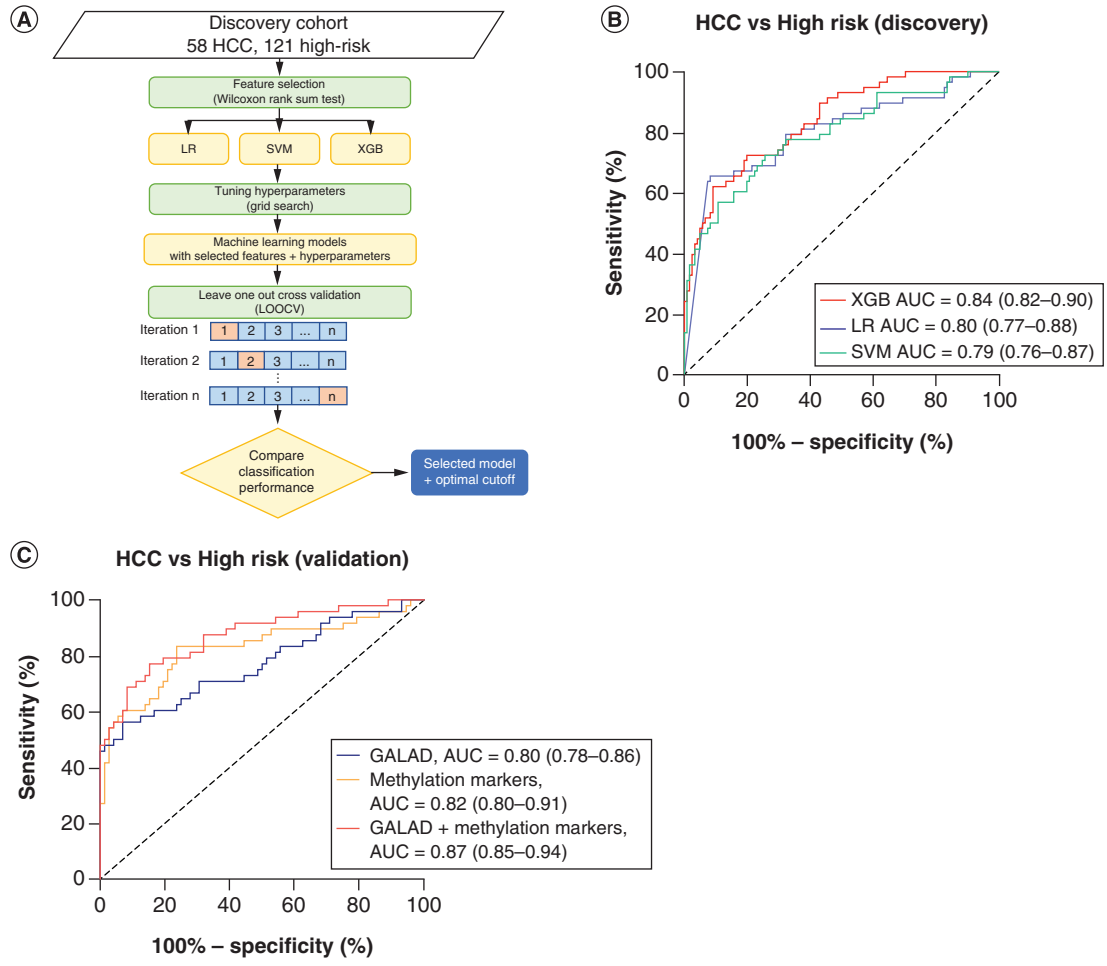


Figure 3. Methylation-based machine learning model improves accuracy of the μTASWako test in discriminating hepatocellular carcinoma patients from cirrhosis and hepatitis patients. (A) Model construction strategy. Differentially methylated regions that could differentiate hepatocellular carcinoma (HCC) patients from non-HCC patients were used to build three different machine learning algorithms including logistic regression (LR), support vector machine (SVM) and XGBoost (XGB). For each algorithm, hyperparameters were tuned using grid-search (scikit-learn packages), followed by evaluation of model performance using a leave-one-out crossvalidation. **(B)** Receiver operating characteristic curves showing area under the curve values and accuracy of three different classifiers (LR, SVM and XGB) in discriminating HCC patients from high-risk patients as a whole group (cirrhosis and hepatitis patients) during leave-one-out crossvalidation (LOOCV). **(C)** Receiver operating characteristic curves showing analysis of methylation-based XGB model, GALAD model alone or the combination of GALAD score with methylation markers in discriminating HCC (n = 48) from high-risk patients (n = 72) in the validation cohort.
AUC: Area under the curve; GALAD: Lectin-reactive AFP, AFP, des-γ-carboxy prothrombin, gender and age.

are widely adopted for binary classification [41]. For each algorithm, we utilized the grid-search function to tune the hyperparameters (Figure 3A). The classification performance of each algorithm with selected hyperparameters was tested using leave-one-out crossvalidation. During this process, we used ROC curves and AUC values to evaluate the accuracy of each model in classification of HCC patients from high-risk patients. The three tested models yielded comparable AUC values for discriminating HCC patients from high-risk patients in the discovery cohort (XGB, AUC: 0.84; 95% CI: 0.82–0.90; LR, AUC: 0.80; 95% CI: 0.77–0.88; SVM, AUC: 0.79; 95% CI: 0.76–0.87; Figure 3B). Although not statistically significant ($p > 0.05$, DeLong's test), XGB showed a slightly higher accuracy for discriminating nonmetastatic HCC patients from high-risk patients than the SVM and LR models; therefore we selected this model for subsequent validation studies. To assess the predictive power of our methylation-based assay in early detection of HCC patients, we performed ROC analysis for stage I and stage II HCC patients. We found that our methylation-based assay achieved an AUC of 0.81 (95% CI: 0.78–0.85; Supplementary Figure 2) for discriminating stage I or II HCC patients from high-risk individuals, which was comparable to the overall

predictive performance for HCC patients with all tested stages in the discovery cohort (AUC: 0.84; 95% CI: 0.82–0.90; Supplementary Figure 2). Thus our results showed that the methylation-based assay enabled detection of HCC patients with very early stage disease (stage I and II).

The XGB model serves as a 'black box' classifier, which computes a probability value (methylation score) for a given sample to estimate the probability of the patient having HCC. By imposing an optimal cut-off on the methylation score, we can decide whether a sample is HCC or not. The selection of optimal cut-off points was based on the benefits, cost and probability of developing disease [38,39]. For an early cancer detection assay, high specificity is recommended to minimize the rate of false positives which lead to overdiagnosis and unnecessary diagnostic workups. As such, we selected the optimal cut-offs to achieve specificity of >90 and >95%. The assay achieved sensitivities of 62.1 (95% CI: 55.2–71.4) and 55.2% (95% CI: 51.7–58.8) at specificities of 90.9 (95% CI: 89.3–94.9) and 95.0% (95% CI: 92.6–97.7), respectively, for discriminating HCC patients from high-risk patients.

We next locked the model parameters and applied the model to predict the samples in an independent validation cohort consisting of 48 HCC patients and 72 high-risk patients (23 liver cirrhosis and 49 hepatitis) recruited from another hospital. The methylation-based model showed consistent performance between the discovery and validation cohorts, with AUC of 0.82 (95% CI: 0.80–0.91; Figure 3C) for discriminating HCC patients from high-risk patients. These results demonstrated that the methylation-based assay could achieve consistently high accuracy in distinguishing HCC patients from at-risk patients during the independent validation phase.

Circulating DNA methylation markers enhance the accuracy of the μTASWako assay in for detecting HCC patients

Patients diagnosed with HCC, liver cirrhosis or hepatitis in the validation cohort were also advised to perform a commercial serological test, μTASWako, which is currently used in clinic to measure three different serum biomarkers (AFP, AFP-L3 and DCP) [42]. It has been reported that the incorporation of the three μTASWako biomarkers with patients' age and gender (collectively known as the GALAD score) could enhance diagnostic power [13]. Consistent with previous studies, in our study the GALAD model could distinguish HCC patients from high-risk patients, with AUC of 0.80 (95% CI: 0.78–0.86, Figure 3C). Interestingly, the combination of our methylation-based assay with the GALAD score provided additive diagnostic performance, with AUCs of 0.87 (95% CI: 0.85–0.94; Figure 3C & Table 2). Furthermore, the statistical analysis by DeLong's test showed that the combination of GALAD scores and methylation markers significantly increased the AUC value compared with the GALAD model alone ($p = 0.0480$). Compared with the GALAD score alone, the combination of GALAD and methylation scores showed comparable specificity of 95.8% (95% CI: 88.9–98.9; Table 2), but higher sensitivity: 68.8 (95% CI: 64.5–80.1) versus 56.3% (95% CI: 52.1–69.3) for GALAD score alone (Table 2) for classification of HCC patients from those with liver cirrhosis and hepatitis. Thus these findings suggest that the HCC-specific methylation markers can be added to the currently available serological tests to improve the performance of these tests for the early detection of HCC.

Association between the methylation scores & patients' clinical features

It has been shown that differences in patients' gender, age and viral infection status can introduce biases to the analysis of methylation signatures [43–45]. Therefore we examined the effect of those parameters on the prediction results of our assay in the discovery cohort. We found that high-risk patients had significantly lower methylation scores than HCC patients, irrespective of their gender ($p < 0.001$; Figure 4A). Likewise, we did not observe any significant difference in the predicted methylation scores between elderly patients aged over the median (57 years) and younger patients (age <57 years) among the two groups (Figure 4B). Moreover, the differences in methylation scores between HCC patients and high-risk individuals were not dependent on HBV or HCV infection status (Figure 4C), indicating that the assay could distinguish HCC patients from high-risk patients regardless of their viral infection status. Notably, when the median tumor diameter (3.4 cm) was used as a cut-off value, HCC patients with tumor diameter greater than 3.4 cm showed significantly higher methylation scores than those with tumor diameter less than 3.4 cm ($p < 0.001$; Figure 4D). Our findings demonstrated that HCC tumors with diameter greater than 3 cm are more likely to be detected by our methylation-based model than those with diameter less than 3 cm. These results suggest that the performance of our methylation assay for detecting HCC patients might not be confounded by the differences in patients' gender, age or viral infection status but might be influenced by tumor size.

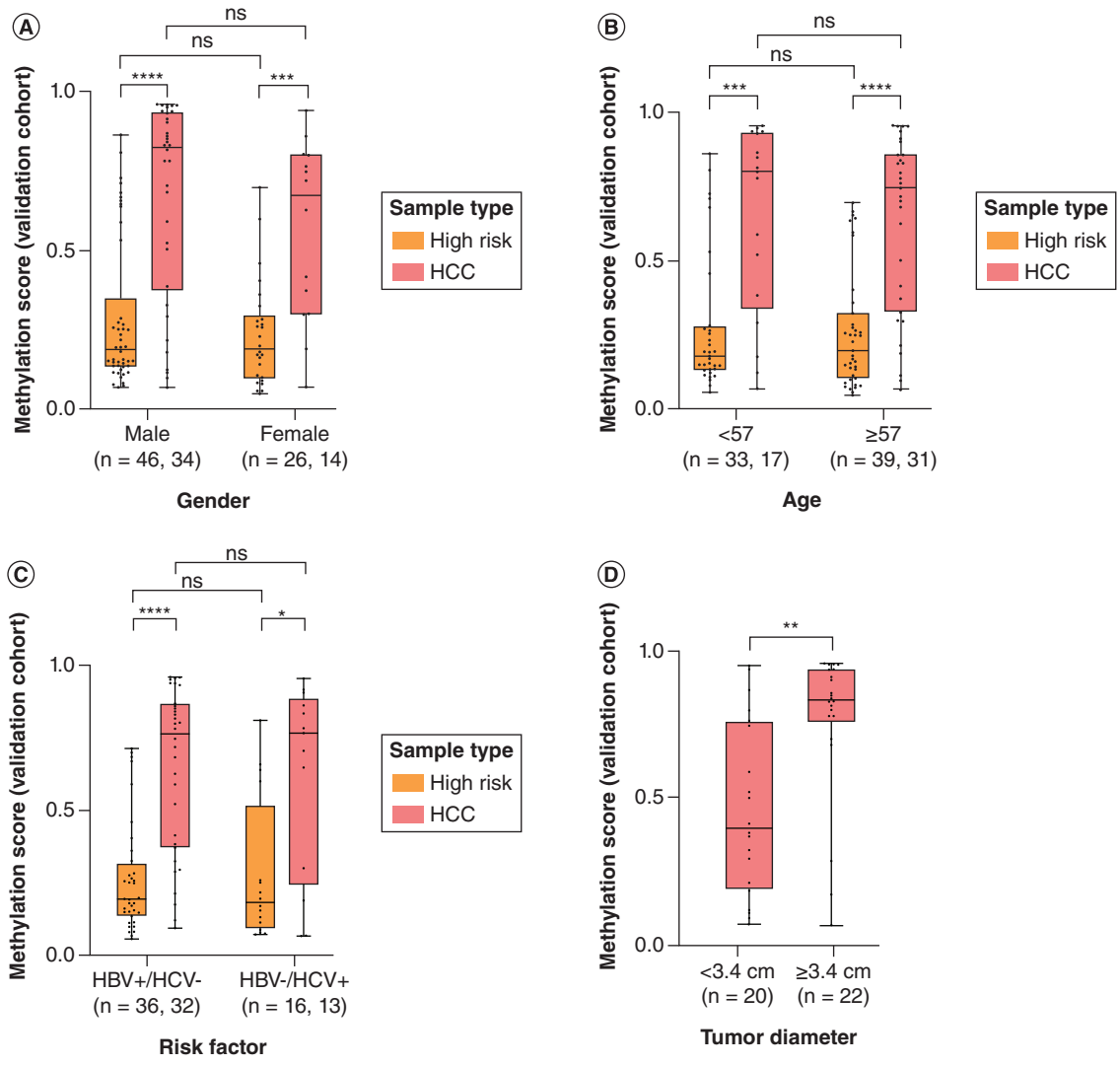


Figure 4. Association between patients' clinical features and the prediction performance. (A & B) Box plots showing methylation assay scores stratified by **(A)** gender and **(B)** age of high-risk (n = 72) and hepatocellular carcinoma (HCC) patients (n = 48) in the validation cohort. The median age of all 120 patients in this cohort was selected as a cut-off value for comparisons. **(C)** Box plots showing methylation assay scores by viral infection status of HCC and high-risk patients in the validation cohort. **(D)** Box plots showing methylation assay scores by tumor size. The median tumor diameter of HCC patients in the validation cohort was selected as a cut-off value for comparison. Wilcoxon's rank sum test was performed to compare the methylation scores across different patient groups.
*$p < 0.05$; **$p < 0.001$; ***$p < 0.0001$; ****$p < 0.00001$.
HBV: Hepatitis B virus; HCC: Hepatocellular carcinoma; HCV: Hepatitis C virus; ns: Not significant.

Discussion

According to the current guidelines from the American Association for the Study of Liver Diseases [46], screening for patients with very early (stage I and II) or nonmetastatic (stage IIIA) HCC in populations with a high prevalence of HBV infection is of significance to the improvement of survival rates. None of the currently used blood-based biomarkers provides sufficient sensitivity and specificity for the early detection of HCC. Thus the combination of serum biomarkers with imaging tests such as liver ultrasonography has been explored and shown to increase sensitivity, but this also increases false-positive rates [8,47].

Alterations in DNA methylation, which are associated with dysregulation of tumor-associated genes, occur in the early stages of carcinogenesis and thus potentially represent unique tumor 'signatures' for cancer early detection [48]. In this study we profiled the methylation status of a panel of 450 target regions, based on previous publications [23,26,30–33], in plasma samples from HCC patients versus high-risk patients with chronic liver cirrhosis

and hepatitis, identifying 295 DMRs. The majority of these DMRs were hypermethylated, consistent with previous findings of widespread hypermethylation of tumor suppressor genes in HCC [49–51] and hepatobiliary cancer [52], possibly associated with overexpression of specific DNA methyltransferases [24,25,53]. Of the DMR-associated genes, we identified enrichment of numerous cancer-related signaling pathways (Supplementary Table 5). Of particular interest were those related to the Wnt pathway, a signal cascade well implicated in HCC to play a role in transition from chronic cirrhosis or hepatitis to hepatocellular adenomas and HCC [54–56].

We next exploited the identified DMRs to build three different binary classifiers (LR, SVM and XGB) to segregate HCC patients from high-risk patients. The XGB model had greater AUC values than LR and SVM for distinguishing early-stage HCC from high-risk patients, and was therefore selected for the following validation in an independent cohort of 120 patients (48 HCC patients and 72 high-risk individuals). During the validation phase, the methylation-based XGB model showed comparable performance to its performance in the discovery cohort for distinguishing HCC from high-risk patients, confirming effective modeling and minimal overfitting.

To further evaluate the capacity of our assay to distinguish HCC patients from high-risk patients, we compared the predictive performance of our methylation assay with the GALAD model, which incorporates the three serum biomarkers used in a commercial HCC diagnostic test, μTASWako, with patients' gender and age. While this model has exhibited high sensitivity and specificity, it has been hypothesized that each serum marker may be derived from a different HCC subtype [57]. For example, AFP-L3 associates with aggressive HCC [58], while DCP is a more ideal marker of intrahepatic metastasis [59]. Serum AFP can also be elevated by hepatic inflammation from chronic viral hepatitis and nonalcoholic fatty liver disease [60,61]. These various confounders could limit the use of the serum biomarker panel for detecting early HCC [62–64], although a machine learning classifier (using gradient boosting) for serum AFP, AFP-L3 and DCP reduced misclassification by 50% compared with the use of each marker alone [65]. In this study we demonstrated that our methylation-based model could distinguish HCC patients from those with cirrhosis and hepatitis, with classification performance comparable to the combination of all three μTASWako markers in the GALAD model. Interestingly, we found that the combination of our assay and the μTASWako biomarkers could provide additive power for classifying HCC patients from high-risk individuals by ~7% in terms of AUC, suggesting that circulating methylation signatures could serve as useful biomarkers to develop a blood test for early HCC detection. Among 48 HCC cases in the validation cohort, there were 33 true positives and 15 false negatives. The two groups were indistinguishable in their clinical characteristics (gender ratio, median age and viral infection status; Supplementary Table 6). We noted that false-negative cases had significantly lower levels of serum biomarkers (AFP, AFP-L3 and DCP; $p < 0.05$), GALAD scores ($p < 0.0001$) and methylation scores ($p < 0.0001$) than true positive cases (Supplementary Table 6), possibly attributed to the misclassification of those cases by the model combining our methylation and GALAD biomarkers. Of 72 high-risk cases, only three were misidentified as cancer cases, and they had different clinical characteristics. Thus it was not possible to draw conclusions on the contribution of clinical characteristics to the misprediction of those false-positive cases.

Although our model achieved comparable performance to that of the methylation-based assay reported by Xu *et al.* [26], our performance was reported on a cohort of nonmetastatic HCC and high-risk individuals, in contrast to the cohort of metastatic late-stage HCC used by Xu *et al.* In addition to methylation signatures, multiple signatures of circulating tumor DNA, including fragment-length patterns and end motifs, have been shown to serve as potential biomarkers that can differentiate HCC patients from non-HCC individuals. Jiang *et al.* conducted an analysis on fragment lengths of plasma DNA in HCC patients and showed that HCC patients had markedly higher ratios of short (<150 bp) to long (>150 bp) cfDNA fragments [66]. In a recent study, the same team examined the 4-mer motifs at the 5′ end of cfDNA fragments in HCC patients and healthy individuals and found significant differences in frequencies of certain motifs between the two groups [67]. By using these signatures to construct classification models, they could distinguish HCC patients from non-HCC individuals with an AUC of 0.86 [67]. These findings provide a rationale for exploiting these signatures in combination with our methylation signatures to enhance accuracy in detecting HCC. This topic is currently being explored by our group.

There are several limitations to this study. First, this was a retrospective study, with small sample sizes for each group in both the discovery and validation cohorts. Thus our current study might be considered an exploratory analysis; future studies with a larger prospective cohort are required for robust validation of the clinical performance of our identified methylation markers. Nonetheless, our data clearly show that DNA methylation markers could be combined with other serum biomarkers to improve the accuracy of early HCC detection. Second, cancer patients and high-risk patients showed significant differences in gender ratios and median ages in the validation cohort, consistent with previous reports [12,50]. However, we did not observe any significant association between these

patients' characteristics (e.g., gender, age or HBV infection status) and prediction scores (Figure 4), suggesting that our model might not be confounded by these factors. Finally, tumor staging records were not available for some HCC patients in the discovery and validation cohorts.

Conclusion

In conclusion, we showed that liver cancer-specific methylation signatures detected in cfDNA could be exploited to build machine learning classifiers for segregating nonmetastatic HCC patients from at-risk individuals (i.e., those with cirrhosis or hepatitis). Our study also provides a rationale for using cfDNA methylation signatures in combination with serum biomarkers to maximize the accuracy and feasibility of early, routine HCC screening for at-risk individuals, which remains a major clinical challenge. Furthermore, our findings also have major implications for combining those blood-based biomarkers with imaging tests to achieve greater efficacy for early HCC detection. Additionally, there may be a future clinical trial to ascertain whether the DNA methylation profiles in combination with the GALAD biomarkers (AFP, AFP-L3 and DCP) can identify early-stage HCC that can be imaged and treated.

Summary points

- Liver cancer was the third-most lethal malignancy in 2020. The most common liver cancer is hepatocellular carcinoma (HCC).
- The 5-year survival rates of HCC patients detected at a very early stage (Barcelona Clinic Liver Cancer 0/A, or American Joint Committee on Cancer tumor node metastasis [TNM] stage I) are 70%, but drop to 3% for those diagnosed in advanced stages with metastatic tumors.
- A commercial diagnostic test, μTASWako® i30, which assesses serum levels of AFP, lectin-reactive AFP and des-γ-carboxy prothrombin, demonstrated up to 78% sensitivity but 62% specificity for early HCC detection, which is not efficient in differentiating HCC patients from high-risk patients (patients with cirrhosis/hepatitis).
- The use of circulating tumor DNA and its unique features as a noninvasive and highly sensitive marker for HCC detection has now been evaluated by several studies.
- Of those features of circulating tumor DNA, changes in DNA methylation patterns occur early in carcinogenesis and thus represent promising markers for early cancer screening.
- We examined differences in circulating methylated DNA profiles between HCC patients (early stages [TNM stage I and II] and nonmetastatic stage [TNM stage IIIA]) and high-risk individuals with liver cirrhosis or chronic hepatitis, using a panel of 450 target regions consisting of 18,000 CpG sites.
- The machine learning model built from those differentially methylated regions achieved high accuracy on a validation cohort. When combined with serum biomarkers, a sensitivity of 68.8% at 95.8% specificity was reached.
- HCC-specific circulating DNA methylation markers may be added to the currently available assay to improve the early detection of HCC.

Supplementary data

To view the supplementary data that accompany this paper please visit the journal website at: www.futuremedicine.com/doi/suppl/10.2217/fon-2022-1218

Author contributions

Conceptualization: DK Truong, H Giang, HN Nguyen, MD Phan, LS Tran. Sampling and clinical consulting: TH Phan, TTT Pham, VC Nguyen, TD Ho, THT Tran, ML Duong, HPT Bach, VV Kim, TA Pham, BT Nguyen, TNV Nguyen, TD Nguyen, DTB Phu, BHH Phan, DS Nguyen, TTT Do. Methodology: VTC Nguyen, TMQ Pham, NDK Le. Software: TH Nguyen, TD Nguyen, H Giang. Formal analysis: VTC Nguyen, TMQ Pham, NDK Le. Data curation: TD Nguyen, H Giang. Writing – original draft preparation: LS Tran. Writing – review and editing: HN Nguyen, LS Tran. Supervision: HN Nguyen, LS Tran, MD Phan. All authors have read and agreed to the published version of the manuscript.

Acknowledgments

The authors thank all participants who agreed to take part in this study, all the clinics and hospitals who assisted in patient consultation and sample collection, and the laboratory technicians at the Medical Genetics Institute in Ho Chi Minh City for performing the assays. The authors thank C Balch of Bioscience Advising, MI, USA, for proofreading the manuscript.

Financial & competing interests disclosure

This study is funded by Gene Solutions. The funder Gene Solutions provided support in the form of salaries for authors VTC Nguyen, TTT Pham, TMQ Pham, TD Nguyen, NDK Le, TH Nguyen, H Giang, HN Nguyen, MD Phan and LS Tran. We also confirm that this does not alter our adherence to Future Oncology journal policies on sharing data and materials. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Ethical conduct of research

The study was conducted in accordance with the Declaration of Helsinki, and approved by the National Cancer Institutes, Vietnam and MEDIC Medical Center, Vietnam. Informed consent was obtained from all people involved in the study.

Data sharing statement

The data presented in this study are available on request from the corresponding author. The data are not publicly available due to ethical regulation.

References

Papers of special note have been highlighted as: • of interest; •• of considerable interest

1. Sung H, Ferlay J, Siegel RL *et al.* Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J. Clin.* 71(3), 209–249 (2021).
2. Dasgupta P, Henshaw C, Youlden DR, Clark PJ, Aitken JF, Baade PD. Global trends in incidence rates of primary adult liver cancers: a systematic review and meta-analysis. *Front Oncol.* 10, 171 (2020).
3. Brar G, Greten TF, Graubard BI *et al.* Hepatocellular carcinoma survival by etiology: a SEER-Medicare database analysis. *Hepatol. Commun.* 4(10), 1541–1551 (2020).

• **Database presenting the 5-year survival rates of different types of cancer stratified by stage at diagnosis.**

4. Llovet JM, Kelley RK, Villanueva A *et al.* Hepatocellular carcinoma. *Nat. Rev. Dis. Primers* 7(1), 6 (2021).

• **Review on different risk factors and pathophysiology for hepatocellular carcinoma (HCC).**

5. National Cancer Institute Surveillance, Epidemiology, and End Results Program. *The SEER Program Code Manual* (3rd Edition) (2003).
6. Tsuchiya N, Sawada Y, Endo I, Saito K, Uemura Y, Nakatsura T. Biomarkers for the early diagnosis of hepatocellular carcinoma. *World J. Gastroenterol.* 21(37), 10573–10583 (2015).
7. Singal A, Volk ML, Waljee A *et al.* Meta-analysis: surveillance with ultrasound for early-stage hepatocellular carcinoma in patients with cirrhosis. *Aliment. Pharmacol. Ther.* 30(1), 37–47 (2009).
8. Giannini EG, Cucchetti A, Erroi V, Garuti F, Odaldi F, Trevisani F. Surveillance for early diagnosis of hepatocellular carcinoma: how best to do it? *World J. Gastroenterol.* 19(47), 8808–8821 (2013).
9. Samant H, Amiri HS, Zibari GB. Addressing the worldwide hepatocellular carcinoma: epidemiology, prevention and management. *J. Gastrointest. Oncol.* 12(Suppl. 2), S361–S373 (2021).
10. Masuzaki R, Karp SJ, Omata M. New serum markers of hepatocellular carcinoma. *Semin. Oncol.* 39(4), 434–439 (2012).
11. Liebman HA, Furie BC, Tong MJ *et al.* Des-gamma-carboxy (abnormal) prothrombin as a serum marker of primary hepatocellular carcinoma. *N. Engl. J. Med.* 310(22), 1427–1431 (1984).
12. Marrero JA, Feng Z, Wang Y *et al.* Alpha-fetoprotein, des-gamma carboxyprothrombin, and lectin-bound alpha-fetoprotein in early hepatocellular carcinoma. *Gastroenterology* 137(1), 110–118 (2009).
13. Best J, Bilgi H, Heider D *et al.* The GALAD scoring algorithm based on AFP, AFP-L3, and DCP significantly improves detection of BCLC early stage hepatocellular carcinoma. *Z. Gastroenterol.* 54(12), 1296–1305 (2016).
14. Adeniji N, Dhanasekaran R. Current and emerging tools for hepatocellular carcinoma surveillance. *Hepatol. Commun.* 5(12), 1972–1986 (2021).
15. Johnson AM, Dudek JM, Edwards DK *et al.* Analytical validation of a novel multi-target blood-based test to detect hepatocellular carcinoma. *Expert Rev. Mol. Diagn.* 21(11), 1245–1252 (2021).
16. Cohen JD, Li L, Wang Y *et al.* Detection and localization of surgically resectable cancers with a multi-analyte blood test. *Science* 359(6378), 926–930 (2018).

17. Bettegowda C, Sausen M, Leary RJ *et al.* Detection of circulating tumor DNA in early- and late-stage human malignancies. *Sci. Transl. Med.* 6(224), 224ra224 (2014).
18. Phallen J, Sausen M, Adleff V *et al.* Direct detection of early-stage cancers using circulating tumor DNA. *Sci. Transl. Med.* 9(403), eaan2415 (2017).
19. Lo YMD, Han DSC, Jiang P, Chiu RWK. Epigenetics, fragmentomics, and topology of cell-free DNA in liquid biopsies. *Science* 372(6538), eaaw3616 (2021).
20. Roy D, Tiirikainen M. Diagnostic power of DNA methylation classifiers for early detection of cancer. *Trends Cancer* 6(2), 78–81 (2020).
21. Warren JD, Xiong W, Bunker AM *et al.* Septin 9 methylated DNA is a sensitive and specific blood test for colorectal cancer. *BMC Med.* 9, 133 (2011).
22. Galardi F, Luca F, Romagnoli D *et al.* Cell-free DNA-methylation-based methods and applications in oncology. *Biomolecules* 10(12), 1677 (2020).
23. Peng JL, Wu JZ, Li GJ *et al.* Association of *RASSF1A* hypermethylation with risk of HBV/HCV-induced hepatocellular carcinoma: a meta-analysis. *Pathol. Res. Pract.* 216(10), 153099 (2020).
24. Zhou Y, Wang XB, Qiu XP, Shuai Z, Wang C, Zheng F. *CDKN2A* promoter methylation and hepatocellular carcinoma risk: a meta-analysis. *Clin. Res. Hepatol. Gastroenterol.* 42(6), 529–541 (2018).
25. Wang J, Qin Y, Li B, Sun Z, Yang B. Detection of aberrant promoter methylation of *GSTP1* in the tumor and serum of Chinese human primary hepatocellular carcinoma patients. *Clin. Biochem.* 39(4), 344–348 (2006).
26. Xu RH, Wei W, Krawczyk M *et al.* Circulating tumour DNA methylation markers for diagnosis and prognosis of hepatocellular carcinoma. *Nat. Mater.* 16(11), 1155–1161 (2017).

•• **A useful workflow for finding differential methylation regions between HCC and healthy plasma samples.**

27. Liu MC, Oxnard GR, Klein EA, Swanton C, Seiden MV. Sensitive and specific multi-cancer detection and localization using methylation signatures in cell-free DNA. *Ann. Oncol.* 31(6), 745–759 (2020).
28. Francica G, Borzio M. Status of, and strategies for improving, adherence to HCC screening and surveillance. *J. Hepatocell. Carcinoma* 6, 131–141 (2019).
29. Amin MB, Greene FL, Edge SB *et al.* The Eighth Edition AJCC Cancer Staging Manual: Continuing to build a bridge from a population-based to a more 'personalized' approach to cancer staging. *CA Cancer J. Clin.* 67(2), 93–99 (2017).
30. Nguyen HN, Cao NT, Van Nguyen TC *et al.* Liquid biopsy uncovers distinct patterns of DNA methylation and copy number changes in NSCLC patients with different *EGFR*-TKI resistant mutations. *Sci. Rep.* 11(1), 16436 (2021).
31. Chen X, Gole J, Gore A *et al.* Non-invasive early detection of cancer four years before conventional diagnosis using a blood test. *Nat. Commun.* 11(1), 3475 (2020).

•• **Demonstrated the use of targeted methylation profiling in the detection of multiple types of cancer.**

32. Devos T, Tetzner R, Model F *et al.* Circulating methylated *SEPT9* DNA in plasma is a biomarker for colorectal cancer. *Clin. Chem.* 55(7), 1337–1346 (2009).
33. Shirahata A, Hibi K. Serum vimentin methylation as a potential marker for colorectal cancer. *Anticancer Res.* 34(8), 4121–4125 (2014).
34. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15), 2114–2120 (2014).
35. Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for bisulfite-seq applications. *Bioinformatics* 27(11), 1571–1572 (2011).
36. Raudvere U, Kolberg L, Kuzmin I *et al.* g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). *Nucleic Acids Res.* 47(W1), W191–w198 (2019).
37. Pedregosa F, Varoquaux G, Gramfort A *et al.* Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* 12(null), 2825–2830 (2011).
38. Baker SG. Improving the biomarker pipeline to develop and evaluate cancer screening tests. *J. Natl Cancer Inst.* 101(16), 1116–1119 (2009).
39. Ignatiadis M, Sledge GW, Jeffrey SS. Liquid biopsy enters the clinic – implementation issues and future challenges. *Nat. Rev. Clin. Oncol.* 18(5), 297–312 (2021).
40. Delong ER, Delong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* 44(3), 837–845 (1988).
41. Liu J, Wu J, Liu S, Li M, Hu K, Li K. Predicting mortality of patients with acute kidney injury in the ICU using XGBoost model. *PLOS ONE* 16(2), e0246306 (2021).
42. Chaiteerakij R, Zhang X, Addissie BD *et al.* Combinations of biomarkers and Milan criteria for predicting hepatocellular carcinoma recurrence after liver transplantation. *Liver Transpl.* 21(5), 599–606 (2015).
43. Hall E, Volkov P, Dayeh T *et al.* Sex differences in the genome-wide DNA methylation pattern and impact on gene expression, microRNA levels and insulin secretion in human pancreatic islets. *Genome Biol.* 15(12), 522 (2014).
44. Bocklandt S, Lin W, Sehl ME *et al.* Epigenetic predictor of age. *PLOS ONE* 6(6), e14821 (2011).

45. Rongrui L, Na H, Zongfang L, Fanpu J, Shiwen J. Epigenetic mechanism involved in the HBV/HCV-related hepatocellular carcinoma tumorigenesis. *Curr. Pharm. Des.* 20(11), 1715–1725 (2014).
46. Sherman M. Liver cancer screening in high-risk populations. *Hepat. Oncol.* 3(1), 13–18 (2016).
47. Tzartzeva K, Obi J, Rich NE *et al.* Surveillance imaging and alpha fetoprotein for early detection of hepatocellular carcinoma in patients with cirrhosis: a meta-analysis. *Gastroenterology* 154(6), 1706–1718.e1701 (2018).
48. Constancio V, Nunes SP, Henrique R, Jeronimo C. DNA methylation-based testing in liquid biopsies as detection and prognostic biomarkers for the four major cancer types. *Cells* 9(3), 624 (2020).

• **Review summarizing the clinical use of circulating DNA for pan-cancer analysis.**

49. Sceusi EL, Loose DS, Wray CJ. Clinical implications of DNA methylation in hepatocellular carcinoma. *HPB (Oxford)* 13(6), 369–376 (2011).
50. Villanueva A, Portela A, Sayols S *et al.* DNA methylation-based prognosis and epidrivers in hepatocellular carcinoma. *Hepatology* 61(6), 1945–1956 (2015).
51. Wang X, Cheng Y, Yan LL, An R, Wang XY, Wang HY. Exploring DNA methylation profiles altered in cryptogenic hepatocellular carcinomas by high-throughput targeted DNA methylation sequencing: a preliminary study for cryptogenic hepatocellular carcinoma. *Onco Targets Ther.* 13, 9901–9916 (2020).

• **Overview of the predictive performance of current serum biomarkers in HCC diagnosis.**

52. Hamilton JP. Epigenetic mechanisms involved in the pathogenesis of hepatobiliary malignancies. *Epigenomics* 2(2), 233–243 (2010).
53. Park HJ, Yu E, Shim YH. DNA methyltransferase expression and DNA hypermethylation in human hepatocellular carcinoma. *Cancer Lett.* 233(2), 271–278 (2006).
54. He S, Tang S. WNT/beta-catenin signaling in the development of liver cancers. *Biomed. Pharmacother.* 132, 110851 (2020).
55. Khalaf AM, Fuentes D, Morshid AI *et al.* Role of Wnt/beta-catenin signaling in hepatocellular carcinoma, pathogenesis, and clinical significance. *J. Hepatocell. Carcinoma* 5, 61–73 (2018).
56. Russell JO, Monga SP. Wnt/beta-catenin signaling in liver development, homeostasis, and pathobiology. *Annu. Rev. Pathol.* 13, 351–378 (2018).
57. Wang X, Zhang Y, Yang N *et al.* Evaluation of the combined application of AFP, AFP-L3%, and DCP for hepatocellular carcinoma diagnosis: a meta-analysis. *Biomed. Res. Int.* 2020, 5087643 (2020).
58. Leerapun A, Suravarapu SV, Bida JP *et al.* The utility of *Lens culinaris* agglutinin-reactive alpha-fetoprotein in the diagnosis of hepatocellular carcinoma: evaluation in a United States referral population. *Clin. Gastroenterol. Hepatol.* 5(3), 394–402 (2007).
59. Marrero JA, Su GL, Wei W *et al.* Des-gamma carboxyprothrombin can differentiate hepatocellular carcinoma from nonmalignant chronic liver disease in american patients. *Hepatology* 37(5), 1114–1121 (2003).
60. Babalı A, Cakal E, Purnak T *et al.* Serum α-fetoprotein levels in liver steatosis. *Hepatol. Int.* 3(4), 551–555 (2009).
61. Chen CH, Lin ST, Kuo CL, Nien CK. Clinical significance of elevated alpha-fetoprotein (AFP) in chronic hepatitis C without hepatocellular carcinoma. *Hepatogastroenterology* 55(85), 1423–1427 (2008).
62. Babali A, Cakal E, Purnak T *et al.* Serum alpha-fetoprotein levels in liver steatosis. *Hepatol. Int.* 3(4), 551–555 (2009).
63. Li D, Mallory T, Satomura S. AFP-L3: a new generation of tumor marker for hepatocellular carcinoma. *Clin. Chim. Acta* 313(1–2), 15–19 (2001).
64. Sterling RK, Jeffers L, Gordon F *et al.* Utility of *Lens culinaris* agglutinin-reactive fraction of alpha-fetoprotein and des-gamma-carboxy prothrombin, alone or in combination, as biomarkers for hepatocellular carcinoma. *Clin. Gastroenterol. Hepatol.* 7(1), 104–113 (2009).
65. Sato M, Morimoto K, Kajihara S *et al.* Machine-learning approach for the development of a novel predictive model for the diagnosis of hepatocellular carcinoma. *Sci. Rep.* 9(1), 7704 (2019).

•• **Study exploiting the combination of multiple markers to increase screening accuracy. The HCC machine learning classifier was built by using extra clinical information as features, including serum biomarker levels, risk factors, age, gender, height and weight.**

66. Jiang P, Chan CW, Chan KC *et al.* Lengthening and shortening of plasma DNA in hepatocellular carcinoma patients. *Proc. Natl Acad. Sci. USA* 112(11), E1317–1325 (2015).
67. Jiang P, Sun K, Peng W *et al.* Plasma DNA end-motif profiling as a fragmentomic marker in cancer, pregnancy, and transplantation. *Cancer Discov.* 10(5), 664–673 (2020).