

# Multiple-level copy number variations in circulating cell-free DNA for prognostic prediction of hepatocellular carcinoma with radical treatments

**Yang Wang**

XiJing Hospital

**Xiangxu Wang**

Xijing Hospital

**Kaixiang Zhou**

Fourth Military Medical University

**Yang Liu**

Fourth Military Medical University

**Zhenyuan Bian**

Xijing Hospital

**Dongna Guo**

Xijing Hospital

**Liping Su**

Fourth Military Medical University

**Kun Liu**

Xijing Hospital

**Xiwen Gu**

Xi'an Jiaotong University

**Xu Guo**

Fourth Military Medical University

**Kaishan Tao**

Xijing Hospital

**Jinliang Xing (✉ [xingjinliang@163.com](mailto:xingjinliang@163.com))**

Fourth Military Medical University

---

## Research

**Keywords:** copy number variations, circulating cell-free DNA, hepatocellular carcinoma, prognosis

**Posted Date:** August 20th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-62262/v1>



# Abstract

## Background

Copy number variations (CNVs) in circulating free DNA (cfDNA) are emerging as minimally invasive prognostic biomarkers for various cancers. However, little has been reported on the multiple-level analysis of cfDNA CNVs for hepatocellular carcinoma (HCC) patients with radical treatments.

## Methods

Here, CNVs at genome-wide, chromosomal-arm and bin levels were analyzed in cfDNA from 117 HCC patients receiving radical treatments via low-coverage whole genome sequencing. 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 based on genome-wide cfDNA CNVs.

## Results

Kaplan-Meier analysis showed that the patients with high TFx, P-score and S-score exhibited a significantly poorer overall survival (OS) and recurrence free survival (RFS) than those with low TFx, P-score and S-score, respectively (All  $P < 0.05$ ). Furthermore, a group of high frequency cfDNA CNVs at chromosomal-arm level including loss of 4q, 17p, 19p and the gain of 8q, 1q clearly predicted prognosis of HCC patients. Finally, a bin-level risk score was constructed based on three most relevant prognostic bin regions identified by a LASSO model. Patients with high bin-score had a significantly poorer OS than those with low bin-score ( $P < 0.001$ ).

## Conclusion

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 WGS may be used as potential prognostic biomarkers of HCC patients.

## 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 clinic 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 contains the fragmented circulating tumor DNA mainly derived from dead tumor cells through necrosis and apoptosis, which can be used as a surrogate source of tumor tissue DNA(10–12). Carrying the comprehensive information of tumor genome profiles, such as methylations, point mutations, and copy number variations (CNVs), cfDNA therefore holds great potential to represent the entire molecular picture of a patient's malignancy. Several studies have reported the significant promise of cfDNA as novel biomarkers 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 ultra-low 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 (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. Then, the relationship between cfDNA CNVs and clinical outcomes was explored. 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 WGS may be used as potential prognostic biomarkers for HCC patients.

## Materials And Methods

### Patients and sample collection

A total of 117 HBV-related hepatocellular carcinoma (HBV-HCC) patients who were subjected to radical treatments were recruited between May 2016 and July 2017 from Department of Hepatobiliary Surgery at Xijing Hospital, Fourth Military Medical University (FMMU), Xi'an, China (Fig. 1). Prior to treatment, 5mL of peripheral blood sample was collected in EDTA tubes (BD) from each patient. Fresh HCC tumor tissue samples were also obtained from available cases. This study was approved by the Ethical Committee of FMMU and written consent was obtained from each patient.

### Plasma processing and cfDNA extraction

To separate plasma, peripheral blood was firstly centrifuged at 1,600 rpm at 4°C for 10 minutes. The resultant plasma supernatant was further purified by centrifugation at 10,000 rpm at 4°C for 15 min to remove residual cells or debris. The plasma was stored at -80°C until DNA extraction. 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).

### **Tumor tissue collection and DNA extraction**

A total of 19 HCC tissue samples were selected for further CNV 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 HCC tumor samples using the E.N.Z.A tissue DNA Kit (Omega) according to the manufacturer's protocol. The quantity and quality of tissue DNA samples were evaluated using NanoDrop 2000 (Thermo) and agilent 2100 bioanalyzer (Agilent).

### **Library preparation and whole genome sequencing**

The construction of whole genome sequencing library for tissue DNA samples 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(19, 20). 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.

### **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 TCGA database was obtained through the UCSC Genome Browser (<http://xena.ucsc.edu/>).

### **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 P-score was calculated as previously reported, which was based on a random forest (RF) model of HBV-HCC and cancer-free HBV patients using the genome-wide CNV profiles(21). This P-score was further used for prognostic assessment. Thirdly, we defined a novel S-score, which was intended to reflect the extent of CNV deviation from diploid genome and was calculated using the following formula:

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

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

## Chromosomal-arm level CNV analysis

First, the genome was divided into non-overlapping bins of 1-megabase 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% regions of 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 / total patient number and used for subsequent analysis.

## Construction of bin-level risk score

A univariable Cox proportional hazard regression analysis was performed on all 1-megabase 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=10000). The LASSO regression was analyzed using 'glmnet' package from R (version 3.6.2) (22). Next, the multivariate Cox proportional hazard regression model was used to construct the bin-level risk score, which was defined as below:

$$\text{Bin score} = \sum_i^n (\beta_i \times \text{Cnv}_i)$$

Where  $\beta$  represents the coefficient value and  $\text{Cnv}$  represents the bin's CNV level defined by ichorCNA algorithm.

## Statistical analysis

Pearson's correlation coefficients were used to determine the correlation of CNV frequencies among the cfDNA and tissue DNA in our cohort and tissue DNA from TCGA. Spearman correlation coefficients were used to evaluate the correlations among genome-wide CNV indicators. 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 Kaplan–Meier log-rank test. Univariable and multivariable Cox proportional-hazards 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  $<0.05$  was considered as statistically significant and all probabilities were two-tailed.

# Results

## Clinical characteristics of 117 HBV-related HCC patients

As shown in Table 1, the median age of patients was 55 years old (ranging from 31–79 years), and 102 (87%) patients were male. In this cohort, tumor size was smaller than 3cm in 30 (26%) patients, 3-5 cm in 50 (42%) patients, and larger than 5cm in 37 (32%) patients. Microvascular invasion (MVI) was observed in 74 (63%) patients. There were 61 (52%), 44 (38%) and 12 (10%) patients with AJCC stage I, stage II and stage III diseases, respectively, while 98 (84%) and 19 (16%) of the patients with Barcelona Clinic Liver Cancer stage

(BCLC) 0-A and B-C. The liver function in nearly all the patients (n=112, 96%) was distributed at Child-Pugh Class (CTP class) A. 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.

### **Multiple-level CNV analysis of cfDNA**

Whole-genome sequencing 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 ichorCNA was used for subsequent CNV analysis. TFX, P-score and S-score were calculated based on genome-wide cfDNA CNVs. The different distribution of TFX, P-score, S-score and cfDNA concentration was shown in Fig. S1A. Significant positive correlation was observed among TFX, P-score and S-score. The top 4 most frequent gains of 20p, 8q, 1q, 20q and the top 4 most frequent losses of 17p, 4q, 19p and 16q were identified in 117 HBV-HCC patients (Fig. S1C). Furthermore, significant negative correlation was found between gain and loss frequencies (Pearson  $r=-0.804$ ,  $P<0.001$ , Fig. S1C), indicating that CNVs at chromosomal-arm level preferred to gain or loss, but rarely both. In addition, the bin-level CNVs were obtained from ichorCNA and the bin frequencies altered across genome among HBV-HCC patients were exhibited in Fig. S1D.

### **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 The Cancer Genome Atlas (TCGA) database. Similar patterns of CNVs at both bin and chromosomal-arm level were observed between cfDNA and tissue DNA (Fig. 2A and 2C). 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:  $r=0.649$ ,  $P<0.001$ ; Loss:  $r=0.856$ ,  $P<0.001$ .  $n=19$ ; Fig. 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:  $r=0.779$ ,  $P<0.001$ ; Loss:  $r=0.912$ ,  $P<0.001$ . Fig. 2D). In addition, the CNV frequency in TCGA tissue DNA were significantly correlated with the CNV frequency in both tissue DNA (Gain:  $r=0.739$ ;  $P<0.001$ , Loss:  $r=0.692$ ;  $P<0.001$ . Fig. 2D) and cfDNA (Gain:  $r=0.718$ ;  $P<0.001$ , Loss:  $r=0.666$ ;  $P<0.001$ . Fig. 2D) 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.

### **Association of cfDNA CNVs with clinical characteristics in HCC patients.**

The association of demographic and clinical characteristics with CNVs detected at different levels and concentration of cfDNA was presented in Table 1 and Table S2-5, respectively. Patients were categorized into two groups by the cut-off 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 ( $\geq 5$ cm) than those with low TFX, P-score and S-score, respectively (All  $P<0.05$ ). In addition, an obvious trend was observed that TFX, P-score and S-score were associated with MVI and BCLC stage, although only P-score attained statistical significance ( $P<0.05$ ). 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  $P < 0.05$ ). The similar results were observed in patients with gains of 20p, 8q, 1q and 20q (Table S2), losses of 17p, 4q, 19p and 16q (Table S3), high bin-score (Table S4) and high cfDNA concentration (Table S5).

### **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 overall survival (OS) and recurrence free survival (RFS) of HCC patients receiving radical treatment, which is greatly consistent with previous reports(23). Furthermore, HCC patients were respectively stratified by cut-off 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  $P < 0.05$ . Fig.3). Similar results were observed when median value of TFx, P-score and S-score was used as cut-off values (All  $P < 0.05$ . Fig. S2A, S2B and S2C). However, no significant difference was observed between patients with high and low cfDNA concentration in OS ( $P = 0.117$ ) and RFS ( $P = 0.731$ . Fig. S2D) when median value was used as the cut-off 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 HR of 4.02 (95% 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). The 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).

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 survival 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, no matter of 3-year survival or 1-year survival (Fig. 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.

### **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 4 most frequent gains or loss were used. The Kaplan-Meier survival analysis (Fig. 5 and Fig. S3) showed that patients with loss of 4q, 17p, 19p 16q and the gain of 8q, 1q, 20q exhibited significantly poorer OS and RFS than those without corresponding chromosomal-arm loss and gain, respectively (All  $P < 0.05$ ). Moreover, 20p gain was only related to recurrence ( $P = 0.002$ ) but not death ( $P = 0.105$ ). Univariable Cox regression analysis also indicated the gain or loss of these chromosomal arms was significantly associated with OS and RFS in HCC patients (Table S6). Multivariate Cox regression analysis showed that loss of 4q, 17p, 19p and the gain of 8q, 1q were independent prognostic factors for OS and RFS. The HR of death were 3.55 (95%CI, 1.66-7.56), 4.31 (95%CI 2.02-9.20), 3.49 (95%CI 1.59-7.68) for loss of 4q, 17p, 19p and 3.20(95%CI 1.49-6.87), 2.55(95%CI 1.16-5.59) for gain of 8q, 1q, respectively (Table 3). The HR of recurrence were 1.81 (95%CI,

1.04-3.14), 2.74(95%CI 1.59-4.73), 3.27 (95%CI 1.82-5.85) for loss of 4q, 17p, 19p, and 2.49 (95%CI 1.44-4.31), 1.84 (95%CI 1.05-3.21) for gain of 8q, 1q, respectively (Table 3). The 20p gain and 20q gain were only observed to be associated with recurrence risk, with HR of 1.76 (95%CI 1.03-3.01) and 1.96 (95%CI 1.13-3.39), respectively (All  $P < 0.05$ . Table 3). ROC curve analysis showed that these gain and loss at chromosomal-arm level showed similar AUCs either for 3-year survival or 1-year survival, with a slightly better performance for 1-year survival (Fig.6). These results suggest that high frequency cfDNA CNVs at chromosomal-arm level may be used as prognostic biomarkers for HCC patients.

### **A bin-level risk score improved the ability of CNVs in predicting prognosis**

To investigate the prognostic value of cfDNA CNVs at bin (1M) 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 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 (Fig. 7A, 7B and Fig. S4B). 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 were shown in Fig. S4A. Two of three bins were located on the chr8q (gain) and chr17q (loss), respectively. Another one was located at chr6q12, which could be found in TCGA copy number dataset from cBioPortal (Fig. S4B). 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  $P < 0.001$ . Fig. 7C). Similarly, the patients with high bin-score were more likely to suffer recurrence than those with low bin-score ( $P = 0.002$ . Fig. S4C). The multivariable Cox regression analysis confirmed that the bin-score was a prognosis biomarker independent of gender, age, BCLC stage, MVI and AFP (OS:  $P < 0.001$ , HR=1.09, 95%CI, 1.04-1.15; RFS:  $P = 0.004$ , 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 0.8202 and 0.7457, respectively (Fig. 7D), 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 (Fig. S4D). These findings suggest that the prediction ability of cfDNA CNVs for HCC prognosis may be improved by screening the most relevant CNV regions.

## **Discussion**

Here, we comprehensively presented the genomic characterization of plasma cfDNA from a large cohort of HCC patients receiving radical treatments at genome-wide, chromosomal-arm and bin levels. The remarkably similar CNV profiles were observed between cfDNA and tumor tissue DNA of HCC patients, indicating the creditable implication of liquid biopsy for tumor detection. Using NGS-based approach, we demonstrated that indicators based on cfDNA CNVs at different levels could be used as independent non-invasive prognostic biomarkers for HCC patients with radical treatments.

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 those 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)(24). Further comparison of CNV profiles between cfDNA and WBC DNA will broaden our understanding on the influence of normal cfDNA on the detection of CNVs(25).

To analyze the association between cfDNA CNVs and HCC prognosis, genome-wide CNVs was 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(25). Based on an extended sample size, similar result was also observed in the present study. In our previous study, a random forest 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, cfDNA concentration could distinguish patients with different prognosis only when the cut-off value was determined by Youden index. The cfDNA concentration may be influenced by experimental operation errors in cfDNA extraction and quantification or by the patients' status at the time of blood extraction(26, 27). And variations in cfDNA concentration have been commonly found in different laboratories(17, 28). In general, all the genome-wide CNVs 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(29–32). Our results highlighted the top 4 frequent gain arms (20p, 8q, 1q, 20q) and the top 4 frequent loss arms (17p, 4q, 19p, 16q) in cfDNA of HCC patients. Several regions, such as gain of 1q, 8q24, 20q13.12-13.33 and loss of 17p, have been reported to be related with the prognosis of HCC patients(31, 33–35). 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 receiving radical treatments, among which gain of 20p and 20q could only be served 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. The cfDNA might also provide recurrence prediction information, which could not be obtained from primary tumor tissue DNA. The biological significance of these CNVs at chromosomal arm level detected in cfDNA should be investigated in further studies.

Finally, the prognostic value of cfDNA CNVs was evaluated at bin level. Three bins were selected to construct a bin-level risk score by using the LASSO and multivariate Cox method. Our result indicated that the bin-score was an independent prognostic factor and superior to other prognostic factors, including 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. These findings suggest that screening the most relevant prognostic regions will improve the predicting capacity of cfDNA CNVs. Moreover, the detection of focused regions would further reduce the cost of the assay, which makes it more suitable for future applications in clinical settings.

## Conclusion

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 CNVs analysis. 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.

## Abbreviations

CNVs: Copy number variations; cfDNA: circulating free DNA; HCC: Hepatocellular carcinoma; Tfx: Tumor fraction; P-score: Prediction score; S-score: stability score; OS: Overall survival; RFS: Recurrence free survival; AFP: Alpha fetoprotein; CTP class: Child-Pugh Class; TCGA: The Cancer Genome Atlas; ROC: Receiver operating characteristic; AUCs: areas under the ROC curve; LASSO: Least absolute shrinkage and selection operator; WBC: white blood cell

## Declarations

## Availability of data and materials

The data that support the findings of this study are available from the corresponding author upon reasonable request.

### Ethics approval and consent to participate

Informed consent was obtained in accordance with the Declaration of Helsinki. This study was approved by the Ethical Committees of FMMU.

### Consent for publication

Not applicable.

### Competing interests

The authors have declared no competing interests.

# Funding

This work was supported by funding from the National Natural Science Foundation of China (81830070) and the Key Research and Development Program of Shaanxi Province, China (2018ZDXM-SF-061).

# Authors' contributions

YW, XW and KZ carried out the sample collection, performed the data analysis, and drafted the manuscript. YL, ZB and LS participated in the bioinformatics analyses. DG and KL performed the laboratory experiments. XG and XG participated in the design of the study and performed the draft revision. KT and JX conceived of the study, and participated in its design and coordination and helped to revise the manuscript. All authors read and approved the final manuscript.

# Acknowledgements

The authors thank Shanshan Guo and Deyang Li in Department of Physiology and Pathophysiology for ongoing support and discussion.

# References

1. Ferlay J, Colombet M, Soerjomataram I, Mathers C, Parkin DM, Piñeros M, et al. Estimating the global cancer incidence and mortality in 2018: GLOBOCAN sources and methods. *International journal of cancer*. 2019;144(8):1941–53.
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. *Cancer J Clin*. 2018;68(6):394–424.
3. Cainap C, Qin S, Huang WT, Chung IJ, Pan H, Cheng Y, et al. Linifanib versus Sorafenib in patients with advanced hepatocellular carcinoma: results of a randomized phase III trial. *Journal of clinical oncology: official journal of the American Society of Clinical Oncology*. 2015;33(2):172–9.
4. Díaz-González Á, Reig M, Bruix J. Treatment of Hepatocellular Carcinoma. *Digestive diseases (Basel, Switzerland)*. 2016;34(5):597–602.
5. Tang ZY, Ye SL, Liu YK, Qin LX, Sun HC, Ye QH, et al. A decade's studies on metastasis of hepatocellular carcinoma. *J Cancer Res Clin Oncol*. 2004;130(4):187–96.
6. Desai JR, Ochoa S, Prins PA, He AR. Systemic therapy for advanced hepatocellular carcinoma: an update. *Journal of gastrointestinal oncology*. 2017;8(2):243–55.
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(2):113–25.
8. Horwitz E, Stein I, Andreozzi M, Nemeth J, Shoham A, Pappo O, et al. Human and mouse VEGFA-amplified hepatocellular carcinomas are highly sensitive to sorafenib treatment. *Cancer discovery*. 2014;4(6):730–43.

9. Hann HW, Fu X, Myers RE, Hann RS, Wan S, Kim SH, 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. *European journal of cancer (Oxford, England: 1990)*. 2012;48(15):2319–27.
10. Stover DG, Parsons HA, Ha G, Freeman SS, Barry WT, Guo H, et al. Association of Cell-Free DNA Tumor Fraction and Somatic Copy Number Alterations With Survival in Metastatic Triple-Negative Breast Cancer. *Journal of clinical oncology: official journal of the American Society of Clinical Oncology*. 2018;36(6):543–53.
11. Weiss GJ, Beck J, Braun DP, Bornemann-Kolatzki K, Barilla H, Cubello R, et al. Tumor Cell-Free DNA Copy Number Instability Predicts Therapeutic Response to Immunotherapy. *Clinical cancer research: an official journal of the American Association for Cancer Research*. 2017;23(17):5074–81.
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(2):115–20.
13. Xu RH, Wei W, Krawczyk M, Wang W, Luo H, Flagg K, et al. Circulating tumour DNA methylation markers for diagnosis and prognosis of hepatocellular carcinoma. *Nat Mater*. 2017;16(11):1155–61.
14. Yang JD, Liu MC, Kisiel JB. Circulating Tumor DNA and Hepatocellular Carcinoma. *Semin Liver Dis*. 2019;39(4):452–62.
15. Molparia B, Oliveira G, Wagner JL, Spencer EG, Torkamani A. A feasibility study of colorectal cancer diagnosis via circulating tumor DNA derived CNV detection. *PloS one*. 2018;13(5):e0196826.
16. Adalsteinsson VA, Ha G, Freeman SS, Choudhury AD, Stover DG, Parsons HA, et al. Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors. *Nature communications*. 2017;8(1):1324.
17. Oh CR, Kong SY, Im HS, Kim HJ, Kim MK, Yoon KA, 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(1):292.
18. Cai Z, Chen G, Zeng Y, Dong X, Li Z, Huang Y, et al. Comprehensive Liquid Profiling of Circulating Tumor DNA and Protein Biomarkers in Long-Term Follow-Up Patients with Hepatocellular Carcinoma. *Clinical cancer research: an official journal of the American Association for Cancer Research*. 2019;25(17):5284–94.
19. Tao K, Bian Z, Zhang Q, Guo X, Yin C, Wang Y, 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.
20. Yin C, Li DY, Guo X, Cao HY, Chen YB, Zhou F, et al. NGS-based profiling reveals a critical contributing role of somatic D-loop mtDNA mutations in HBV-related hepatocarcinogenesis. *Annals of oncology: official journal of the European Society for Medical Oncology*. 2019;30(6):953–62.
21. Li YR, Glessner JT, Coe BP, Li J, Mohebnasab M, Chang X, et al. Rare copy number variants in over 100,000 European ancestry subjects reveal multiple disease associations. *Nature communications*. 2020;11(1):255.
22. Tibshirani R. The lasso method for variable selection in the Cox model. *Statistics in medicine*. 1997;16(4):385–95.

23. Wang XH, Liao B, Hu WJ, Tu CX, Xiang CL, Hao SH, et al. Novel Models Predict Postsurgical Recurrence and Overall Survival for Patients with Hepatitis B Virus-Related Solitary Hepatocellular Carcinoma  $\leq 10$  cm and Without Portal Venous Tumor Thrombus. *The oncologist*. 2020.
24. Siravegna G, Bardelli A. Genotyping cell-free tumor DNA in the blood to detect residual disease and drug resistance. *Genome biology*. 2014;15(8):449.
25. Leal A, van Grieken NCT, Palsgrove DN, Phallen J, Medina JE, Hruban C, et al. White blood cell and cell-free DNA analyses for detection of residual disease in gastric cancer. *Nature communications*. 2020;11(1):525.
26. Breitbach S, Sterzing B, Magallanes C, Tug S, Simon P. Direct measurement of cell-free DNA from serially collected capillary plasma during incremental exercise. *Journal of applied physiology (Bethesda, Md: 1985)*. 2014;117(2):119 – 30.
27. De Vlaminc I, Martin L, Kertesz M, Patel K, Kowarsky M, Strehl C, et al. Noninvasive monitoring of infection and rejection after lung transplantation. *Proc Natl Acad Sci USA*. 2015;112(43):13336–41.
28. Liao W, Yang H, Xu H, Wang Y, Ge P, Ren J, et al. Noninvasive detection of tumor-associated mutations from circulating cell-free DNA in hepatocellular carcinoma patients by targeted deep sequencing. *Oncotarget*. 2016;7(26):40481–90.
29. Guan XY, Fang Y, Sham JS, Kwong DL, Zhang Y, Liang Q, et al. Recurrent chromosome alterations in hepatocellular carcinoma detected by comparative genomic hybridization. *Genes Chromosom Cancer*. 2000;29(2):110–6.
30. 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. *British journal of cancer*. 2005;92(5):935–41.
31. Kim TM, Yim SH, Shin SH, Xu HD, Jung YC, Park CK, et al. Clinical implication of recurrent copy number alterations in hepatocellular carcinoma and putative oncogenes in recurrent gains on 1q. *International journal of cancer*. 2008;123(12):2808–15.
32. Hashimoto K, Mori N, Tamesa T, Okada T, Kawauchi S, Oga A, et al. Analysis of DNA copy number aberrations in hepatitis C virus-associated hepatocellular carcinomas by conventional CGH and array CGH. *Modern pathology: an official journal of the United States and Canadian Academy of Pathology. Inc*. 2004;17(6):617–22.
33. Dong F, Yang Q, Wu Z, Hu X, Shi D, Feng M, et al. Identification of survival-related predictors in hepatocellular carcinoma through integrated genomic, transcriptomic, and proteomic analyses. *Biomedicine & pharmacotherapy = Biomedecine & pharmacotherapie*. 2019;114:108856.
34. Qin LX, Tang ZY. The prognostic molecular markers in hepatocellular carcinoma. *World journal of gastroenterology*. 2002;8(3):385–92.
35. Wang D, Zhu ZZ, Jiang H, Zhu J, Cong WM, Wen BJ, et al. Multiple genes identified as targets for 20q13.12-13.33 gain contributing to unfavorable clinical outcomes in patients with hepatocellular carcinoma. *Hepatology international*. 2015;9(3):438–46.

## Tables

Table 1. Clinicopathologic characteristics and genome-wide CNV indicators of HCC patients										
	All	S-score			P-score			TFx		
	patients	Low	High		Low	High		Low	High	
Variables	n=117	n=73	n=44	<i>P</i>	n=69	n=48	<i>P</i>	n=68	n=49	<i>P</i>
<b>Age</b>				0.255			0.262			0.093
<55	58 (50)	33 (45)	25 (57)		31 (45)	27 (56)		29 (43)	29 (59)	
≥55	59 (50)	40 (55)	19 (43)		38 (55)	21 (44)		39 (57)	20 (41)	
<b>Gender</b>				0.783			0.585			0.782
Male	102 (87)	63 (86)	39 (89)		59 (86)	43 (90)		60 (88)	42 (86)	
Female	15 (13)	10 (14)	5 (11)		10 (14)	5 (10)		8 (12)	7 (14)	
<b>AJCC stage</b>				0.020*			<0.001*			0.020*
I	61 (52)	46 (63)	15 (34)		47 (68)	14 (29)		43 (63)	18 (37)	
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)	
<b>AFP</b>				0.703			0.353			0.712
<20 ng/mL	55 (47)	35 (48)	19 (43)		35 (51)	20 (42)		33 (49)	22 (45)	
≥20 ng/mL	62 (53)	38 (52)	25 (57)		34 (49)	28 (58)		35 (51)	27 (55)	
<b>Tumor size</b>				0.001*			<0.001*			<0.001*
<3 cm	30 (26)	26 (35)	4 (9)		26 (37)	4 (8)		26 (38)	4 (8)	
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)	
<b>MVI</b>				0.103			0.041*			0.232
No	74 (63)	50 (69)	24 (55)		48 (70)	26 (54)		45 (66)	29 (59)	
Yes	37 (32)	19	18		16	21		18	19	

		(26)	(41)		(23)	(44)	(27)	(39)
Unknown	6 (5)	4 (5)	2 (4)		5 (7)	1 (2)	5 (7)	1 (2)
<b>BCLC stage</b>				0.069			0.069	0.046*
0/A	98 (84)	65 (89)	33 (75)		62 (89)	36 (75)	61 (90)	37 (76)
B/C	19 (16)	8 (11)	11 (25)		7 (11)	12 (25)	7 (10)	12 (24)
<b>CTP class</b>				1.000			1.000	1.000
A	112 (96)	70 (96)	42 (95)		66 (96)	46 (96)	65 (96)	47 (96)
B	5 (4)	3 (4)	2 (5)		3 (4)	2 (4)	3 (4)	2 (4)
<b>Recurrence</b>				<0.001*			0.001*	<0.001*
No	51 (44)	42 (57)	9 (20)		39 (56)	12 (25)	41 (60)	10 (20)
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)
<b>Vital status</b>				<0.001*			<0.001*	<0.001*
Alive	82 (70)	62 (85)	20 (45)		58 (84)	24 (50)	58 (85)	24 (49)
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 cut-off points of S-score, P-score and TFx were determined by the maximizing Youden index of classifying one-year recurrence. Abbreviations: HCC, hepatocellular carcinoma; S-score, stability score; P-score, predict score; TFx, tumor fraction; AFP, alpha-fetoprotein; MVI, Microvascular invasion; AJCC, American Joint Committee on Cancer Staging								

<b>Table 2. Multivariable Cox proportional hazard regression analysis of genome-wide CNV indicators for OS and RFS</b>				
<b>Variables</b>	<b>HR of death (95%CI)<sup>a</sup></b>	<b><i>P</i></b>	<b>HR of recurrence (95%CI)<sup>a</sup></b>	<b><i>P</i></b>
<b>S-score</b>				
<0.04	Ref		Ref	
≥0.04	4.02(1.87-8.64)	<0.001*	2.55(1.52-4.30)	<0.001*
<b>P-score</b>				
<0.74	Ref		Ref	
≥0.74	3.70(1.67-8.17)	0.001*	2.06(1.21-3.51)	<0.008*
<b>TFx</b>				
<0.02	Ref		Ref	
≥0.02	3.72(1.70-8.14)	0.001*	2.78(1.64-4.70)	<0.001*
<b>Concentration</b>				
<48.49 ng/mL	Ref		Ref	
≥48.49 ng/mL	3.95(1.89-8.27)	<0.001*	1.97(1.08-3.59)	0.027*
<sup>a</sup> Adjusted by age, gender, MVI, AFP and BCLC stage.				
Abbreviations: HR, Hazard Ratio; CI, confidence interval.				

<b>Table 3. Multivariable Cox proportional hazard regression analysis of CNVs at chromosomal-arm level for OS and RFS</b>				
<b>Variables</b>	<b>HR of death (95%CI)<sup>a</sup></b>	<b><i>P</i></b>	<b>HR of recurrence (95%CI)<sup>a</sup></b>	<b><i>P</i></b>
<b>4q loss</b>				
No	Ref		Ref	
Yes	3.55(1.66-7.56)	0.001*	1.81(1.04-3.14)	0.035*
<b>17p loss</b>				
No	Ref		Ref	
Yes	4.31(2.02-9.20)	<0.001*	2.74(1.59-4.73)	<0.001*
<b>19p loss</b>				
No	Ref		Ref	
Yes	3.49(1.59-7.68)	0.002*	3.27(1.82-5.85)	<0.001*
<b>16q loss</b>				
No	Ref		Ref	
Yes	1.89(0.87-4.17)	0.109	1.76(0.99-3.15)	0.056
<b>20p gain</b>				
No	Ref		Ref	
Yes	1.74(0.83-3.64)	0.143	1.76(1.03-3.01)	0.039*
<b>8q gain</b>				
No	Ref		Ref	
Yes	3.20(1.49-6.87)	0.003*	2.49(1.44-4.31)	0.001*
<b>1q gain</b>				
No	Ref		Ref	
Yes	2.55(1.16-5.59)	0.019*	1.84(1.05-3.21)	0.033*
<b>20q gain</b>				
No	Ref		Ref	
Yes	1.69(0.80-3.56)	0.169	1.96(1.13-3.39)	0.017*
<sup>a</sup> Adjusted by age, gender, MVI, AFP and BCLC stage.				
Abbreviations: HR, Hazard Ratio; CI, confidence interval.				

## Figures

Figure 1

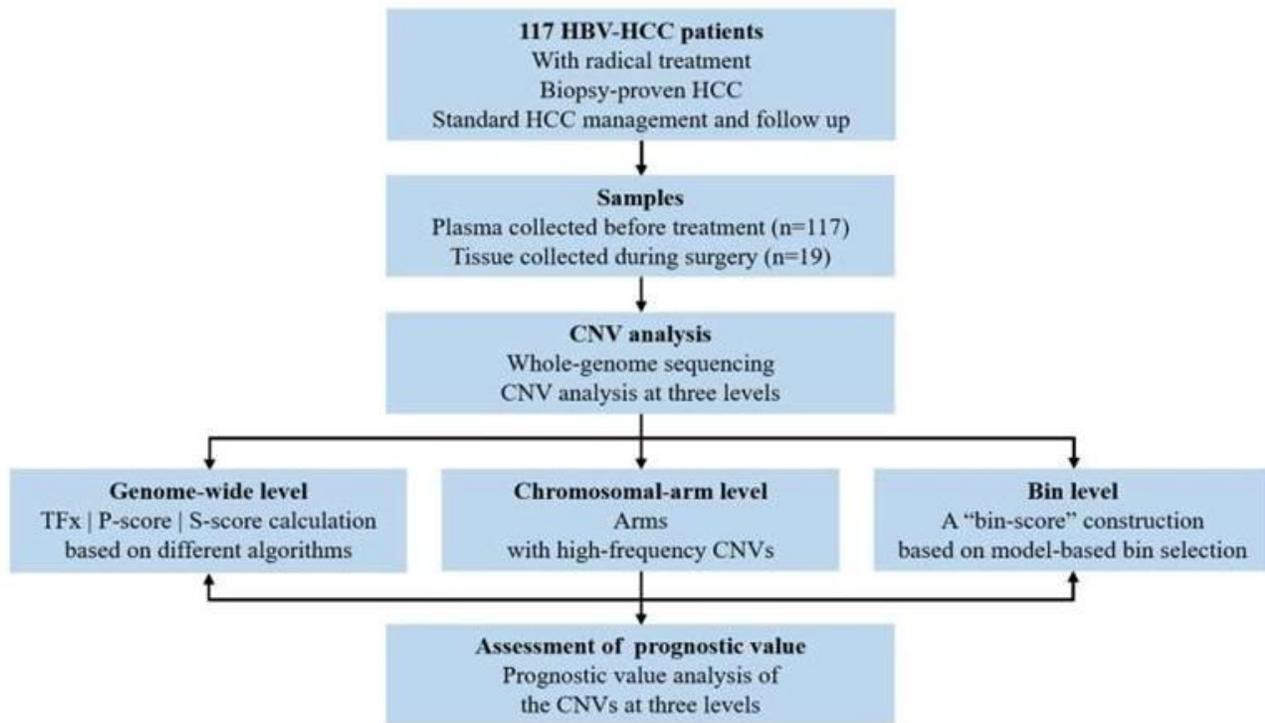


Figure 1

Flow diagram of the study design.

Figure 2

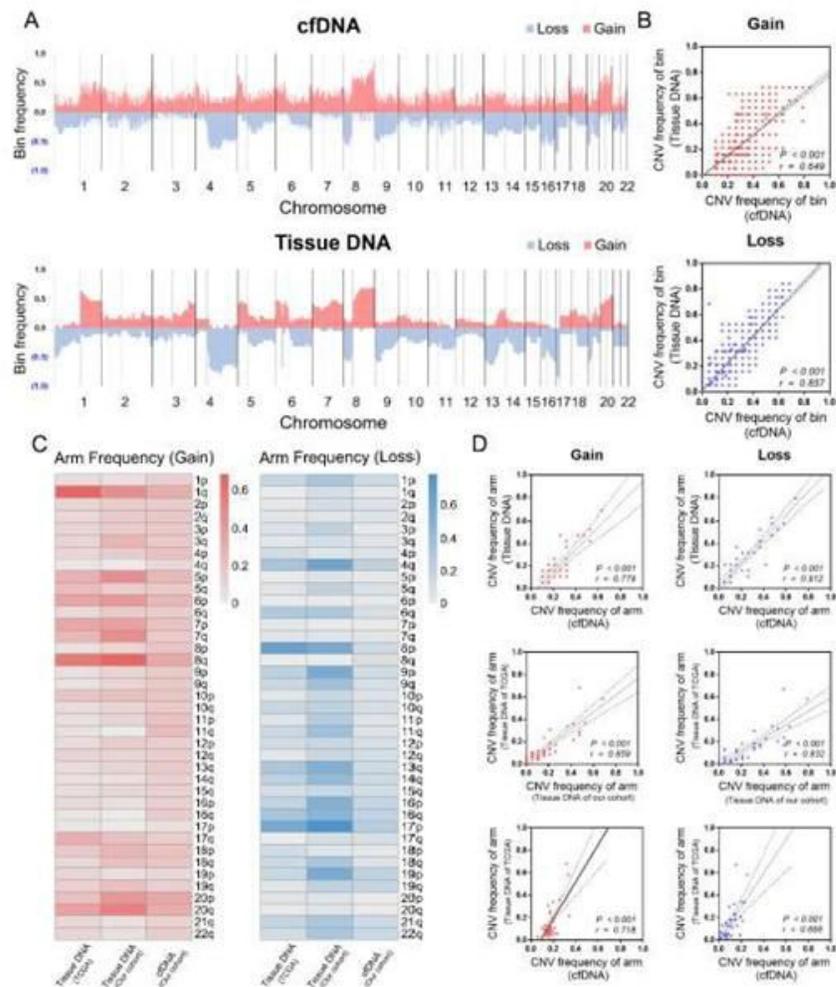


Figure 2

Consistency analysis of CNV frequencies between cfDNA and tumor tissue DNA. (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 that in tissue DNA ( $n=19$ , pair-matched.  $r$ , Pearson correlation). (C) The distribution of the arm-level CNVs of tissue DNA from TCGA ( $n=370$ ), tissue DNA from our cohort ( $n=19$ ) and cfDNA from our cohort ( $n=117$ ). The arm frequency of copy number gains (red) and losses (blue) altering among patients are exhibited. (D) The correlation of CNV frequency of each arm between cfDNA and tissue DNA from our cohort, tissue DNA from our cohort and TCGA, cfDNA and tissue DNA from TCGA ( $r$ , Pearson correlation).

Figure 3

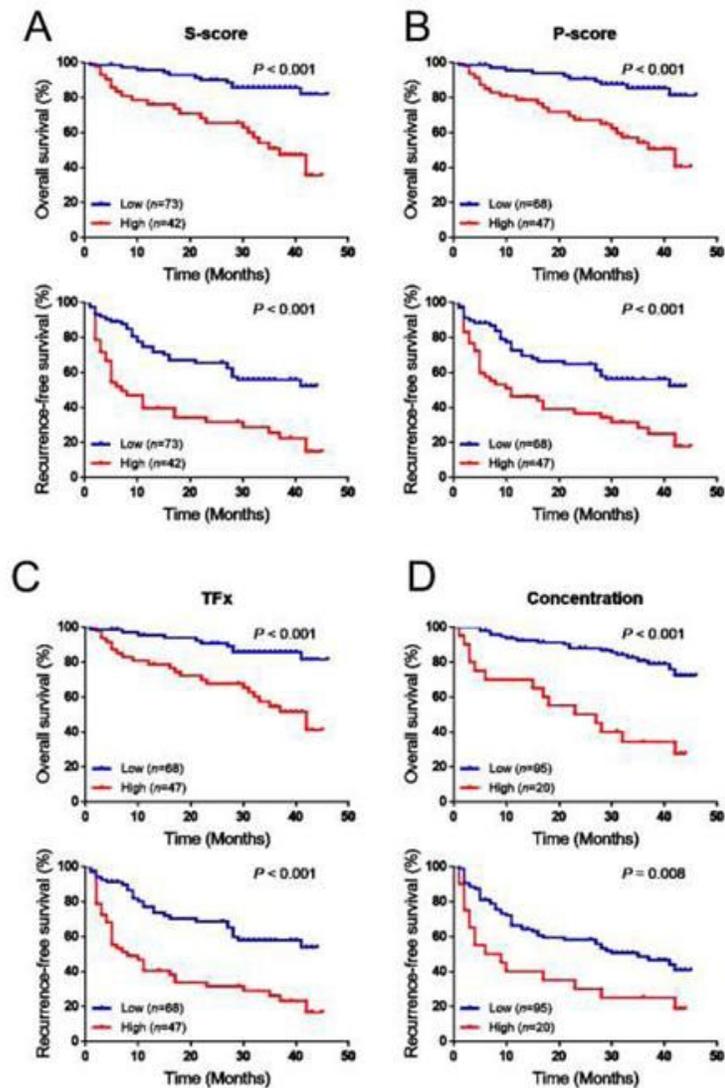


Figure 3

Prognostic values of genome-wide CNV indicators and concentration for HCC patients. Kaplan-Meier curves for low group and high group of HCC patients according to the genome-wide CNV indicators (A) S-score, (B) P-score, (C) TFx, (D) Concentration. Cases are stratified by cut-off value determined by maximizing Youden index. The Mantel-Cox log-rank significance value is shown for each. Cut off value: TFx: 0.02, P-score: 0.74, S-score: 0.04, Concentration: 48.49 ng/mL.

Figure 4

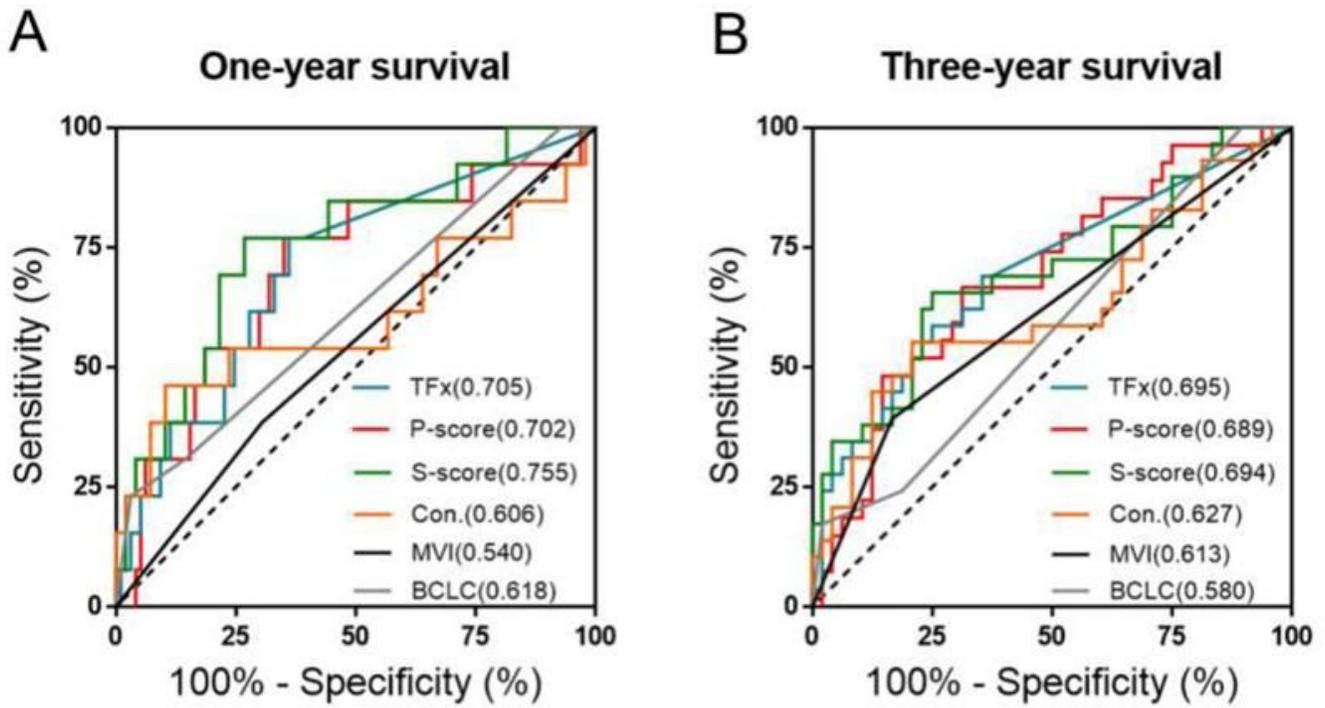


Figure 4

ROC curve for (A) One-year survival prediction and (B) Three-year survival prediction using TFX, P-score, S-score, Concentration (Con.), MVI and BCLC. Area under the curves (AUCs) is shown in corresponding brackets.

Figure 5

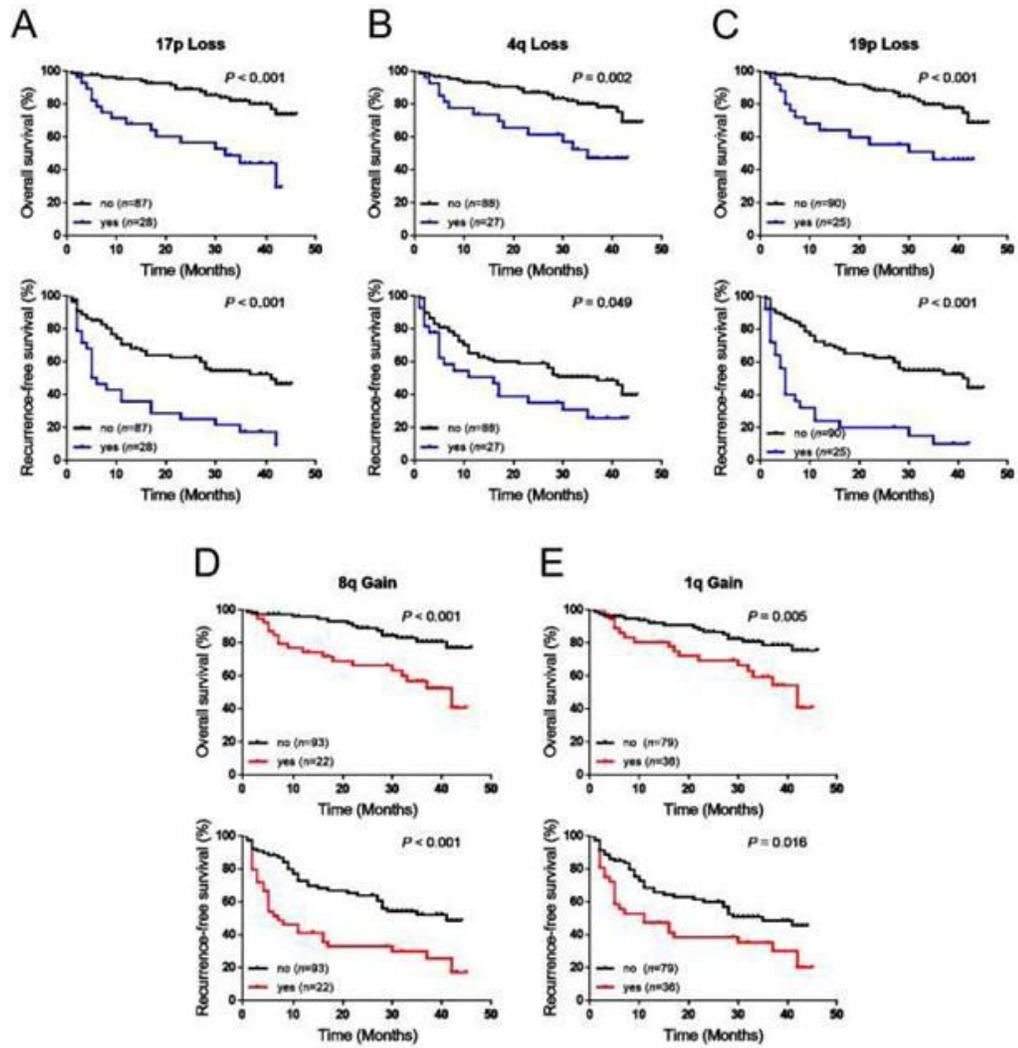


Figure 5

Prognostic values of chromosomal-arm CNVs. Kaplan–Meier curves for HCC patients stratified by harboring or do not harboring (A) 17p loss, (B) 4q loss, (C) 19p loss, (D) 8q gain, and (E) 1q gain. The Mantel-Cox log-rank significance value is shown for each.

Figure 6

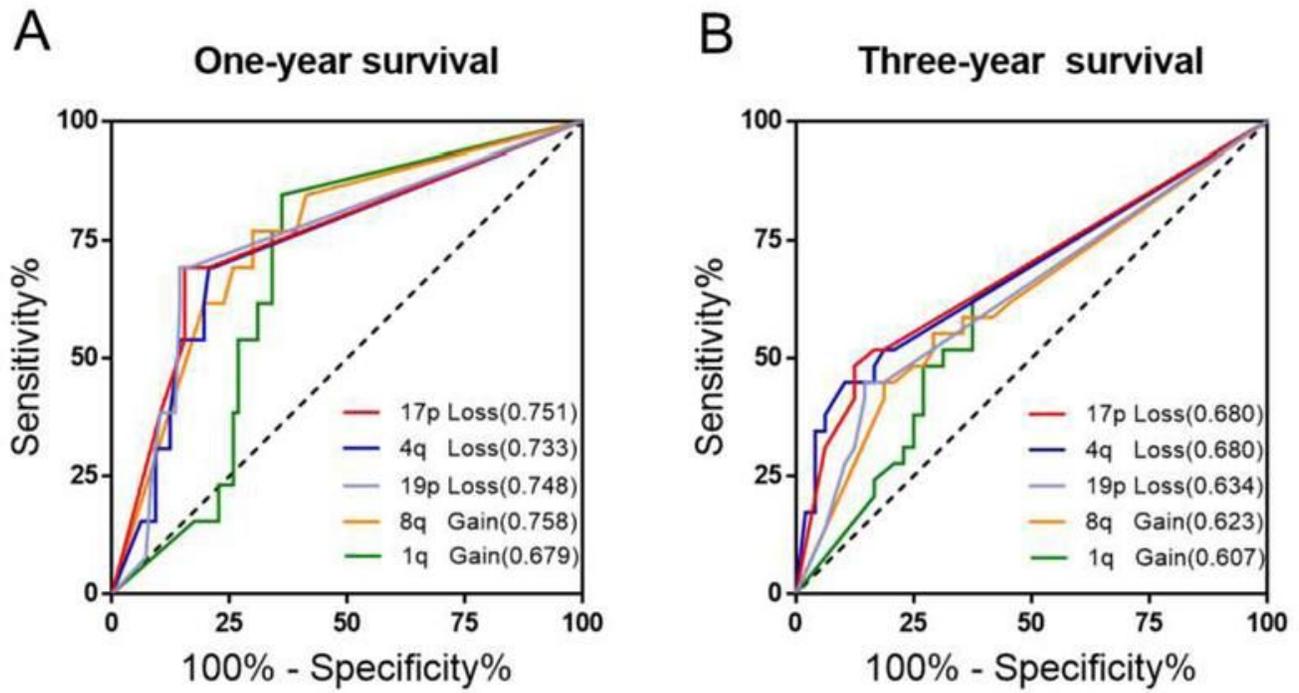


Figure 6

ROC curve for (A) One-year survival prediction, (B) Three-year survival predicted by CNVs at chromosomal-arm level. Area under the curves (AUCs) is shown in corresponding brackets.

Figure 7

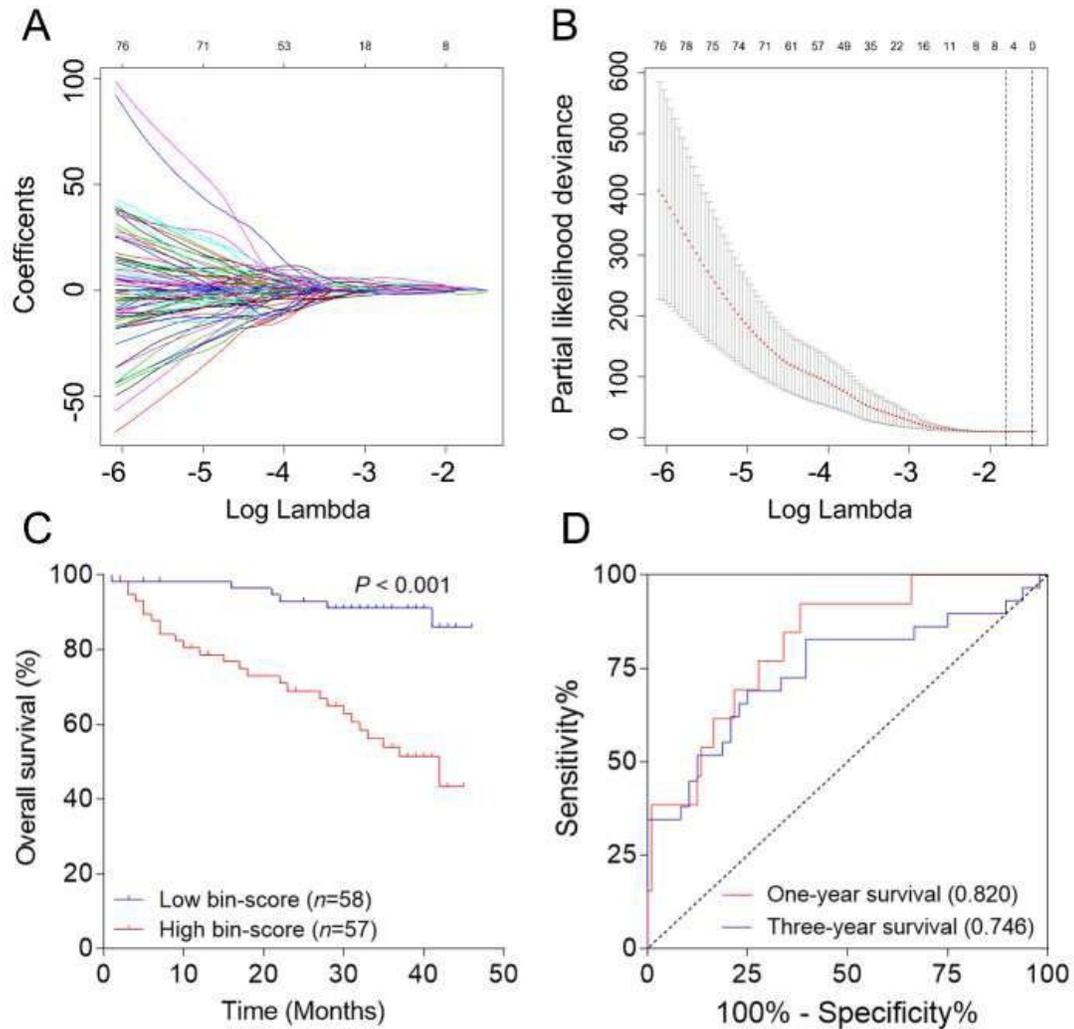


Figure 7

Bin selection via the LASSO regression model and prognostic value estimation of bin-score for HCC patients. (A) The LASSO coefficients profile of the 2475 bins for OS against the log lambda sequence. (B) Tenfold cross-validated error (first vertical line equals the minimum error, whereas the second vertical line shows the cross-validated error within 1 standard error of the minimum). (C) Kaplan–Meier curve for HCC patients stratified by low bin-score and high bin-score according to the median value. (D) ROC and corresponding AUCs for one-year survival and three-year survival predicted by bin-score.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [TableS7.docx](#)
- [TableS7.docx](#)
- [TableS6.docx](#)
- [TableS6.docx](#)

- TableS5.docx
- TableS5.docx
- TableS4.docx
- TableS4.docx
- TableS3.docx
- TableS3.docx
- TableS2.docx
- TableS2.docx
- TableS1.docx
- TableS1.docx
- FigureS1.png
- FigureS1.png
- FigureS2.png
- FigureS2.png
- FigureS3.png
- FigureS3.png
- FigureS4.png
- FigureS4.png