RESEARCH ARTICLE   **Open Access**

Identification by TCGA database search of five genes that are aberrantly expressed and involved in hepatocellular carcinoma potentially via DNA methylation changes

Junya Matsushita[1,2], Takehiro Suzuki[1], Kazuyuki Okamura[1], Gaku Ichihara[2] and Keiko Nohara[1*]

Abstract

Background: Various treatments for hepatocellular carcinoma (HCC) are utilized in clinical practice; however, the prognosis is still poor on account of high recurrence rates. DNA methylation levels of CpG islands around promoters (promoter CpGis) inversely regulate gene expression and closely involved in carcinogenesis. As a new strategy, several chemicals globally inhibiting DNA methylation have been developed aiming at reducing DNA methylation levels and maintaining the expression of tumor suppressor genes. On the other hand, since these drugs nonspecifically modify DNA methylation, they can cause serious adverse effects. In order to ameliorate the methods by targeting specific CpGs, information of cancer-related genes that are regulated by DNA methylation is required.

Methods: We searched candidate genes whose expressions were regulated by DNA methylation of promoter CpGi and which are involved in HCC cases. To do so, we first identified genes whose expression were changed in HCC by comparing gene expressions of 371 HCC tissues and 41 non-tumor tissues using the Cancer Genome Atlas (TCGA) database. The genes were further selected for poor prognosis by log-rank test of Kaplan-Meier plot and for cancer relevance by Pubmed search. Expression profiles of upregulated genes in HCC tissues were assessed by Gene Ontology (GO) analysis. Finally, using DNA methylation data of TCGA database, we selected genes whose promoter DNA methylation levels were inversely correlated with gene expression.

Results: We found 115 genes which were significantly up- or downregulated in HCC tissues and were associated with poor prognosis and cancer relevance. The upregulated genes were significantly enriched in cell division, cell cycle, and cell proliferation. Among the upregulated genes in HCC, we identified hypomethylation of CpGis around promoters of *FANCB*, *KIF15*, *KIF4A*, *ERCC6L*, and *UBE2C*. In addition, TCGA data showed that the tumor suppressor gene *P16* is unexpectedly overexpressed in many types of cancers.

Conclusions: We identified five candidate genes whose expressions were regulated by DNA methylation of promoter CpGi and associate with cancer cases and poor prognosis in HCC. Modification of site-specific DNA methylation of these genes enables a different approach for HCC treatment with higher selectivity and lower adverse effects.

Keywords: Hepatocellular carcinoma, DNA methylation, Gene expression, TCGA

---

* Correspondence: keikon@nies.go.jp
[1]Center for Health and Environmental Risk Research, National Institute for Environmental Studies, Tsukuba, Japan
Full list of author information is available at the end of the article

Background

Hepatocellular carcinoma (HCC) is the main type of liver cancer and has poor prognosis and low survival rates [1–3]. The GLOBOCAN database estimates that HCC is the sixth most commonly diagnosed cancer, and the fourth ranked contributor to the cause of cancer-related deaths [4]. Worldwide, approximately 841,000 new cancer cases and 782,000 deaths were reported to occur because of HCC in 2018 [4]. For inhibiting tumor-specific biological reactions, there are a variety of medical techniques, such as surgical resection, liver transplantation, and medication [2]. However, due to the high recurrence rates and metastasis, the prognosis of HCC patients is still poor, and the 5-year disease-free survival rate was 50.2% [4]. Hence, it is necessary to develop different approaches for the diagnosis and therapy of HCC.

Epigenetic modifications, such as DNA methylation, histone modifications, and non-coding RNAs (ncRNA) are pivotal factors of gene expression regulation without alteration of the DNA sequence [5]. Among them, DNA methylation is an extensively characterized epigenetic mechanism of gene regulation in mammals. DNA methylation levels of CpG islands (CpGis) around the transcription start sites (promoter CpGis) conversely regulate gene expression [6] and such alterations are closely involved in various cancers including HCC [7–9]. The tumor suppressor gene *P16* (cyclin-dependent kinase inhibitor 2A) is a well-known target that suffers DNA methylation silencing in human cancers [10, 11]. Suppression of DNA methylation changes at such specific targets are expected to provide new approaches to cancer treatment [8, 12].

Several drugs targeting aberrant DNA methylation, such as 5-aza-2′-deoxycytidine (5-aza-dC), have already been utilized for the therapy of refractory or relapsed cancer patients [12, 13]. 5-aza-dC exerts anticancer activity by effectively reducing DNA methylation levels and restoring the expression of *P16* [14, 15]. However, since this type of drug globally reduces DNA methylation, severe side effects were observed in clinical use [16]. To reduce this risk, it is desirable to restore the site-specific DNA methylation levels involved in cancer progression. For the purpose, recent techniques, such as the Crispr-Cas9 system, enable manipulation of DNA methylation/demethylation at specific sites [17–19]. On the other hand, fewer DNA methylation targets are identified in HCC compared to colon or gastric cancers [20].

The Cancer Genome Atlas (TCGA) database (https://tcga-data.nci.nih.gov/tcga/) provides valuable information about not only gene expression but also DNA methylation levels in various cancers from patients in multi-stages. Using the TCGA database, the present study aimed to identify DNA methylation changes which regulate the cancer-related gene expressions in HCC tissues to propose candidate targets for treatment.

For the purpose, we searched for cancer-related genes whose expression levels are significantly altered in HCC and which are associated with poor survival rates using data of 371 HCC tissues and 41 non-tumor tissues compiled in the TCGA database. They were further selected by the presence of promoter CpGis and significant changes of CpGi methylation levels in HCC tissues. Among them, we found 5 genes whose expressions are inversely correlated with DNA methylation and related to poor prognosis. Overall, we identified cancer-related genes whose expressions are associated with the DNA methylation of promoter CpGis in HCC tissues. The method described in this study is applicable to other types of cancers to identify candidates of genes that are regulated by DNA methylation.

Materials and methods

Gene expression and DNA methylation data for HCC and non-tumor tissues

Gene expression and DNA methylation data of HCC patients were downloaded from TCGA (https://cancergenome.nih.gov/) using UCSC Xena (http://xena.ucsc.edu/, version 10-04-2017). We selected total 371 HCC samples and 41 non-tumor samples which are opened for both DNA methylation and gene expression data. The clinical data of HCC patients are shown in Table 1 and sample ID lists are shown in Table S1. Gene Expression Profiling Interactive Analysis (GEPIA, http://gepia.cancer-pku.cn/) was utilized to obtain the expression data of *P16* in various cancers.

Survival analysis

The relationships between the expression of 929 genes and overall survival time of HCC patients were found by a log-rank test of Kaplan-Meier analysis using OncoLnc (http://www.oncolnc.org/). In the analysis of OncoLnc, 360 HCC data were available. For the analysis of each gene, HCC patients (n = 360) were divided into the top 20% expression group (n = 72) and bottom 20% expression group (n = 72). Kaplan-Meier analysis for *P16* was also carried out using data of top and bottom 25% expression groups (n = 90). For the analysis of 5-year survival rate, survival rate data for the top and bottom 20% expression group were retrieved from TCGA using OncoLnc and analyzed using EZR (Saitama Medical Center, Jichi Medical University, Saitama, Japan, version 1.42) [21]. The hazard ratios (HRs) and 95% confidence intervals (CIs) of HCC were estimated with reference to the bottom 20% expression group using a Cox proportional hazard regression model that included several covariates (gender, age, history, histological grade, and AJCC pathological stage).

Table 1 Clinical features of patients

		Number of patients (rates)	
		Non-tumor tissue	HCC
Races	White	26 (63.4%)	184 (49.6%)
	American Indian or Alaska native	0 (0%)	2 (0.5%)
	Black or African American	7 (17.1%)	17 (4.6%)
	Asian	5 (12.2%)	158 (42.6%)
	Not reported	3 (7.3%)	10 (2.7%)
Age (year)		20–78	16–85
Gender	Male	18 (43.9%)	250 (67.4%)
	Female	23 (56.1%)	121 (32.6%)
History of HCC factor	None	14 (34.1%)	91 (24.5%)
	Alcohol consumption	6 (14.6%)	68 (18.3%)
	Hepatitis B	6 (14.6%)	75 (20.2%)
	Hepatitis C	5 (12.2%)	32 (8.6%)
	Non-alcoholic fatty liver disease	2 (4.9%)	11 (3.0%)
	Hemochromatosis	0 (0%)	5 (1.3%)
	Complex	1 (2.4%)	58 (15.6%)
	Other	4 (9.8%)	12 (3.2%)
	No data	3 (7.3%)	19 (5.1%)
Histologic grade	G1	5 (12.2%)	55 (14.8%)
	G2	20 (48.8%)	177 (47.7%)
	G3	14 (34.1%)	122 (32.9%)
	G4	2 (4.9%)	12 (3.2%)
	No data	0 (0%)	5 (1.3%)
AJCC TNM staging system[a] (presence or absence of distant metastasis)	M0	27 (65.9%)	266 (71.7%)
	M1	1 (2.4%)	4 (1.1%)
	MX[b]	13 (31.7%)	101 (27.2%)
AJCC TNM staging system (lymph node involvement)	N0	25 (61.0%)	252 (67.9%)
	N1	1 (2.4%)	4 (1.1%)
	NX[c]	14 (34.1%)	114 (30.7%)
	No data	1 (2.4%)	1 (0.3%)
AJCC TNM staging system (tumor size and extent of tumors)	T1	19 (46.3%)	181 (48.8%)
	T2	10 (24.4%)	94 (25.3%)
	T3	9 (22.0%)	80 (21.6%)
	T4	3 (7.3%)	13 (3.5%)
	Other[d]	0 (0%)	3 (0.8%)
AJCC pathological stage	Stage I	17 (41.5%)	171 (46.1%)
	Stage II	7 (17.1%)	86 (23.2%)
	Stage III	8 (19.5%)	85 (22.9%)
	Stage IV	1 (2.4%)	5 (1.3%)
	No data[e]	8 (19.5%)	22 (5.9%)
	Other[e]	0 (0%)	2 (0.5%)

[a]AJCC TNM staging system: tumor-node-metastasis staging system which was creased and is updated by the American Joint Committee on Cancer (AJCC) (https://www.cancer.gov/about-cancer/diagnosis-staging/staging)
[b]MX: metastasis cannot be measured
[c]NX: cancer in nearby lymph nodes cannot be measured
[d]Other includes TX (main tumor cannot be measured), [discrepancy], and no data
[e]Other includes [discrepancy]
No data, data are not available

Pubmed search

The cancer relevancy of the 124 genes shown in Fig. 1 was found by using the PubMed database (version 10-14-7) with the search term "cancer." The search details were "gene" [All Fields] AND ("neoplasms"[MeSH Terms] OR "neoplasms"[All Fields] OR "cancer"[All Fields]).

Identification of promoter CpGis and calculation of methylation value (β value)

The positions of CpGis and TSSs (transcription start sites) in the reference genome hg38 were obtained from the University of California, Santa Cruz (UCSC) genome browser website (https://genome.ucsc.edu/). The promoter CpGis that are overlapping TSSs were detected by bedtools 2.25.0 (bedtools.readthedocs.io) in the Cygwin64 software. The average CpGi methylation β value (methylated array intensity/(methylated + unmethylated array intensity)) of the top 20% expression-changed HCC ($n = 74$) and that of control livers ($n = 41$) were compared for each gene.

Functional enrichment analysis using the Database for Annotation, Visualization and Integrated Discovery (DAVID)

Gene Ontology (GO) enrichment analysis including the biological process cellular component and molecular function was performed for upregulated genes shown in Table S2 by using DAVID v6.8 (https://david.ncifcrf.gov/) [22, 23].
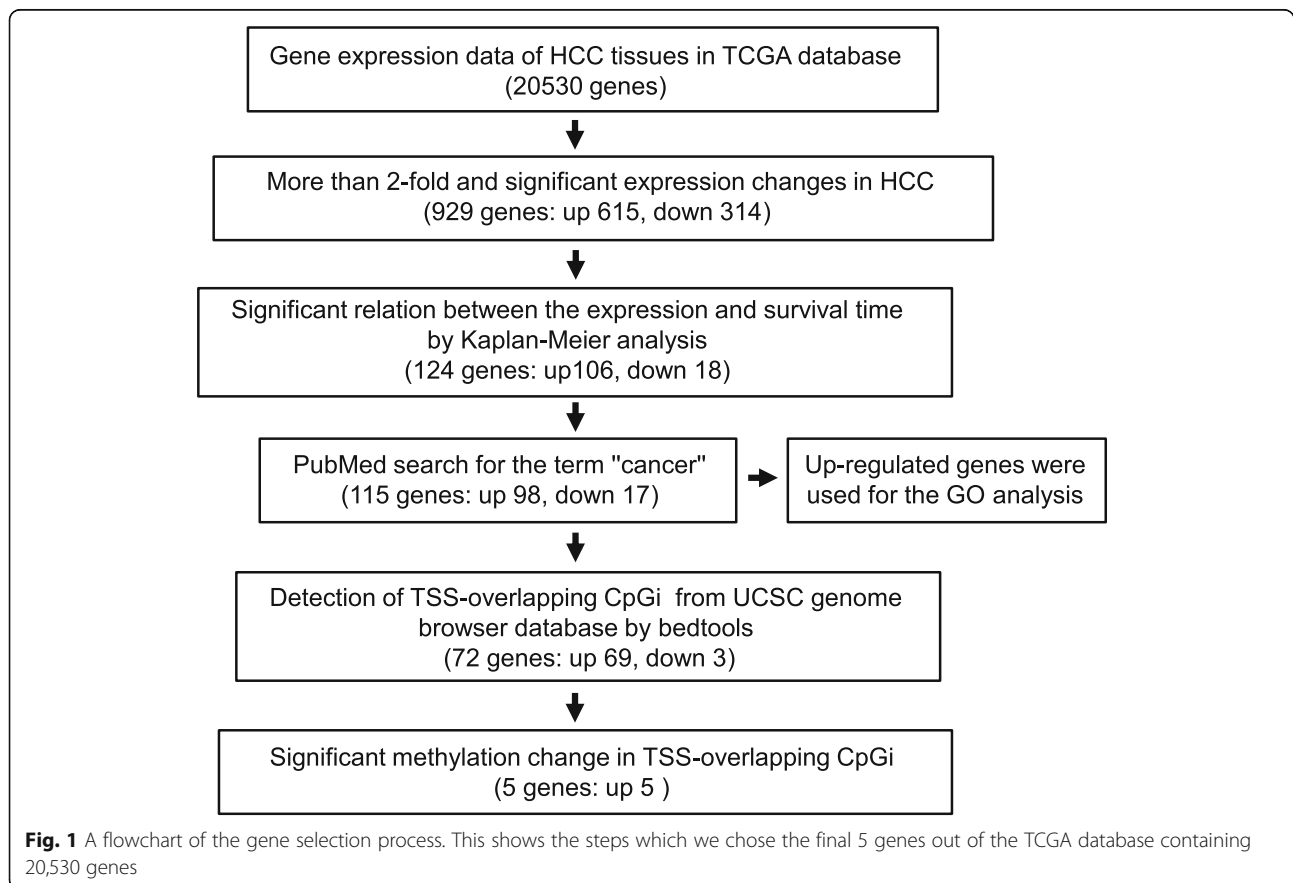
Statistical analysis

The differences in methylation β values of CpGs and gene expression levels between HCC tissues and non-tumor tissues were analyzed by a two-tailed paired Student's *t* test. The methylation β values of CpGis were analyzed by a Wilcoxon rank sum test using an online program (http://www.gen-info.osaka-u.ac.jp/MEPHAS/wilc1.html). The β values of *P16* in HCC tissues of pathological stages I, II, III, and IV and non-tumor tissues were analyzed by one-way ANOVA followed by Turkey-Kramer test as a post hoc comparison. A *p* value < 0.05 was considered statistically significant. In the GO analysis, *p* values were controlled for false discovery rate (FDR) using the Benjamini-Hochberg test. *p* values of < 0.05 were considered statistically significant.

Results

Selection of cancer-related genes whose expressions were up- or downregulated and associated with poor prognosis in human hepatocellular carcinoma

To identify cancer-related genes whose expressions were inversely correlated with DNA methylation in HCC

Fig. 1 A flowchart of the gene selection process. This shows the steps which we chose the final 5 genes out of the TCGA database containing 20,530 genes

tissues, we analyzed data of 371 HCC tissues and 41 non-tumor tissues compiled in the TCGA database using the algorithm depicted in Fig. 1. The clinical features of 371 patients are summarized in Table 1. HCC is reported to be more common in males, and hepatitis B (HVB) infection is the most common etiological factor of HCC in the world. Hepatitis C virus (HVC) infection, alcohol abuse, and non-alcoholic fatty liver disease are other major risk factors [3, 24]. The clinical features shown in Table 1 were similar to the previous findings, although "none" was the most abundant feature in the history in the TCGA database.

After eliminating low expression genes by the criteria of expression values < 1.0, we selected 929 genes which are significantly up- or downregulated more than 2-fold in HCC tissues compared with non-tumor tissues.

For the 929 genes, we performed Kaplan-Meier survival analysis between the top 20% expression group and bottom 20% expression group, and found that 124 genes (106 and 18 genes up- or downregulated in HCC, respectively) have a significant relevance between overall survival time and the expression. From the 124 genes, we shortlisted 115 genes (98 upregulated and 17 downregulated in HCC) which appeared in a Pubmed search for the term "cancer" and "neoplasm" (Table S2).

Upregulated genes in HCC were associated with cell cycle, cell division and cell proliferation

To assess the roles of 98 genes that were upregulated more than 2-fold in HCC tissues (listed in Table S2), we performed GO analysis using the DAVID database. Nine out of 43 terms were significantly enriched in the field of biological processes as shown in Fig. 2a. The majority of the terms were related to cell division, cell cycle, and proliferation. In the field of cellular components (Fig. 2b), genes were particularly enriched in the terms related to the nucleus. In the molecular function field, the major parts of genes were categorized in the terms related to protein binding (Fig. 2c). We further found that 19 out of 98 genes were not only involved in cell division, cell cycle, or cell proliferation, but also located in the nucleus and had molecular functions of protein binding. We omitted GO analysis for downregulated genes due to the small number of them (17 genes listed in Table S2).

Analysis of methylation status of promoter CpGis

Since DNA methylation levels in promoter CpGis are closely correlated with gene expression, we examined whether 115 genes shown in Table S2 have promoter CpGis and whether they are hypo- or hyper-methylated in HCC. We found that 72 out of 115 genes have CpGis overlapping TSSs (Fig. 1).

We compared the top 20% mean β values of patients (n = 74) who showed higher or lower expression in HCC tissues and mean β values of non-tumor tissues (n = 41). Five genes, *FANCB* (FA complementation group B), *KIF15* (kinesin family member 15), *KIF4A* (kinesin family member 4A), *ERCC6L* (ERCC excision repair 6 like), and *UBE2C* (ubiquitin conjugating enzyme E2 C), were significantly hypomethylated in HCC tissues compared with non-tumor tissues (p value < 0.05 or 0.01, Fig. 3a). Furthermore, multiple CpG sites in the CpGis of these genes were significantly hypomethylated in HCC tissues (Fig. 3b–f).

## *FANCB*, *KIF15*, *KIF4A*, *ERCC6L*, and *UBE2C* were significantly upregulated and their expressions were associated with low survival rates in HCC

The expression levels of *FANCB* (a), *KIF15* (b), *KIF4A* (c), *ERCC6L* (d), and *UBE2C* (e) in the individual 41 non-tumor tissues and 371 HCC tissues are plotted in Fig. 4. The expression of non-tumor tissues and HCC tissues were significantly different with p < 0.001 for all genes.

As the expression of these five genes were shown to be closely associated with poor prognosis in early time by Kaplan-Meier analysis as shown in the selection process of 124 genes, we focused on the 5-year survival rate of them. The results of Kaplan-Meier survival analysis for these five genes are shown in Fig. 5a–e (left panels). The prognosis of the higher expression groups of all the five genes were significantly poor compared with their low expression groups, and this poor prognosis persisted even after adjustment for several confounders (Fig. 5a–e, right panels). AJCC pathological stage was the significant covariate with the expression of *KIF15*, *ERCC6L*, and *UBE2C*. *FANCB* expression and gender were associated with poor prognosis.

The CpGs of *P16* were hyper-methylated but upregulated in HCC tissues

P16 is well-recognized as a tumor suppressor gene which is silenced by hyper-methylation of CpGi in human cancers and restores expression by DNA methyltransferase inhibitors [10, 14, 25, 26]. We investigated the DNA methylation and gene expression of *P16* using the TCGA database. Consistent with previous studies, 3 CpG sites in CpGi were significantly hyper-methylated in HCC tissues (Fig. 6a). The methylation level of CpG site at + 187 bp from TSS significantly increased from the early stage of HCC (Fig. 6b). On the other hand, the expression of *P16* is upregulated in HCC tissues compared with non-tumor tissues (Fig. 6c) and the upregulation occurred from the early stage of HCC (Fig. 6d).
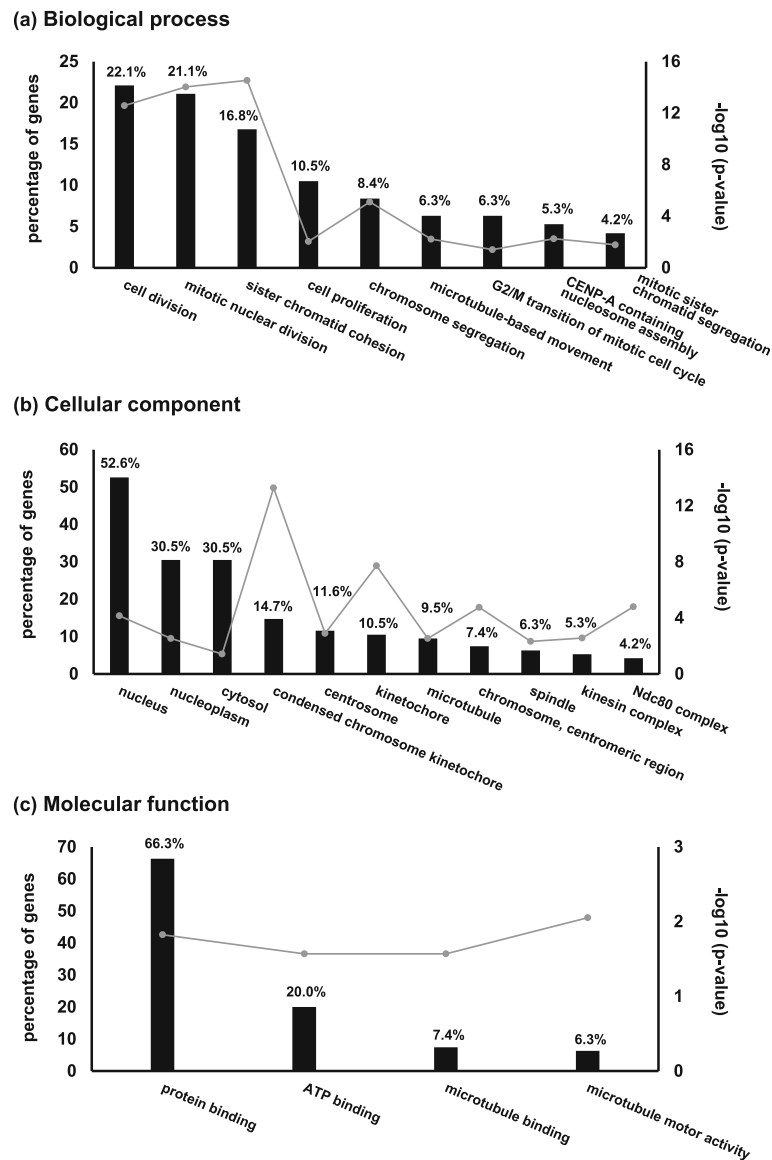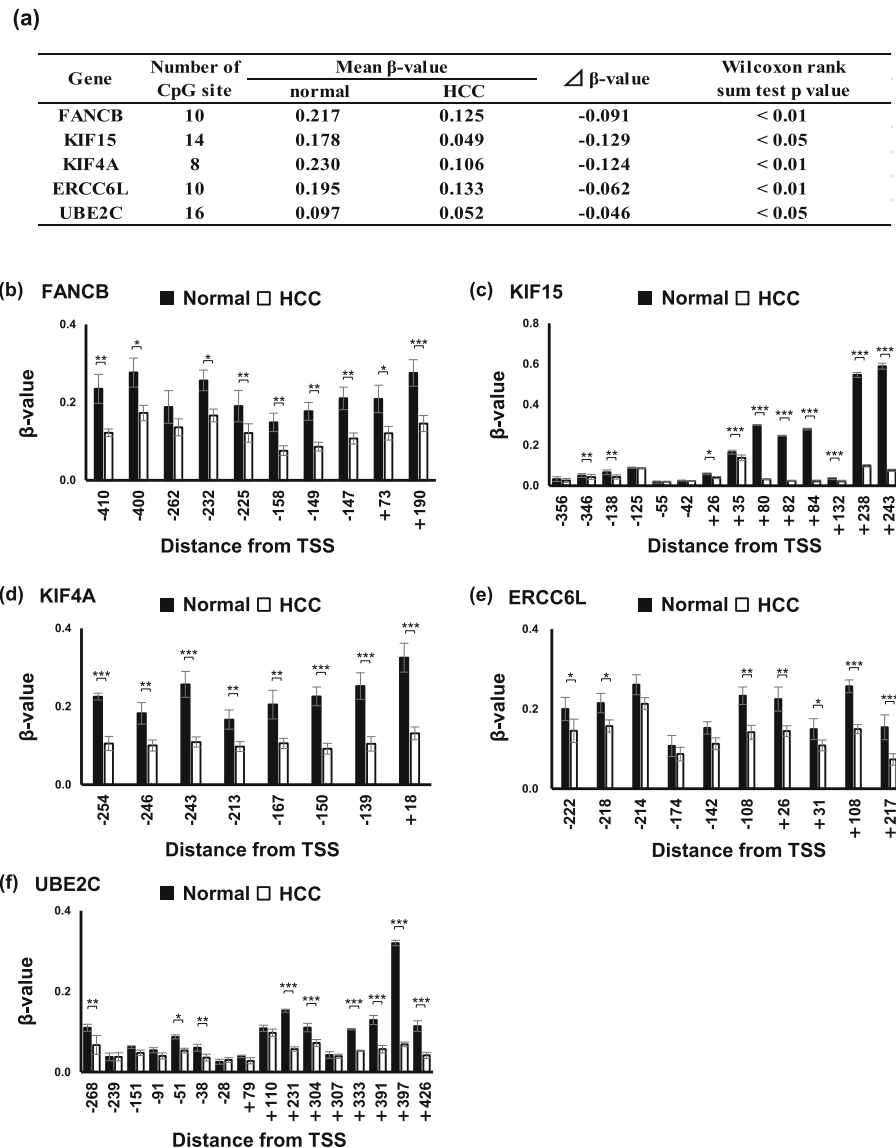
Fig. 2 Gene Ontology enrichment analysis of the upregulate genes shown in Table S2. Bar graphs represent the percentages of genes involved in each term of biological process (BP) (**a**), cellular component (CC) (**b**), and molecular function (MF) (**c**). Line graphs represent the Benjamini corrected *p* values (significant difference, *p* < 0.05). Significantly enriched GO terms were shown in these graphs

Kaplan-Meier analysis revealed that the top 25% expression group of *P16* showed significant poor prognosis compared to the bottom 25% expression group in HCC (Fig. 6e). The prognosis of the higher and lower expression group was not significantly different by Cox regression analysis (data not shown). The results showed that *P16* is upregulated and is relevant to poor prognosis; however, its expression is not inversely correlated with DNA methylation in the HCC tissues. Further analysis showed that *P16* is upregulated in various cancer types including breast, colon, lung, and kidney (Fig. 6 f.).

Discussion

In the present study, using the TCGA database, we searched for the candidates of cancer-related genes whose expressions are regulated by DNA methylation of CpGis and involved in poor prognosis in HCC tissues. We performed GO analysis for the 98 genes which are upregulated and associated with poor prognosis in HCC and are cancer relevant (Fig. 1). The result showed that they were significantly involved in the G2/M transition of the mitotic cell cycle, mitotic nuclear division, mitotic sister chromatid segregation, cell division, and cell

(a)

Gene	Number of CpG site	Mean β-value		⊿ β-value	Wilcoxon rank sum test p value
		normal	HCC		
FANCB	10	0.217	0.125	-0.091	< 0.01
KIF15	14	0.178	0.049	-0.129	< 0.05
KIF4A	8	0.230	0.106	-0.124	< 0.01
ERCC6L	10	0.195	0.133	-0.062	< 0.01
UBE2C	16	0.097	0.052	-0.046	< 0.05

Fig. 3 CpGis of *FANCB*, *KIF15*, *KIF4A*, *ERCC6L*, and *UBE2C* were hypomethylated in HCC tissues. **a** Mean methylation *β* values were calculated using data from 74 HCC tissues highly expressing each gene and data from 41 non-tumor tissues. Wilcoxon rank sum test was conducted using an online program and *p* value < 0.05 was considered statistically significant. The *β* values of CpG in CpGi of *FANCB* (**b**), *KIF15* (**c**), *KIF4A* (**d**), *ERCC6L* (**e**), and *UBE2C* (**f**) in 74 HCC tissues highly expressing each gene were compared with the *β* values of 41 non-tumor tissues. Statistical significance between the two groups was analyzed by Student' s *t* test. *. **. *** Significantly different at *p* < 0.05, *p* < 0.01, and *p* < 0.001, respectively

proliferation (Fig. 2). These terms were related to cell malignancy, and their abnormalities or incompletions may lead to carcinogenesis [27–29]. In addition, more than half of the genes are located in the nuclei (Fig. 2b) where early events of cell division occur and mainly have a molecular function of protein binding, which suggest that they belong to transcription regulators controlling cell division and proliferation.

Among the 98 genes, we identified five genes (*FANCB, KIF15, KIF4A, ERCC6L*, and *UBE2C*) whose expressions were inversely correlated with DNA methylation (Figs. 3,

4, and 5). Recently, Sun et al. reported on the correlation analysis between DNA methylation and gene expression in HCC also using the TCGA database [30]. As their analysis did not adopt inverse correlation between expression and DNA methylation, they assigned different genes from those we identified in the present study. FANCB, one of the Fanconi anemia proteins, is involved in the repair of DNA lesions and its upregulation is suggested to be required for the survival of colon cancer [31]. FANCB is also reported to be associated with other types of cancers [32, 33]. KIF family proteins, including
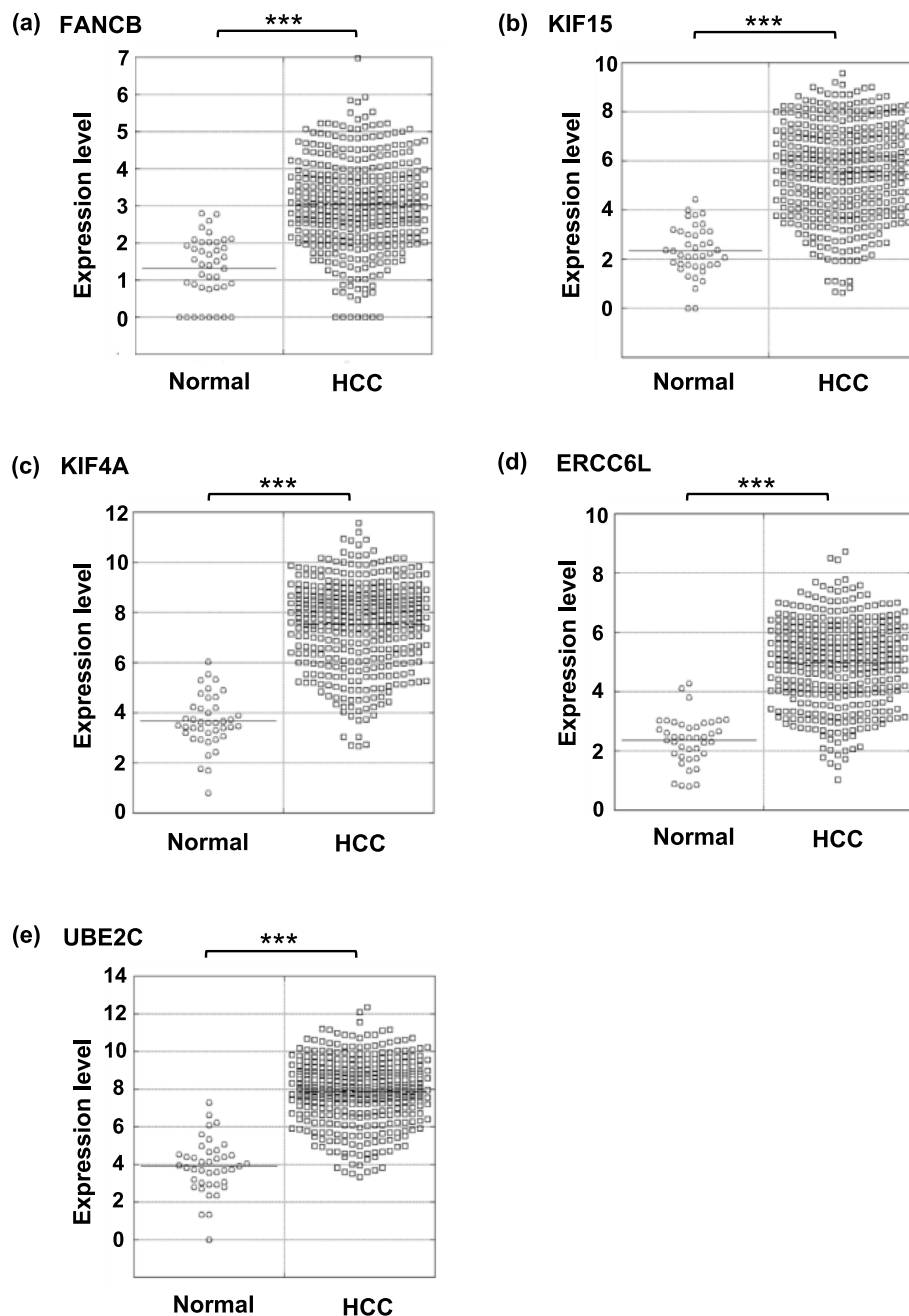
Fig. 4 The expression levels of *FANCB* (**a**), *KIF15* (**b**), *KIF4A* (**c**), *ERCC6L* (**d**), and *UBE2C* (**f**). Statistical significance between the two groups was analyzed by Student's *t* test. *** Significantly different at $p < 0.001$

KIF4A and KIF15 encode kinesin-related proteins which are molecular motor proteins that travel along microtubule tracks, play multiple roles in intracellular transport and cell division [34]. Kinesins are reported to have oncogenic functions such as progression and development of cancers [35]. Knockdown using siRNA and overexpression of *KIF4A* resulted in attenuation and

promotion of proliferation of HCC cell lines, respectively [36]. Knockdown of *KIF15* by shRNA suppressed proliferation of HCC cell lines in vitro and in mice [37]. Recent studies reported that an increased expression of *KIF4A* and *KIF15* are potential prognostic factors in prostate cancer [38] and lung adenocarcinoma [39], respectively. *ERCC6L* encodes a newly discovered DNA

(a) FANCB

| Characteristics		HR	CI	p-value
FANCB expression (High / Low)		2.693	1.258 - 5.766	0.011*
Gender (Female / Male)		0.650	0.277 - 1.529	0.324
Age		1.004	0.977 - 1.031	0.787
History (/ None)	Hepatitis B	0.239	0.057 - 1.010	0.052
	Hepatitis C	1.254	0.410 - 3.837	0.692
	Alcohol consumption	0.527	0.180 - 1.544	0.243
	Complex	0.261	0.068 - 0.994	0.049*
	other	1.301	0.373 - 4.533	0.680
Histologic grade (G4+G3 / G1+G2)		1.460	0.702 - 3.037	0.312
Pathological stage (Stage III+IV / I+II)		1.614	0.764 - 3.413	0.210

(b) KIF15

Characteristics		HR	CI	p-value
KIF15 expression (High / Low)		2.286	1.112 - 4.700	0.025*
Gender (Female / Male)		1.416	0.635 - 3.158	0.396
Age		0.983	0.959 - 1.009	0.196
History (/ None)	Hepatitis B	0.365	0.095 - 1.400	0.142
	Hepatitis C	1.324	0.402 - 4.360	0.644
	Alcohol consumption	1.466	0.553 - 3.884	0.442
	Complex	0.518	0.182 - 1.474	0.218
	other	1.143	0.397 - 3.297	0.804
Histologic grade (G4+G3 / G1+G2)		1.074	0.558 - 2.067	0.832
Pathological stage (Stage III+IV / I+II)		3.128	1.607 - 6.089	< 0.001***

(c) KIF4A

Characteristics		HR	CI	p-value
KIF4A expression (High / Low)		3.942	1.819 - 8.543	< 0.001***
Gender (Female / Male)		1.785	0.784 - 4.066	0.168
Age		0.996	0.972 - 1.020	0.754
History (/ None)	Hepatitis B	0.609	0.153 - 2.424	0.482
	Hepatitis C	1.905	0.583 - 6.224	0.286
	Alcohol consumption	0.938	0.351 - 2.509	0.899
	Complex	0.404	0.107 - 1.525	0.181
	other	0.958	0.288 - 3.191	0.945
Histologic grade (G4+G3 / G1+G2)		1.171	0.583 - 2.352	0.657
Pathological stage (Stage III+IV / I+II)		1.784	0.860 - 3.702	0.120

(d) ERCC6L

Characteristics		HR	CI	p-value
ERCC6L expression (High / Low)		2.781	1.251 - 6.182	0.012*
Gender (Female / Male)		1.504	0.625 - 3.620	0.362
Age		0.989	0.965 - 1.015	0.405
History (/ None)	Hepatitis B	1.000	0.252 - 3.965	0.100
	Hepatitis C	2.855	0.734 - 11.100	0.130
	Alcohol consumption	1.342	0.482 - 3.742	0.574
	Complex	0.345	0.065 - 1.815	0.209
	other	1.175	0.340 - 4.062	0.799
Histologic grade (G4+G3 / G1+G2)		1.457	0.680 - 3.122	0.333
Pathological stage (Stage III+IV / I+II)		4.398	1.970 - 9.816	< 0.001***

(e) UBE2C

Characteristics		HR	CI	p-value
UBE2C expression (High / Low)		4.569	2.113 - 9.880	< 0.001***
Gender (Female / Male)		1.064	0.469 - 2.416	0.882
Age		0.990	0.969 - 1.012	0.381
History (/ None)	Hepatitis B	0.438	0.137 - 1.398	0.163
	Hepatitis C	1.333	0.441 - 4.031	0.616
	Alcohol consumption	0.825	0.302 - 2.252	0.707
	Complex	0.457	0.132 - 1.580	0.216
	other	2.755	0.584 - 12.990	0.200
Histologic grade (G4+G3 / G1+G2)		0.989	0.515 - 1.899	0.973
Pathological stage (Stage III+IV / I+II)		2.236	1.171 - 4.270	0.015*

Fig. 5 Kaplan-Meier analysis and Cox regression analysis of overall survival of *FANCB* (**a**), *KIF15* (**b**), *KIF4A* (**c**), *ERCC6L* (**d**), and *UBE2C* (**e**). Left figures and right tables were the results of Kaplan-Meier analysis and Cox regression analysis for 5-year survival data, respectively. Data from 72 HCC patients with higher or lower expression for each gene were selected in this study. A *p* value < 0.05 was considered statistically significant by log-rank test. In the Cox regression analysis, statistical significance is marked with the star symbol: **p* < 0.05, ***p* < 0.01, ****p* < 0.001. HR hazard ratio, CI confidence interval. Complex: including more than two histories. Other: including non-alcoholic fatty liver disease, hemochromatosis, and "other"

helicase that is highly expressed in almost all cancers [40]. *ERCC6L* is known to be an oncogenic protein of solid tumors, since the high expression leads to cancer cell proliferation and tumor growth [40]. *ERCC6L* knockdown was demonstrated to result in

downregulation of PLK1 which serves an important role in the control of the proliferation and cell cycle in cancer cells [41]. *UBE2C* encodes a member of the E2 ubiquitin-conjugating enzyme family and is required for the destruction of mitotic cyclin and for cell cycle
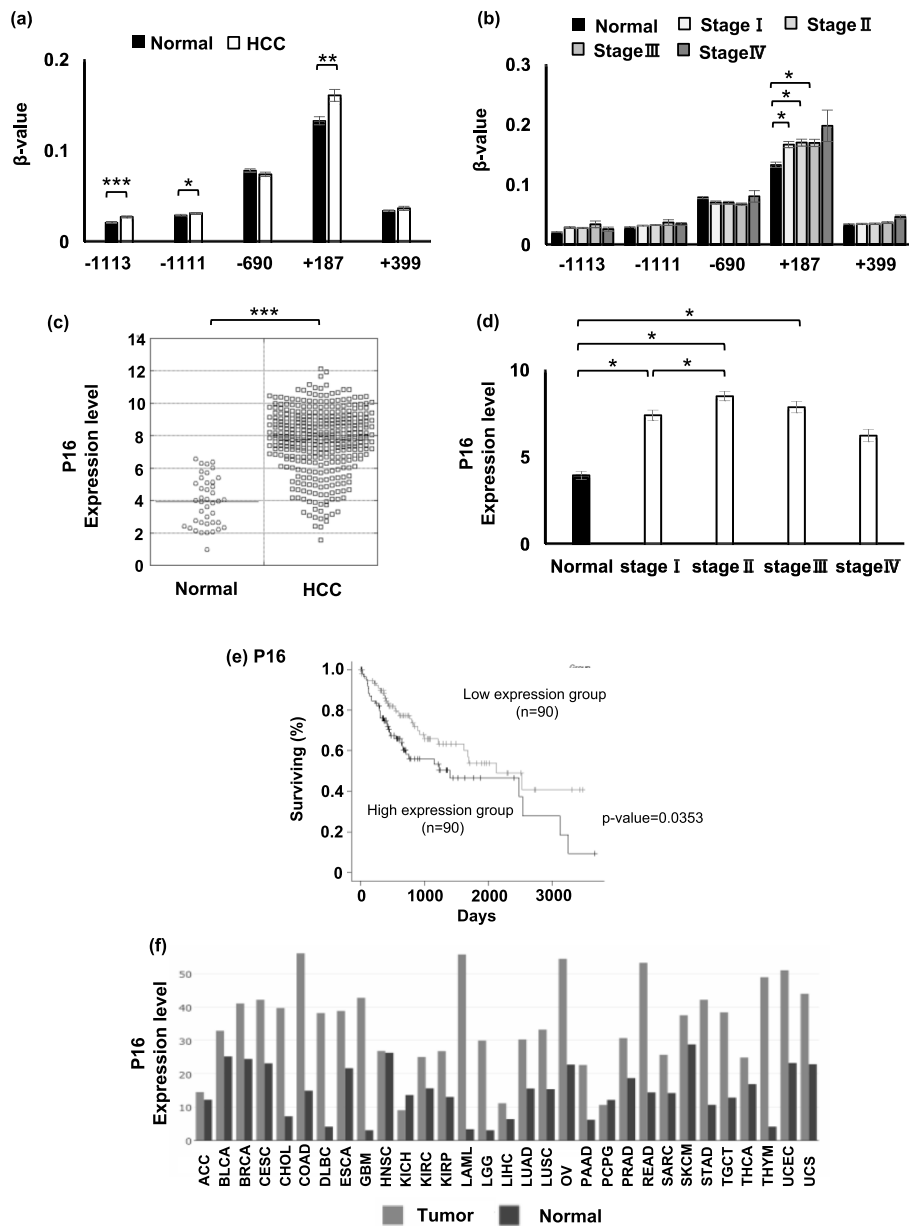
Fig. 6 DNA methylation, gene expression, and survival analysis of *P16* gene. **a** The *β* values in the 74 HCC tissues highly expressing *P16* and in 41 non-tumor tissues. Statistical significance between the two groups was analyzed by Student's *t* test. *. **. *** Significantly different at $p < 0.05$, $p < 0.01$, and $p < 0.001$ respectively. **b** The *β* values of *P16* in HCC tissues of pathological stages I, II, III, and IV and non-tumor tissues. Statistical significance was analyzed by one-way ANOVA followed by Turkey-Kramer test as a post hoc comparison. * Significantly different at $p < 0.05$. **c** The expression levels of *P16* in HCC tissues and non-tumor tissues. **d** The expression levels of *P16* in HCC tissues of pathological stages I, II, III, and IV and non-tumor tissues. **e** Kaplan-Meier analyzes for 90 HCC patients with higher and lower expression of *P16*. **f** The expression levels of *P16* in various types of cancers. Light black and dark black bars indicate expression in tumor tissues and non-tumor tissues, respectively. ACC adenoid cystic carcinoma, BLCA bladder carcinoma, BRCA breast invasive carcinoma, CESC cervical squamous cell carcinoma, CHOL cholangiocarcinoma, COAD colon adenocarcinoma, DLBC diffuse large B cell lymphoma, ESCA esophageal carcinoma, GBM glioblastoma, HNSC head-neck squamous cell carcinoma, KICH kidney chromophobe, KIRC kidney renal clear cell carcinoma, KIRP kidney renal papillary cell carcinoma, LAML acute myeloid leukemia, LGG low-grade glioma, LIHC liver hepatocellular carcinoma, LUAD lung adenocarcinoma, LUSC lung squamous cell carcinoma, OV ovarian cancer, PAAD pancreatic adenocarcinoma, PCPG pheochromocytoma and paraganglioma, PRAD prostate adenocarcinoma, READ rectum adenocarcinoma, SAC sarcoma, SKCM skin cutaneous melanoma, STAD stomach adenocarcinoma, TGCT testicular germ cell tumors, THCA thyroid cancer, THYM thymoma, UCEC uterine corpus endometrial carcinoma, UCS uterine carcinosarcoma

progression. *UBE2C* expression is upregulated in various cancers including the liver [42] and abnormal expression of *UBE2C* promotes cell cycle progression [43]. The deletion of UBE2C notably reduced the level of phosphorylated aurora kinase A via Wnt/β−catenin and PI3K/Akt and results in inhibition of the cancer growth and metastasis [44]. In addition, this gene is the target of miRNAs leading to the inhibition of cancer cell growth and survival in vitro and in vivo [44].

The present study showed promoter hypomethylation of these genes is associated with increased expression in HCC patients. Newly developed technologies using genome editing, such as CRISPR/Cas9, have enabled not only site-specific DNA demethylation but also methylation in vitro and even in vivo [17–19]. Thus, manipulations of the altered methylation sites of these genes might be promising targets of HCC therapy. As described above, the five genes (*FANCB*, *KIF15*, *KIF4A*, *ERCC6L*, and *UBE2C*) are all involved in fundamental cellular functions and found in many types of cancers. Thus, manipulations of the altered methylation sites of these genes might be promising targets of therapy of HCC and possibly of other types of cancers. The relations between methylation changes of specific gene promoters and cancer etiology are yet to be investigated.

Silencing of the tumor suppressor gene *P16* by DNA methylation is known to lead to development of cancer cells [45]. The analysis of the TCGA database in the present study, however, indicated upregulation of *P16* in many types of cancers (Fig. 6f). *P16* takes a part as an early gatekeeper against cancer and the silencing begins at preinvasive stages of a variety of cancers [10, 46]. On the other hand, recent studies reported that cellular senescence, an irreversible cell cycle arrest, becomes rather a promoting factor of cancer exacerbation through acquisition of the senescence-associated secretory phenotype (SASP) and P*16* upregulation associates with cellular senescence [47–49]. Upregulation of *P16* in many types of cancers (Fig. 6f) may reflect the stage when *P16* downregulation is no longer a factor of cancer progression but rather *P16* associates with cellular senescence.

Conclusion

In the present study, we clarified that expression of *FANCB*, *KIF15*, *KIF4A*, *ERCC6L*, and *UBE2C* are regulated by DNA methylation changes of their promoter CpGis and closely associated with tumor prognosis in HCC using the TCGA database. The manipulation of these methylations is a promising novel approach for treatment of HCC and possibly of other cancers with higher site specificity and less side effects. Furthermore, the protocol that we presented in the present study is applicable to other cancers and has great potential to drive new strategies of epigenetic cancer treatment forward.

Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1186/s12199-020-00871-8.

Additional file 1: Table S1. Sample ID of HCC and normal liver tissues in TCGA database

Additional file 2: Table S2. List of up/down regulated genes in HCC tissues compared with normal liver tissues

---

Abbreviations
HCC: Hepatocellular carcinoma; ncRNA: non-coding RNA; CpGi: CpG island; TSS: Transcription start site; 5-aza-dC: 5-aza-2′-deoxycytidine; TCGA: The Cancer Genome Atlas; GO: Gene Ontology

Authors' contributions
JM and KN designed the study and wrote the manuscript. TS, KO, and GI gave instructions and were involved in analysis. All authors read and approved the final version of the manuscript.

Availability of data and materials
The datasets analyzed in the current study are available from the Cancer Genome Atlas (TCGA) database.

Ethics approval and consent to participate
Not applicable

Consent for publication
Not applicable

Competing interests
All authors declare that they have no competing interests in this manuscript.

Author details
[1]Center for Health and Environmental Risk Research, National Institute for Environmental Studies, Tsukuba, Japan. [2]Graduate School of Pharmaceutical Sciences, Tokyo University of Science, Noda, Japan.

References
1.   Llovet JM, Zucman-Rossi J, Pikarsky E, Sangro B, Schwartz M, Sherman M, et al. Hepatocellular carcinoma. Nat Rev Dis Primers. 2016;2:16018.
2.   Villanueva A. Hepatocellular Carcinoma. N Engl J Med. 2019;380:1450–62.
3.   Waller LP, Deshpande V, Pyrsopoulos N. Hepatocellular carcinoma: A comprehensive review. World J Hepatol. 2015;7:2648–63.
4.   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.

5. Skvortsova K, Iovino N, Bogdanovic O. Functions and mechanisms of epigenetic inheritance in animals. Nat Rev Mol Cell Biol. 2018;19:774–90.
6. Tirado-Magallanes R, Rebbani K, Lim R, Pradhan S, Benoukraf T. Whole genome DNA methylation: beyond genes silencing. Oncotarget. 2017;8: 5629–37.
7. Hassler MR, Egger G. Epigenomics of cancer - emerging new concepts. Biochimie. 2012;94:2219–30.
8. Khan FS, Ali I, Afridi UK, Ishtiaq M, Mehmood R. Epigenetic mechanisms regulating the development of hepatocellular carcinoma and their promise for therapeutics. Hepatol Int. 2017;11:45–53.
9. Villanueva A, Portela A, Sayols S, Battiston C, Hoshida Y, Mendez-Gonzalez J, et al. DNA methylation-based prognosis and epidrivers in hepatocellular carcinoma. Hepatology. 2015;61:1945–56.
10. Jones PA, Baylin SB. The epigenomics of cancer. Cell. 2007;128:683–92.
11. Zhao R, Choi BY, Lee MH, Bode AM, Dong Z. Implications of Genetic and Epigenetic Alterations of CDKN2A (p16(INK4a)) in Cancer. EBioMedicine. 2016;8:30–9.
12. Egger G, Liang G, Aparicio A, Jones PA. Epigenetics in human disease and prospects for epigenetic therapy. Nature. 2004;429:457–63.
13. Ma YY, Zhao M, Liu Y, Zhao DF, Wang LX, Chen XP, et al. Use of decitabine for patients with refractory or relapsed acute myeloid leukemia: a systematic review and meta-analysis. Hematology. 2019;24:507–15.
14. Cheng JC, Weisenberger DJ, Gonzales FA, Liang G, Xu GL, Hu YG, et al. Continuous zebularine treatment effectively sustains demethylation in human bladder cancer cells. Mol Cell Biol. 2004;24:1270–8.
15. Issa JP. DNA methylation as a therapeutic target in cancer. Clin Cancer Res. 2007;13:1634–7.
16. Garcia JS, Jain N, Godley LA. An update on the safety and efficacy of decitabine in the treatment of myelodysplastic syndromes. Onco Targets Ther. 2010;3:1–13.
17. Lei Y, Huang YH, Goodell MA. DNA methylation and de-methylation using hybrid site-targeting proteins. Genome Biol. 2018;19:187.
18. Lei Y, Zhang X, Su J, Jeong M, Gundry MC, Huang YH, et al. Targeted DNA methylation in vivo using an engineered dCas9-MQ1 fusion protein. Nat Commun. 2017;8:16026.
19. McDonald JI, Celik H, Rois LE, Fishberger G, Fowler T, Rees R, et al. Reprogrammable CRISPR/Cas9-based system for inducing site-specific DNA methylation. Biol Open. 2016;5:866–74.
20. Shin SH, Kim BH, Jang JJ, Suh KS, Kang GH. Identification of novel methylation markers in hepatocellular carcinoma using a methylation array. J Korean Med Sci. 2010;25:1152–9.
21. Kanda Y. Investigation of the freely available easy-to-use software 'EZR' for medical statistics. Bone Marrow Transplant. 2013;48:452–8.
22. Huangda W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009;4:44–57.
23. Gene Ontology Consortium. The Gene Ontology (GO) project in 2006. Nucleic Acids Res. 2006;34:D322–6.
24. Ghouri YA, Mian I, Rowe JH. Review of hepatocellular carcinoma: Epidemiology, etiology, and carcinogenesis. J Carcinog. 2017;16:1.
25. Arya AK, Bhadada SK, Singh P, Sachdeva N, Saikia UN, Dahiya D, et al. Promoter hypermethylation inactivates CDKN2A, CDKN2B and RASSF1A genes in sporadic parathyroid adenomas. Sci Rep. 2017;7:3123.
26. Herman JG, Merlo A, Mao L, Lapidus RG, Issa JP, Davidson NE, et al. Inactivation of the CDKN2/p16/MTS1 gene is frequently associated with aberrant DNA methylation in all common human cancers. Cancer Res. 1995;55:4525–30.
27. Collins K, Jacks T, Pavletich NP. The cell cycle and cancer. Proc Natl Acad Sci U S A. 1997;94:2776–8.
28. Dominguez-Brauer C, Thu KL, Mason JM, Blaser H, Bray MR, Mak TW. Targeting Mitosis in Cancer: Emerging Strategies. Mol Cell. 2015;60:524–36.
29. Williams GH, Stoeber K. The cell cycle and cancer. J Pathol. 2012;226:352–64.
30. Sun XJ, Wang MC, Zhang FH, Kong X. An integrated analysis of genome-wide DNA methylation and gene expression data in hepatocellular carcinoma. FEBS Open Bio. 2018;8:1093–103.
31. Alix-Panabieres C, Cayrefourcq L, Mazard T, Maudelonde T, Assenat E, Assou S. Molecular Portrait of Metastasis-Competent Circulating Tumor Cells in Colon Cancer Reveals the Crucial Role of Genes Regulating Energy Metabolism and DNA Repair. Clin Chem. 2017;63:700–13.
32. Glaas MF, Wiek C, Wolter LM, Roellecke K, Balz V, Okpanyi V, et al. Mutational and Functional Analysis of FANCB as a Candidate Gene for Sporadic Head and Neck Squamous Cell Carcinomas. Anticancer Res. 2018;38:1317–25.
33. Chen T, Yang S, Xu J, Lu W, Xie X. Transcriptome sequencing profiles of cervical cancer tissues and SiHa cells. Funct Integr Genomics. 2020;20:211–21.
34. Myers SM, Collins I. Recent findings and future directions for interpolar mitotic kinesin inhibitors in cancer therapy. Future Med Chem. 2016;8:463–89.
35. Liu X, Gong H, Huang K. Oncogenic role of kinesin proteins and targeting kinesin therapy. Cancer Sci. 2013;104:651–6.
36. Hou G, Dong C, Dong Z, Liu G, Xu H, Chen L, et al. Upregulate KIF4A Enhances Proliferation, Invasion of Hepatocellular Carcinoma and Indicates poor prognosis Across Human Cancer Types. Sci Rep. 2017;7:4148.
37. Sun YF, Wu HL, Shi RF, Chen L, Meng C. KIF15 Promotes Proliferation and Growth of Hepatocellular Carcinoma. Anal Cell Pathol (Amst). 2020;2020: 6403012.
38. Gao H, Chen X, Cai Q, Shang Z, Niu Y. Increased KIF4A expression is a potential prognostic factor in prostate cancer. Oncol Lett. 2018;15:7941–7.
39. Qiao Y, Chen J, Ma C, Liu Y, Li P, Wang Y, et al. Increased KIF15 Expression Predicts a Poor Prognosis in Patients with Lung Adenocarcinoma. Cell Physiol Biochem. 2018;51:1–10.
40. Pu SY, Yu Q, Wu H, Jiang JJ, Chen XQ, He YH, et al. ERCC6L, a DNA helicase, is involved in cell proliferation and associated with survival and progress in breast and kidney cancers. Oncotarget. 2017;8:42116–24.
41. Xie Y, Yu J, Wang F, Li M, Qiu X, Liu Y, et al. ERCC6L promotes cell growth and invasion in human colorectal cancer. Oncol Lett. 2019;18:237–46.
42. Ieta K, Ojima E, Tanaka F, Nakamura Y, Haraguchi N, Mimori K, et al. Identification of overexpressed genes in hepatocellular carcinoma, with special reference to ubiquitin-conjugating enzyme E2C gene expression. Int J Cancer. 2007;121:33–8.
43. Wagner KW, Sapinoso LM, El-Rifai W, Frierson HF, Butz N, Mestan J, et al. Overexpression, genomic amplification and therapeutic potential of inhibiting the UbcH10 ubiquitin conjugase in human carcinomas of diverse anatomic origin. Oncogene. 2004;23:6621–9.
44. Zhang Y, Tian S, Li X, Ji Y, Wang Z, Liu C. UBE2C promotes rectal carcinoma via miR-381. Cancer Biol Ther. 2018;19:230–8.
45. Cui C, Gan Y, Gu L, Wilson J, Liu Z, Zhang B, et al. P16-specific DNA methylation by engineered zinc finger methyltransferase inactivates gene transcription and promotes cancer metastasis. Genome Biol. 2015;16:252.
46. Pfeifer GP. Defining Driver DNA Methylation Changes in Human Cancer. Int J Mol Sci. 2018;19.
47. Horne DJ, Jones BE, Kamada A, Fukushima K, Winthrop KL, Siegel SAR, et al. Multicenter study of QuantiFERON((R))-TB Gold Plus in patients with active tuberculosis. Int J Tuberc Lung Dis. 2018;22:617–21.
48. Inoue K, Fry EA. Aberrant expression of p16(INK4a) in human cancers - a new biomarker? Cancer Rep Rev. 2018;2.
49. Romagosa C, Simonetti S, Lopez-Vicente L, Mazo A, Lleonart ME, Castellvi J, et al. p16(Ink4a) overexpression in cancer: a tumor suppressor gene associated with senescence and high-grade tumors. Oncogene. 2011;30: 2087–97.

Publisher's Note