

Circulating tumour DNA methylation markers for diagnosis and prognosis of hepatocellular carcinoma

Rui-hua Xu^{1*†}, Wei Wei^{1,2†}, Michal Krawczyk^{2†}, Wenqiu Wang^{2†}, Huiyan Luo^{1,2†}, Ken Flagg², Shaohua Yi², William Shi², Qingli Quan³, Kang Li³, Lianghong Zheng⁴, Heng Zhang⁵, Bennett A. Caughey², Qi Zhao¹, Jiayi Hou², Runze Zhang², Yanxin Xu³, Huimin Cai^{3,4}, Gen Li^{3,4}, Rui Hou⁴, Zheng Zhong², Danni Lin², Xin Fu², Jie Zhu², Yaou Duan², Meixing Yu³, Binwu Ying⁶, Wengeng Zhang³, Juan Wang⁷, Edward Zhang², Charlotte Zhang², Oulan Li², Rongping Guo¹, Hannah Carter², Jian-kang Zhu⁵, Xiaoke Hao⁷ and Kang Zhang^{2,3,8*}

An effective blood-based method for the diagnosis and prognosis of hepatocellular carcinoma (HCC) has not yet been developed. Circulating tumour DNA (ctDNA) carrying cancer-specific genetic and epigenetic aberrations may enable a noninvasive ‘liquid biopsy’ for diagnosis and monitoring of cancer. Here, we identified an HCC-specific methylation marker panel by comparing HCC tissue and normal blood leukocytes and showed that methylation profiles of HCC tumour DNA and matched plasma ctDNA are highly correlated. Using cfDNA samples from a large cohort of 1,098 HCC patients and 835 normal controls, we constructed a diagnostic prediction model that showed high diagnostic specificity and sensitivity ($P < 0.001$) and was highly correlated with tumour burden, treatment response, and stage. Additionally, we constructed a prognostic prediction model that effectively predicted prognosis and survival ($P < 0.001$). Together, these findings demonstrate in a large clinical cohort the utility of ctDNA methylation markers in the diagnosis, surveillance, and prognosis of HCC.

Hepatocellular carcinoma (HCC) is a leading cause of cancer deaths worldwide¹. As with many cancers, HCC found at an early stage carries much-improved prognosis compared to advanced stage disease², in part due to the relative efficacy of local treatments compared with systemic therapy. Thus, early detection has significant potential for reducing the mortality of HCC. Unfortunately, there has been little success in developing effective blood-based methods to screen for HCC. Alpha fetal protein (AFP) is the only currently available blood test for detection and surveillance of HCC; however, its clinical utility is limited by low sensitivity³.

Circulating tumour DNA (ctDNA) consists of extracellular nucleic acid fragments shed into plasma via tumour cell necrosis, apoptosis, and active release of DNA⁴. Recent research demonstrates that ctDNA has the potential to revolutionize screening, diagnosis, and treatment of cancer by enabling a noninvasive ‘liquid biopsy’—that is, a blood test that enables molecular testing of solid malignancies^{5,6}. Compared to tissue biopsy, cell-free DNA (cfDNA) sequencing has some obvious advantages. First, the collection of peripheral blood to obtain cfDNA is minimally invasive compared with tumour biopsy, regardless of site. Second, blood can be taken at any time during therapy, allowing for real-time and dynamic monitoring of molecular changes in tumours rather than depending on the challenges of invasive biopsy or even imaging. Furthermore,

monitoring of cfDNA may detect tumour that is not apparent or is indeterminate on imaging (for example, residual tumour post-resection). Finally, ctDNA may represent the entire molecular picture of a patient’s malignancy, while a tumour biopsy may be affected by intra-tumour heterogeneity.

DNA methylation is an epigenetic regulator of gene expression that usually results in gene silencing⁷. Increased methylation of tumour suppressor genes is an early event in many tumours, suggesting that altered DNA methylation patterns could be one of the first detectable neoplastic changes associated with tumorigenesis^{8–10}. ctDNA-bearing cancer-specific methylation patterns have been investigated as feasible biomarkers in cancers¹¹; however, currently there are few validated methylation markers available, such as *SEPT9* in colorectal cancer¹². DNA methylation profiling offers several advantages over somatic mutation analysis for cancer detection, including higher clinical sensitivity and dynamic range, many methylation target regions in diseases, and multiple altered CpG sites within each targeted genomic region. Further, each methylation marker is present in both cancer tissue and cfDNA, whereas only a fraction of mutations present in cancer tissue may be detected in cfDNA¹³.

Obtaining reliable and quantitative measurements of methylation values in a minimum amount of cfDNA remain challenging;

¹State Key Laboratory of Oncology in South China, Collaborative Innovation Center of Cancer Medicine, Sun Yat-sen University Cancer Center, Guangzhou 510060, China. ²Moore’s Cancer Center and Institute for Genomic Medicine, University of California, San Diego, La Jolla, California 92093, USA. ³Molecular Medicine Research Center, West China Hospital, Sichuan University, Chengdu 610041, China. ⁴Guangzhou Youze Biological Pharmaceutical Technology Company Ltd., Guangzhou 510005, China. ⁵Shanghai Center for Plant Stress Biology, Shanghai Institute for Biological Sciences, Chinese Academy of Sciences, Shanghai 210602, China. ⁶Department of Clinical Laboratory Medicine, West China Hospital, Sichuan University, Chengdu 610041, China. ⁷Department of Clinical Laboratory Medicine, Xijing Hospital, the Fourth Military Medical University, Xi’an, Shaanxi 710032, China. ⁸Veterans Administration Healthcare System, San Diego, California 92093, USA. [†]These authors contributed equally to this work. *e-mail: xurh@susucc.org.cn; k5zhang@ad.ucsd.edu

more sensitive assays need to be developed. It is hypothesized that adjacent CpG sites in the same DNA strand may be modified by a methyltransferase or demethylase together¹⁴. These adjacent stretches of CpG methylation, which we refer to as a methylation correlated blocks (MCBs), are similar in concept to haplotype blocks of adjacent single nucleotide polymorphisms (SNPs) in DNA sequence variations and have the potential to enhance the accuracy of methylation allele calling.

In this study, to evaluate the potential of ctDNA methylation markers in diagnosis and prognosis of HCC, we compared differential methylation profiles of HCC tissues and blood leukocytes in normal individuals by analysing 485,000 CpG markers, and identified a methylation marker panel enriched in HCC. After validation of this panel in matched HCC tumour DNA and plasma cfDNA within the same patients, we employed multiple statistical methods to develop diagnostic and prognostic prediction models with selected methylation markers. We further compared the efficacy of methylation marker-based models and current available approaches, such as AFP and TNM staging classification, in the diagnosis and prognosis of HCC in 1098 HCC and 885 normal samples. These results show that ctDNA methylation analysis may be reliable biomarkers in the diagnosis, surveillance, and prognosis of HCC.

Patient and sample characteristics

Clinical characteristics and molecular profiling including methylation data for comparison between HCC and blood lymphocytes were assembled from sources including 377 HCC tumour samples from The Cancer Genome Atlas (TCGA) and 754 blood leukocyte samples of healthy control individuals from a data set used in our previous methylation study on ageing (GSE40279)¹⁵. To study ctDNA in HCC, plasma samples were obtained from Chinese patients with HCC and randomly selected healthy controls undergoing routine health care maintenance, resulting in a training cohort of 715 HCC patients and 560 normal healthy controls and a validation cohort of 383 HCC patients and 275 healthy controls. All participants provided written informed consent. Clinical characteristics of all patients and controls are listed in Supplementary Table 1.

Methylation markers for differentiating HCC and blood

We hypothesized that CpG markers with a maximal difference in methylation between HCC and blood leukocytes in normal individuals would be most likely to demonstrate detectable methylation differences in the cfDNA of HCC patients when compared to that of normal controls. We used the ‘moderated *t*-statistics’ method with Empirical Bayes for shrinking the variance¹⁶, and the Benjamini–Hochberg procedure¹⁷ to control the false discovery rate (FDR) at a significance level of 0.05 to identify the top 1,000 markers with the most significantly different rates of methylation (that is, those with the lowest *p* values) between HCC and blood. Unsupervised hierarchical clustering of these top 1,000 markers was able to distinguish between HCC and blood leukocytes in normal individuals (Supplementary Fig. 1). We designed molecular-inversion (padlock) probes corresponding to these 1,000 markers and tested them in 28 pairs of HCC tissue DNA and matched plasma ctDNA from the same patient. The methylation profiles in HCC tumour DNA and matched plasma ctDNA were consistent (Supplementary Fig. 2a,b). 401 markers with a good experimental amplification profile and dynamic methylation range were selected for further analysis.

Methylation block structure for allele-calling accuracy

We employed the well-established concept of genetic linkage disequilibrium (LD block) to study the degree of co-methylation among different DNA strands^{18,19}, with the underlying assumption that DNA sites in close proximity are more likely to be co-methylated than distant sites. We used paired-end Illumina sequencing reads to identify each individual methylation block

(mBlock). We applied a Pearson correlation method to quantify co-methylation or mBlock²⁰. We compiled all common mBlocks of a region by calculating different mBlock fractions (see Methods). We then partitioned the genome into blocks of tightly co-methylated CpG sites we termed methylation correlated blocks (MCBs), using an r^2 cutoff of 0.5. We then surveyed MCBs in cfDNA of 500 normal samples and found that MCBs are highly consistent. We next determined methylation levels within an MCB in the cfDNA from 500 HCC samples. We found a highly consistent methylation pattern in MCBs when comparing normal versus HCC cfDNA samples, which significantly enhanced allele-calling accuracy (Supplementary Fig. 3). This technique was employed in all subsequent sequencing analysis.

cfDNA diagnostic prediction for HCC

The methylation values of the 401 selected markers that showed good methylation ranges in cfDNA samples were analysed by Random Forest and Least Absolute Shrinkage and Selection Operator (LASSO) methods to further reduce the number of markers by modelling them in 715 HCC ctDNA and 560 normal cfDNA samples (Fig. 1, see Methods). We obtained 24 markers using the Random-Forest analysis. We also obtained 30 markers using a LASSO analysis in which we required selected markers to appear over 450 times out of a total of 500 repetitions. There were ten overlapping markers between these two methods (Table 1). Using a logistic regression method, we constructed a diagnostic prediction model with these ten markers. Applying the model yielded a sensitivity of 85.7% and specificity of 94.3% for HCC in the training data set of 715 HCC and 560 normal samples (Fig. 2a) and a sensitivity of 83.3% and specificity of 90.5% in the validation data set of 383 HCC and 275 normal samples (Fig. 2b). We also demonstrated this model could differentiate HCC from normal controls both in the training data set (AUC = 0.966) and the validation data set (AUC = 0.944) (Fig. 2c,d). Unsupervised hierarchical clustering of these ten markers was able to distinguish HCC from normal controls with high specificity and sensitivity (Fig. 2e,f and Supplementary Fig. 4).

We next assessed a combined diagnostic score (cd-score) of the model for differentiating between liver diseases (hepatitis B virus/hepatitis C virus (HBV/HCV) infection, and fatty liver) and HCC, since these liver diseases are known major risk factors for HCC. We found that the cd-score could differentiate HCC patients from those with liver diseases or healthy controls (Fig. 3a). These results were consistent and comparable with those predicted by AFP levels (Supplementary Fig. 5a).

Methylation markers predicted clinical outcomes

We next studied the utility of the cd-score in assessing treatment response, the presence of residual tumour following treatment, and staging of HCC. Clinical and demographic characteristics, such as age, gender, race, and American Joint Committee on Cancer (AJCC) stage were included in the analysis. The cd-scores of patients with detectable residual tumour following treatment ($n = 828$) were significantly higher than those with no detectable tumour ($n = 270$), and both were significantly greater than normal controls ($n = 835$) ($p < 0.0001$, Fig. 3b). Similarly, cd-scores were significantly higher in patients before treatment ($n = 109$) or with progression ($n = 381$) compared to those with treatment response ($n = 248$) ($p < 0.0001$, Fig. 3c). In addition, cd-scores were significantly lower in patients with complete tumour resection after surgery ($n = 170$) compared with those before surgery ($n = 109$), yet were higher in patients with recurrence ($n = 155$) ($p < 0.0001$, Fig. 3d). Furthermore, there is good correlation between the cd-scores and tumour stage. Patients with early stage disease (I, II) had substantially lower cd-scores compared to those with advanced stage disease (III, IV) ($p < 0.05$, Fig. 3e). Collectively, these results suggest that the cd-score (that is, the amount of ctDNA in plasma) correlates well with tumour burden

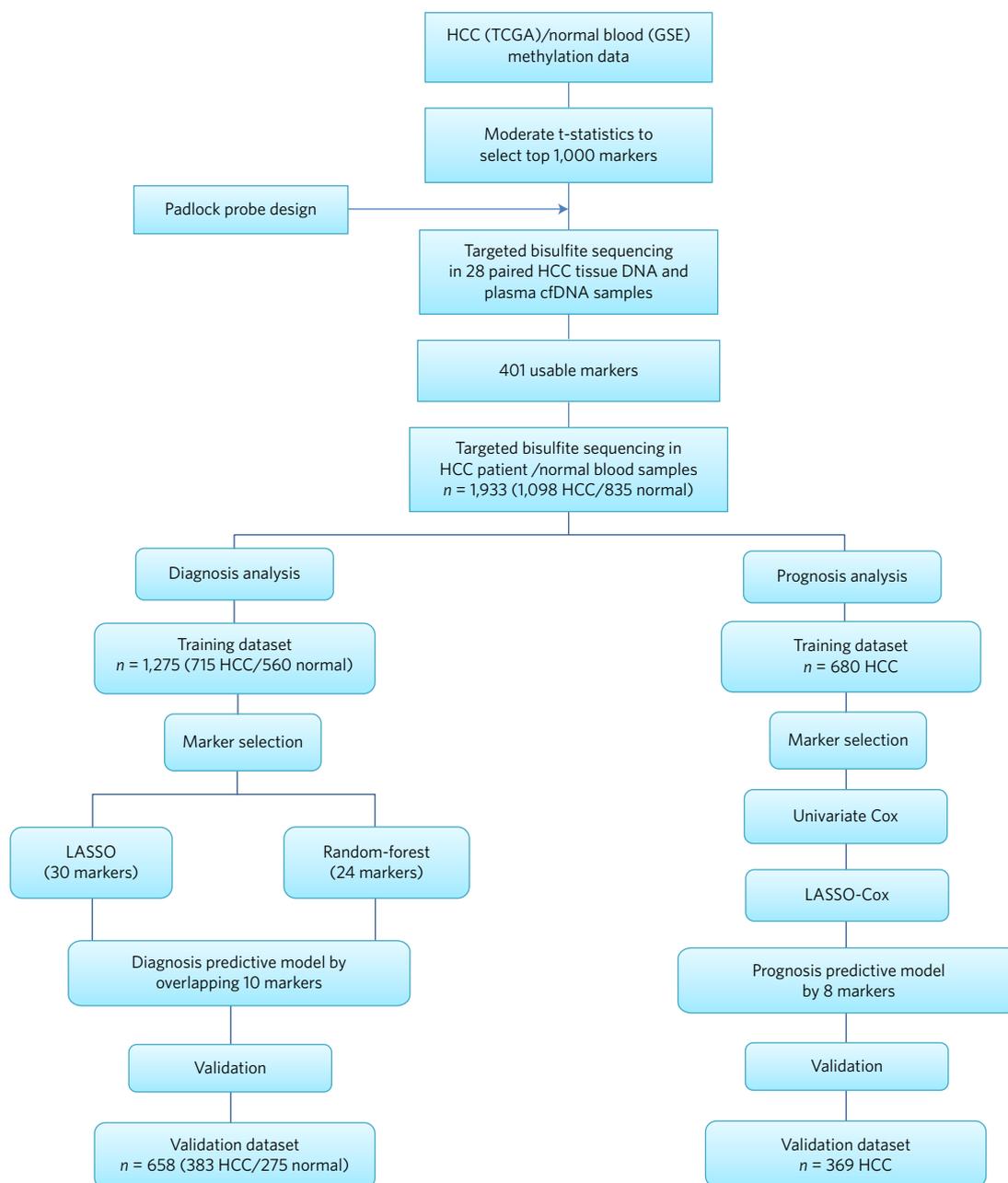


Figure 1 | Workflow chart of data generation and analysis. Whole genome methylation data on HCC and normal lymphocytes were used to identify 401 candidate markers. Left panel: diagnostic marker selection: LASSO and random-forest analyses were applied to a training cohort of 715 HCC and 560 normal patients to identify a final selection of ten markers. These ten markers were applied to a validation cohort of 383 HCC and 275 normal patients. Right panel: prognostic marker selection: Univariate-Cox and LASSO-Cox were applied to a training cohort of 680 HCC patients with survival data to identify a final selection of eight markers. These eight markers were applied to a validation cohort of 369 HCC with survival data.

and may have utility in predicting tumour response and surveillance for recurrence.

Utility of ctDNA diagnostic prediction and AFP

Currently, the only blood biomarker for risk assessment and surveillance of HCC is serum AFP levels. However, its low sensitivity makes it inadequate to detect all patients that will develop HCC and severely limits its clinical utility. In fact, many cirrhotic patients develop HCC without any increase in AFP levels. Strikingly, 40% patients of our HCC study cohort have a normal serum AFP ($<25 \text{ ng ml}^{-1}$).

In biopsy-proven HCC patients, the cd-score demonstrated superior sensitivity and specificity than AFP for HCC diagnosis

(AUC 0.969 versus 0.816, Fig. 3f). In patients with treatment response, tumour recurrence, or progression, cd-score showed more significant changes compared to testing at initial diagnosis than AFP (Supplementary Fig. 5b,c). In patients with serial samples, those with a positive treatment response had a concomitant significant decrease in cd-score compared to that prior to treatment, and there was an even further decrease in patients after surgery. By contrast, our patients with progressive or recurrent disease all had an increase in cd-score (Supplementary Fig. 6). By comparison, AFP was less sensitive for assessing treatment efficacy in individual patients (Supplementary Fig. 7). In addition, while cd-score correlated well with tumour stage (Supplementary Fig. 5d), particularly among patients with stage I, II and III, there was no significant

Table 1 | Characteristics of ten methylation markers and their coefficients in HCC diagnosis.

Markers	Ref Gene	Coefficients	SE	z value	p value
cg10428836	BMP1A	15.595	2.395	6.513	<0.001
cg26668608	PSD	11.543	0.885	-13.040	<0.001
cg25754195	ARHGAP25	4.557	0.889	5.129	<0.001
cg05205842	KLF3	2.519	0.722	3.487	<0.001
cg11606215	PLAC8	-3.612	0.954	-3.785	<0.001
cg24067911	PLAC8	6.865	1.095	6.271	<0.001
cg24067911	ATXN1	-5.439	0.868	-6.265	<0.001
cg18196829	Chr 6:170	-9.078	1.355	-6.698	<0.001
cg23211949	Chr 6:3	-5.209	1.081	-4.819	<0.001
cg17213048	ATAD2	6.660	1.422	4.683	<0.001
cg25459300	Chr 8:20	1.994	1.029	1.938	0.053

SE: standard errors of coefficients; z value: Wald z-statistic value.

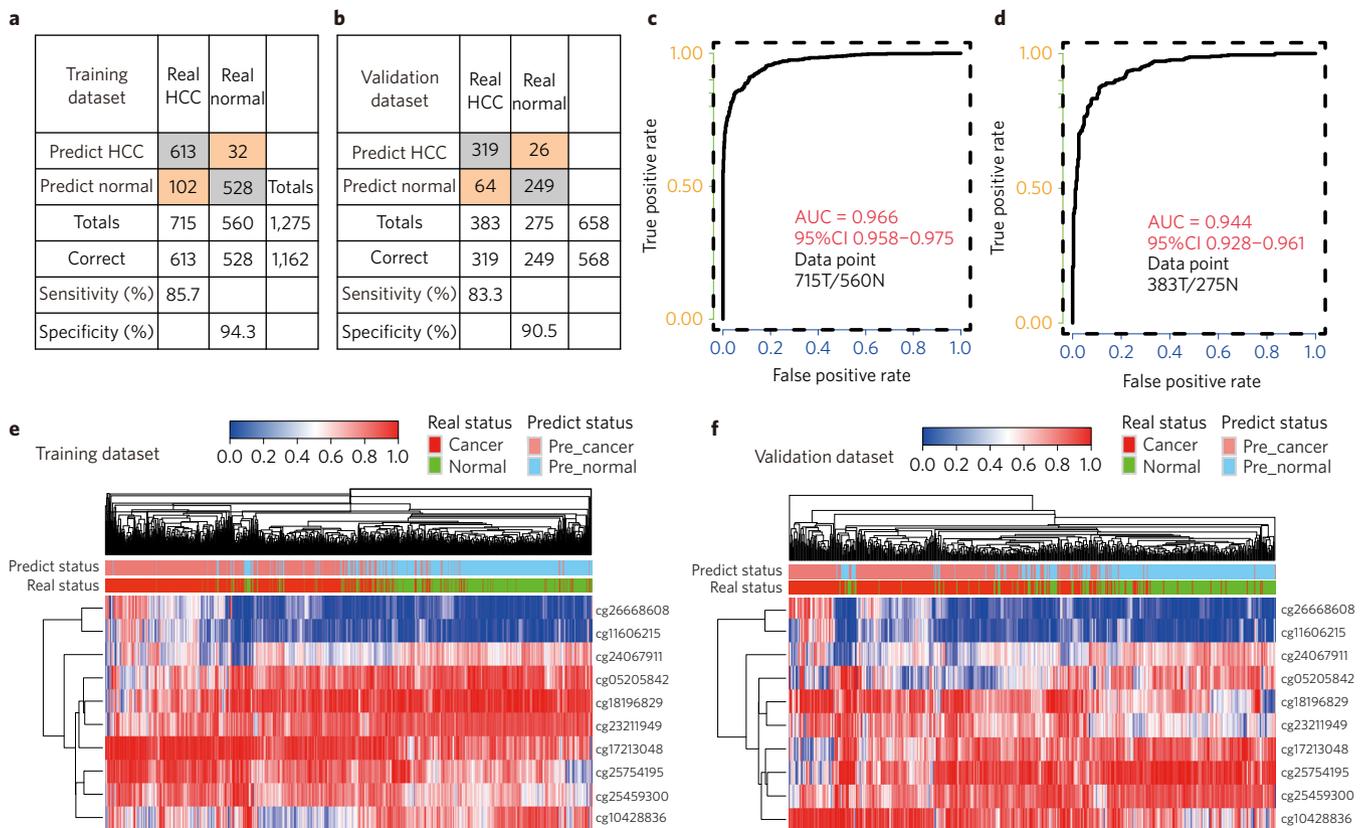


Figure 2 | cfDNA methylation analysis of HCC diagnosis. **a, b**, Confusion tables of binary results of the diagnostic prediction model in the training (**a**) and validation data sets (**b**). **c, d**, ROC of the diagnostic prediction model with methylation markers in the training (**c**) and validation data sets (**d**). **e, f**, Unsupervised hierarchical clustering of ten methylation markers selected for use in the diagnostic prediction model in the training (**e**) and validation data sets (**f**).

difference in AFP values in patients with different stages, except between patients with stage III and IV (Supplementary Fig. 5e), indicating an advantage of cd-score over AFP in differentiation of early stage HCC.

ctDNA prognostic prediction for HCC

We then investigated the potential of using methylation markers in ctDNA for prediction of prognosis in HCC in combination with clinical and demographic characteristics including age, gender, race, and AJCC stage. We randomly split the 1049 HCC patients with complete survival information into training and validation data sets with an allocation of 2:1. We implemented UniCox and LASSO-Cox methods to reduce the dimensionality and constructed a Cox-model

to predict prognosis with an 8-marker panel (Table 2). We generated Kaplan–Meier curves in training and validation data sets using a combined prognosis score (cp-score) with these markers. The high-risk group (cp-score > -0.24) had 341 observations with 53 events in the training data set and 197 observations with 26 events in the validation data set; and the low-risk group (cp-score ≤ -0.24) has 339 observations with 7 events in the training data set and 172 observations with 9 events in the validation data set. Median survival was significantly different in both the training set ($p < 0.0001$) and the validation set ($p = 0.0014$) by log-rank test (Fig. 4a,b).

Multivariate variable analysis showed that the cp-score was significantly correlated with risk of death both in the training

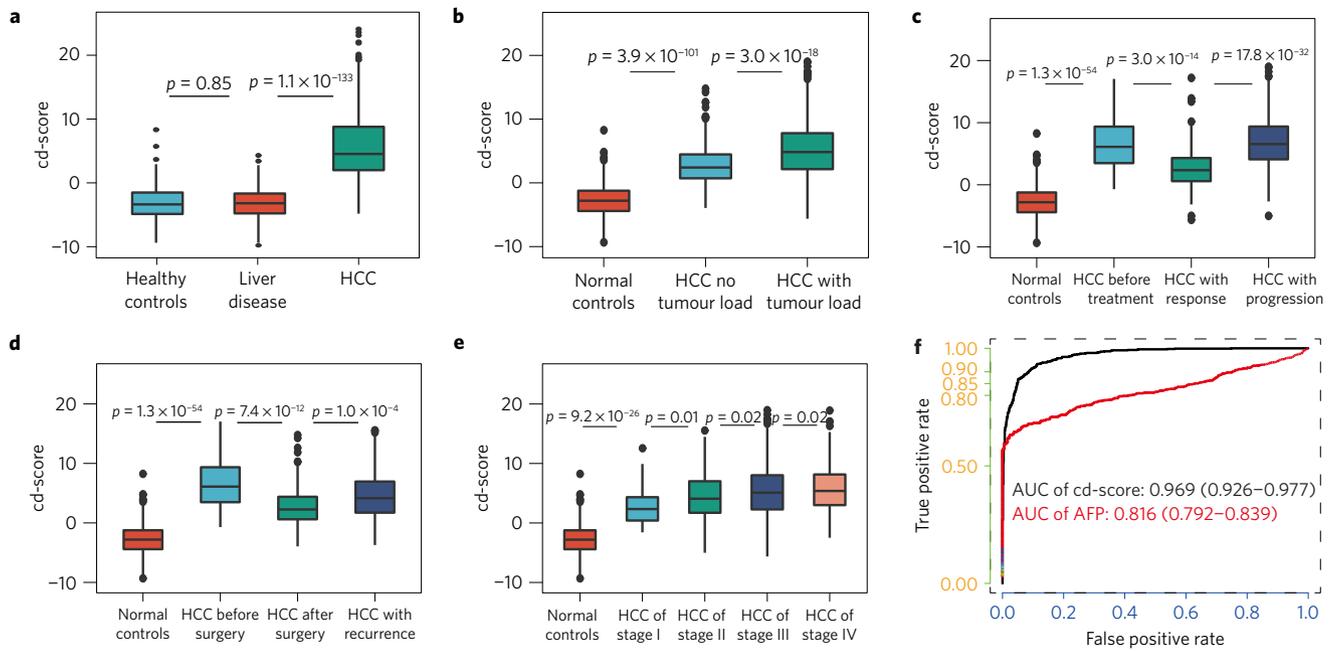


Figure 3 | cfDNA methylation analysis and tumour burden, treatment response, and staging. **a**, The combined diagnosis score (cd-score) in healthy controls, individuals with liver diseases (HBV/HCV infection, and fatty liver) and HCC patients. **b**, cd-score in normal controls and HCC patients with and without detectable tumour burden. **c**, cd-score in normal controls, HCC patients before treatment, with treatment response, and with progression. **d**, cd-score in normal controls and HCC patients before surgery, after surgery, and with recurrence. **e**, cd-score in normal controls and HCC patients from stage I-IV. **f**, The ROC of cd-score and AFP for HCC diagnosis in whole HCC cohort.

Table 2 | Characteristics of eight methylation markers and their coefficients in HCC prognosis prediction.

Markers	Ref Gene	Coefficients	HR	CI	SE	z value	p value
cg23461741	SH3PXD2A	-1.264	0.282	0.024-3.340	1.2604	-1.003	0.316
cg06482904	C11orf9	-0.247	0.781	0.067-9.100	1.2530	-0.197	0.844
cg25574765	PPFIA1	1.026	2.790	0.488-15.900	0.8894	1.153	0.249
cg07459019	Chr 17:78	-8.156	0.000	0.000-0.012	1.9112	-4.267	<0.001
cg20490031	SERPINB5	6.082	438.000	13.200-14,600.000	1.7885	3.400	0.001
cg01643250	NOTCH3	-5.368	0.005	0.000-0.140	1.7357	-3.093	0.002
cg11397370	GRHL2	1.497	4.470	1.030-19.400	0.7506	1.994	0.046
cg11825899	TMEM8B	2.094	8.120	0.957-68.900	1.0909	1.920	0.055

HR: Hazard Ratio; CI: 95.0% confidence interval; SE: standard errors of coefficients; z value: Wald z-statistic value.

and validation data set and that the cp-score was an independent risk factor of survival (hazard ratio [HR]: 2.512; 95% confidence interval [CI]: 1.966–3.210; $p < 0.001$ in the training set; HR: 1.553, CI: 1.240–1.944; $p < 0.001$ in the validation set, Supplementary Table 2). Interestingly, AFP was no longer significant as a risk factor when cp-score and other clinical characteristics were taken into account (Supplementary Table 2).

As expected, TNM stage predicted the prognosis of patients in our training and validation data set (Supplementary Fig. 8a,b). However, the combination of cp-score and TNM staging significantly improved our ability to predict prognosis in both the training (AUC 0.7935, Fig. 4c) and validation data sets (AUC 0.7586, Fig. 4d). Kaplan–Meier curves also showed that patients separated by both cp-score and staging have significantly different prognosis ($p < 0.0001$, Fig. 4e). These results demonstrate that ctDNA methylation analysis may contribute to risk stratification and prediction of prognosis in patients with HCC. However, this application merits further investigation in an HCC population with longer clinical follow-up than we had access to for our study.

The finding that tumours shed nucleic acids (DNA and RNA) into the blood and can be used as a surrogate source of tumour DNA

has opened an exciting new avenue in cancer diagnosis and prognosis^{21,22}. Despite substantial variability in the somatic mutations of individual tumours (with some notable exceptions), methylation patterns turn out to be remarkably consistent. Methylation patterns detected in cfDNA therefore have the potential to be more reliable discriminatory tools for the detection and diagnosis of malignancy.

In this study, we first determined differentially methylated CpG sites between HCC tumour samples and blood leukocytes in normal individuals for an HCC-specific panel. We then constructed a diagnostic prediction model using a 10-methylation marker panel (cd-score) for use in cfDNA; the cd-score effectively discriminated patients with HCC from individuals with HBV/HCV infection, cirrhosis, and fatty liver as well as healthy controls. Given that patients with these liver diseases are the target screening population under current guidelines, it is essential that a serum test reliably distinguish these disease states from HCC. In our study, the sensitivity of the cd-score for HCC is comparable to liver ultrasound²³, the current standard for HCC screening, markedly superior to AFP, and may represent a more cost-effective and less resource-intensive approach. Prospective clinical evaluation is warranted to compare or potentially combine ultrasound screening with cd-score. Further-

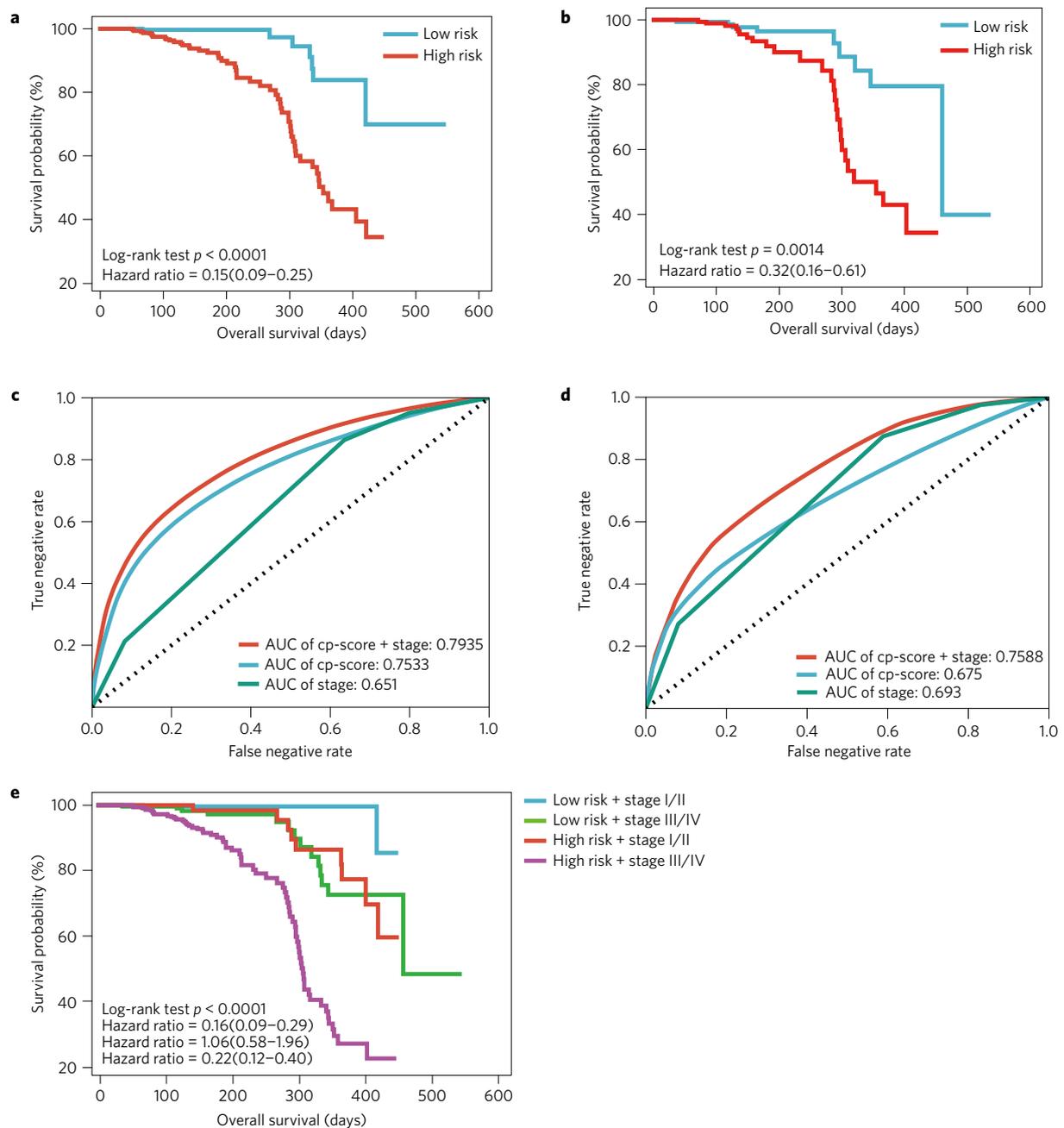


Figure 4 | cfDNA methylation analysis for prognostic prediction HCC survival. **a, b**, Overall survival curves of HCC patients with low or high risk, according to the combined prognosis score (cp-score) in the training (**a**) and validation data sets (**b**). **c, d**, The ROC for the cp-score, stage, and cp-score combined with stage in the training (**c**) and validation data sets (**d**). **e**, Survival curves of HCC patients with combinations of cp-score risk and stage in the whole HCC cohort.

more, the cd-score of our model showed high correlation with HCC tumour burden, treatment response, and stage, and is superior to the performance of AFP in our cohort. The cd-score may therefore be particularly useful for assessment of treatment response and surveillance for recurrence.

Additionally, we constructed a prognostic prediction model with an independent 8-marker panel and generated a combined prognosis score system (cp-score). The cp-score, which effectively distinguished HCC patients with significantly different prognosis, was validated as an independent prognostic risk factor in a multivariable analysis in our cohort and was again superior to AFP. This type of analysis may assist in the identification of patients for whom more or less aggressive treatment and surveillance is warranted. However, our study was limited by a relatively short clinical follow-up period.

Further study is warranted with longer clinical surveillance, in particular to fully assess whether this score can meaningfully contribute to clinical decision making for patients.

By sequencing of bisulfite converted cfDNA, we identified many previously unknown CpG markers differentially methylated in cancer versus normal plasma. Specifically, we employed a direct sequencing approach that captured the methylation status of adjacent CpG markers and found that the methylation of many adjacent markers is highly correlated with the initially targeted CpG, forming an MCB. A similar concept has been proposed before in which multiple adjacent CpG sites share a similar methylation pattern^{14,24–27}. This information allowed us to identify additional markers and improve the accuracy of sequencing for determining significant methylation differences.

Oncologists currently evaluate treatment response of HCC by imaging and AFP. Even with the modified Response Evaluation Criteria in Solid Tumours (mRECIST)²⁸, there are often difficult cases in which data is inconsistent and determining response and prognosis of patients is challenging. AFP is a useful serum marker in many patients, but is limited by its poor sensitivity and has proven to be a less than ideal surrogate for monitoring treatment response of HCC²⁹, as demonstrated by others and consistent with our study. In contrast, our results showed that methylation markers of ctDNA have high sensitivity and specificity that correlate with tumour burden, stage, treatment response, and prognosis of HCC patients. Furthermore, it is possible for relatively rapid adjustment of the treatment plan based on cfDNA due to its relatively short half-life (about 2 h)³⁰.

Some recent studies have reported that monitoring the somatic alterations in ctDNA can provide the earliest measure of treatment response in some solid cancers, including lung, colorectal and breast cancer^{31–34}. Unlike these studies, an advantage of our methylation markers is that we do not first need identification of somatic mutations in an individual patient. Furthermore, based on targeted sequencing of specific markers, our method can avoid the high cost of deep sequencing, which may make for its more routine and cost-effective application. Alternatively, it is intriguing to imagine the identification of a broad ‘pan-cancer’ methylation panel for use in cfDNA, possibly in synergy with somatic mutation analysis, that would allow pan screening for malignancy. Collectively, our study demonstrates the utility of cfDNA methylation analysis in the diagnosis, treatment evaluation, and prognosis of HCC, and represents a proof of concept for its use in solid malignancies broadly beyond HCC.

Methods

Methods, including statements of data availability and any associated accession codes and references, are available in the [online version of this paper](#).

Received 26 July 2016; accepted 30 August 2017;
published online 9 October 2017

References

- Torre, L. A. *et al.* Global cancer statistics, 2012. *CA Cancer J. Clin.* **65**, 87–108 (2015).
- Bruix, J. & Sherman, M. AASLD Practice Guideline: Management of hepatocellular carcinoma. *Hepatology* **42**, 1208–1236 (2005).
- Johnson, P. Role of alpha - fetoprotein in the diagnosis and management of hepatocellular carcinoma. *J. Gastroenterol. Hepatol.* **14**, S32–S36 (1999).
- Stroun, M. *et al.* The origin and mechanism of circulating DNA. *Ann. NY Acad. Sci.* **906**, 161–168 (2000).
- Bettegowda, C. *et al.* Detection of circulating tumor DNA in early- and late-stage human malignancies. *Sci. Trans. Med.* **6**, 224ra224 (2014).
- Newman, A. M. *et al.* An ultrasensitive method for quantitating circulating tumor DNA with broad patient coverage. *Nat. Med.* **20**, 548–554 (2014).
- Esteller, M. Epigenetics in cancer. *N. Engl. J. Med.* **358**, 1148–1159 (2008).
- Baylin, S. B. & Jones, P. A. Epigenetic determinants of cancer. *Cold Spring Harb. Perspect. Biol.* **8**, a019505 (2016).
- Irizarry, R. A. *et al.* The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat. Genet.* **41**, 178–186 (2009).
- Baylin, S. B. & Jones, P. A. A decade of exploring the cancer epigenome—biological and translational implications. *Nat. Rev. Cancer* **11**, 726–734 (2011).
- Board, R. E. *et al.* DNA methylation in circulating tumour DNA as a biomarker for cancer. *Biomarker Insights* **2**, 307–319 (2008).
- Warren, J. D. *et al.* Septin 9 methylated DNA is a sensitive and specific blood test for colorectal cancer. *BMC Med.* **9**, 133 (2011).
- Pishvaian, M. J. *et al.* A pilot study evaluating concordance between blood-based and patient-matched tumor molecular testing within pancreatic cancer patients participating in the Know Your Tumor (KYT) initiative. *Oncotarget* **7**, 13225 (2016).

- Lehmann-Werman, R. *et al.* Identification of tissue-specific cell death using methylation patterns of circulating DNA. *Proc. Natl Acad. Sci. USA* **113**, 201519286 (2016).
- Hannum, G. *et al.* Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol. Cell* **49**, 359–367 (2013).
- Smyth, G. K. *Bioinformatics and Computational Biology Solutions Using R and Bioconductor* 397–420 (Springer, 2005).
- Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Statist. Soc. B* **57**, 289–300 (1995).
- Reich, D. E. *et al.* Linkage disequilibrium in the human genome. *Nature* **411**, 199–204 (2001).
- Ardlie, K. G., Kruglyak, L. & Seielstad, M. Patterns of linkage disequilibrium in the human genome. *Nat. Rev. Genet.* **3**, 299–309 (2002).
- Hao, X. *et al.* DNA methylation markers for diagnosis and prognosis of common cancers. *Proc. Natl Acad. Sci. USA* **114**, 7414–7419 (2017).
- Aravanis, A. M., Lee, M. & Klausner, R. D. Next-generation sequencing of circulating tumor DNA for early cancer detection. *Cell* **168**, 571–574 (2017).
- Snyder, M. W., Kircher, M., Hill, A. J., Daza, R. M. & Shendure, J. Cell-free DNA comprises an *in vivo* nucleosome footprint that informs its tissues-of-origin. *Cell* **164**, 57–68 (2016).
- Singal, A. *et al.* Meta-analysis: surveillance with ultrasound for early-stage hepatocellular carcinoma in patients with cirrhosis. *Aliment Pharmacol. Ther.* **30**, 37–47 (2009).
- Butcher, L. M. *et al.* Non-CG DNA methylation is a biomarker for assessing endodermal differentiation capacity in pluripotent stem cells. *Nat. Commun.* **7**, 10458 (2016).
- Libertini, E. *et al.* Information recovery from low coverage whole-genome bisulfite sequencing. *Nat. Commun.* **7**, 11306 (2016).
- Burger, L., Gaidatzis, D., Schubeler, D. & Stadler, M. B. Identification of active regulatory regions from DNA methylation data. *Nucleic Acids Res.* **41**, e155 (2013).
- Guo, S. *et al.* Identification of methylation haplotype blocks aids in deconvolution of heterogeneous tissue samples and tumor tissue-of-origin mapping from plasma DNA. *Nat. Genet.* **49**, 635–642 (2017).
- Lencioni, R. & Llovet, J. M. Modified RECIST (mRECIST) assessment for hepatocellular carcinoma. *Semin. Liver Dis.* **30**, 52–60 (2010).
- Raoul, J. L. *et al.* Using modified RECIST and alpha-fetoprotein levels to assess treatment benefit in hepatocellular carcinoma. *Liver Cancer* **3**, 439–450 (2014).
- Diehl, F. *et al.* Circulating mutant DNA to assess tumor dynamics. *Nat. Med.* **14**, 985–990 (2008).
- Mok, T. *et al.* Detection and dynamic changes of EGFR mutations from circulating tumor DNA as a predictor of survival outcomes in NSCLC patients treated with first-line intercalated erlotinib and chemotherapy. *Clin. Cancer Res.* **21**, 3196–3203 (2015).
- Diaz, L. A. Jr. *et al.* The molecular evolution of acquired resistance to targeted EGFR blockade in colorectal cancers. *Nature* **486**, 537–540 (2012).
- Misale, S. *et al.* Emergence of KRAS mutations and acquired resistance to anti-EGFR therapy in colorectal cancer. *Nature* **486**, 532–536 (2012).
- Dawson, S. J. *et al.* Analysis of circulating tumor DNA to monitor metastatic breast cancer. *N. Engl. J. Med.* **368**, 1199–1209 (2013).

Acknowledgements

The results published here are in part based upon data generated by the TCGA Research Network: <http://cancergenome.nih.gov>. We thank staff at Kang Zhang and Ruihua Xu laboratories for technical assistance. This study was funded by Richard Annesser Fund, Michael Martin Fund, Dick and Carol Hertzberg Fund, SYSUCC, Xijing Hospital, and West China Hospital.

Author contributions

W.Wei, M.K., W.Wang, H.L., K.F., W.S., S.Y., L.Z., H.Z., R.Z., Y.X., K.L., H.Cai, G.L., L.Z., R.-h.X., Z.Z., D.L., E.Z. and C.Z. performed the experiments; M.K. W.Wang, H.L., K.F., B.A.C., Q.Q., Q.Z., L.Z., R.-h.X., J.Z., X.F., J.-k.Z., Y.D., H.Carter, M.Y., W.Z., R.G. and X.H. collected and analysed the data. K.Z. and R.-h.X. conceived the project, designed the experiments, and wrote the manuscript; All authors discussed the results and reviewed the manuscript.

Additional information

Supplementary information is available in the [online version of the paper](#). Reprints and permissions information is available online at www.nature.com/reprints. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to R.-h.X. or K.Z.

Competing financial interests

The authors declare no competing financial interests.

Methods

Patient data. Tissue DNA methylation data was obtained from The Cancer Genome Atlas (TCGA). Complete clinical, molecular, and histopathological data sets are available at the TCGA website: <https://tcga-data.nci.nih.gov/docs/publications/tcga>. Individual institutions that contributed samples coordinated the consent process and obtained informed written consent from each patient in accordance to their respective institutional review boards.

A second independent Chinese cohort consisted of HCC patients at the Sun Yat-sen University Cancer Center in Guangzhou, Xijing Hospital in Xi'an and the West China Hospital in Chengdu, China. Those who presented with HCC from stage I–IV were selected and enrolled in this study. Patient characteristics and tumour features are summarized in Supplementary Table 1. The TNM staging classification for HCC is according to the 7th edition of the AJCC cancer staging manual³⁵. The TNM Staging System is one of the most commonly used tumour staging systems. This system was developed and is maintained by the American Joint Committee on Cancer (AJCC) and adopted by the Union for International Cancer Control (UICC). The TNM classification system was developed as a tool for oncologists to stage different types of cancer based on certain standard criteria. The TNM Staging System is based on the extent of the tumour (T), the extent of spread to the lymph nodes (N), and the presence of metastasis (M). This project was approved by the Institutional Review Boards (IRBs) of Sun Yat-sen University Cancer Center, Xijing Hospital, and West China Hospital. Informed consent was obtained from all patients. Tumour and normal tissues were obtained as clinically indicated for patient care and were retained for this study. Human blood samples were collected by venipuncture and plasma samples were obtained by taking supernatant after centrifugation and stored at -80°C before cfDNA extraction.

Cell-free DNA extraction from plasma samples. We used minimal 1.5 ml plasma samples throughout our study by investigating the minimal volume of plasma that will give a consistent cfDNA recovery and reliable sequencing coverage defined as more than 20 reads for a target cg marker. EliteHealth cfDNA extraction Kit (EliteHealth, Guangzhou Youze, China) was used for cell-free DNA extraction. More detailed information is described in the Supplementary Information.

Bisulfite conversion of genomic DNA. 10–15 ng of cfDNA was converted to bis-DNA using EZ DNA Methylation-Lightning Kit (Zymo Research) according to the manufacturer's protocol. The efficiency of bisulfite conversion was $>99.8\%$, as verified by deep sequencing of bis-DNA and analysing the ratio of C to T conversion of CH (non-CG target-captured) dinucleotides.

Determination of DNA methylation levels by deep sequencing of bis-DNA target-captured with molecular-inversion (padlock) probes. CpG markers whose methylation levels significantly differed in any of the comparisons between any cancer tissue and any normal tissue in TCGA data set were used to design padlock probes for capture and sequencing of cfDNA. Padlock capture of bis-DNA was based on the technique on published methods with modifications^{36–38}. We used a two-step approach wherein the first step is to identify optimal cg markers with the largest methylation beta value difference between HCC tissue and normal blood leukocytes; the second step to validate these top cg markers using cfDNA from plasma sample of HCC and normal patients. Because of a relatively modest total size of captured regions/cg markers, this approach offers much lower cost of sequencing than any current methods, including whole methylome-wide sequencing, therefore enabling us to evaluate a large number of samples. Furthermore, our direct targeted sequencing approach offers digital readout, and requires much less starting cfDNA material (10–15 ng) than more traditional recent methods based on hybridization on a chip (for example, Infinium, Illumina) or target-enrichment by hybridization (for example, SureSelect, Agilent). This approach is also less sensitive to unequal amplification as it utilizes unique molecular identifiers (UMIs).

Probe design and synthesis. Padlock probes were designed using the ppDesigner software³⁷. The average length of the captured region was 100 bp, with the CpG marker located in the central portion of the captured region. Linker sequence between arms contained binding sequences for amplification primers separated by a variable stretch of Cs to produced probes of equal length. We incorporated a 6-bp UMI sequence in probe design to allow for the identification of unique individual molecular capture events and accurate scoring of DNA methylation levels. Padlock probe sequence information on the final ten diagnostic markers and eight prognostic markers are listed in Supplementary Table 4.

Probes were synthesized as separate oligonucleotides using standard commercial synthesis methods (ITD). For capture experiments, probes were mixed, *in vitro* phosphorylated with T4 PNK (NEB) according to manufacturer's recommendations, and purified using P-30 Micro Bio-Spin columns (Bio-Rad).

Sequencing data analysis. Mapping of sequencing reads was done using the software tool bisReadMapper with some modifications³⁷. First, UMIs were extracted from each sequencing read and appended to read headers within FASTQ files using a custom script. Reads were on-the-fly converted as if all C were non-methylated and mapped to in-silico converted DNA strands of the human genome, also as if all C were non-methylated, using Bowtie2 (ref. 39). Original reads were merged and filtered for a single UMI—that is, reads carrying the same UMI were discarded, leaving a single, unique read. Methylation frequencies were calculated for all CpG dinucleotides contained within the regions captured by padlock probes by dividing the numbers of unique reads carrying a C at the interrogated position by the total number of reads covering the interrogated position.

Identification of methylation correlated blocks (MCBs). Pearson correlation coefficients between methylation frequencies of each pair of CpG markers separated by no more than 200 bp were calculated separately across 50 cfDNA samples from each of the two diagnostic categories—that is, normal health blood and HCC. A value of Pearson's $r < 0.5$ was used to identify transition spots (boundaries) between any two adjacent markers indicating uncorrelated methylation. Markers not separated by a boundary were combined into MCBs. This procedure identified a total of $\sim 1,550$ MCBs in each diagnostic category within our padlock data, combining between 2 and 22 CpG positions in each block. Methylation frequencies for entire MCBs were calculated by summing up the numbers of Cs at all interrogated CpG positions within an MCB and dividing by the total number of C+Ts at those positions.

Data availability. Raw beta value data for ten diagnostic markers are listed in Supplementary Table 5 (Pages 15–81); raw beta value data for eight prognostic markers are listed in Supplementary Table 6 (Pages 82–118). Key raw data were also verified and uploaded onto the Research Data Deposit public platform (www.researchdata.org.cn) with an approval number RDDB2017000132.

References

- Edge, S. B. & Compton, C. C. The American Joint Committee on Cancer: the 7th edition of the AJCC cancer staging manual and the future of TNM. *Ann. Surg. Oncol.* **17**, 1471–1474 (2010).
- Porreca, G. J. *et al.* Multiplex amplification of large sets of human exons. *Nat. Methods* **4**, 931–936 (2007).
- Diep, D. *et al.* Library-free methylation sequencing with bisulfite padlock probes. *Nat. Methods* **9**, 270–272 (2012).
- Deng, J. *et al.* Targeted bisulfite sequencing reveals changes in DNA methylation associated with nuclear reprogramming. *Nat. Biotechnol.* **27**, 353–360 (2009).
- Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).