

Research Paper

# Aberrant AFP expression characterizes a subset of hepatocellular carcinoma with distinct gene expression patterns and inferior prognosis

Wei Chen\*, Jianjun Peng\*, Jinning Ye, Weigang Dai, Guanghua Li<sup>✉</sup>, Yulong He<sup>✉</sup>

Department of Gastrointestinal Surgery, First Affiliated Hospital of Sun Yat-sen University, 510080 Guangzhou, Guangdong province, People's Republic of China.

\* Wei Chen and Jianjun Peng contributed equally to this work.

✉ Corresponding author: Guanghua Li. Email: ligh26@mail.sysu.edu.cn; Yulong He. Email: ylonghe@126.com. Center of Gastrointestinal Surgery, First Affiliated Hospital of Sun Yat-sen University, 510080, No. 58, Zhongshan 2nd street, Guangzhou, Guangdong, China. Tel: 020-28823389; Fax: 020-28823389

© The author(s). This is an open access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>). See <http://ivyspring.com/terms> for full terms and conditions.

Received: 2018.11.12; Accepted: 2019.08.26; Published: 2020.01.01

## Abstract

**Background** Serum tumor markers are ubiquitously used in the clinic for cancer screening. However, the mechanisms accounting for the elevated levels of the serum tumor markers remain to be explored.

**Methods** We performed a pan-cancer analysis of serum alpha-fetoprotein (AFP), carcinoembryonic antigen (CEA) and prostate-specific antigen (PSA). The relation between concentration of serum tumor markers and the expression of their coding genes was assessed. Then the expression of AFP and its genomic background in hepatocellular carcinoma (liver cancer) was studied.

**Results** High expression of AFP mRNA was found mainly in liver cancer. In gastric cancer, breast cancer and lung cancer, high AFP mRNA expression was also discovered occasionally. In liver cancer patients, serum AFP levels correlated significantly with AFP mRNA expression in cancer tissues ( $r = 0.72$ ,  $p = 1.6e-45$ ). Whole transcriptome analysis revealed that serum AFP levels clearly separated liver cancer into two classes with distinct expression profiles according to PCA analysis. Gene co-expression analysis revealed that AFP expression was connected to a module enriched with genes accounting for cell cycle and cell proliferation regulation. In addition, high AFP expression was associated with the molecular classification of liver cancer, including iCluster (Chi-square: 16.86,  $P = 0.0002$ ). Methylation analysis revealed de-methylation of AFP promoter occurred in some liver cancer tissues, which was significantly related to AFP mRNA expression. Survival analysis indicated high serum AFP levels was prognostic of poorer survival of the liver cancer patients (Log-rank test:  $p = 0.046$ ). This was confirmed by an independent dataset in which liver cancer patients with high serum AFP also had poorer survival (Log-rank test:  $p = 0.024$ ).

**Conclusion** High expression of AFP defined a subtype of liver cancer with distinct gene expression profiles and clinical features. De-methylation of cytosine from CpG di-nucleotides in AFP promoter may be the cause of AFP re-expression in adult human liver cancer tissue.

Key words: AFP expression, liver cancer, De-methylation

## Introduction

Serum tumor markers are essential noninvasive tools for the screening and diagnosis of various cancer types. Many serum tumor markers have been studied

and approved for use in the clinic. For example, alpha-fetoprotein (AFP) is widely used for the diagnosis of hepatocellular carcinoma (liver cancer).

In addition, positive AFP could also be encountered in other cancer types<sup>[1-4]</sup>. The carcinoembryonic antigen (CEA) is used to assist in the diagnosis of cancers that arise from gastrointestinal tract<sup>[5]</sup>. The prostate-specific antigen (PSA), which is a prostate differentiation marker, is exclusively used for screening of prostate cancer. Other markers, including CA15-3, CA125, CA19.9, et.al. are also widely used for the screening of various cancer types<sup>[6]</sup>.

AFP is the main component of mammalian fetal serum. It is synthesized by visceral endoderm of the yolk sac and by fetal liver. After birth AFP level decreases dramatically in blood. However, AFP synthesis can be reactivated in liver tumors and germinogeneous teratoblastomas. As a result, serum AFP is an important marker for liver tumors and is widely used in clinical practice<sup>[7]</sup>. Although aberrant AFP expression is mainly found in liver cancer, it could also be found occasionally in gastric cancer<sup>[4,8]</sup>, breast cancer<sup>[2]</sup>, lung cancer<sup>[1,3]</sup>, pancreatic cancer<sup>[9]</sup>. This raises an interesting question as why AFP is activated in such tumors. The mechanisms leading to aberrant expression of AFP in liver cancer and other types of tumors are not clear yet.

Carcinoembryonic antigen (CEA) is a glycoprotein of about 200,000 Daltons in molecular weight. It is expressed in significant amounts during embryonic life, especially by the large intestine, and postnatally by carcinomas arising from this site. Many tumors of epithelial origin at other sites may also express CEA and are associated with elevated CEA levels in blood circulation<sup>[5]</sup>. CEA level is often monitored in the management of colorectal cancer and its measurement before surgery is recommended by American Society of Clinical Oncology(ASCO). CEA level can be measured prior to surgery to predict prognosis. Besides, CEA can also be employed to assess response to treatment or after completion of therapy to monitor for recurrence<sup>[6]</sup>.

PSA is most frequently detected in prostate cancer. Positive correlation has been confirmed between the PSA level and tumor stage and volume<sup>[10]</sup>. Besides, PSA is an indicator for prostate cancer survival and recurrence<sup>[11,12]</sup>. Additionally, a serial test of PSA is valuable for monitoring the treatment efficacy<sup>[13]</sup>.

Although serum tumor markers like AFP, CEA and PSA have been extensively studied and used in the clinic, questions regarding the mechanisms leading to their expression in cancer have not been well answered. This is exemplified by AFP, for which the molecular mechanisms leading to re-expression of AFP in adult human liver cancer and other tumor types is still unclear<sup>[7]</sup>. Knowing the source of these serum tumor markers may be helpful for us to

understand the molecular mechanisms accounting for their aberrant expression, and to uncover the advantages as well as the limitations related to each of them. In addition, it may also be helpful in finding new serum markers for cancer diagnosis.

In this study, we tried to address the questions mentioned above using genomic data from large cohort of cancer patients. We compared the expression of the coding genes of AFP, CEA and PSA, and found that only AFP expression correlated with its serum levels in liver cancer patients. Further analysis revealed aberrant AFP re-expression actually represented a shift of cell transcriptome that involved many genes, especially those related to cell proliferation regulation. We then identified de-methylation of a CpG locus located at the promoter of AFP that may be accountable for the AFP re-activation.

## Method

### Patients and samples

Totally 4,666 tumor samples were included in the pan-cancer gene expression analysis of tumor markers. There were 421 liver cancer samples, 326 colon cancer samples and 549 prostate cancer samples. The details of patient and sample information could be found in the publications associated with each study<sup>[14-16]</sup>. The abbreviations for different datasets used are as follows: BRCA (Breast invasive carcinoma), STAD (Stomach adenocarcinoma), PAAD (Pancreatic adenocarcinoma), COAD (Colon adenocarcinoma), READ (Rectum adenocarcinoma), LIHC (Liver hepatocellular carcinoma), LUAD (Lung adenocarcinoma), LUSC (Lung squamous cell carcinoma), OV (Ovarian serous cystadenocarcinoma), and PRAD (Prostate adenocarcinoma).

For hepatocellular carcinoma dataset from TCGA (The Cancer Genome Atlas) study, all patients received surgical resection of their tumors. No other treatment, for example ablation, chemotherapy, or radiotherapy had been given to the patients before surgery. Surgical specimens of primary hepatocellular carcinoma and matched blood samples were collected for each patient. Totally 371 patients with full genomic data were included in this study. For clinical data analysis, patient information from the TCGA hepatocellular carcinoma publication was used, which contained 196 samples<sup>[16]</sup>. In another dataset GSE14520, hepatocellular carcinoma patients who underwent radical resection were enrolled. Tumor samples and adjacent normal liver tissues were subjected to microarray gene expression analysis. Only samples tested by Affymetrix HT Human

Genome U133A Array were included in our analysis. Totally, there were 220 normal liver and 225 liver cancer tissues in this dataset.

### Data availability

Genomics data from TCGA project were downloaded using RTCGA package in R. The most recent version of RTCGA dataset was used (2016-01-28). For mRNA sequencing analysis, the RNASeq v2 data was used. For methylation analysis of TCGA data, the beta values of each CpG locus derived from Human Methylation 450 array platform was used. For survival analysis, a recently updated follow-up data processed by Liu J et.al. was used<sup>[17]</sup>. An independent data set profiling the whole methylome and transcriptome of three fetal liver and three adult liver tissues were downloaded from Gene Expression Omnibus (GEO) with accession IDs GSE69852 and GSE69713. Another dataset profiling a large cohort of liver cancer tissues and non-cancer tissues was also downloaded from GEO (GSE14520).

### Gene co-expression network analyses

Gene co-expression network analyses was performed using Co-Expression Modules identification Tool (CEMiTool), an R package that can identify and analyze co-expression modules in a fully automated manner. CEMiTool is featured by unsupervised gene filtering, automated parameter selection for identifying modules, enrichment and module functional analyses, as well as integration with interactome data to generate molecular interactions<sup>[18]</sup>.

### Gene methylation analysis

The whole genome methylation analysis was performed using Human Methylation 450 array platform. There are 4 probes targeting AFP gene, including one probe mapped onto the promoter region (cg10778295), two probes mapped onto around -5,000 from the transcription start site (cg02199826, cg03874137), and another probe mapped onto the interior of the gene (cg02387745). However, the probe inside AFP has no readings. Pearson correlation analysis was performed to assess relationship between methylation status of each probe with AFP expression.

### Statistical analysis

Statistical analysis was performed using R software. Human genome assembly hg19 was used as the reference genome. Liver cancer patients were separated into three groups based on serum AFP levels: high ( $\geq 300$  ng/ml), middle ( $< 300$  ng/ml,  $\geq 6$  ng/ml), low ( $< 6$  ng/ml). Survival plot was estimated using Kaplan-Meier method and groups

were compared with log-rank test. For correlation analysis between the levels of markers in serum and tumor, serum marker levels were added by 1 and then log<sub>2</sub> transformed. Pearson correlation analysis was then performed. For comparison of means of gene expression or methylation between two groups, t test was used.

## Results

### Expression of AFP, CEA and KLK3 across ten tumor types

To unveil the roots of serum tumor markers, we selected AFP, CEA (encoded by CEACAM5)<sup>[19]</sup> and PSA (encoded by KLK3)<sup>[20]</sup>, which are widely used in hepatocellular carcinoma (LIHC), colorectal cancer (COAD, READ) and prostate cancer (PRAD), respectively. We assessed the expression of these markers using RNA sequencing data across 10 major tumor types from TCGA<sup>[21]</sup>. AFP was highly expressed in a subset of liver cancer compared to normal liver. Apart from liver cancer, high expression of AFP was also observed in breast cancer (BRCA), gastric cancer (STAD) and lung cancer (LUAD, LUSC) sometimes (Figure 1A). The CEA coding gene CEACAM5 was highly expressed by normal colorectal epithelial and intermediately expressed by normal gastric and lung tissue. Colorectal cancer expressed similar levels of CEACAM5 as normal colorectal tissue. However, some gastric cancer, pancreatic cancer (PAAD), lung cancer and breast cancer also showed elevated CEACAM5 expression (Figure 1B). The PSA coding gene KLK3 was expressed at similar high levels in normal and tumor prostate tissue, but was never seen in other normal or tumor tissues (Figure 1C). These data suggested higher expression of AFP, CEACAM5 and KLK3 in tumors was a prerequisite for their presence in the serum of patients with specific cancer types.

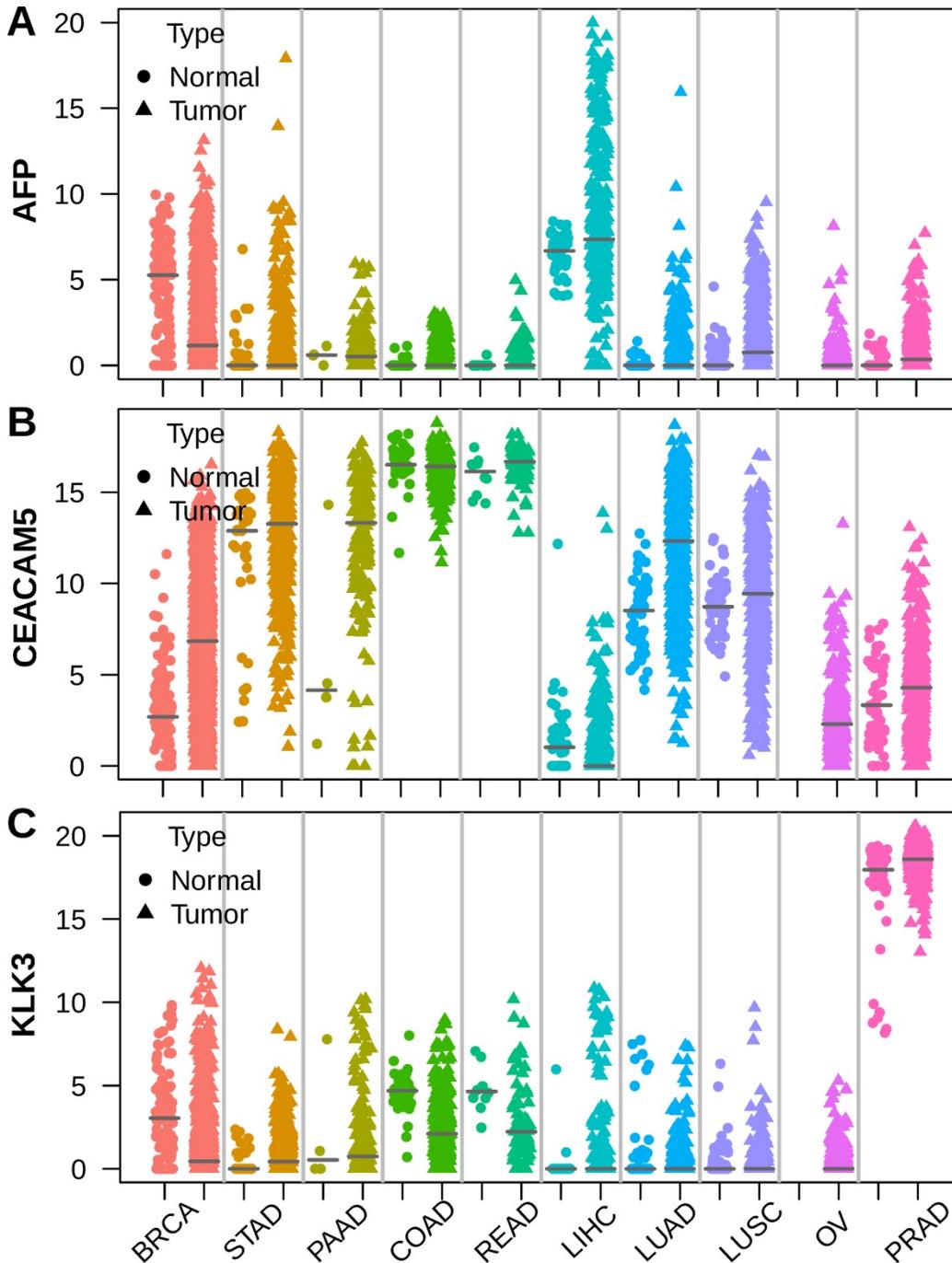
### Serum AFP levels correlated with AFP gene expression in liver tumor

Next, we asked whether the presence of tumor markers in the bloodstream was resulted from high expression of these markers by tumors. We compared the expression of CEA, AFP and PSA in tumor tissue with their serum levels. In liver cancer, a strong correlation between AFP gene expression and serum levels was observed ( $r = 0.72$ ,  $p = 1.6e-45$ ) (Figure 2A). In colorectal cancer, CEACAM5 expression showed no correlation with its serum levels ( $r = 0.09$ ,  $p = 0.25$ ) (Figure 2B). In prostate cancer, KLK3 expression showed significant correlation with PSA levels from the serum of prostate cancer patients ( $r = -0.21$ ,  $P = 1.1e-05$ ) (Figure 2C). However, the coefficient of

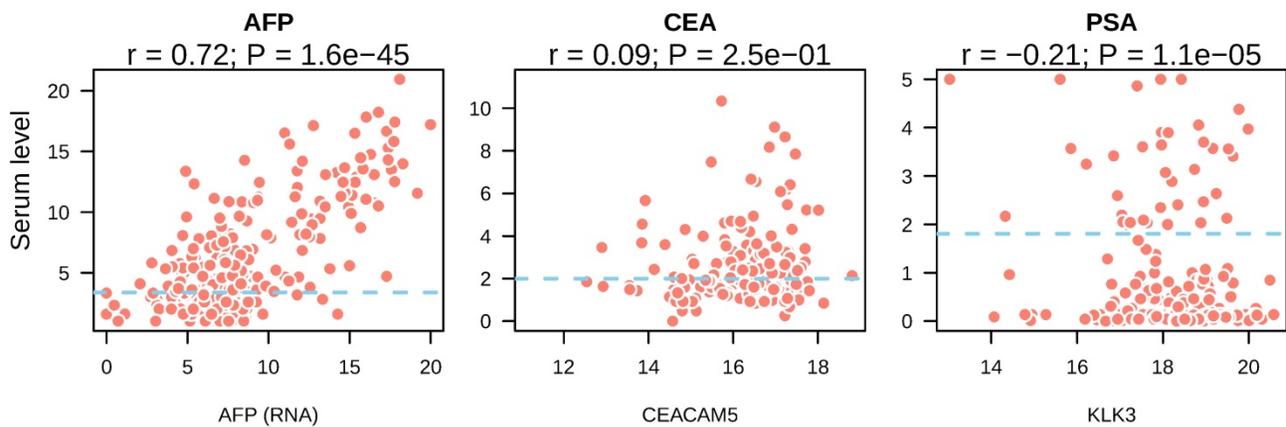
Pearson correlation was negative and p value was small, suggesting it may be a result of stochastic cause. These data suggested that high serum AFP in liver cancer patients was caused by re-activation and expression of AFP in liver tumors.

To validate these findings, we chose an independent dataset containing a large cohort of samples. Similarly, the expression of AFP were significantly higher in liver cancer than non-tumor

liver tissue (t-test:  $p = 3.78e-24$ ) (Supplement Figure 1). In addition, patients with higher serum AFP levels also had significantly higher AFP expression in the primary liver tumors (t-test:  $p = 7.92e-23$ ) (Supplement Figure 1). These data confirmed that aberrant high AFP expression was observed in some liver cancer, and it was closely related to the high serum AFP levels in these patients.



**Figure 1.** Stripchart showing the expression of AFP, CEA (CEACAM5) and PSA (KLK3) in tumor and adjacent normal tissues across ten selected tumor types. Dot represents normal tissue and triangle represents tumor tissue. Horizontal bar indicates median of gene expression. (A) AFP was expressed in both normal liver and liver cancer, with some liver cancer showing much higher AFP expression. (B) CEACAM5 was expressed at very high levels in both normal and tumor colorectal tissue. (C) KLK3 was expressed at very high levels in both normal prostate tissue and prostate cancer.



**Figure 2.** Correlation between serum tumor marker levels and the expression of their coding genes in tumor tissues. The mRNA expression of tumor marker coding genes in each tumor type was depicted in the x axis and serum tumor marker levels were shown in y axis. Blue dotted lines indicated the reference cutoff values of each tumor marker used in the clinic for cancer diagnosis. Both serum marker levels and mRNA expression levels were  $\log_2$  transformed.

### PCA analysis of liver cancer with gene expression data

Principal component analysis (PCA) was performed to assess the inner structure of liver cancer expression data. The distribution of liver cancer samples was plotted using the top three principal components and patients are colored by their serum AFP levels. A clear separation between high and low serum AFP patients could be seen (Figure 3A). This data suggested that AFP high expression may represent a subtype of liver cancer as defined by distinct transcriptome profiles. In total, the top three principal components in PCA accounted for 31.2% of total variance (16.5%, 8.6% and 6.1% for each) (Figure 3B). AFP contributed to 8.3%, 11.8% and 21.7% of top three components, respectively (Figure 3C). In addition, AFP was positively correlated with PC1, but negatively correlated with PC2 and PC3 (Figure 3D). In another dataset GSE14520, liver cancer patients with high and low serum AFP were also clearly separable as indicated by PCA analysis (Supplement Figure 2). These data revealed that AFP high expression liver cancer had distinct gene expression profiles compared with AFP low expression liver cancer.

### Gene co-expression analysis in liver cancer

To identify key genomic modules co-regulated with AFP, we performed gene co-expression analysis using mRNA sequencing data of liver cancer. Five gene modules were revealed for liver cancer and AFP belonged to M4 module. M4 and M5 module were significantly related to AFP high expressing liver cancer (Figure 4A). We then integrated co-expression information with protein-protein interaction data to identify main regulators and hubs in the module M4. Ten hub genes based on protein-protein interaction network were revealed and five more were

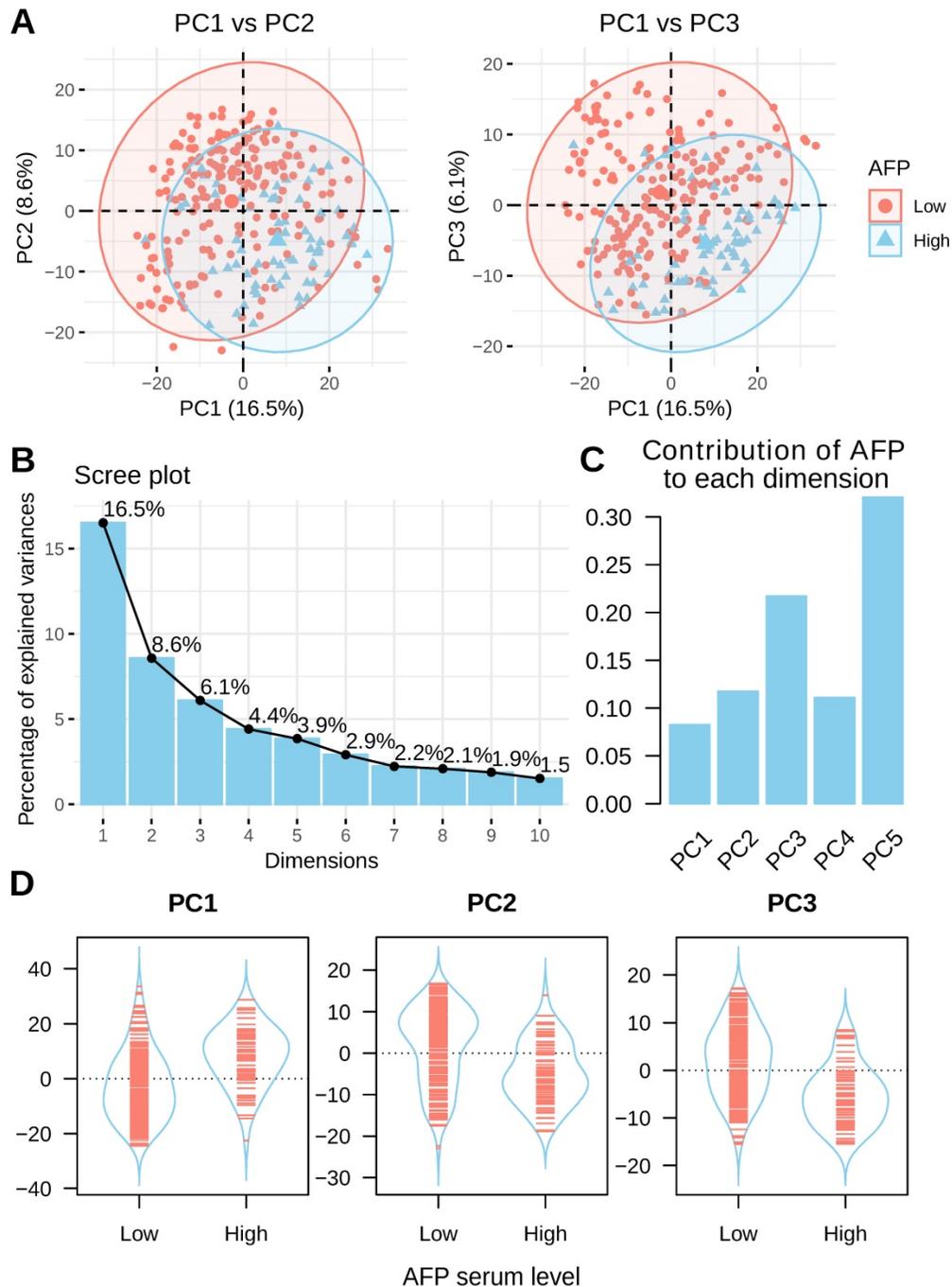
discovered by gene co-expression analysis (Figure 4B). Notably, many of those hub genes were related to cell proliferation. For example, CDC25C directs dephosphorylation of cyclin B-bound CDC2 and triggers entry into mitosis<sup>[22]</sup>. CCNB2 (Cyclin B2) is a member of the B-type cyclins, which are essential components of the cell cycle regulatory machinery<sup>[23]</sup>. PTP4A1 is a cell signaling molecule that plays regulatory roles in a variety of cellular processes, including cell proliferation and migration<sup>[24]</sup>. BIRC5 is a member of the inhibitor of apoptosis (IAP) gene family, which encode negative regulatory proteins that prevent apoptotic cell death<sup>[25]</sup>. Consistently, gene-set over-representation analysis revealed that cell proliferation related gene-sets were significantly enriched in M4 module (Figure 4C). These data suggested high expression of AFP may be related to aberrant cell proliferation control of liver cancer.

### AFP re-expression was associated with de-methylation of AFP promoter

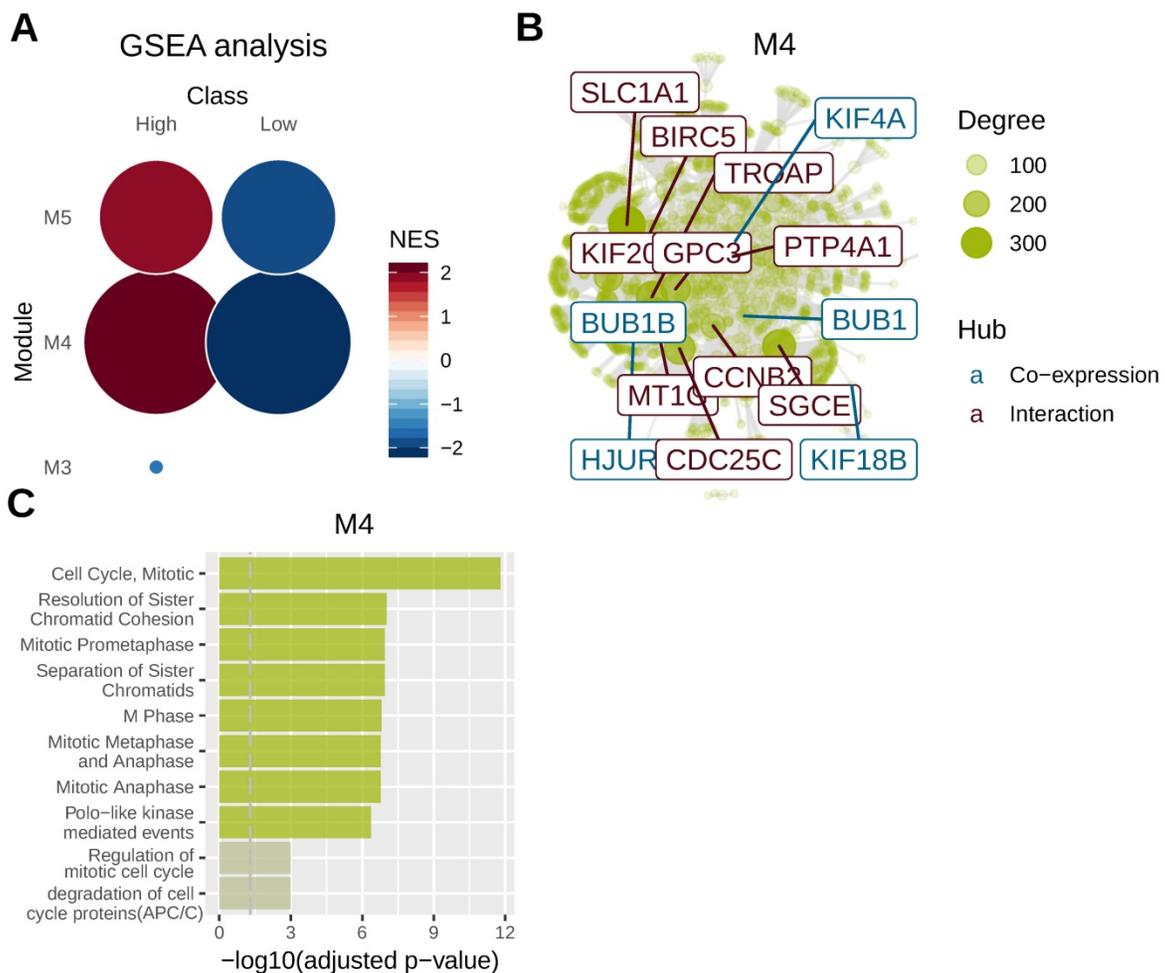
Next, we tried to find out what led to AFP re-expression in cancer. As hypo-methylation of promoter is a common mechanism for gene activation in cancer, we checked methylation of AFP in normal liver and liver cancer. In Infinium Human Methylation 450K Beadchip platform, which is widely used for whole genome methylation study, there are three probes mapped onto CpGs upstream of AFP transcription start site. All these three CpGs were invariably highly methylated in normal liver tissue (average beta values were cg10778295:  $0.92 \pm 0.02$ ; cg02199826:  $0.89 \pm 0.03$ ; cg03874137:  $0.82 \pm 0.05$ ) (Figure 5A). In addition, methylation of these CpGs had no correlation with AFP mRNA expression (Pearson correlation: all  $P > 0.1$ ). In liver cancer, some tissues showed de-methylation in these three CpGs (average beta values were cg10778295:  $0.83 \pm 0.13$ ; cg02199826:

0.80±0.12; cg03874137: 0.75±0.15) (Figure 5A). All three CpGs showed significant de-methylation in liver cancer (t test:  $P < 1e-9$ ). More importantly, methylation of cg10778295, a CpG located in the promoter of AFP (-824 from transcription start site), correlated very well with AFP mRNA expression (Pearson correlation:  $r = -0.55$ ,  $P = 1.2e-22$ ) (Figure 5A). We then validated this finding using fetal liver tissue, which normally expression AFP. In an independent data set, the whole methylome and

transcriptome analysis were conducted in three fetal liver and three adult liver tissues (GSE69852 and GSE69713). Similar to the findings in liver cancer, fetal liver showed high methylation in cg10778295 and lower expression of AFP ( $t = 5.56$ ,  $P = 0.01$ ;  $t = -6.91$ ,  $P = 0.02$ ) (Figure 5B). These data demonstrated that de-methylation of AFP occurred in some liver cancer tissue, which may be accountable for re-activation of AFP transcription.



**Figure 3.** PCA analysis of liver cancer using whole transcriptome profiling data. (A) PCA analysis of liver cancer using top 1000 most variable genes across all liver cancer tissues. Patients with high serum AFP were labeled in blue and patients with low serum AFP were labeled in red. (B) The contribution of top 10 principals to total variance of all samples. (C) The contribution of AFP expression to top five components. (D) The coordinates of top three components for each sample as grouped by serum AFP levels of the patients.



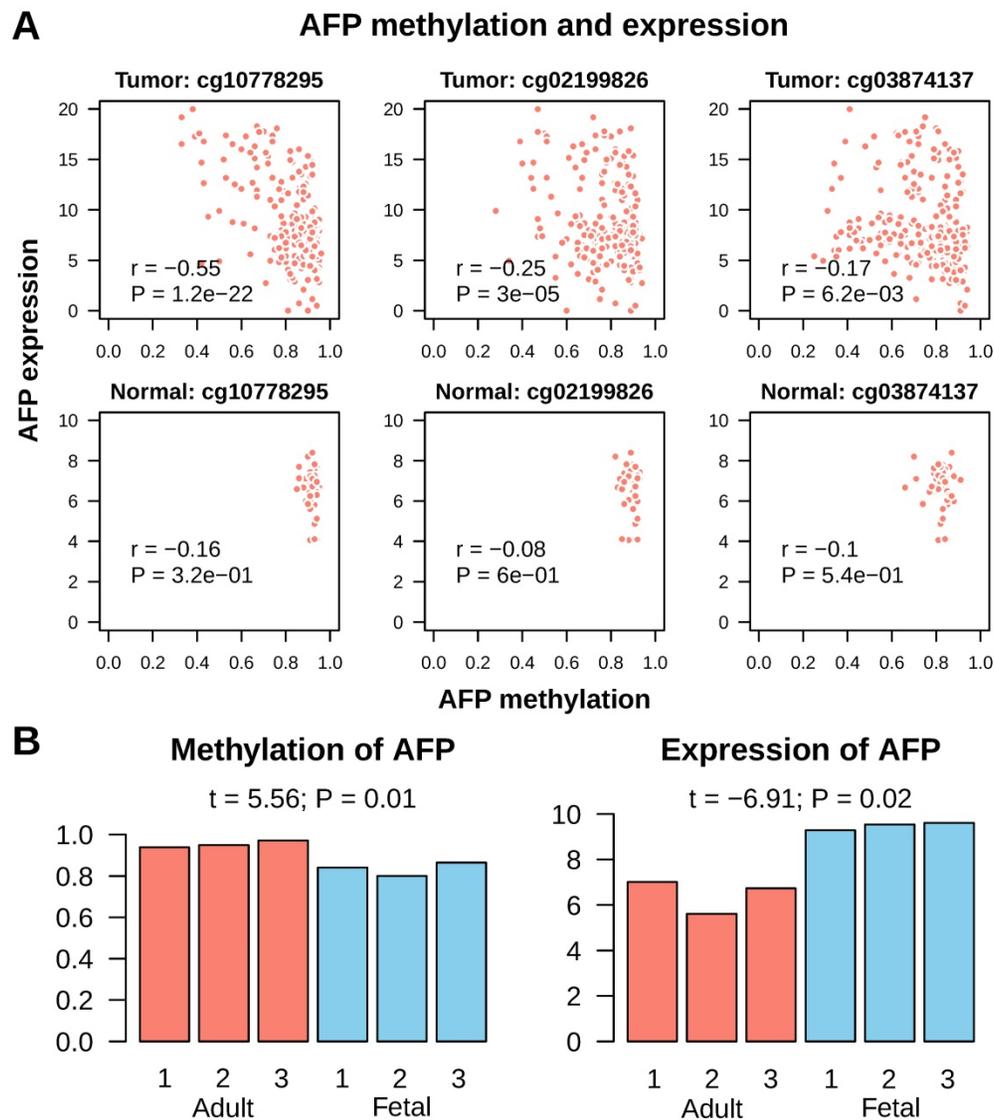
**Figure 4.** Gene co-expression analysis in liver cancer. Whole transcriptome data from TCGA liver cancer study was used for gene co-expression analysis. Tumor samples were grouped according to serum AFP levels. (A) Geneset enrichment analysis (GSEA) revealed M4 and M5 modules were enriched in AFP high expressing samples. (B) protein-protein interaction network analysis of genes in M4 module. Ten hub genes based on protein-protein interaction network (brown) and five hub genes based on co-expression analysis (blue) were revealed. The size of the node is proportional to its degree. (C) Geneset over-representation analysis of genes in M4 module.

### Clinicopathologic and genomic relevance of AFP activation in cancer

We then asked whether AFP expression had clinical significance. Survival analysis indicated poorer survival in liver cancer patient with middle and high serum AFP levels (Log-rank test:  $P = 0.046$ ) (Figure 6). Similarly, in an independent dataset GSE14520, liver cancer patients with high serum AFP levels also had significantly worse overall survival than patients with low serum AFP levels (Log-rank test:  $p = 0.024$ ) (Supplement Figure 3). To further assess the implications of high expression of AFP in liver cancer, we performed a thorough correlation study between serum AFP and clinicopathological features of the patients from TCGA study. Higher serum AFP level was associated with higher tumor grade (Chi-square: 8.07,  $P = 0.018$ ) (Table 1). It also tended to correlate with more advanced tumor stage (Chi-square: 5.76,  $P = 0.056$ ) (Table 1). Serum AFP

level was not associated with gene mutations frequently found in liver cancer, including TP53 mutation, CTNNB1 mutation and TERT promoter mutation (Table 1).

Recent high throughput genomic studies classified liver cancer into several molecularly distinct subtypes. We thus asked whether AFP activation is related to the molecular classification of liver cancer. Higher serum AFP level was significantly associated with iCluster (Chi-square: 16.86,  $P = 0.0002$ ) of liver cancer. The iCluster clustering of liver cancer was derived by integrating multiple whole genome platforms, which is supposed to be the most comprehensive and unbiased classification of cancer [16]. Specifically, 20 out of 31 liver cancers patients with high serum AFP levels belonged to iClust1, which ranked lowest in prognosis of all three subtypes. These data suggest AFP re-expression is part of systematic changes in liver cancer that define different subtype of tumors.



**Figure 5.** Methylation of AFP in normal liver and liver cancer. A, Correlation between AFP methylation and expression in normal adult liver and liver cancer tissue from TCGA study. Three CpGs located upstream of AFP transcription starting site were studied. Statistics in each figure are derived from Pearson correlation analysis. B, AFP methylation and expression in normal adult versus fetal liver. Data was downloaded from GEO datasets GSE69852 and GSE69713. Group means are compared using t test.

## Discussion

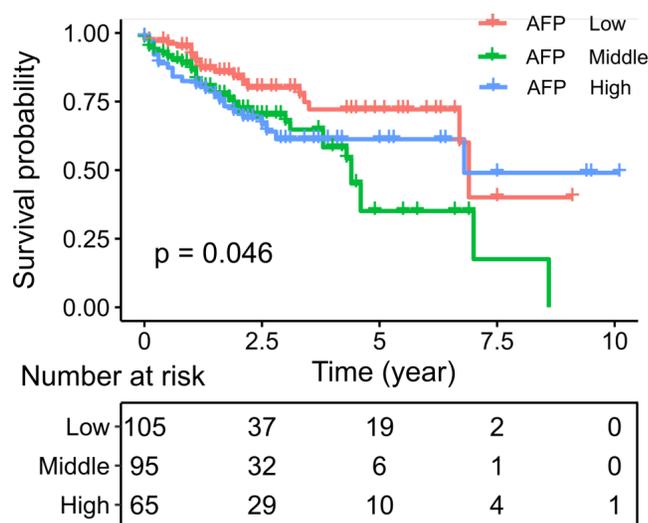
Aiming at identifying the molecular mechanisms underlying the elevated levels of the serum tumor markers, we performed a pan-cancer analysis of AFP, CEA and PSA to assess the relationship between concentration of serum tumor markers and the expression of their coding genes. Only AFP expression in tumor tissue correlated with its serum levels. In addition, high expression of AFP defined a subtype of liver cancer with distinct gene expression profiles and clinical features. Methylation analysis revealed de-methylation of AFP promoter occurred in some liver cancer tissues, and was significantly related to AFP mRNA expression.

Previous studies suggested that serum AFP levels are mainly controlled at transcriptional level in fetal liver. For example, in normal development of mouse liver, a parallel accumulation of both AFP and albumin mRNAs before birth, followed by a selective nonreciprocal decrease in AFP mRNA after birth, was observed<sup>[26]</sup>. An interesting finding in our study is that the spectrum of tumor types that produce AFP according to literature matches tumor types with high AFP mRNA expression in TCGA dataset perfectly. This is quite different from other popular serum tumor markers like CEA and PSA, which were expressed at equally high levels in tumor tissue as their normal counter parts. These data confirmed transcriptional control as a key mechanism regulating AFP levels in the serum of the cancer patients.

**Table 1:** Correlation between AFP expression and clinical factors

Variable	Total Cases*	Levels	AFP low Count1/Mean1	AFP high Count2/Mean2	t/chisq	P
<b>General information</b>						
Gender	138	female	34	16	3.28	0.07
		male	73	15		
Age	138		61.5±13.43	58.65±16.03	0.9	0.37
Child pugh	98	A	68	15	1	0.32
		B	10	5		
HBV	138	negative	86	25	0	1
		positive	21	6		
HCV	138	negative	82	28	2	0.16
		positive	25	3		
<b>Pathology</b>						
Leukocyte			0.19±0.13	0.2±0.19	-0.34	0.74
Grade	137	G1	16	1	8.07	0.018
		G2	59	13		
		G3	31	17		
Stage	131	I	52	8	5.76	0.056
		II	23	10		
		III/IV	26	12		
Purity	126		0.73±0.21	0.7±0.22	0.67	0.51
<b>Mutation</b>						
TP53 mutation	135	no	75	20	0.35	0.56
		yes	29	11		
CTNNB1 mutation	135	no	73	27	2.73	0.099
		yes	31	4		
TERT mutation	136	no	56	21	1.48	0.22
		yes	49	10		
<b>Molecular classification</b>						
Paradigm clusters	132	C1	30	3	19.76	0.00019
		C2	16	14		
		C3	21	11		
		C4	34	3		
iCluster	130	1	27	20	16.86	0.00022
		2	37	2		
		3	35	9		

\*Note. Only 138 out of 196 patients had information of serum AFP levels and were used in the correlation study. Some patients also lack other clinical information and the total counts of patients may be less than 138.



**Figure 6.** Kaplan-Meier plot showing the survival of liver cancer patients grouped by serum AFP levels. Survival rates were compared by log-rank test.

Liver cancer is a heterogeneous disease. Many molecular subtypes have been proposed for liver cancer [16,27-29], of which the iCluster defined by TCGA is a popular one. The iCluster was derived from a joint

analysis of five platforms of genomic data and three major subtypes were derived (iClust1, iClust2, iClust3). We found that AFP was significantly enriched in iClust1 subtype of liver cancer. iClust1 tumors are characterized by younger age, Asian ethnicity, female gender and high tumor grade. In addition, iClust1 patients showed poorest survival among the three subtypes, which is consistent to our analysis showing poor prognosis of AFP high patients[16].

Most previous studies investigated AFP alone. What we found in this study suggested aberrant activation of AFP was not an isolated event. It was actually part of the systematic diversion of the transcriptome in the liver cancer cells. More importantly, this diversion in transcription also has clinical significance. The most obvious features of AFP high tumors is its poorer prognosis compared with serum AFP normal tumors. Actually, the inferior prognosis in serum AFP positive tumors has been reported by many other studies already[30,31].

Although the transcriptional control of AFP expression is evident, it's unclear how liver cell cease

expressing AFP after birth, and how AFP expression is re-activated in liver cancer. *In vivo* experimental using mouse models revealed that a 7 kb regulatory region upstream of AFP gene was related to the regulation of AFP expression [32]. Within this region a tissue-specific promoter, three independent enhancers, and a silencer that is at least partially responsible for AFP gene expression decrease in adult liver have been defined[7]. Of note, the sequence from -200 to the transcription start site of mouse AFP is characterized by tissue-specific promoter activity and contains multiple overlapping binding sites for ubiquitous and tissue-specific transcription factors [33]. Our data revealed that de-methylation of a CpG locus in the AFP promoter occurred in some liver cancer tissues, and was significantly related to AFP mRNA expression. Thus, demethylation AFP promoter may be a reason for aberrant AFP expression in some liver cancers.

Evidence suggests that AFP may not only be a biomarker for cancer diagnosis, but also play functional roles in tumorigenesis. AFP can promote cancer cell proliferation through binding AFP receptor (AFPR), activating PI3K/AKT and many other cancer related genes [34-38]. Another study found that binding of AFP/AFPR led to Ca<sup>2+</sup> influx, prompting DNA synthesis and enhancing tumor cell proliferation [39]. Interestingly, our gene co-expression analysis assigned AFP into M4 gene module, which was significantly enriched with genes related to cell proliferation control.

In conclusion, this study indicates that AFP re-activation is a result of the systematic transcriptome change, which collectively define a molecular subtype of liver cancer. More importantly, high AFP expression is associated with over activation of cell growth or cell cycle control genes. In addition, high AFP expression is related to poor survival in the patients. These data highlights AFP as a biomarker for both liver cancer classification and prognosis.

## Supplementary Material

Supplementary figures and tables.

<http://www.jcancer.org/v11p0403s1.pdf>

## Acknowledgements

### Funding

This work was supported by the Natural Science Foundation of Guangdong Province, China (No. 2016A030310155, 2016A020213002) and National Natural Science Foundation of China (No. 81602049).

The funders had no role in the design of the study and collection, analysis, and interpretation of

data and in writing the manuscript.

## Authors' Contributions

Ye, jinning and Dai, Weigang extract data. Peng, Jianjun and Chen, Wei performed the analysis. Li, Guanghua and He, Yulong designed the project.

All authors have read and approved the final manuscript.

## Availability of data and materials

The datasets used and/or analyzed in the current study are available from the corresponding author on reasonable request.

## Competing Interests

The authors have declared that no competing interest exists.

## References

- [1] Yasunami R, Hashimoto Z, Ogura T, et al. Primary lung cancer producing alpha-fetoprotein: a case report[J]. *Cancer*, 1981, 47(5): 926-9.
- [2] Villacampa M J, Moro R, Naval J, et al. Alpha-fetoprotein receptors in a human breast cancer cell line[J]. *Biochem Biophys Res Commun*, 1984, 122(3): 1322-7.
- [3] Okunaka T, Kato H, Konaka C, et al. Primary lung cancer producing alpha-fetoprotein[J]. *Ann Thorac Surg*, 1992, 53(1): 151-2.
- [4] Chang Y C, Nagasue N, Abe S, et al. [The characters of AFP-producing early gastric cancer][J]. *Nihon Geka Gakkai Zasshi*, 1990, 91(10): 1574-80.
- [5] Goldenberg D M, Neville A M, Carter A C, et al. CEA (carcinoembryonic antigen): its role as a marker in the management of cancer[J]. *J Cancer Res Clin Oncol*, 1981, 101(3): 239-42.
- [6] Sturgeon C. Practice guidelines for tumor marker use in the clinic[J]. *Clin Chem*, 2002, 48(8): 1151-9.
- [7] Lazarevich N L. Molecular mechanisms of alpha-fetoprotein gene expression[J]. *Biochemistry (Mosc)*, 2000, 65(1): 117-33.
- [8] Yamanaka H, Nakane Y, Tanaka K, et al. [AFP-producing gastric cancer—analyses of 17 cases][J]. *Gan No Rinsho*, 1986, 32(15): 1934-40.
- [9] Kawamoto S, Hiraoka T, Kanemitsu K, et al. Alpha-fetoprotein-producing pancreatic cancer—a case report and review of 28 cases[J]. *Hepatogastroenterology*, 1992, 39(3): 282-6.
- [10] Carvalhal G F, Daudi S N, Kan D, et al. Correlation between serum prostate-specific antigen and cancer volume in prostate glands of different sizes[J]. *Urology*, 2010, 76(5): 1072-6.
- [11] Furuya Y, Nagakawa O, Fuse H. Prognostic significance of changes in short-term prostate volume and serum prostate-specific antigen after androgen withdrawal in men with metastatic prostate cancer[J]. *Urol Int*, 2003, 70(3): 195-9.
- [12] Small E J, Roach M, 3rd. Prostate-specific antigen in prostate cancer: a case study in the development of a tumor marker to monitor recurrence and assess response[J]. *Semin Oncol*, 2002, 29(3): 264-73.
- [13] Lilja H, Ulmert D, Vickers A J. Prostate-specific antigen and prostate cancer: prediction, detection and monitoring[J]. *Nat Rev Cancer*, 2008, 8(4): 268-78.
- [14] Cancer Genome Atlas N. Comprehensive molecular characterization of human colon and rectal cancer[J]. *Nature*, 2012, 487(7407): 330-7.
- [15] Cancer Genome Atlas Research N. The Molecular Taxonomy of Primary Prostate Cancer[J]. *Cell*, 2015, 163(4): 1011-25.
- [16] Cancer Genome Atlas Research Network. Electronic Address W B E, Cancer Genome Atlas Research N. Comprehensive and Integrative Genomic Characterization of Hepatocellular Carcinoma[J]. *Cell*, 2017, 169(7): 1327-1341 e23.
- [17] Liu J, Lichtenberg T, Hoadley K A, et al. An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics[J]. *Cell*, 2018, 173(2): 400-416 e11.
- [18] Russo P S T, Ferreira G R, Cardozo L E, et al. CEMiTool: a Bioconductor package for performing comprehensive modular co-expression analyses[J]. *BMC Bioinformatics*, 2018, 19(1): 56.
- [19] Pavlopoulou A, Scorilas A. A comprehensive phylogenetic and structural analysis of the carcinoembryonic antigen (CEA) gene family[J]. *Genome Biol Evol*, 2014, 6(6): 1314-26.
- [20] Stenman U H, Leinonen J, Zhang W M, et al. Prostate-specific antigen[J]. *Semin Cancer Biol*, 1999, 9(2): 83-93.
- [21] Hutter C, Zenklusen J C. The Cancer Genome Atlas: Creating Lasting Value beyond Its Data[J]. *Cell*, 2018, 173(2): 283-285.
- [22] Goh W C, Manel N, Emerman M. The human immunodeficiency virus Vpr protein binds Cdc25C: implications for G2 arrest[J]. *Virology*, 2004, 318(1): 337-49.

- [23] Gong D, Ferrell J E, Jr. The roles of cyclin A2, B1, and B2 in early and late mitotic events[J]. *Mol Biol Cell*, 2010, 21(18): 3149-61.
- [24] Liu L Z, He Y Z, Dong P P, et al. Protein tyrosine phosphatase PTP4A1 promotes proliferation and epithelial-mesenchymal transition in intrahepatic cholangiocarcinoma via the PI3K/AKT pathway[J]. *Oncotarget*, 2016, 7(46): 75210-75220.
- [25] Pavlyukov M S, Antipova N V, Balashova M V, et al. Survivin monomer plays an essential role in apoptosis regulation[J]. *J Biol Chem*, 2011, 286(26): 23296-307.
- [26] Tilghman S M, Belayew A. Transcriptional control of the murine albumin/alpha-fetoprotein locus during development[J]. *Proc Natl Acad Sci U S A*, 1982, 79(17): 5254-7.
- [27] Minguez B, Hoshida Y, Villanueva A, et al. Gene-expression signature of vascular invasion in hepatocellular carcinoma[J]. *J Hepatol*, 2011, 55(6): 1325-31.
- [28] Finn R S, Aleshin A, Dering J, et al. Molecular subtype and response to dasatinib, an Src/Abl small molecule kinase inhibitor, in hepatocellular carcinoma cell lines *in vitro*[J]. *Hepatology*, 2013, 57(5): 1838-46.
- [29] Deshmukh M, Hoshida Y. Genomic profiling of cell lines for personalized targeted therapy for hepatocellular carcinoma[J]. *Hepatology*, 2013, 58(6): 2207.
- [30] Peng S Y, Chen W J, Lai P L, et al. High alpha-fetoprotein level correlates with high stage, early recurrence and poor prognosis of hepatocellular carcinoma: significance of hepatitis virus infection, age, p53 and beta-catenin mutations[J]. *Int J Cancer*, 2004, 112(1): 44-50.
- [31] Huse S M, Gruppuso P A, Boekelheide K, et al. Patterns of gene expression and DNA methylation in human fetal and adult liver[J]. *BMC Genomics*, 2015, 16: 981.
- [32] Krumlauf R, Hammer R E, Tilghman S M, et al. Developmental regulation of alpha-fetoprotein genes in transgenic mice[J]. *Mol Cell Biol*, 1985, 5(7): 1639-48.
- [33] Hammer R E, Krumlauf R, Camper S A, et al. Diversity of alpha-fetoprotein gene expression in mice is generated by a combination of separate enhancer elements[J]. *Science*, 1987, 235(4784): 53-8.
- [34] Yamaguchi R, Yano H, Iemura A, et al. Expression of vascular endothelial growth factor in human hepatocellular carcinoma[J]. *Hepatology*, 1998, 28(1): 68-77.
- [35] Li M, Li H, Li C, et al. Alpha-fetoprotein: a new member of intracellular signal molecules in regulation of the PI3K/AKT signaling in human hepatoma cell lines[J]. *Int J Cancer*, 2011, 128(3): 524-32.
- [36] Zhang L, He T, Cui H, et al. Effects of AFP gene silencing on apoptosis and proliferation of a hepatocellular carcinoma cell line[J]. *Discov Med*, 2012, 14(75): 115-24.
- [37] Zhu M, Guo J, Xia H, et al. Alpha-fetoprotein activates AKT/mTOR signaling to promote CXCR4 expression and migration of hepatoma cells[J]. *Oncoscience*, 2015, 2(1): 59-70.
- [38] Lu Y, Zhu M, Li W, et al. Alpha fetoprotein plays a critical role in promoting metastasis of hepatocellular carcinoma cells[J]. *J Cell Mol Med*, 2016, 20(3): 549-58.
- [39] Li M S, Li P F, He S P, et al. The promoting molecular mechanism of alpha-fetoprotein on the growth of human hepatoma Bel7402 cell line[J]. *World J Gastroenterol*, 2002, 8(3): 469-75.