ORIGINAL ARTICLE

Cancer Science WILEY

Multiple-level copy number variations in cell-free DNA for prognostic prediction of HCC with radical treatments

Yang Wang[1] | Kaixiang Zhou[2] | Xiangxu Wang[3] | Yang Liu[2] | Dongnan Guo[4] | Zhenyuan Bian[1] | Liping Su[2] | Kun Liu[1] | Xiwen Gu[5] | Xu Guo[2] | Lin Wang[1] | Hongmei Zhang[3] | Kaishan Tao[1] | Jinliang Xing[2] (iD)

[1]Department of Hepatobiliary Surgery, Xijing Hospital, Fourth Military Medical University, Xi'an, China

[2]Department of Physiology and Pathophysiology, State Key Laboratory of Cancer Biology, Fourth Military Medical University, Xi'an, China

[3]Department of Oncology, Xijing Hospital, Fourth Military Medical University, Xi'an, China

[4]School of Pharmacy, Health Science Center, Xi'an Jiaotong University, Xi'an, China

[5]Key Laboratory of Shaanxi Province for Craniofacial Precision Medicine Research, College of Stomatology, Xi'an Jiaotong University, Xi'an, China

Correspondence
Jinliang Xing, Department of Physiology and Pathophysiology, State Key Laboratory of Cancer Biology, Fourth Military Medical University, 169 Changle West Road, Xi'an 710032, China.
Email: xingjl@fmmu.edu.cn

Kaishan Tao, Department of Hepatobiliary Surgery, Xijing Hospital, Fourth Military Medical University, 127 Changle West Road, Xi'an 710032, China.
Email: taokaishan0686@163.com

Abstract

Copy number variations (CNVs) in cell-free DNA (cfDNA) are emerging as noninvasive biomarkers for various cancers. However, multiple-level analysis of cfDNA CNVs for hepatocellular carcinoma (HCC) patients with radical treatments remains uninvestigated. Here, CNVs at genome-wide, chromosomal-arm, and bin levels were analyzed in cfDNA from 117 HCC patients receiving radical treatments. Then, the relationship between cfDNA CNVs and clinical outcomes was explored. Our results showed that a concordant profile of CNVs was observed between cfDNA and tumor tissue DNA. Three genome-wide CNV indicators including tumor fraction (TFx), prediction score (P-score), and stability score (S-score) were calculated and demonstrated to exhibit significant correlation with poorer overall survival (OS) and recurrence-free survival (RFS). Furthermore, the high-frequency cfDNA CNVs at chromosomal-arm level including the loss of 4q, 17p, and 19p and the gain of 8q and 1q clearly predicted HCC prognosis. Finally, a bin-level risk score was constructed to improve the ability of CNVs in predicting prognosis. Altogether, our study indicates that the multiple-level cfDNA CNVs are significantly associated with OS and RFS in HCC patients with radical treatments, suggesting that cfDNA CNVs detected by low-coverage whole-genome sequencing (WGS) may be used as potential prognostic biomarkers of HCC patients.

Yang Wang, Kaixiang Zhou, and Xiangxu Wang contributed equally to this work.

- -

1 | INTRODUCTION

Hepatocellular carcinoma (HCC) is one of the most common cancers and a leading cause of death worldwide.[1,2] More than 75% HCC patients are etiologically associated with chronic hepatitis B virus (HBV) infection in the Asia-Pacific region.[3] Radical treatment, including surgery and radio frequency ablation, is the mainstay curative treatment for early-stage HCC patients.[4] Unfortunately, tumor recurrence and metastasis still greatly contribute to the poor prognosis of HCC, partially due to the lack of effective biomarkers.[5,6] Therefore, it is of primary importance to identify prognostic biomarkers for guiding treatment decisions, especially in patients with radical treatments. Currently, some new biomarkers, such as miR-181a-5p[7] and VEGFA,[8] have been reported as potential prognostic indicators of HCC patients. However, these biomarkers still need further validation for clinical applications. To date, the only clinically available blood-based biomarker for HCC surveillance is alpha fetoprotein (AFP), whereas its clinical utility is severely limited by low sensitivity and specificity.[9] Hence, there remains an urgent unmet need to develop novel prognostic biomarkers for HCC patients.

Cell-free DNA (cfDNA) in tumor patients containing the fragmented circulating tumor DNA, which carries the comprehensive information of tumor genome profiles, such as methylations, point mutations, and copy number variations (CNVs), can be used as a surrogate source of tumor tissue DNA.[10-12] Several studies have reported the significant promise of cfDNA as a novel biomarker for HCC by assessing its methylation status and point mutations.[13,14] However, the high cost and complex detection pipeline limit the wide application of methylation-based assay. Meanwhile, only a small fraction of mutations presented in both tumor tissue and plasma, which greatly hinders their clinical utility. CNV profiling offers several advantages over methylation or mutation analysis for cancer detection, including larger genomic region span and the ability for detecting structural genomic alterations.[15] Moreover, the detection of cfDNA CNVs by ultralow coverage sequencing makes it a cost-effective approach for potential clinical applications.[16] Previous studies have suggested the prognostic utility of cfDNA CNVs in HCC patients treated with Sorafenib.[17] Furthermore, Cai et al have reported that plasma CNV indicator (tumor fraction, TFx) at genome-wide level are dynamically correlated with tumor burden in a relatively small cohort (n = 34) of HCC patients receiving surgical resection.[18] Recent multiple-level analysis of CNVs in triple-negative breast cancer patients shed light on a more appropriate and comprehensive way in the analysis of cfDNA CNVs.[10] Nevertheless, little has been addressed on cfDNA CNVs at multiple levels for HCC patients who received radical treatments.

Here, CNVs at genome-wide, chromosomal-arm, and bin levels were analyzed in plasma cfDNA from 117 HBV-HCC patients by low-coverage whole-genome sequencing (WGS). Then, the relationship between cfDNA CNVs and clinical outcomes was explored. Our study indicates that the multiple-level cfDNA CNVs are significantly associated with overall survival (OS) and recurrence-free survival (RFS) in HCC patients with radical treatments, suggesting that cfDNA CNVs detected by low-coverage WGS may be used as potential prognostic biomarkers for HCC patients.

2 | MATERIALS AND METHODS

2.1 | Patients and sample collection

A total of 117 HBV-HCC patients were recruited between May 2016 and July 2017 from the Department of Hepatobiliary Surgery at Xijing Hospital, Fourth Military Medical University, Xi'an, China (Figure 1A). Only HBV-related HCC patients were enrolled. None of the patients had received any therapy prior to radical treatment. About 5 mL of peripheral blood sample was collected in EDTA tubes (BD) from each patient before treatment. Fresh HCC tumor tissue samples were also obtained from available cases. All patients were followed up every 6 months after the radical treatment, ending in February, 2020. Follow-up information was obtained from 115 patients, with a median follow-up time of 32 months (ranging from 1 to 45 months). During the follow-up period, 33 patients died of HCC and 66 patients developed recurrence.

2.2 | Plasma processing and cfDNA extraction

To separate plasma, peripheral blood was firstly centrifuged at 240 g at 4°C for 10 minutes. The resultant plasma supernatant was further purified by centrifugation at 9391 g at 4°C for 15 minutes. The plasma was stored at −80°C until use. Plasma cfDNA was extracted using the QIAamp Circulating Nucleic Acid Kit (Qiagen) according to the manufacturer's protocol. The quantity and quality of cfDNA samples were evaluated using Qubit3.0 (Thermo) and Agilent 2100 bioanalyzer (Agilent).

2.3 | Tumor tissue collection and DNA extraction

A total of 19 HCC tissue samples were selected for further analysis. H&E staining was carefully reviewed by a pathologist to ensure that the cancer cell content was >70%. Tissue DNA was extracted from frozen tumor samples using the ENZA tissue DNA kit (Omega) according to the manufacturer's protocol. The quantity and quality of tissue DNA samples were evaluated using NanoDrop 2000 and Agilent 2100 bioanalyzer (Agilent).
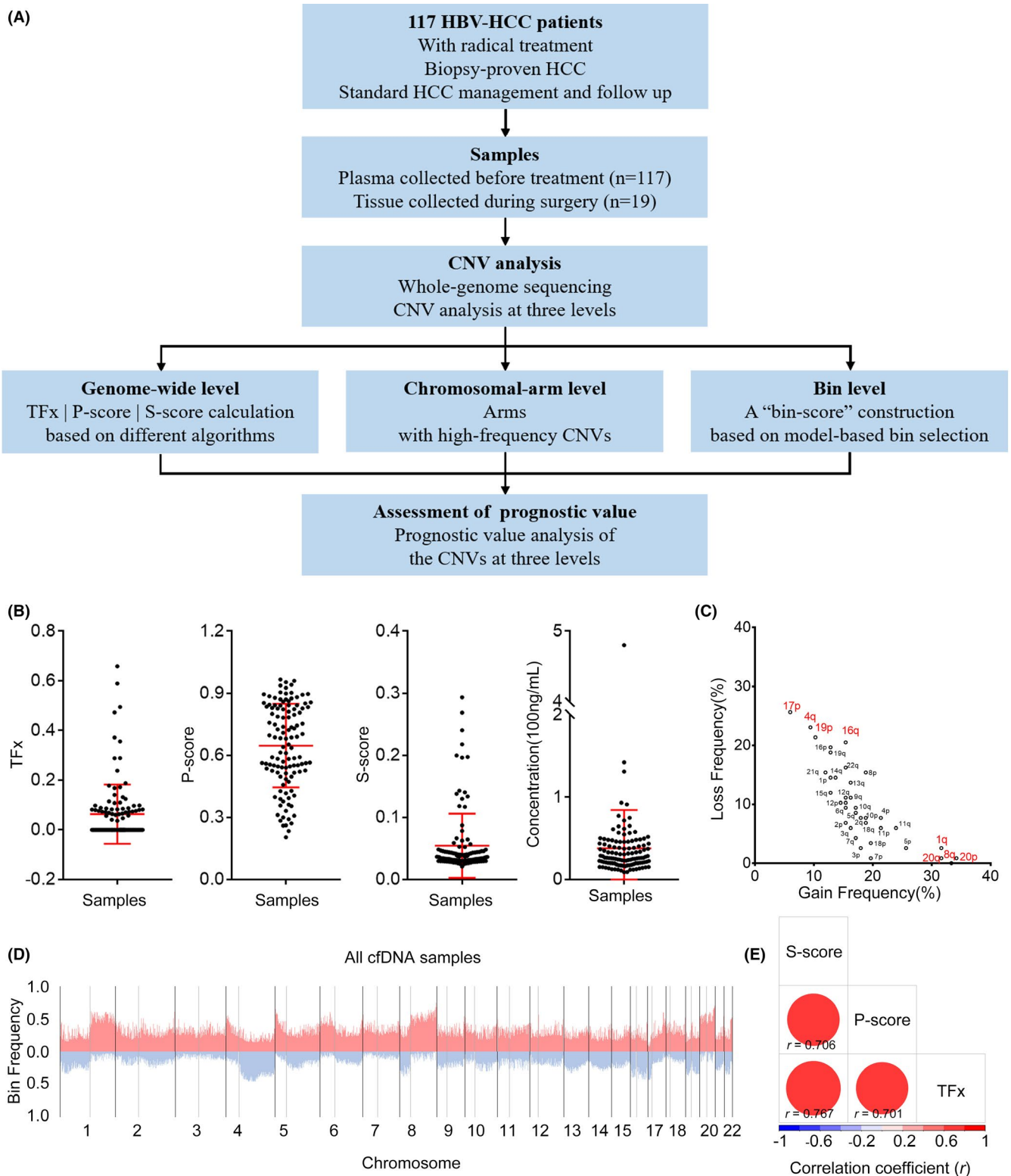
FIGURE 1 Overview of the study. A, Flow diagram of the study design. HBV, hepatitis B virus; HCC, hepatocellular carcinoma. B, The distribution of tumor fraction (TFx), P-score, S-score, and concentration (100 ng/mL) (n = 117). C, A scatter plot of the chromosomal-arm level copy number variation (CNV) frequency measured in circulating free DNA (cfDNA) (n = 117). The top four most frequent chromosomal arms harboring gain or loss are marked as red. D, The distribution of bin frequency of copy number gains (red) and losses (blue) alter among patients (n = 117) across the genome. E, Correlation analysis of genome-wide CNV indicators (*r*, Spearman correlation). The size of the circles and the darkness of the color are proportional to the strength of the correlation. Red and blue colors represent the positive and negative direction, respectively

2.4 | Library preparation and WGS

The construction of WGS library for tissue DNA and plasma cfDNA samples was performed using the NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB) according to the manufacturer's protocol as previously described.[15,19] The resulting libraries were sequenced to about 5× genome-wide depths for cfDNA and about 40× genome-wide depths for tissue DNA on an Illumina HiSeq X Ten platform in 150-bp paired-end mode.

2.5 | Sequencing data analysis

The raw paired-end reads were mapped to the human reference genome GRCh37 with BWA-0.7.4 and the BAM files were analyzed using Picard tools-1.9.2 (https://broadinstitute.github.io/picard/) to mark and remove the duplicate reads. Quality control was performed using QPLOT to analyze the genome-level statistics. Then, ichorCNA was used to profile the CNVs across the genome for further multiple-level analysis using default algorithm parameters (available at https://github.com/broadinstitute/ichorCNA).[16] In addition, the copy number data of 370 HCC patients from The Cancer Genome Atlas (TCGA) database was obtained through the UCSC Genome Browser (http://xena.ucsc.edu/).

2.6 | Calculation of genome-wide CNV indicators

To quantify the genome-wide CNVs in plasma cfDNA, three indicators were calculated based on three different algorithms. TFx was directly estimated by ichorCNA algorithm with default parameters.[16] In addition, a prediction score (P-score) was calculated as previously reported, which was based on the random forest (RF) model of HBV-HCC and cancer-free HBV patients using the genome-wide CNV profiles.[15] This P-score was further used for prognostic assessment. Thirdly, we defined a novel stability score (S-score), which was intended to reflect the extent of CNV deviation from diploid genome and was calculated using the following formula:

$$S\_score = \sqrt{\frac{\sum_i^n (Xi - 0)^2}{n}}$$

where Xi means the $\log_2$ ratio of the bin-specific value calculated by ichorCNA, 0 represents a desirable $\log_2$ ratio value of diploid genome, and n is the number of bins ($n = 2475$).

2.7 | Chromosomal-arm level CNV analysis

First, the genome was divided into nonoverlapping bins of 1-Mb size. Bins with corrected copy number greater than 2 were defined as bin gain, and bins with corrected copy number smaller than 2 were defined as bin loss. Chromosomal-arm gain or loss was defined when more than 50% region of the entire chromosome arm was identified as bin gain or bin loss. The frequency of each chromosome arm with gain or loss was calculated as the number of patients with specific chromosomal-arm gain or loss divided by the total patient number and used for subsequent analysis.

2.8 | Construction of bin-level risk score

A univariable Cox proportional hazard regression analysis was performed on all 1-Mb bins to select the significant bin candidates, and the least absolute shrinkage and selection operator (LASSO) regression model was used to further screen the candidate bins with prognostic significance (iteration = 10 000). The LASSO regression was analyzed using "glmnet" package from R (version 3.6.2).[20] Next, the multivariate Cox proportional hazard regression model was used to construct the bin-level risk score, which was defined as below:

$$\text{Bin score} = \sum_i^n (\beta i \times Cnvi)$$

where βi represents the coefficient value, and $Cnv\,i$ represents the $\log_2$ ratio of each bin calculated by ichorCNA algorithm.

2.9 | Statistical analysis

Correlations were evaluated using Spearman's correlation coefficients. Consistency between paired cfDNA and tumor DNA samples was estimated by calculating the proportion of paired samples with consistent CNVs at each bin or chromosomal arm. Associations between clinical variables and multiple-level CNV indicators were assessed using either chi-square test or Fisher's exact test. Survival analysis was performed using the Kaplan-Meier log-rank test. Univariable and multivariable Cox proportional hazard models were used to examine the association between the CNV indicators and death or recurrence in HCC patients. The ROC curve was adopted to assess the performance of the CNV-related indicators using "survival ROC" package. All above statistical analyses were carried out in R package (version 3.6.2). The P-value of <.05 was considered statistically significant, and all probabilities were two-tailed.

3 | RESULTS

3.1 | Clinical characteristics of 117 HBV-related HCC patients

The clinical characteristics of the HCC patients are shown in Table 1. Briefly, the median age of patients was 55 years old, and 102 patients were male. There were 61, 44, and 12 patients with American Joint Committee on Cancer (AJCC) stage I, stage II, and stage III diseases,

respectively, while 98 and 19 of the patients with Barcelona Clinic Liver Cancer (BCLC) stage 0-A and B-C, respectively.

3.2 | Multiple-level CNV analysis of cfDNA

The WGS was carried out at a mean depth of 4.01× (range, 2.78 to 7.86) for plasma cfDNA samples and 29.51× (range, 18.42 to 49.23) for tissue DNA samples. The sequencing data information was summarized in Table S1, showing good sequencing quality. The TFx, P-score, and S-score were calculated based on CNV analysis. The different distribution of TFx, P-score, S-score, and cfDNA concentration was shown in Figure 1B. Statistically significant positive correlation was observed among TFx, P-score, and S-score (Figure 1E). The top four most frequent gains of 20p, 8q, and 1q, 20q and the top four most frequent losses of 17p, 4q, 19p, and 16q were identified in 117 HBV-HCC patients (Figure 1C). Furthermore, significant negative correlation was found between gain and loss frequencies (Spearman $r = -0.862$, $P < .001$, Figure 1C), indicating that CNVs at chromosomal-arm level preferred either gain or loss, but rarely both. In addition, the bin-level (1 Mb) CNVs were obtained from ichorCNA, and the bin frequencies altered across the genome among HBV-HCC patients are exhibited in Figure 1D.

3.3 | Concordant profile of CNVs was observed between cfDNA and tumor tissue DNA

To estimate the concordance between cfDNA and tissue DNA, the frequency of CNVs at bin level or chromosomal-arm level was analyzed among the cfDNA, tissue DNA in our cohort, and tissue DNA from TCGA database. Similar patterns of CNVs at both bin and chromosomal-arm level were observed between cfDNA and tissue DNA (Figure 2A and Figure S1A). CNV signature of tumor DNA and cfDNA obtained from the same patients was analyzed. Four representative cases are shown in Figure S1B. The altered segments in tumor tissues were detected in cfDNA with the overall sensitivity and specificity at 0.76 and 0.94. The sensitivity and specificity of CNV gain was 0.73 and 0.93. The sensitivity and specificity of CNV loss was 0.79 and 0.95. Moreover, our data indicated that the frequency of CNVs at bin level in cfDNA was significantly correlated with that in matched tumor tissue DNA (gain: $r = 0.649$, $P < .001$; loss: $r = 0.856$, $P < .001$. n = 19; Figure 2B). When analyzed at chromosomal-arm level, the frequency of CNVs in cfDNA was also significantly correlated with that in matched tumor tissue DNA (gain: $r = 0.779$, $P < .001$; loss: $r = 0.912$, $P < .001$. Figure 2C). In addition, the CNV frequency in TCGA tissue DNA was significantly correlated with the CNV frequency in both tissue DNA (gain: $r = 0.739$; $P < .001$, loss: $r = 0.692$; $P < .001$. Figure 2C) and cfDNA (gain: $r = 0.718$; $P < .001$, loss: $r = 0.666$; $P < .001$. Figure 2C) in our cohort. Overall, our data indicated the concordant profile of CNVs between cfDNA and tissue DNA, suggesting the possibility of cfDNA as a biopsy surrogate in HCC-related clinical practices.

3.4 | Association of cfDNA CNVs with clinical characteristics in HCC patients

The association of demographic and clinical characteristics with CNVs detected at different levels and concentrations of cfDNA is presented in Table 1 and Tables S2-S5, respectively. Patients were categorized into two groups by the cutoff values of TFx, P-score, S-score, and cfDNA concentration determined by maximizing Youden index (Table 1). The results indicated that patients with high TFx, P-score, and S-score exhibited a significantly higher proportion at late stage (stage III) and with large tumor size (≥5 cm) than those with low TFx, P-score, and S-score, respectively (All $P < .05$). In addition, an obvious trend was observed that TFx, P-score, and S-score were associated with microvascular invasion (MVI) and BCLC stage, although only P-score attained statistical significance ($P < .05$). Importantly, patients with high TFx, P-score, and S-score exhibited a significantly higher recurrence and death risk than those with low TFx, P-score, and S-score, respectively (All $P < .05$). Similar results were observed in patients with gains of 8q and 1q (Table S2), losses of 17p, 19p, and 16q (Table S3), high bin score (Table S4), and high cfDNA concentration (Table S5).

3.5 | Genome-wide indicators of cfDNA CNVs were significantly associated with prognosis of HCC patients receiving radical treatments

As shown in Table S6, the univariable Cox regression analysis indicated that BCLC stage, AJCC stage, and tumor size were associated with OS and RFS of HCC patients receiving radical treatment, which is greatly consistent with previous reports.[21] Furthermore, HCC patients were respectively stratified by cutoff value of three genome-wide CNV indicators determined by maximizing Youden index. Kaplan-Meier analysis showed that the patients with high TFx, P-score, S-score, and cfDNA concentration exhibited a significantly poorer OS and RFS than those with low TFx, P-score, S-score, and cfDNA concentration, respectively (All $P < .05$. Figure 3). Similar results were observed when median values of TFx, P-score, and S-score were used as cutoff values (all $P < .05$. Figure S2A-C). However, no significant difference was observed between patients with high and low cfDNA concentration in OS ($P = .117$) and RFS ($P = .731$, Figure S2D) when the median value was used as the cutoff for stratification. After adjusting for gender, age, BCLC stage, MVI, and AFP, the multivariable Cox proportional hazard regression analysis indicated that the three genome-wide CNV indicators were significantly associated with death risk of HCC patients receiving radical treatments, with hazard ratio (HR) of 4.02 (95% confidence interval [CI], 1.87-8.64), 3.70 (95% CI, 1.67-8.17), 3.72 (95% CI, 1.70-8.14), and 3.95 (95% CI, 1.89-8.27) for S-score, P-score, TFx, and cfDNA concentration, respectively (Table 2). Similar results were observed in the associations of the cfDNA indicators with recurrence risk, with HR of 2.55 (95% CI, 1.52-4.30), 2.06 (95% CI, 1.21-3.51), 2.78 (95% CI, 1.64-4.70), and 1.97 (95% CI, 1.08-3.59) for S-score, P-score, TFx, and cfDNA concentration, respectively (Table 2).

TABLE 1 Clinicopathologic characteristics and genome-wide copy number variation (CNV) indicators of hepatocellular carcinoma (HCC) patients

Variables	All patients n = 117	S-score Low n = 73	S-score High n = 44	P	P-score Low n = 69	P-score High n = 48	P	TFx Low n = 68	TFx High n = 49	P
Age										
<55	58 (50)	33 (45)	25 (57)	.255	31 (45)	27 (56)	.262	29 (43)	29 (59)	.093
≥55	59 (50)	40 (55)	19 (43)		38 (55)	21 (44)		39 (57)	20 (41)	
Gender										
Male	102 (87)	63 (86)	39 (89)	.783	59 (86)	43 (90)	.585	60 (88)	42 (86)	.782
Female	15 (13)	10 (14)	5 (11)		10 (14)	5 (10)		8 (12)	7 (14)	
AJCC stage										
I	61 (52)	46 (63)	15 (34)	.020*	47 (68)	14 (29)	<.001*	43 (63)	18 (37)	.020*
II	44 (38)	24 (33)	20 (46)		20 (29)	24 (50)		23 (34)	21 (43)	
III	12 (10)	3 (4)	9 (20)		2 (3)	10 (21)		3 (3)	10 (20)	
AFP										
<20 ng/mL	55 (47)	35 (48)	19 (43)	.703	35 (51)	20 (42)	.353	33 (49)	22 (45)	.712
≥20 ng/mL	62 (53)	38 (52)	25 (57)		34 (49)	28 (58)		35 (51)	27 (55)	
Tumor size										
<3 cm	30 (26)	26 (35)	4 (9)	.001*	26 (37)	4 (8)	<.001*	26 (38)	4 (8)	<.001*
3-5 cm	50 (43)	35 (48)	15 (34)		31 (45)	19 (40)		30 (44)	20 (41)	
≥5 cm	37 (32)	12 (17)	25 (57)		12 (18)	25 (52)		12 (18)	25 (51)	
MVI										
No	74 (63)	50 (69)	24 (55)	.103	48 (70)	26 (54)	.041*	45 (66)	29 (59)	.232
Yes	37 (32)	19 (26)	18 (41)		16 (23)	21 (44)		18 (27)	19 (39)	
Unknown	6 (5)	4 (5)	2 (4)		5 (7)	1 (2)		5 (7)	1 (2)	
BCLC stage										
0/A	98 (84)	65 (89)	33 (75)	.069	62 (89)	36 (75)	.069	61 (90)	37 (76)	.046*
B/C	19 (16)	8 (11)	11 (25)		7 (11)	12 (25)		7 (10)	12 (24)	
CTP class										
A	112 (96)	70 (96)	42 (95)	1.000	66 (96)	46 (96)	1.000	65 (96)	47 (96)	1.000
B	5 (4)	3 (4)	2 (5)		3 (4)	2 (4)		3 (4)	2 (4)	
Recurrence										
No	51 (44)	42 (57)	9 (20)	<.001*	39 (56)	12 (25)	.001*	41 (60)	10 (20)	<.001*
Yes	64 (55)	31 (43)	33 (75)		29 (43)	35 (73)		27 (40)	37 (76)	
Lost	2 (2)	0 (0)	2 (5)		1 (1)	1 (2)		0 (0)	2 (4)	
Vital status										
Alive	82 (70)	62 (85)	20 (45)	<.001*	58 (84)	24 (50)	<.001*	58 (85)	24 (49)	<.001*
Deceased	33 (28)	11 (15)	22 (50)		10 (15)	23 (48)		10 (15)	23 (47)	
Lost	2 (2)	0 (0)	2 (5)		1 (1)	1 (2)		0 (0)	2 (4)	

Note: Data presented as No. (%) unless otherwise noted. The cutoff points of S-score, P-score, and TFx were determined by the maximizing Youden index of classifying 1-y recurrence.

Abbreviations: AFP, alpha-fetoprotein; AJCC, American Joint Committee on Cancer; BCLC, Barcelona Clinic Liver Cancer; CTP class, Child-Turcotte-Pugh class; MVI, microvascular invasion; P-score, prediction score; S-score, stability score; TFx, tumor fraction.
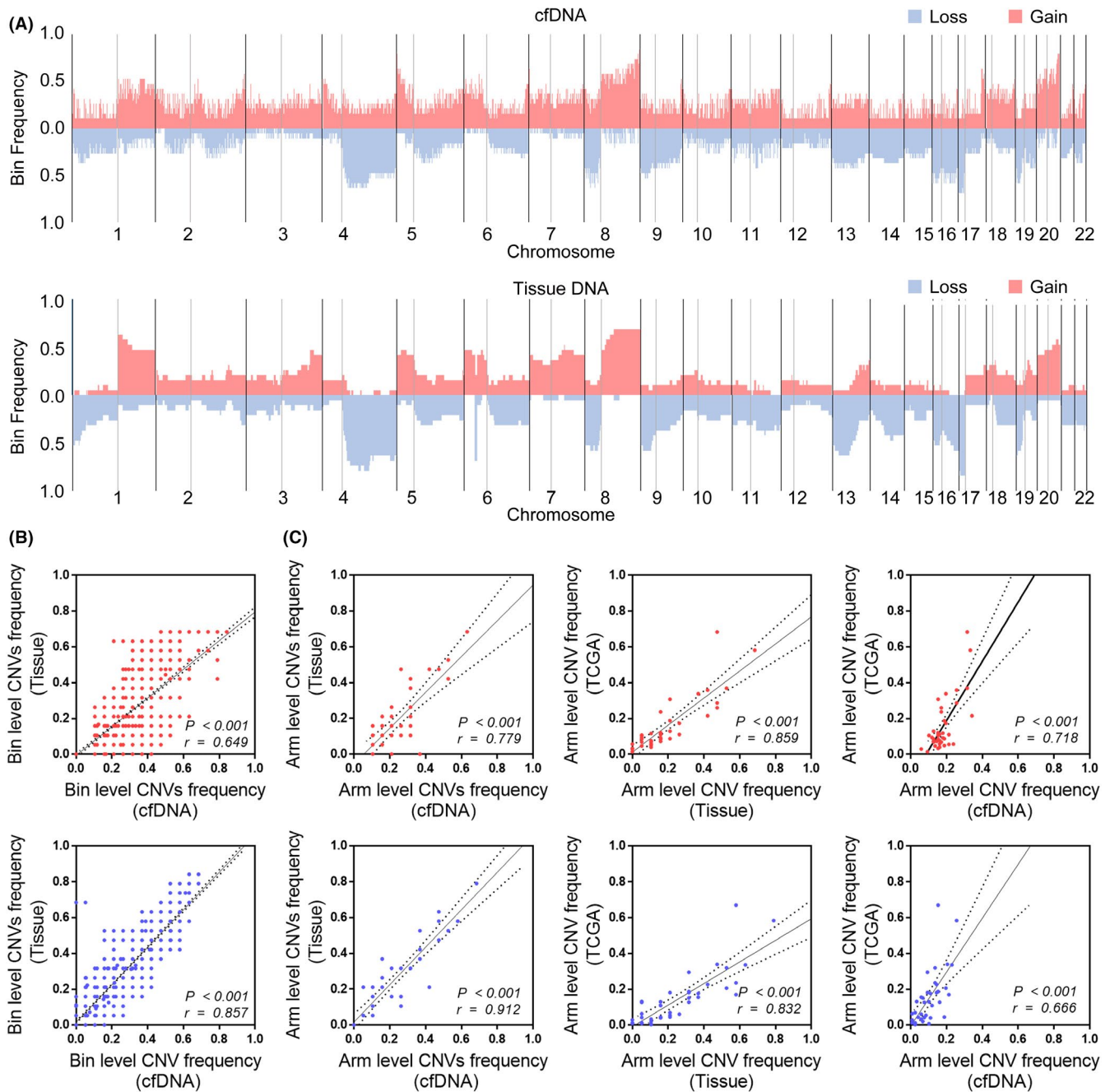
* P value has statistical significane (< 0.05).

FIGURE 2 Comparison of copy number variations (CNVs) between circulating free DNA (cfDNA) and tumor tissue DNA samples. A, The distribution of the bin-level CNVs across the genome of cfDNA (top) and paired tumor tissue DNA (bottom). The bin frequency of copy number gains (red) and losses (blue) altering among patients are exhibited. B, The correlation between CNV frequency of each bin in cfDNA and in tissue DNA (n = 19, pair-matched. r, Spearman correlation). C, The correlation of CNV frequency of each arm between cfDNA and tissue DNA from our cohort, tissue DNA from our cohort and The Cancer Genome Atlas (TCGA), and cfDNA and tissue DNA from TCGA (r, Spearman correlation)

Furthermore, the receiver operating characteristic (ROC) curves were used to characterize the discrimination potential of TFx, P-score, S-score, cfDNA concentration, MVI, and BCLC stage for 3-year or 1-year survival. The results showed that the TFx, S-score, and P-score had similar areas under the ROC curve (AUCs) and exhibited a slightly better performance than cfDNA concentration, BCLC, and MVI, irrespective of 3-year survival or 1-year survival (Figure 3). It should be noted that the ROC curves of the concentration lay across the diagonal line, which may be due to an unstable estimation using concentration.

FIGURE 3 Prognostic values of genome-wide copy number variation (CNV) indicators and concentration for hepatocellular carcinoma (HCC) patients. Kaplan-Meier curves for the low group and high group of HCC patients according to the genome-wide CNV indicators: (A) S-score, (B) P-score, (C) tumor fraction (TFx), and (D) concentration. Cases are stratified by cutoff value determined by maximizing Youden index. The Mantel-Cox log-rank significance value is shown for each. Cutoff value: TFx: 0.02, P-score: 0.74, S-score: 0.04, concentration: 48.49 ng/mL. Receiver operating characteristic (ROC) curve for (E) 1-y survival and (F) 3-y survival predicted by TFx, P-score, S-score, concentration (Con.), MVI, and Barcelona Clinic Liver Cancer (BCLC) stage. The AUCs are shown in corresponding brackets
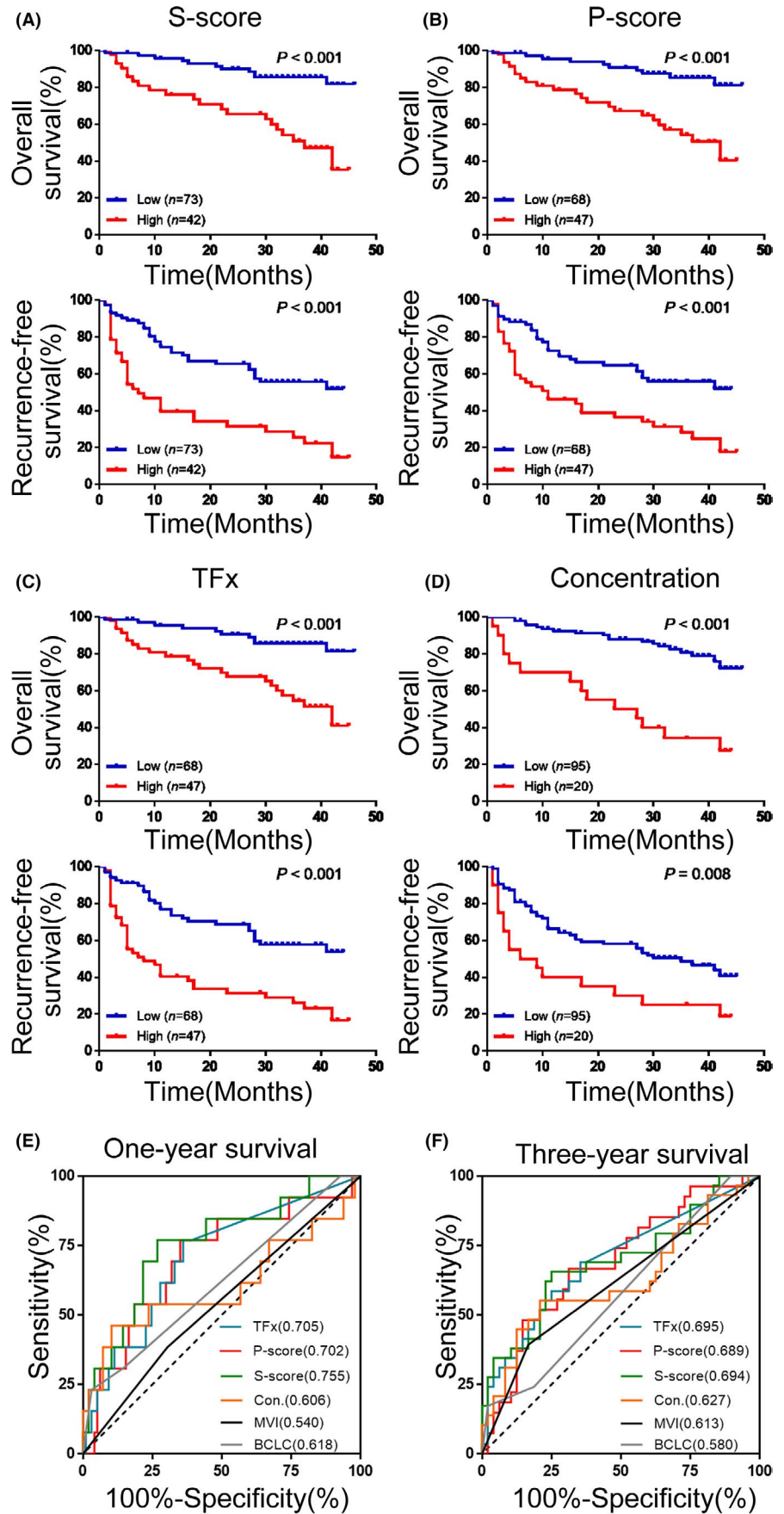
Variables	HR of death (95%CI)[a]	P	HR of recurrence (95%CI)[a]	P
S-score				
<0.04	Ref	<.001*	Ref	<.001*
≥0.04	4.02 (1.87-8.64)		2.55 (1.52-4.30)	
P-score				
<0.74	Ref	.001*	Ref	<.008*
≥0.74	3.70 (1.67-8.17)		2.06 (1.21-3.51)	
TFx				
<0.02	Ref	.001*	Ref	<.001*
≥0.02	3.72 (1.70-8.14)		2.78 (1.64-4.70)	
Concentration				
<48.49 ng/mL	Ref	<.001*	Ref	.027*
≥48.49 ng/mL	3.95 (1.89-8.27)		1.97 (1.08-3.59)	

TABLE 2 Multivariable Cox proportional hazard regression analysis of genome-wide copy number variation (CNV) indicators for overall survival (OS) and recurrence-free survival (RFS)

Abbreviations: CI, confidence interval; HR, Hazard Ratio; P-score, prediction score; S-score, stability score; TFx, tumor fraction.

[a]Adjusted by age, gender, microvascular invasion (MVI), alpha fetoprotein (AFP), and Barcelona Clinic Liver Cancer (BCLC) stage.

* P value has statistical significane (< 0.05).

3.6 | High frequency cfDNA CNVs at chromosomal-arm level predicted prognosis of HCC patients

To evaluate the prognostic value of cfDNA CNVs at chromosomal-arm level in HCC patients receiving radical treatments, the top four most frequent gains or losses were used. The Kaplan-Meier survival analysis (Figure S3) showed that patients with the loss of 4q, 17p, 19p, and 16q and the gain of 8q, 1q, and 20q exhibited significantly poorer OS and RFS than those without corresponding chromosomal-arm loss and gain, respectively (All $P < .05$). Moreover, 20p gain was only related to recurrence ($P = .002$) but not death ($P = .105$). Multivariate Cox regression analysis showed that the loss of 4q, 17p, and 19p and the gain of 8q and1q were independent prognostic factors for OS and RFS. The 20p gain and 20q gain were only observed to be associated with recurrence risk (all $P < .05$. Table 3). Further, as shown in Figure S4, the combination of 8q gain and 17p loss significantly improved the efficiency for prognosis prediction. These results suggest that high-frequency cfDNA CNVs at chromosomal-arm level may be used as prognostic biomarkers for HCC patients, and the combination of indicators at chromosomal-arm level may be a better strategy.

3.7 | A bin-level risk score improved the ability of CNVs in predicting prognosis

To investigate the prognostic value of cfDNA CNVs at bin (1 M) level in HCC patients, 1406 prognosis-associated candidate bins were selected from a total of 2475 bins by a univariable Cox regression analysis for further screening based on the LASSO model and multivariable Cox regression model. Finally, three bins were chosen to construct the bin-level risk score (simplified as bin score) according to the coefficients from the multivariable Cox regression analysis (Figures 4A, B and S5B). The distribution of patients with the low and high bin score, survival status of patients, and heat map of CNVs of the three hub bins are shown in Figure S5A. Two of three bins were located on chr8q (gain) and chr17q (loss), respectively. Another one was located on chr6q12, which could be found in TCGA copy number dataset from cBioPortal (Figure S5B). When HCC patients were stratified into two groups by the median value of bin score, the results showed that patients with high bin score had a significantly poorer OS than those with low bin score (log rank $P < .001$. Figure 4C). Similarly, the patients with high bin score were more likely to suffer recurrence than those with low bin score ($P = .002$. Figure S5C). The multivariable Cox regression analysis confirmed that the bin score was a prognosis biomarker independent of gender, age, BCLC stage, MVI, and AFP (OS: $P < .001$, HR = 1.09, 95%CI, 1.04-1.15; RFS: $P = .004$, HR = 2.16, 95%CI, 1.26-3.70. Table S7). The AUCs of bin score in predicting 1-year and 3-year survival were 0.820 and 0.746, respectively (Figure 4D), which was superior to the genome-wide cfDNA CNV indicators, high-frequency cfDNA CNVs at chromosomal-arm level, BCLC stage, and MVI. Furthermore, the bin score could also predict the prognosis in the patients whose TFx value was zero (Figure S5D). These findings suggest that the prediction ability of cfDNA CNVs for HCC prognosis may be improved by screening the most relevant CNV regions.

4 | DISSCUSSION

Here, we comprehensively presented the genomic characterization of plasma cfDNA from a large cohort of HCC patients receiving radical treatments at multiple levels. Remarkably similar CNV

TABLE 3 Multivariable Cox proportional hazard regression analysis of copy number variations (CNVs) at chromosomal-arm level for overall survival (OS) and recurrence-free survival (RFS)

Variables	HR of death (95%CI)[a]	P	HR of recurrence (95%CI)[a]	P
4q loss				
No	Ref	.001*	Ref	.035*
Yes	3.55 (1.66-7.56)		1.81 (1.04-3.14)	
17p loss				
No	Ref	<.001*	Ref	<.001*
Yes	4.31 (2.02-9.20)		2.74 (1.59-4.73)	
19p loss				
No	Ref	.002*	Ref	<.001*
Yes	3.49 (1.59-7.68)		3.27 (1.82-5.85)	
16q loss				
No	Ref	.109	Ref	.056
Yes	1.89 (0.87-4.17)		1.76 (0.99-3.15)	
20p gain				
No	Ref	.143	Ref	.039*
Yes	1.74 (0.83-3.64)		1.76 (1.03-3.01)	
8q gain				
No	Ref	.003*	Ref	.001*
Yes	3.20 (1.49-6.87)		2.49 (1.44-4.31)	
1q gain				
No	Ref	.019*	Ref	.033*
Yes	2.55 (1.16-5.59)		1.84 (1.05-3.21)	
20q gain				
No	Ref	.169	Ref	.017*
Yes	1.69 (0.80-3.56)		1.96 (1.13-3.39)	

Abbreviations: CI, confidence interval; HR, hazard ratio.

[a]Adjusted by age, gender, microvascular invasion (MVI), alpha fetoprotein (AFP), and Barcelona Clinic Liver Cancer (BCLC) stage.

* P value has statistical significane (< 0.05).

profiles were observed between cfDNA and tissue DNA of HCC patients, indicating the creditable implication of liquid biopsy for tumor detection. Using an NGS-based approach, indicators at genome-wide, chromosomal-arm, and bin levels were developed. Genome-wide CNV indicators including TFx, P-score, and S-score contain comprehensive information of variations at genome-wide level. However, CNV indicators at chromosomal-arm and bin level can identify the prognosis-predicting CNVs in a more concentrated area: chromosomal-arm and 1-Mb bin. We demonstrated that indicators based on cfDNA CNVs at different levels could be used as independent noninvasive prognostic biomarkers for HCC patients with radical treatments. Moreover, the bin-level indicator could also further improve the prognostic prediction ability of HCC. Therefore, it is more feasible and cost effective to analyze the prognostic value of cfDNA CNVs at bin level.

The consistency of genomic copy number between the cfDNA and tumor tissue DNA has been demonstrated in previous studies.[10,16] In this study, the consistency of CNV profiles between the tissue DNA of our cohort and TCGA indicated the reliability of our

sequencing data. Moreover, our results exhibited remarkably similar CNV profiles between cfDNA and tissue DNA at both bin level and chromosomal-arm level, suggesting that the cfDNA was an alternative sample source of tissue DNA for sequencing. It should be noted that the frequency of CNVs in cfDNA was significantly lower than that in tissue DNA samples. A possible explanation is that the detectable CNVs were diluted by predominantly normal cfDNA as background from white blood cell (WBC).[22] Further comparison of CNV profiles between cfDNA and WBC DNA will broaden our understanding on the influence of normal cfDNA in detecting CNVs.[23]

To analyze the association between cfDNA CNVs and HCC prognosis, genome-wide CNVs were taken into consideration firstly. As reported in previous studies, TFx could be used as a prognostic biomarker for HCC patients who received surgical resection and adjuvant therapies.[18] Based on an extended sample size, similar result was also observed in the present study. In our previous study, a RF model has been constructed for the detection of early-stage HCC patients. The current study further estimated the prognosis prediction capacity of the P-score calculated based on the RF model. As
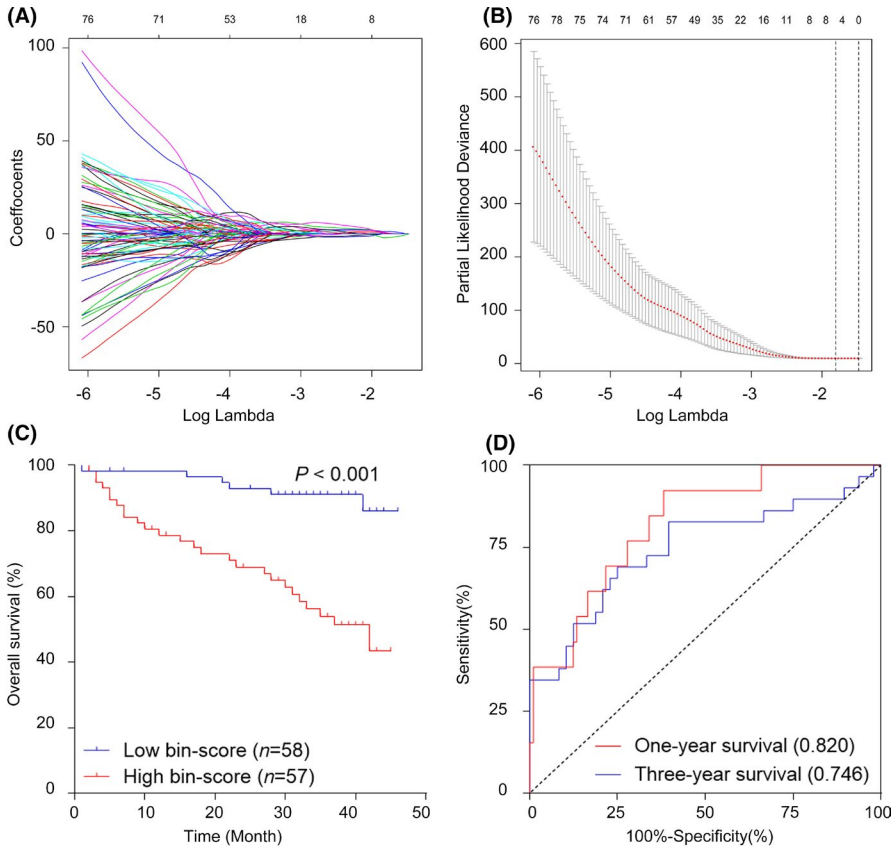
expected, the high P-score was found to be significantly associated with poor outcomes in HCC patients with radical treatments. The S-score, newly developed to estimate the genome stability, was also shown as an independent prognostic biomarker for HCC patients. Furthermore, no significant difference was observed among three genome-wide CNV indicators in prognosis prediction, possibly due to the high correlations among them, although they were calculated using different algorithms. In comparison, the predictive ability of cfDNA concentration was observed not very stable in distinguished patients with different prognosis. This is presumably due to the fact that the cfDNA concentration was influenced by experimental operation errors in cfDNA extraction and quantification or by the patients' status at the time of blood extraction.[24,25] In addition, variations in cfDNA concentration have been commonly found in different laboratories.[17,26] In general, all the genome-wide CNV indicators were shown to be independent prognostic factors, exhibiting more superior efficacy to the cfDNA concentration and current evaluation variables.

Next, the cfDNA CNVs at chromosomal-arm level were investigated. A series of previous studies have reported that CNVs at chromosomal-arm level are commonly observed in HCC tissues. Copy number gains at 1q, 6p, 8q, and 20q and copy number losses at 16q, 17p, 19p, 19q, 4q, 1p, 8p, 13q, and 22q have commonly been identified in HCC tissue samples.[27-30] Our results highlighted the top four frequent gain arms (20p, 8q, 1q, 20q) and the top four frequent loss arms (17p, 4q, 19p, 16q) in cfDNA of HCC patients. Several regions, such as gain of 1q, 8q24, and 20q13.12-13.33

and loss of 17p, have been reported to be related with the prognosis of HCC patients.[29,31-33] In the present study, we focused on CNVs at chromosomal-arm level. These arms characterized by high-frequency CNVs were observed to be associated with prognosis in HCC patients, among which gain of 20p and 20q could only serve as independent prognostic factors for predicting recurrence. In addition, the 20p gain was easier to be detected in cfDNA than in primary tissue DNA, which further highlighted the complementary role of cfDNA in the detection of HCC-related CNVs. A series of studies have demonstrated that the analysis of CNV loss or gain contributes to clarifying the potential molecular mechanisms in hepatocarcinogenesis. Gain of 1q has been suggested as an early genomic lesion during HCC progression. ALC1 gene, which is located on 1q, has been reported to promote the proliferation of HCC cells by downregulating p53 expression.[34] Moreover, overexpression of PTK2 and EIF3S3, which are located on Chr8q, has been reported to result in HCC progression.[35] Previous studies have also described the PRDM5 gene located on 4q as a major microenvironment-dependent and epigenetically regulated lineage-commitment factor in HCC.[36] Furthermore, underlying mechanisms how the loss or gain of CNV affects the prognosis of hepatocellular carcinoma need to be further explored.

Finally, the prognostic value of cfDNA CNVs was evaluated at bin level. Three bins were selected to construct a bin score. Our result indicated that the bin score was an independent prognostic factor and superior to clinical factors, genome-wide CNV indicators and chromosomal-arm CNVs, which provided a new way for developing

the cfDNA CNVs as new prognostic biomarkers. The bin score could also distinguish the patients with undetectable TFx into high and low risk of death and recurrence, indicating the wide application of the bin score in prognosis prediction of HCC patients. Moreover, we also analyzed the potential candidate genes in these regions. Briefly, CYP5D2, a member of the cytochrome P450 family, was located at 17p13.2. Previous studies have demonstrated the downregulation of CYP5D2 is associated with breast cancer progression, suggesting that CYP5D2 might be a potential cancer suppressor gene.[37] Another candidate gene, PREX2, was located at chr8q13.2, which has previously been reported to be associated with poor survival of HCC patients.[38] PTP4A1, located at chr6q12, was also identified as candidate gene, which has previously been reported to enhance the migration and invasion of HCC cells.[39] These findings suggest that screening the most relevant prognostic regions may improve the predicting capacity of cfDNA CNVs. Moreover, the detection of focused regions can further help select the potential candidate genes related to the prognosis of HCC.

In practical application, each of these "levels" was demonstrated to have its own advantage. Among them, our present study suggests that the CNV indicator at bin level may be most valuable for the prognostic prediction of HCC patients. On the one hand, it exhibited the greatest AUC among all CNV indicators. On the other hand, the CNV indicator at bin level can also predict the prognosis of the patients whose TFx value was zero, while other CNV indicators cannot. As there may be partial information overlaps between these levels, it may be not appropriate to perform the combination across different levels. However, we have observed an increased efficacy by combining the indicators at chromosomal-arm level, which suggests that the combination of indicators at chromosomal-arm level may be an optimal approach. Our study is also to some extent limited by a relatively small sample size. An independent data set is needed to validate the performance of cfDNA CNVs at three levels in predicting prognosis of HCC patients.

Taken together, a framework of multiple-level cfDNA CNV analysis is illustrated in this study. Our results provide a comprehensive genomic profile of cfDNA CNVs from HCC patients. We demonstrate that cfDNA CNV indicators at different levels provide important prognosis information for HCC patients with radical treatments beyond clinicopathologic factors and cfDNA concentration. This approach is helpful for broadening the applicable strategy to reveal clinically useful biomarkers based on cfDNA CNV analysis.

CONFLICTS OF INTEREST
The authors declare that they have no competing interests.

AUTHORS' CONTRIBUTIONS
YW, XW, and KZ performed experiments, analyzed data, and wrote the manuscript. DG and YL revised the manuscript. ZB and KL carried out the sample collection and performed experiments. LS analyzed data. XG, LW, HZ, and XG analyzed and discussed data. JX and KT conceived and designed the study, discussed data, and revised the manuscript. All authors approved the final version of the manuscript.

ETHICAL APPROVAL AND CONSENT TO PARTICIPATE
This research was approved by the Ethical Committee of the Fourth Military Medical University. All recruited patients provided written informed consent before enrolment.

DATA AVAILABILITY STATEMENT
The raw sequencing data are available in BIG Sub with access number PRJCA006319 (https://bigd.big.ac.cn/gsub/).

ORCID
Jinliang Xing [iD] https://orcid.org/0000-0002-7010-1822

REFERENCES
1. Ferlay J, Colombet M, Soerjomataram I, et al. Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. *Int J Cancer*. 2019;144:1941-1953.
2. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2018;68:394-424.
3. Cainap C, Qin S, Huang WT, et al. Linifanib versus Sorafenib in patients with advanced hepatocellular carcinoma: results of a randomized phase III trial. *J Clin Oncol*. 2015;33:172-179.
4. Díaz-González Á, Reig M, Bruix J. Treatment of Hepatocellular Carcinoma. *Dig Dis*. 2016;34:597-602.
5. Tang ZY, Ye SL, Liu YK, et al. A decade's studies on metastasis of hepatocellular carcinoma. *J Cancer Res Clin Oncol*. 2004;130:187-196.
6. Desai JR, Ochoa S, Prins PA, He AR. Systemic therapy for advanced hepatocellular carcinoma: an update. *J Gastrointest Oncol*. 2017;8:243-255.
7. Nishida N, Arizumi T, Hagiwara S, Ida H, Sakurai T, Kudo M. MicroRNAs for the prediction of early response to sorafenib treatment in human hepatocellular carcinoma. *Liver Cancer*. 2017;6:113-125.
8. Horwitz E, Stein I, Andreozzi M, et al. Human and mouse VEGFA-amplified hepatocellular carcinomas are highly sensitive to sorafenib treatment. *Cancer Discov*. 2014;4:730-743.
9. Hann HW, Fu X, Myers RE, et al. Predictive value of alpha-fetoprotein in the long-term risk of developing hepatocellular carcinoma in patients with hepatitis B virus infection–results from a clinic-based longitudinal cohort. *Eur J Cancer*. 2012;48:2319-2327.
10. Stover DG, Parsons HA, Ha G, et al. Association of Cell-Free DNA Tumor Fraction and Somatic Copy Number Alterations With Survival in Metastatic Triple-Negative Breast Cancer. *J Clin Oncol*. 2018;36:543-553.
11. Weiss GJ, Beck J, Braun DP, et al. Tumor Cell-Free DNA Copy Number Instability Predicts Therapeutic Response to Immunotherapy. *Clin Cancer Res*. 2017;23:5074-5081.
12. Oellerich M, Schütz E, Beck J, Walson PD. Circulating cell-free DNA-diagnostic and prognostic applications in personalized cancer therapy. *Ther Drug Monit*. 2019;41:115-120.
13. Xu RH, Wei W, Krawczyk M, et al. Circulating tumour DNA methylation markers for diagnosis and prognosis of hepatocellular carcinoma. *Nat Mater*. 2017;16:1155-1161.
14. Yang JD, Liu MC, Kisiel JB. Circulating Tumor DNA and Hepatocellular Carcinoma. *Semin Liver Dis*. 2019;39:452-462.

15. Tao K, Bian Z, Zhang Q, et al. Machine learning-based genome-wide interrogation of somatic copy number aberrations in circulating tumor DNA for early detection of hepatocellular carcinoma. *EBioMedicine*. 2020;56:102811.

16. Adalsteinsson VA, Ha G, Freeman SS, et al. Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors. *Nat Commun*. 2017;8:1324.

17. Oh CR, Kong SY, Im HS, et al. Genome-wide copy number alteration and VEGFA amplification of circulating cell-free DNA as a biomarker in advanced hepatocellular carcinoma patients treated with Sorafenib. *BMC Cancer*. 2019;19:292.

18. Cai Z, Chen G, Zeng Y, et al. Comprehensive liquid profiling of circulating tumor DNA and protein biomarkers in long-term follow-up patients with hepatocellular carcinoma. *Clin Cancer Res*. 2019;25:5284-5294.

19. Yin C, Li DY, Guo X, et al. NGS-based profiling reveals a critical contributing role of somatic D-loop mtDNA mutations in HBV-related hepatocarcinogenesis. *Ann Oncol*. 2019;30:953-962.

20. Tibshirani R. The lasso method for variable selection in the Cox model. *Stat Med*. 1997;16:385-395.

21. Wang XH, Liao B, Hu WJ, et al. Novel models predict postsurgical recurrence and overall survival for patients with hepatitis B virus-related solitary hepatocellular carcinoma ≤10 cm and without portal venous tumor thrombus. *Oncologist*. 2020;25(10):e1552-e1561.

22. Siravegna G, Bardelli A. Genotyping cell-free tumor DNA in the blood to detect residual disease and drug resistance. *Genome Biol*. 2014;15:449.

23. Leal A, van Grieken NCT, Palsgrove DN, et al. White blood cell and cell-free DNA analyses for detection of residual disease in gastric cancer. *Nat Commun*. 2020;11:525.

24. Breitbach S, Sterzing B, Magallanes C, Tug S, Simon P. Direct measurement of cell-free DNA from serially collected capillary plasma during incremental exercise. *J Appl Physiol*. 1985;2014(117):119-130.

25. De Vlaminck I, Martin L, Kertesz M, et al. Noninvasive monitoring of infection and rejection after lung transplantation. *Proc Natl Acad Sci U S A*. 2015;112:13336-13341.

26. Liao W, Yang H, Xu H, et al. Noninvasive detection of tumor-associated mutations from circulating cell-free DNA in hepatocellular carcinoma patients by targeted deep sequencing. *Oncotarget*. 2016;7:40481-40490.

27. Guan XY, Fang Y, Sham JS, et al. Recurrent chromosome alterations in hepatocellular carcinoma detected by comparative genomic hybridization. *Genes Chromosomes Cancer*. 2000;29:110-116.

28. Moinzadeh P, Breuhahn K, Stützer H, Schirmacher P. Chromosome alterations in human hepatocellular carcinomas correlate with aetiology and histological grade–results of an explorative CGH meta-analysis. *Br J Cancer*. 2005;92:935-941.

29. Kim TM, Yim SH, Shin SH, et al. Clinical implication of recurrent copy number alterations in hepatocellular carcinoma and putative oncogenes in recurrent gains on 1q. *Int J Cancer*. 2008;123:2808-2815.

30. Hashimoto K, Mori N, Tamesa T, et al. Analysis of DNA copy number aberrations in hepatitis C virus-associated hepatocellular carcinomas by conventional CGH and array CGH. *Mod Pathol*. 2004;17:617-622.

31. Dong F, Yang Q, Wu Z, et al. Identification of survival-related predictors in hepatocellular carcinoma through integrated genomic, transcriptomic, and proteomic analyses. *Biomed Pharmacother*. 2019;114:108856.

32. Qin LX, Tang ZY. The prognostic molecular markers in hepatocellular carcinoma. *World J Gastroenterol*. 2002;8:385-392.

33. Wang D, Zhu ZZ, Jiang H, et al. Multiple genes identified as targets for 20q13.12-13.33 gain contributing to unfavorable clinical outcomes in patients with hepatocellular carcinoma. *Hepatol Int*. 2015;9:438-446.

34. Ma NF, Hu L, Fung JM, et al. Isolation and characterization of a novel oncogene, amplified in liver cancer 1, within a commonly amplified region at 1q21 in hepatocellular carcinoma. *Hepatology*. 2008;47:503-510.

35. Okamoto H, Yasui K, Zhao C, Arii S, Inazawa J. PTK2 and EIF3S3 genes may be amplification targets at 8q23-q24 and are associated with large hepatocellular carcinomas. *Hepatology*. 2003;38:1242-1249.

36. Seehawer M, Heinzmann F, D'Artista L, et al. Necroptosis microenvironment directs lineage commitment in liver cancer. *Nature*. 2018;562:69-75.

37. Kimura I, Nakayama Y, Konishi M, et al. Neuferricin, a novel extracellular heme-binding protein, promotes neurogenesis. *J Neurochem*. 2010;112:1156-1167.

38. Li CH, Yen CH, Chen YF, et al. Characterization of the GNMT-HectH9-PREX2 tripartite relationship in the pathogenesis of hepatocellular carcinoma. *Int J Cancer*. 2017;140:2284-2297.

39. Wei-Ya L, Yu-Qing D, Yang-Chun M, Xin-Hua L, Ying M, Wang RL. Exploring the cause of the inhibitor 4AX attaching to binding site disrupting protein tyrosine phosphatase 4A1 trimerization by molecular dynamic simulation. *J Biomol Struct Dyn*. 2019;37:4840-4851.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.