Identification of Flap endonuclease 1 as a potential core gene in hepatocellular carcinoma by integrated bioinformatics analysis

Chuanfei Li[1], Feng Qin[2], Hao Hong[3], Hui Tang[4], Xiaoling Jiang[5], Shuangyan Yang[4], Zhechuan Mei[1], Di Zhou[6]

[1] Department of Gastroenterology, The Second Affiliated Hospital of Chongqing Medical University, Chongqing, China
[2] Department of Infectious Diseases, The People's Hospital of Shi Zhu, Chongqing, China
[3] Department of Orthopaedics, The Second Affiliated Hospital of Chongqing Medical University, Chongqing, China
[4] Department of Infectious Diseases, Institute for Viral Hepatitis, The Key Laboratory of Molecular Biology for Infectious Diseases, Chinese Ministry of Education, The Second Affiliated Hospital of Chongqing Medical University, Chongqing, China
[5] Tongnan District People' Hospital, The First Affiliated Hospital of Chongqing Medical University, Chongqing, China
[6] Department of Radiology, The First Affiliated Hospital of Chongqing Medical University, Chongqing, China

Corresponding Author:
Di Zhou
No.1 Youyi Road, Yuzhong District, Chongqing, 400016, China
Email address: zhoudi@cqmu.edu.cn

ABSTRACT

Hepatocellular carcinoma (HCC) is a common yet deadly form of malignant cancer. However, the specific mechanisms involved in HCC diagnosis have not yet fully elucidated. Herein, we screened four publically available Gene Expression Omnibus (GEO) expression profiles (GSE14520, GSE29721, GSE45267 and GSE60502), and used them to identify 409 differentially expressed genes (DEGs), including 142 and 267 up- and down-regulated genes, respectively. The DAVID database was used to look for functionally enriched pathways among DEGs, and the STRING database and Cytoscape platform were used to generate a protein-protein interaction (PPI) network for these DEGs. The cytoHubba plug-in was utilized to detect 185 hub genes, and three key clustering modules were constructed with the MCODE plug-in. Gene functional enrichment analyses of these three key clustering modules were further performed, and nine core genes including BIRC5, DLGAP5, DTL, FEN1, KIAA0101, KIF4A, MCM2, MKI67, and RFC4, were identified in the most critical cluster. Subsequently, the

39 hierarchical clustering and expression of core genes in TCGA liver cancer tissues were analyzed
40 using the UCSC Cancer Genomics Browser, and whether elevated core gene expression was
41 linked to a poor prognosis in HCC patients was assessed using the GEPIA database. The PPI of
42 the nine core genes revealed an interaction between FEN1, MCM2, RFC4, and BIRC5.
43 Furthermore, the expression of FEN1 was positively correlated with that of three other core
44 genes in TCGA liver cancer tissues. FEN1 expression in HCC and other tumor types was
45 assessed with the FIREBROWSE and ONCOMINE databases, and results were verified in HCC
46 samples and hepatoma cells. FEN1 levels were also positively correlated with tumor size, distant
47 metastasis and vascular invasion. In conclusion, we identified nine core genes associated with
48 HCC development, offering novel insight into HCC progression. In particular, the aberrantly
49 elevated FEN1 may represent a potential biomarker for HCC diagnosis and treatment.
50
51 **Subjects** Bioinformatics, Gastroenterology and Hepatology
52 **Keywords** Hepatocellular carcinoma, Core genes, Bioinformatics analysis, Flap endonuclease 1
53

INTRODUCTION

55 Hepatocellular carcinoma remains among the most common and deadly forms of cancer
56 globally, posing a significant threat to human life (Forner et al. 2018). HCC and other tumors
57 develop as the result of the long-term accumulation of genetic mutations. Although a large
58 number of biomarkers for the diagnosis of HCC have been identified (C et al. 2018), the specific
59 molecular mechanisms related to the onset, recurrence and treatment of HCC remain obscure.
60 Therefore, it is essential to identify and exploit novel biomarkers involved in HCC onset and
61 progression to better understand the pathogenesis of HCC.
62   Between human genome sequencing efforts and the rapid development of gene sequencing
63 technologies, precision medicine has risen to prominence and been widely employed in the field
64 of oncology (IR et al. 2017; R et al. 2019). Precision medicine relies upon initially exploring
65 potential therapeutic targets via high-throughput sequencing technologies (PN, 2012), as these
66 technologies allow for the large-scale investigation of altered gene expression in the context of
67 disease (Chen et al. 2010). However, sequencing results are often limited and inconsistent owing
68 to the heterogeneity of samples in independent studies, and due to the fact that most studies focus
69 on one cohort. As such, this study sought to analyze genes involved in liver cancer development
70 using a range of available liver cancer-related gene chip datasets, with the goal of identifying
71 potential novel molecular targets for liver cancer treatment and diagnosis.
72   For the purposes of this study, four HCC related Gene Expression Omnibus (GEO) database
73 datasets were downloaded: GSE14520 (S et al. 2012; S et al. 2010), GSE29721 (B et al. 2011),
74 GSE45267 (CL et al. 2018) and GSE60502 (YH et al. 2014). By analyzing these four datasets,
75 we identified 409 DEGs including 142 up-regulated and 267 down-regulated genes. The DAVID
76 database was then used for a functional enrichment analysis of these DEGs, while the STRING
77 database and Cytoscape were utilized to generate a protein-protein interaction (PPI) network, and
78 three clustering modules were filtered out with the MCODE plug-in, among which clustering

79   module 1 was most associated with HCC. In addition, nine core genes including BIRC5,
80   DLGAP5, DTL, FEN1, KIAA0101, KIF4A, MCM2, MKI67, and RFC4, were identified within
81   clustering module 1, and these corresponded to the hub genes in our PPI network. Using the
82   GEPIA database, we performed survival analyses of patients based on expression of these nine
83   core genes, revealing that their overexpression was linked to a poorer prognosis in HCC patients.
84   When we surveyed the literature surrounding these genes, we found that all except for FEN1 had
85   previously been confirmed to play a vital role in HCC. Existing studies have shown that FEN1 is
86   highly expressed in various cancers, such as brain (Nikolova et al. 2009), lung (He et al. 2017;
87   Zhang et al. 2018), breast (Abdel-Fatah et al. 2014), gastric (Wang et al. 2014), prostate (Lam et
88   al. 2006) and pancreatic cancer (Isohookana et al. 2018), but its expression and role in HCC
89   remains unclear. Our previous data indicated that FEN1 expression was elevated in HCC tumors,
90   and this was confirmed upon comparing HCC samples and hepatoma cells to appropriate
91   controls. Together, the results of these analyses suggest that FEN1 may be a core gene
92   orchestrating the progression of HCC.
93
94   **MATERIALS & METHODS**
95   **Data collection and DEG validation**
96   Four liver cancer-related datasets (GSE14520, GSE29721, GSE45267 and GSE60502), were
97   downloaded from GEO (http://www.ncbi.nlm.nih.gov/geo/). These datasets contained a total of
98   299 tumor samples and 289 non-tumor samples. GSE14520 contained 225 liver cancer samples
99   and 220 adjacent controls; GES29721 contained 10 pairs of liver cancer samples and adjacent
100 control tissue; GSE45267 contained 46 liver cancer samples and 41 adjacent controls; and
101 GSE60502 contained 18 pairs of liver cancer tissues and adjacent controls. The inclusive criteria
102 for datasets were as follows: (1) the samples were from human HCC tissues and paired adjacent
103 or non-tumor tissues; (2) gene expression profiling of mRNA; (3) each dataset included no less
104 than ten paired samples; (4) the datasets were from 2009-2019. The exclusive criteria for datasets
105 were as follows: (1) the HCC samples were induced by specific diseases, including chronic
106 hepatitis B, hepatitis C, hepatitis D, alcoholic liver disease, nonalcoholic fatty liver disease and
107 cholangiocarcinoma; (2) Details of the samples were not available; (3) the datasets could not be
108 analyzed by GEO2R; (4) No gene symbol. The specific platform information for these four
109 datasets is compiled in Table 1. GEO2R was utilized to identify DEGs in these studies, using the
110 screening criteria: │ log FC (fold change) │ ⩾ 1, P < 0.05, and adjusted P < 0.05. In addition,
111 genes with multiple probe set or probe sets lacking matched gene symbols were removed or
112 averaged, respectively. Then, the overall DEGs, as well as those that were up- and down-
113 regulated in the four datasets were intersected and visualized using Funrich (v 3.0,
114 http://funrich.org/index.html).
115 **Functional enrichment analyses**
116 Gene ontology analyses focuses on three domains: biological processes (BP), cellular
117 components (CC), and molecular functions (MF), and such analyses are commonly used to
118 understand the biological functions, pathways, or localization of DEGs. The Kyoto Encyclopedia

119 of Genes and Genomes (KEGG) pathway analysis database surveys as a valuable resource for
120 assessing how particular DEGs may be involved in or influenced by specific signaling pathways
121 and disease states. The DAVID (https://david.ncifcrf.gov/) was utilized for functional enrichment
122 analyses, with a P < 0.05 cutoff for significance.
123

PPI network analysis

125 The STRING database (v10.5; https://stringdb.org/) was used to generate a DEG PPI, with a
126 minimum interaction score cutoff of 0.4. Cytoscape (v 3.4.0, https://cytoscape.org/) was used for
127 network visualization, and the CytoHubba plug-in was used to identify hub genes with the
128 criteria of filtering degree ⩾ 10. The MCODE plug-in was used to construct key clustering
129 modules (MCODE score > 10, degree cut-off = 2, node score cut-off = 0.2, Max depth = 100 and
130 k-score = 2).
131

Validation of core genes

133 Hub genes among overall DEGs and the most critical clustering module were identified through
134 the CytoHubba plug-in, and intersecting core genes were identified. The UCSC Cancer
135 Genomics Browser (https://genome-cancer.ucsc.edu/) was then used for hierarchical clustering
136 of these core genes. In addition, the expression profiles of these core genes in 421 TCGA liver
137 cancer tissues, including 50 solid normal tissues and 371 primary tumors, were determined by
138 analyzing available datasets. A core gene PPI network was constructed with the cBioportal
139 online database (http://www.cbioportal.org/). Furthermore, correlations between the expression
140 of FEN1 and MCM2, RFC4, or BIRC5 in TCGA liver cancer tissues were investigated. Finally,
141 we used the ONCOMINE (https://www.oncomine.org/resource/login.html) and FIREBROWSE
142 online database (http://firebrowse.org/) to investigate the FEN1 expression in various cancers
143 including HCC.
144

Survival analysis

146 The GEPIA database (http://gepia.cancer-pku.cn/) was used to conduct survival analyses based
147 on core gene expression, with hazard ratios (HRs) and 95% confidence intervals being
148 calculated, and logrank P value < 0.05 being the threshold of statistical significance.
149

Clinical samples

151 A total of 34 paired HCC tumor tissues as well as corresponding adjacent non-cancerous tissues
152 were obtained from our Hospital's Department of Hepatobiliary Surgery, with all patients
153 providing informed consent. The content of the informed consent includes research purposes,
154 risks and discomfort, benefits, and privacy issues. The form of the informed consent is in written.
155 The study was examined and approved by the Second affiliated Hospital of Chongqing Medical
156 University Ethics Committee (approval number: 2018-60).
157

Cell culture

159   Seven human liver cancer cell lines (SMMC-7721, BEL-7404, HCCLM3, HepG2, MHCC97-H,
160   SK-HEP-1, and Huh-7) as well as normal human liver HL7702 cells were donated by the
161   Institute for Viral Hepatitis, Chongqing Medical University. All cells were grown in high
162   glucose DMEM (Gibco, USA) containing 10% FBS (Corning, USA) at 37°C in a 5% CO_2
163   incubator.
164
165   **Hematoxylin and eosin (H&E) staining**
166   Paraffin-embedded tissues were dewaxed in xylene I, II, and III for 20 minutes each, and then
167   dehydrated in an ethanol gradient (100%, 95%, 90%, 80%, and 70%), 3 minutes per step.
168   Sections were then rinsed with distilled water for 5 minutes, and the nucleus was counterstained
169   with hematoxylin for 3 minutes. Sections were washed again in water, followed by
170   differentiation for 30 seconds in a 75% hydrochloric acid alcohol solution, and blue color was
171   returned by washing with distilled water for 5 minutes. A red dye was then used for
172   counterstaining for 5 minutes, after which samples were dehydrated in 70%, 80%, 95%, and
173   100% ethanol, 1 minute per concentration. Sections were then cleared with xylene and sealed
174   using neutral gum.
175
176   **IHC analysis**
177   IHC staining was conducted as previously described (WG et al. 2016). Briefly, paraffin-
178   embedded sections were incubated at 56°C for 2 h, and then 3% hydrogen peroxide was used for
179   antigen retrieval. Afterwards, the sections were incubated with rabbit anti-human FEN1 (1:100,
180   A1175, ABclonal, China) at 4°C overnight. Then, sections were probed for 1 h using HRP-
181   conjugated secondary antibodies at 37°C, after which a DAB substrate kit was utilized, and
182   hematoxylin was used for nuclear staining. The Image-Pro Plus (IPP) software (Media
183   Cybernetics, Rockville, MD, USA) was used to quantify staining intensity. The frequency of
184   positively-stained cells was determined on a 0 − 100 scale, while staining intensity was scored as
185   follows: 0 = negative; 1 = weak; 2 = moderate; 3 = strong. These two scores were then multiplied
186   together to yield an IHC score between 0 and 300. The final scores were assigned by two
187   independent pathologists. The mean IHC score was used as a cutoff value to separate patients
188   into low- and high-expression groups.
189
190   **RT-qPCR analysis**
191   TRIzol (ThermoFisher Scientific, USA) was used for total RNA extraction, and RNA was then
192   reverse transcribed with the PrimeScript RT-PCR kit (Takara Bio, Dalian, China) based on
193   provided protocols. The SYBR Premix Ex Taq II (Takara, Japan) kit was used to conduct RT-
194   qPCR analysis on a Bio-Rad CFX96 Real-Time System (Bio-Rad, Hercules, CA). The reaction
195   conditions were as follows: pre-denaturation at 95℃ for 30 s, followed by 40 cycles of
196   amplification at 95℃ for 5 s and 60℃ for 30 s, and a final cycle along the melting curve from
197   65℃ to 95℃ in increments of 0.5℃ for 5 s. mRNA levels were determined via the $2^{-\Delta\Delta Ct}$
198   method, with GAPDH used for normalization. Primers used were: GAPDH F:5'-

199 GGTGGTCTCCTCTGACTTCAACA -3' and R:5'- GTTGCTGTAGCCAAATTCGTTGT-3',

200 FEN1 F: 5'- CTGTGGACCTCATCCAGAAGCA -3' and R: 5'-

201 CCAGCACCTCAGGTTCCAAGA -3'.

202

Statistical analysis

204 Statistical analyses and graphing were performed with SPSS v19.0 (SPSS Inc., USA) and Graph

205 Pad Prism v8.0 (Graph Pad Software, USA), respectively. Data are means ± standard deviation

206 (SD). Student's t-tests were used to compare groups. Fisher's exact test was used to assess

207 correlations between the expression of FEN1 and HCC patient clinicopathological features.

208 Spearman's correlation analyses were used to compare the expression of pairs of genes in TCGA

209 liver cancer tissues. $P < 0.05$ was the significance threshold ($*P < 0.05$, $**P < 0.01$).

210

RESULTS

HCC-associated DEG identification

213 In this study, 1088 total DEGs (505 and 583 up- and down-regulated, respectively) in

214 GSE14520, 1449 total DEGs (837 and 612 up- and down-regulated, respectively) in GSE29721,

215 1604 total DEGs (713 and 891 up- and down-regulated, respectively) in GSE45267, and 1533

216 total DEGs (792 and 741 up- and down-regulated, respectively) in GSE60502 were screened.

217 Based on these datasets, a total of 409 overlapping DEGs were identified among these four

218 datasets (142 and 267 up- and down-regulated, respectively), as visualized with the Funrich

219 software (Figure 1). DEGs are listed in Table S1.

220

DEG functional enrichment analyses

222 To explore the biological activities of these DEGs, the DAVID database was used to conduct GO

223 and KEGG enrichment analysis. With respect to BPs, up-regulated DEGs were primarily

224 enriched in processes such as mitotic nuclear division, cell division, cell cycle, DNA replication,

225 and mitotic sister chromatid segregation (Figure S1A), while down-regulated DEGs were

226 primarily enriched in processes such as redox process, the cytochrome 450 pathway, drug

227 metabolism, and negative regulation of growth (Figure S2A). With respect to CCs, up-regulated

228 DEGs were primarily enriched in the nucleoplasm, nucleus, cytoplasm, spindle, and cellular

229 intermediates (Figure S1B), while down-regulated DEGs were mostly enriched in extracellular

230 exosomes, organelle membranes, blood microparticles, extracellular regions, and the

231 mitochondrial matrix (Figure S2B). With respect to MFs, up-regulated DEGs were primarily

232 enriched in functions such as protein binding, ATP binding, DNA helicase activity, protein

233 kinase binding, single-stranded DNA binding, and chromatin binding (Figure S1C), whereas

234 down-regulated DEGs were primarily associated with iron ion binding, oxidoreductase activity,

235 heme binding, monooxygenase activity, and oxygen binding (Figure S2C).

236   A KEGG analysis revealed that up-regulated DEGs were particularly enriched in pathways

237 such as the cell cycle, DNA replication, P53 signaling, and tumor pathways (Figure S1D).

238 However, down-regulated DEGs were mostly associated with metabolic pathways, fatty acid
239 degradation, chemical carcinogenesis, and PPAR signaling (Figure S2D).
240
241 **PPI network and module analyses**
242 To better understand interactions among DEGs, the STRING online database was used to
243 generate a PPI network consisting of 403 nodes and 3502 edges, which was visualized using
244 Cytoscape. Six of the 409 DEGs were not included in this network (Figure 2A). This network
245 was then analyzed using the MCODE plug-in, and three clustering modules were filtered out
246 according to the chosen screening conditions. Clustering module 1 scored 58.492 with 62 nodes
247 and 1784 edges (Figure 2B), clustering module 2 scored 11.529 with 18 nodes and 98 edges
248 (Figure 2C), and clustering module 3 scored 10.917 with 25 nodes and 131 edges (Figure 2D).
249 The genes in clustering module 1 were up-regulated DEGs, whereas those in the other two
250 modules were primarily down-regulated DEGs.
251
252 **Functional enrichment analysis of key clustering modules**
253 The DAVID database was next used to explore the biological functions of genes in these key
254 clustering modules (Tables S2-4). With respect to BPs, clustering module 1 was primarily
255 enriched in cell differentiation, mitotic nuclear division, DNA replication, and DNA helicase
256 activity, while clustering module 2 was primarily enriched in plasminogen activation,
257 coagulation, cytolysis, and complement activation regulation, and clustering module 3 was
258 primarily enriched in steroid metabolism, heterogeneous biomass metabolism, and exogenous
259 drug catabolism. With respect to CCs, clustering module 1 was primarily enriched for the
260 nucleoplasm, nucleus, intermediate, spindle, cytoplasm, nuclear chromosome, while clustering
261 module 2 was primarily enriched for exosomes, extracellular regions, membrane attack
262 complexes, and extracellular vesicles, and clustering module 3 was primarily enriched for
263 organelle membranes, the endoplasmic reticulum membrane, and high-density lipoproteins. With
264 respect to MFs, clustering module 1 was primarily enriched for protein binding, ATP binding,
265 DNA helicase activity, protein kinase binding, and DNA binding, while clustering module 2 was
266 primarily enriched for endopeptidase activity, transcription factor binding, steroid binding, RNA
267 polymerase II transcription factor activity, and enzyme binding activity, and clustering module 3
268 was primarily enriched for aerobic binding, iron ion binding and heme binding.
269 A KEGG analysis revealed that clustering module 1 was primarily enriched in the cell
270 cycle, oocyte meiosis, DNA replication, and p53 signaling, while clustering module 2 was
271 primarily enriched in the complement system, prion disease, and systemic lupus erythematosus,
272 and clustering module 3 was primarily enriched in chemical carcinogenesis, retinol metabolism,
273 P450 drug metabolism, and metabolic pathways.
274
275 **Identification of core genes and analysis of their clinical significance**
276 Next, core genes involved in HCC were identified based on their levels of interaction via
277 analyzing our PPI network using the Cytoscape program. Based on our clustering module

278 analysis, clustering module 1 included 62 genes was found closely related to the progression of
279 HCC. Then, nine total genes (BIRC5, DLGAP5, DTL, FEN1, KIAA0101, KIF4A, MCM2,
280 MKI67, and RFC4) were identified based on the intersecting genes among the top 40 genes
281 derived from 12 different algorithms by the cytoHubba plug-in (Table S5). In addition, 185 hub
282 genes were identified across 403 nodes in our PPI network based on the filtering degree $\geqslant$ 10
283 criteria (Table S6). The nine identified core genes were found belonged to this larger subset of
284 hub genes (Figure 3A).
285   To investigate core gene expression in HCC, a hierarchical clustering analysis was
286 performed using the UCSC Cancer Genomics Browser, revealing that these nine core genes were
287 highly expressed in most liver cancer samples (Figure 3C). Next, the expression profiles of these
288 nine core genes in 421 TCGA liver cancer tissues, including 50 solid normal tissues and 371
289 primary tumors, were downloaded and analyzed, revealing that the expression of these core
290 genes was significantly elevated in HCC (Figure 3B). The correlations between core gene
291 expression levels and patient prognosis in 182 total HCC samples were further assessed,
292 revealing that BIRC5 expression (HR= 2, logrank P= 6.7e-05) was correlated with worse overall
293 survival (OS) for HCC patients, as was that of DLGAP5 (HR= 1.9, logrank P= 0.00039), DTL
294 (HR= 1.7, logrank P= 0.0049), FEN1 (HR= 1.5, logrank P= 0.022), KIAA0101 (HR= 1.7,
295 logrank P= 0.002), KIF4A (HR= 1.8, logrank P= 0.001), MCM2 (HR= 1.7, logrank P= 0.0022),
296 MKI67 (HR= 1.9, logrank P= 0.00045), and RFC4 (HR= 1.7, logrank P= 0.004) (Figure 4).
297 Furthermore, the expression of BIRC5 (HR= 1.6, logrank P= 0.002) was correlated with
298 decreased disease-free survival (DFS) for HCC patients, as was that of DLGAP5 (HR= 1.6,
299 logrank P= 0.0033), DTL (HR= 1.6, logrank P= 0.0016), FEN1 (HR= 1.5, logrank P= 0.0075),
300 KIAA0101 (HR= 1.6, logrank P= 0.0022), KIF4A (HR= 1.6, logrank P= 0.0011), MCM2 (HR=
301 1.6, logrank P= 0.0034), MKI67 (HR= 1.9, logrank P= 4.2e-05), and RFC4 (HR= 1.5, logrank
302 P= 0.011) (Figure 5). High expression of these nine core genes was associated with significantly
303 reduced survival among HCC patients.
304
305 FEN1 may be a key candidate gene in HCC
306 To clarify the PPI network for these nine core genes, the cBioportal online database were
307 explored and identified an interaction between FEN1, MCM2, BIRC5 and RFC4 (Figure 6A).
308 Previous studies have confirmed that MCM2, BIRC5, and RFC4 are abnormally highly
309 expressed in HCC, and that they participate in the regulation of HCC tumor biology. However,
310 the expression of FEN1 and its clinical significance in HCC is unclear. Therefore, the
311 correlations between the expression of FEN1 and these three other genes were further analyzed.
312 The expression of FEN1 in TCGA liver cancer tissues was positively correlated with that of
313 MCM2 (r= 0.853, P= 0.000), BIRC5 (r= 0.809, P= 0.000), and RFC4 (r= 0.852, P= 0.000)
314 (Figure 6B-D), suggesting that FEN1 may play as important a role in the progression of liver
315 cancer as do MCM2, BIRC5, and RFC4. The ONCOMINE and FIREBROWSE databases were
316 further utilized to investigate the expression of FEN1 in various cancers, including HCC. The
317 results revealed that FEN1 expression was clearly elevated in most cancers, including bladder,

318 breast, colorectal, esophageal, lung, and liver cancer (Figure 7A, B). In addition, FEN1 was
319 found overexpressed in three HCC-related datasets (Figure 7C).
320
321 **Experimental validation**
322 In this study, 34 paired HCC and adjacent control tissues were used to verify the expression of
323 FEN1 in HCC. The pathologic diagnosis of HCC and matched adjacent samples were confirmed
324 by H&E staining (Figure 8A, C). FEN1 staining was localized to the nucleus and cytoplasm,
325 with stronger expression in the HCC samples relative to the adjacent controls (Figure 8B, D).
326 Furthermore, IHC analysis indicated that FEN1 levels were significantly higher in HCC relative
327 to adjacent tissues (Figure 8E), with IHC scores being significantly higher in HCC samples
328 (Figure 8F).
329 The correlation between FEN1 expression and HCC patient clinicopathological features was
330 investigated via Fisher's exact test. As shown in Table 2, there were significant correlations
331 between FEN1 expression and tumor size (P = 0.047 < 0.05), metastasis (P =0.013 < 0.05) and
332 vascular invasion (P =0.024 < 0.05). FEN1 expression did not significantly correlated with
333 gender, age, tumor multiplicity, TNM stage, pathological grade, HBsAg, liver cirrhosis or serum
334 alpha-fetoprotein (AFP) (P > 0.05). Furthermore, FEN1 mRNA levels were found significantly
335 elevated in six human hepatoma cell lines relative to that in the normal human liver cell line
336 HL7702 (Figure 8G).
337

DISCUSSION

339 The occurrence and progression of HCC are complex, with multiple cumulative genetic changes
340 ultimately culminating in progressive disease. High-throughput technologies such as gene chips
341 have been widely employed to elucidate the underlying mechanisms, providing an innovative
342 and effective approach to the diagnosis, prevention, and treatment of HCC. Numerous studies
343 have been conducted to clarify genetic changes underlying HCC development, but results to date
344 remain inconclusive of incomplete. As such, there is further need to investigate the molecular
345 mechanisms governing HCC.
346 Functional analysis of DEGs and screening of core genes: In this study, we identified 409
347 total DEGs (142 and 267 up- and down-regulated, respectively) shared among four HCC
348 datasets, and these genes were used for functional enrichment analyses. A GO analysis revealed
349 the up-regulated DEGs to be primarily linked with cell division, the cell cycle, and DNA
350 replication, whereas down-regulated DEGs were mostly associated with redox reactions,
351 cytochrome 450 functionality, and negative growth regulation. A KEGG pathway analysis
352 revealed up-regulated DEGs to mostly be associated with signaling relating to the cell cycle,
353 DNA replication, p53 signaling, and tumor pathways, whereas down-regulated DEGs were
354 mostly linked to metabolic pathways such as fatty acid degradation. These results suggested that
355 the up-regulated DEGs may affect HCC progression via regulating DNA replication and the cell
356 cycle, whereas down-regulated DEGs may be linked to HCC progression through metabolic
357 pathways. Previous studies have indicated that the dysregulation of the cell cycle is a key

358   hallmark of many cancer types (P et al., 2017). As one of the most important tumor suppressor
359   genes, p53 is closely related to tumorigenesis, with at least 50% of cancer patients exhibiting p53
360   mutations or loss of function (C et al. 2013; G & M 2019; MP et al. 2015). P53 signaling
361   dysregulation has repeatedly been confirmed to be linked with cancer development (M & J
362   2016). There is also increasing evidence that metabolism regulates cancer growth and
363   proliferation (J & CB 2019; T et al. 2016). We further used our DEGs to generate a PPI, which
364   consisted of three key clustering modules screened using the MCODE plug-in. Functional
365   enrichment analyses revealed that clustering module 1 was closely related to gene mutations in
366   the progression of HCC, and was enriched in genes linked to the cell cycle, DNA replication, and
367   the p53 signaling pathway. Then nine core genes (BIRC5, DLGAP5, DTL, FEN1, KIAA0101,
368   KIF4A, MCM2, MKI67 and RFC4) in clustering module 1 were screened, as these genes were
369   also hub genes in the overall PPI network. There are numerous studies related to HCC chips that
370   have sought to identify the molecules involved in the onset and progression of HCC. However,
371   due to the different data sets and search strategies included in each study, the key molecules
372   identified are different. Nevertheless, the core genes we identified were similar to those found by
373   other researchers. Cai C (Cai, Wang, & Tu, 2019) identified twelve hub gene including BIRC5
374   and MCM2 in HCC. DTL and MCM2 were also identified as hub genes for HCC in another
375   study (Sang, Wang, Xu, & Zhao, 2018). Shen S (Shen et al., 2019) identified FEN1 and RFC4 as
376   core genes by analyzing four HCC datasets. The above studies indicated the core genes screened
377   and identified in our study were of value. Afterwards, the hierarchical clustering and expression
378   profiles of core genes in TCGA liver cancer tissues were analyzed using the UCSC Cancer
379   Genomics Browser, and survival analyses suggested that aberrantly high expression of these core
380   genes was predictive of a poor HCC patient prognosis, suggesting these core genes may be key
381   molecular biomarkers for HCC diagnosis and treatment.
382   Research progress of core genes in HCC: BIRC5, also known as survivin, is an anti-
383   apoptotic protein reported to function as a potential oncogene in the context of many cancers (MJ
384   et al. 2007). Studies have shown that BIRC5 is highly expressed in the vast majority of tumors,
385   including HCC (C 2016). In addition, elevated BIRC5 levels have been found to be associated
386   with histological grade, tumor size, and TNM stage in HCC patients (W et al. 2019). DLGAP5,
387   also known as HURP, is a cell cycle regulatory gene that has been found to be highly expressed
388   in liver cancer (Chang et al. 2011; Tsou et al. 2003). Weijia Liao (Liao et al. 2013) reported the
389   abnormally high expression of DLGAP5 in HCC, and found that this was related to promoter
390   methylation level, with the silencing of DLGAP5 significantly inhibiting cell proliferation,
391   migration, and colony formation in vitro. DLGAP5 knockdown inhibited the proliferation of
392   hepatoma cells by reducing P53 accumulation (Kuo et al. 2012). DTL is a substrate receptor for
393   the CRL4 ubiquitin ligase, serving as a key regulator of the cell cycle and genomic stability.
394   DTL upregulation in invasive HCC has been found to be positively correlated with tumor grade
395   and patient survival (Chen et al. 2018b; Pan et al. 2006). FEN1, a structurally specific metal
396   nuclease, plays a vital role in DNA damage repair (Lieber 1997) and maintenance of genomic
397   stability (Becker et al. 2018). FEN1 has been found to be expressed at high levels in several

398   cancer types, including those of the lung (He et al. 2017), breast (Abdel-Fatah et al. 2014),
399   gastric (Wang et al. 2014), prostate (Lam et al. 2006) and pancreatic (Isohookana et al. 2018).
400   FEN1 mutations have been associated with the occurrence of gastrointestinal tumors, including
401   HCC (Liu et al. 2012b), suggesting that FEN1 may be important in the development of
402   gastrointestinal tumors. Nevertheless, the expression and biological function of FEN1 in HCC
403   remains unclear. KIAA0101 is a nuclear antigen-associated factor that is present in proliferating
404   cells and which regulates proliferative processes. KIAA0101 has been found to be highly
405   expressed in HCC, representing a potential biomarker for this tumor type that is relevant to
406   disease treatment and patient prognosis (Yuan et al. 2007). KIAA0101 variant 1 can promote the
407   survival of liver cancer cells by regulating p53 functionality, indicating that inhibition of
408   KIAA0101 variant 1 may be a promising therapeutic strategy (Liu et al. 2012a). Numerous
409   studies have revealed that KIF4A is an oncogene in the context of lung, oral, and breast cancer.
410   High KIF4A expression was significantly correlated with tumor stage, tumor differentiation, and
411   metastasis, and may be a biomarker for poor HCC prognosis (Hou et al. 2017). In addition,
412   KIF4A promoted the proliferation and invasion of hepatoma cells via p53 and Akt signaling
413   (Hou et al. 2017; Huang et al. 2018). MCM2 has been shown to be critical for chromatin
414   composition and for the focal formation of p53 binding protein 1 in HepG2 cells (Chen et al.
415   2018a). High MCM2 expression has been associated with poor prognosis in HCC patients (Liu et
416   al. 2018). MKI67 encodes a nuclear antigen expressed during the G1, S and G2-M phases in
417   proliferating cells, and its expression levels are closely related to tumor growth rate (T et al.
418   1995), histological stage (IO et al. 1995), and tumor recurrence for HCC (K et al. 1995). Higher
419   MKI67 expression was associated with both faster HCC progression and a poorer patient
420   prognosis (Y et al. 2015). Consistent with our results, RFC4 has been found to be overexpressed
421   in HCC (M et al. 2009). The silencing of RFC4 expression was able to reduce HepG2 cell
422   proliferation and promote apoptosis, and was also associated with the increased sensitivity of
423   cells to doxorubicin and camptothecin, suggesting RFC4 may be a novel target for liver cancer
424   treatment.
425   The identification of FEN1 and its potential mechanisms in HCC: As these past results
426   show, all of these core genes except for FEN1 have been reported to play a role in HCC
427   progression. To further explore interactions among these nine core genes, a PPI network was
428   generated revealing interactions between FEN1, MCM2, RFC4, and BIRC5. As expected, FEN1
429   expression was significantly positively correlated with that of MCM2, RFC4, and BIRC5 in
430   TCGA liver cancer tissues, suggesting that FEN1 may be as involved in the development of
431   HCC as are MCM2, RFC4, and BIRC5. Previous studies had confirmed that FEN1 was
432   aberrantly up-regulated in various cancers. However, the expression of FEN1 in HCC remain
433   obscure. Therefore, we first utilized the ONCOMINE and FIREBROWSE online databases to
434   investigate the expression of FEN1 in HCC, and found that FEN1 was highly expressed in
435   multiple tumor tissues including those of HCC patients, consistent with previous reports. We
436   then verified this in HCC tissues from 34 patients via IHC staining. Interestingly, a correlation
437   analysis revealed that high expression of FEN1 was significantly correlated with tumor size,

438   metastasis and vascular invasion, suggesting that FEN1 may be involved in the regulation of
439   liver cancer proliferation and metastasis. Similarly, a study had confirmed that the expression of
440   FEN1 in gastric cancer was positively associated with the degree of differentiation, lymphatic
441   metastasis, tumor size and TNM stage (K et al. 2014). In addition, the FEN1 was up-regulated in
442   lung cancer cells and associated with the clinical lung cancer stage (L et al. 2017). Another study
443   also suggested that FEN1 might contribute to nodal metastasis in lung adenocarcinoma and result
444   in a poor prognosis (JC et al. 2015). These findings suggested that FEN1 may be involved in the
445   regulation of multiple tumor biological processes. Mechanistically, FEN1, a structure-specific
446   metallonuclease, is involved in DNA replication, DNA synthesis, DNA damage repair, non-
447   homologous end joining and homologous recombination (Lieber 1997; Wu et al. 1999).
448   Importantly, FEN1 is essential for maintaining genome stability (Becker et al. 2018). The
449   upstream and downstream molecules of FEN1 may be involved in the regulation of various
450   tumors. A study has previously reviewed three underlying mechanisms by which FEN1 may be
451   involved in tumorigenesis (L et al. 2011). First, the interactions between FEN1 and other DNA
452   metabolic proteins are formed. Second, the super-accumulated FEN1 in nucleolus and
453   mitochondrion is thought to play a vital role in maintaining the stability of tandem ribosomal
454   DNA repeats and in mitochondrial DNA replication. Third, FEN1 could be post-translationally
455   modified through acetylation, methylation, or phosphorylation, and this was important for its
456   ability to regulate nuclease activities and protein partner selection. One study reported that FEN1
457   knockdown was able to induce DNA damage and enhance the activation of p53 in lung cancer
458   cells (L et al. 2017). Furthermore, FEN1 was found be regulated by the transcriptional repressor
459   Ying Yang 1 (YY1) in response to DNA damaging agents in the context of breast cancer.
460   Another study demonstrated that FEN1 promoted breast cancer cell proliferation via an
461   epigenetic mechanism whereby FEN1-mediated up-regulation of DNMT1 and DNMT3a (X et al.
462   2019). At present, whether FEN1 participates in HCC progression and what the underlying
463   mechanisms of such participation may be remains unclear. We speculate that it may be related to
464   the above mechanisms. In addition, our results showed that FEN1 was elevated in six human
465   hepatoma cell lines relative to a control cell line via RT-qPCR analysis, which provided some
466   basis for further study of the function of FEN1 in liver cancer. However, the specific biological
467   function of FEN1 in HCC remains to be further investigated.
468

CONCLUSIONS

470   In summary, we screened 409 HCC-associated DEGs, of which 142 and 267 were up- and down-
471   regulated, respectively. We identified nine core genes, including BIRC5, DLGAP5, DTL, FEN1,
472   KIAA0101, KIF4A, MCM2, MKI67, and RFC4, that were up-regulated in HCC and that may
473   play important roles in the development or progression of this cancer. We further confirmed the
474   expression of FEN1 in HCC, suggesting that FEN1 may have potential as a new biomarker for
475   HCC diagnosis, treatment, or prognosis determination. Our results offer significant
476   improvements to current understanding of HCC pathogenesis. Additional studies will be needed
477   to validate our findings, and to confirm whether and how FEN1 regulates HCC.

478
ACKNOWLEDGEMENTS
480 None.

481

REFERENCES

Abdel-Fatah TM, Russell R, Albarakati N, Maloney DJ, Dorjsuren D, Rueda OM, Moseley P, Mohan V, Sun H, Abbotts R, Mukherjee A, Agarwal D, Illuzzi JL, Jadhav A, Simeonov A, Ball G, Chan S, Caldas C, Ellis IO, Wilson DM, 3rd, and Madhusudan S. 2014. Genomic and protein expression analysis reveals flap endonuclease 1 (FEN1) as a key biomarker in breast and ovarian cancer. *Mol Oncol* 8:1326-1338. DOI 10.1016/j.molonc.2014.04.009.

B S, J H, B B, M S, M H, ZG H, and M S. 2011. Definition of the landscape of promoter DNA hypomethylation in liver cancer. *Cancer Res* 71:5891-5903. DOI 10.1158/0008-5472.CAN-10-3823.

Becker JR, Gallo D, Leung W, Croissant T, Thu YM, Nguyen HD, Starr TK, Brown GW, and Bielinsky AK. 2018. Flap endonuclease overexpression drives genome instability and DNA damage hypersensitivity in a PCNA-dependent manner. *Nucleic Acids Res* 46:5634-5650. DOI 10.1093/nar/gky313.

C K, MD M, F V, K Y, B N, C L, M X, Q Z, JF M, MA W, MDM L, CA M, JS W, MJ W, MC W, TJ L, RK W, BJ R, and L D. 2013. Mutational landscape and significance across 12 major cancer types. *Nature* 502:333-339. DOI 10.1038/nature12634.

C S. 2016. Survivin in survival of hepatocellular carcinoma. *Cancer Lett* 379:184-190. DOI 10.1016/j.canlet.2015.06.016.

C T, A S, and RS H. 2018. Cancer genetics, precision prevention and a call to action. *Nat Genet* 50:1212-1218. DOI 10.1038/s41588-018-0202-0.

Cai C, Wang W, and Tu Z. 2019. Aberrantly DNA Methylated-Differentially Expressed Genes and Pathways in Hepatocellular Carcinoma. *J Cancer* 10:355-366. DOI 10.7150/jca.27832.

Chang ML, Lin SM, and Yeh CT. 2011. HURP expression-assisted risk scores identify prognosis distinguishable subgroups in early stage liver cancer. *PLoS One* 6:e26323. DOI 10.1371/journal.pone.0026323.

Chen CF, Hsu EC, Lin KT, Tu PH, Chang HW, Lin CH, Chen YJ, Gu DL, Lin CH, Wu JY, Chen YT, Hsu MT, and Jou YS. 2010. Overlapping high-resolution copy number alterations in cancer genomes identified putative cancer genes in hepatocellular carcinoma. *Hepatology* 52:1690-1701. DOI 10.1002/hep.23847.

Chen Y, Weng C, Zhang H, Sun J, and Yuan Y. 2018a. A Direct Interaction Between P53-Binding Protein 1 and Minichromosome Maintenance Complex in Hepg2 Cells. *Cell Physiol Biochem* 47:2350-2359. DOI 10.1159/000491607.

Chen YC, Chen IS, Huang GJ, Kang CH, Wang KC, Tsao MJ, and Pan HW. 2018b. Targeting DTL induces cell cycle arrest and senescence and suppresses cell growth and colony formation through TPX2 inhibition in human hepatocellular carcinoma cells. *Onco Targets Ther* 11:1601-1616. DOI 10.2147/ott.s147453.

CL C, YS T, YH H, YJ L, YY S, CW S, GY C, YC Y, YS C, JT H, and JC W. 2018. Lymphoid Enhancer Factor 1 Contributes to Hepatocellular Carcinoma Progression Through Transcriptional Regulation of Epithelial-Mesenchymal Transition Regulators and

523   Stemness Genes. *Hepatol Commun* 2:1392-1407. DOI 10.1002/hep4.1229.

524   Forner A, Reig M, and Bruix J. 2018. Hepatocellular carcinoma. *Lancet* 391:1301-1314. DOI
525   10.1016/s0140-6736(18)30010-2.

526   G DO, and M C. 2019. Mutant p53 and Cellular Stress Pathways: A Criminal Alliance That
527   Promotes Cancer Progression. *Cancers (Basel)* 11: E614. DOI 10.3390/cancers11050614.

528   He L, Luo L, Zhu H, Yang H, Zhang Y, Wu H, Sun H, Jiang F, Kathera CS, Liu L, Zhuang Z,
529   Chen H, Pan F, Hu Z, Zhang J, and Guo Z. 2017. FEN1 promotes tumor progression and
530   confers cisplatin resistance in non-small-cell lung cancer. *Mol Oncol* 11:640-654. DOI
531   10.1002/1878-0261.12058.

532   Hou G, Dong C, Dong Z, Liu G, Xu H, Chen L, Liu L, Wang H, and Zhou W. 2017. Upregulate
533   KIF4A Enhances Proliferation, Invasion of Hepatocellular Carcinoma and Indicates poor
534   prognosis Across Human Cancer Types. *Sci Rep* 7:4148. DOI 10.1038/s41598-017-
535   04176-9.

536   Huang Y, Wang H, Lian Y, Wu X, Zhou L, Wang J, Deng M, and Huang Y. 2018. Upregulation
537   of kinesin family member 4A enhanced cell proliferation via activation of Akt signaling
538   and predicted a poor prognosis in hepatocellular carcinoma. *Cell Death Dis* 9:141. DOI
539   10.1038/s41419-017-0114-4.

540   IO N, J N, EC L, ST F, and M N. 1995. Ki-67 antigen expression in hepatocellular carcinoma
541   using monoclonal antibody MIB1. A comparison with proliferating cell nuclear antigen.
542   *Am J Clin Pathol* 104:313-318. DOI 10.1093/ajcp/104.3.313.

543   IR K, O F, G H, E vM, and MV K. 2017. What is precision medicine? *Eur Respir J* 50: 1700391.
544   DOI 10.1183/13993003.00391-2017.

545   Isohookana J, Haapasaari KM, Soini Y, Leppanen J, and Karihtala P. 2018. Proteins of the
546   retinoblastoma pathway, FEN1 and MGMT are novel potential prognostic biomarkers in
547   pancreatic adenocarcinoma. *Pathol Res Pract* 214:840-847. DOI
548   10.1016/j.prp.2018.04.016.

549   J Z, and CB T. 2019. Metabolic regulation of cell growth and proliferation. *Nat Rev Mol Cell
550   Biol* 20:436-450. DOI 10.1038/s41580-019-0123-5.

551   JC H, WW S, HP T, KC H, CM Y, CJ C, HC T, CT H, GS S, JG C, KT Y, and TC L. 2015. The
552   Overexpression of FEN1 and RAD54B May Act as Independent Prognostic Factors of
553   Lung Adenocarcinoma. *PLoS One* 10:e0139435. DOI 10.1371/journal.pone.0139435.

554   K S, M K, S T, T M, T M, and K S. 1995. A long-term survivor of ruptured hepatocellular
555   carcinoma after hepatic resection. *J Gastroenterol Hepatol* 10:351-354.

556   K W, C X, and D C. 2014. Flap endonuclease 1 is a promising candidate biomarker in gastric
557   cancer and is involved in cell proliferation and apoptosis. *Int J Mol Med* 33:1268-1274.
558   DOI 10.3892/ijmm.2014.1682.

559   Kuo TC, Chang PY, Huang SF, Chou CK, and Chao CC. 2012. Knockdown of HURP inhibits
560   the proliferation of hepacellular carcinoma cells via downregulation of gankyrin and
561   accumulation of p53. *Biochem Pharmacol* 83:758-768. DOI 10.1016/j.bcp.2011.12.034.

562   L H, L L, H Z, H Y, Y Z, H W, H S, F J, CS K, L L, Z Z, H C, F P, Z H, J Z, and Z G. 2017.
563   FEN1 promotes tumor progression and confers cisplatin resistance in non-small-cell lung
564   cancer. *Mol Oncol* 11:640-654. DOI 10.1002/1878-0261.12058.

565   L Z, J J, LD F, Z G, C Z, and B S. 2011. Functional regulation of FEN1 nuclease and its link to
566   cancer. *Nucleic Acids Res* 39:781-794. DOI 10.1093/nar/gkq884.

567   Lam JS, Seligson DB, Yu H, Li A, Eeva M, Pantuck AJ, Zeng G, Horvath S, and Belldegrun AS.
568   2006. Flap endonuclease 1 is overexpressed in prostate cancer and is associated with a

569   high Gleason score. *BJU Int* 98:445-451. DOI 10.1111/j.1464-410X.2006.06224.x.

570   Liao W, Liu W, Yuan Q, Liu X, Ou Y, He S, Yuan S, Qin L, Chen Q, Nong K, Mei M, and
571   Huang J. 2013. Silencing of DLGAP5 by siRNA significantly inhibits the proliferation
572   and invasion of hepatocellular carcinoma cells. *PLoS One* 8:e80789. DOI
573   10.1371/journal.pone.0080789.

574   Lieber MR. 1997. The FEN-1 family of structure-specific nucleases in eukaryotic DNA
575   replication, recombination and repair. *Bioessays* 19:233-240. DOI
576   10.1002/bies.950190309.

577   Liu L, Chen X, Xie S, Zhang C, Qiu Z, and Zhu F. 2012a. Variant 1 of KIAA0101,
578   overexpressed in hepatocellular carcinoma, prevents doxorubicin-induced apoptosis by
579   inhibiting p53 activation. *Hepatology* 56:1760-1769. DOI 10.1002/hep.25834.

580   Liu L, Zhou C, Zhou L, Peng L, Li D, Zhang X, Zhou M, Kuang P, Yuan Q, Song X, and Yang
581   M. 2012b. Functional FEN1 genetic variants contribute to risk of hepatocellular
582   carcinoma, esophageal cancer, gastric cancer and colorectal cancer. *Carcinogenesis*
583   33:119-123. DOI 10.1093/carcin/bgr250.

584   Liu Z, Li J, Chen J, Shan Q, Dai H, Xie H, Zhou L, Xu X, and Zheng S. 2018. MCM family in
585   HCC: MCM6 indicates adverse tumor features and poor outcomes and promotes S/G2
586   cell cycle progression. *BMC Cancer* 18:200. DOI 10.1186/s12885-018-4056-8.

587   M A, N K, N I, A H, K H, O M, and M Y. 2009. The knockdown of endogenous replication
588   factor C4 decreases the growth and enhances the chemosensitivity of hepatocellular
589   carcinoma cells. *Liver Int* 29:55-62. DOI 10.1111/j.1478-3231.2008.01792.x.

590   M L, and J Z-R. 2016. Mechanisms of HBV-induced hepatocellular carcinoma. *J Hepatol*
591   64:S84-S101. DOI 10.1016/j.jhep.2016.02.021.

592   MJ D, N OD, DJ B, WM G, and BM R. 2007. Survivin: a promising tumor biomarker. *Cancer*
593   *Lett* 249:49-60. DOI 10.1016/j.canlet.2006.12.020.

594   MP K, Y Z, and G L. 2015. Mutant p53: Multiple Mechanisms Define Biologic Activity in
595   Cancer. *Front Oncol* 5:249. DOI 10.3389/fonc.2015.00249.

596   Nikolova T, Christmann M, and Kaina B. 2009. FEN1 is overexpressed in testis, lung and brain
597   tumors. *Anticancer Res* 29:2453-2459.

598   P H, Y W, A F, X L, V M, T O, YJ C, KE S, JM S, H Y, RL B, S G, H Y, KJ K, J Y, C J, Y L, D
599   F, R S, K N, E S, L A, JJ Z, K P, DG A, C L, and P S. 2017. Cell-Cycle-Targeting
600   MicroRNAs as Therapeutic Tools against Refractory Cancers. *Cancer cell* 31:576-
601   590.e578. DOI 10.1016/j.ccell.2017.03.004.

602   Pan HW, Chou HY, Liu SH, Peng SY, Liu CL, and Hsu HC. 2006. Role of L2DTL, cell cycle-
603   regulated nuclear and centrosome protein, in aggressive hepatocellular carcinoma. *Cell*
604   *Cycle* 5:2676-2687. DOI 10.4161/cc.5.22.3500.

605   PN R. 2012. Deep phenotyping for precision medicine. *Hum Mutat* 33:777-780. DOI
606   10.1002/humu.22080.

607   R N, H J, CJ T, and F C. 2019. Review: Precision medicine and driver mutations: Computational
608   methods, functional assays and conformational principles for interpreting cancer drivers.
609   *PLoS Comput Biol* 15:e1006658. DOI 10.1371/journal.pcbi.1006658.

610   S R, EL L, A B, Y C, X Z, J J, R W, HL J, QH Y, LX Q, ZY T, P H, KW H, SS T, PS M, and
611   XW W. 2012. Integrative genomic identification of genes on 8p associated with
612   hepatocellular carcinoma progression and patient survival. *Gastroenterology* 142:957-
613   966.e912. DOI 10.1053/j.gastro.2011.12.039.

614   S R, HL J, A B, M F, QH Y, JS L, SS T, Z S, ZY T, LX Q, and XW W. 2010. A unique

615   metastasis gene signature enables prediction of tumor relapse in early-stage
616   hepatocellular carcinoma patients. *Cancer Res* 70:10202-10212. DOI 10.1158/0008-
617   5472.CAN-10-2607.
618 Sang L, Wang XM, Xu DY, and Zhao WJ. 2018. Bioinformatics analysis of aberrantly
619   methylated-differentially expressed genes and pathways in hepatocellular carcinoma.
620   *World J Gastroenterol* 24:2605-2616. DOI 10.3748/wjg.v24.i24.2605.
621 Shen S, Kong J, Qiu Y, Yang X, Wang W, and Yan L. 2019. Identification of core genes and
622   outcomes in hepatocellular carcinoma by bioinformatics analysis. *J Cell Biochem*
623   120:10069-10081. DOI 10.1002/jcb.28290.
624 T M, K T, E A, T M, K S, H H, K S, and M T. 1995. Small hepatocellular carcinoma of single
625   nodular type: a specific reference to its surrounding cancerous area undetected
626   radiologically and macroscopically. *J Surg Oncol* 60:75-79.
627 T S, Y I, K T, T S, T M, K T, Y H, M N, T I, T F, M O, K E, T U, Y N, S O, M O, T K, R I, Y T,
628   R O, D Y, K T, T F, J P, MS L, T U, T O, T M, T W, H K, T O, T N, H M, S W, T S, M
629   Y, K T, and M K. 2016. p62/Sqstm1 promotes malignancy of HCV-positive
630   hepatocellular carcinoma through Nrf2-dependent metabolic reprogramming. *Nat*
631   *Commun* 7:12030. DOI 10.1038/ncomms12030.
632 Tsou AP, Yang CW, Huang CY, Yu RC, Lee YC, Chang CW, Chen BR, Chung YF, Fann MJ,
633   Chi CW, Chiu JH, and Chou CK. 2003. Identification of a novel cell cycle regulated gene,
634   HURP, overexpressed in human hepatocellular carcinoma. *Oncogene* 22:298-307. DOI
635   10.1038/sj.onc.1206129.
636 W C, W D, J W, Z W, and X H. 2019. Elevated Expressions of Survivin and Endoglin in Patients
637   with Hepatic Carcinoma. *Cancer Biother Radiopharm* 34:7-12. DOI
638   10.1089/cbr.2018.2539.
639 Wang K, Xie C, and Chen D. 2014. Flap endonuclease 1 is a promising candidate biomarker in
640   gastric cancer and is involved in cell proliferation and apoptosis. *Int J Mol Med* 33:1268-
641   1274. DOI 10.3892/ijmm.2014.1682.
642 WG Z, CF L, M L, XF C, K S, X K, L L, and ZC M. 2016. Aquaporin 9 is down-regulated in
643   hepatocellular carcinoma and its over-expression suppresses hepatoma cell invasion
644   through inhibiting epithelial-to-mesenchymal transition. *Cancer Lett.* 378:111-119. DOI
645   10.1016/j.canlet.2016.05.021.
646 Wu X, Wilson TE, and Lieber MR. 1999. A role for FEN-1 in nonhomologous DNA end joining:
647   the order of strand annealing and nucleolytic processing events. *Proc Natl Acad Sci U S A*
648   96:1303-1308. DOI 10.1073/pnas.96.4.1303.
649 X Z, X Q, C Z, L X, K H, Y L, N Z, J F, S S, L Z, J X, Z G, Y T, and X C. 2019. FEN1 mediates
650   miR-200a methylation and promotes breast cancer cell growth MET and EGFR signaling.
651   *FASEB J* 2:fj201900273R. DOI 10.1096/fj.201900273R.
652 Y L, F R, Y L, Z S, Z T, H X, Y D, and G C. 2015. Clinicopathological and prognostic
653   significance of high Ki-67 labeling index in hepatocellular carcinoma patients: a meta-
654   analysis. *Int J Clin Exp Med* 8:10235-10247.
655 YH W, TY C, TY C, KM C, VP C, and KJ K. 2014. Plasmalemmal Vesicle Associated Protein
656   (PLVAP) as a therapeutic target for treatment of hepatocellular carcinoma. *BMC Cancer*
657   14:815. DOI 10.1186/1471-2407-14-815.
658 Yuan RH, Jeng YM, Pan HW, Hu FC, Lai PL, Lee PH, and Hsu HC. 2007. Overexpression of
659   KIAA0101 predicts high stage, early tumor recurrence, and poor prognosis of
660   hepatocellular carcinoma. *Clin Cancer Res* 13:5368-5376. DOI 10.1158/1078-0432.ccr-

661   07-1113.
662 Zhang K, Keymeulen S, Nelson R, Tong TR, Yuan YC, Yun X, Liu Z, Lopez J, Raz DJ, and
663   Kim JY. 2018. Overexpression of Flap Endonuclease 1 Correlates with Enhanced
664   Proliferation and Poor Prognosis of Non-Small-Cell Lung Cancer. *Am J Pathol* 188:242-
665   251. DOI 10.1016/j.ajpath.2017.09.011.

Figure 1

Venn diagram.

(A) 409 DEGs were identified in 4 datasets (GSE14520, GSE29721, GSE45267, and GSE60502) via FUNRICH. These included 142 upregulated genes. (B) and 267 downregulated genes. (C) Colors correspond to specific datasets, with intersecting areas indicating overlapping gene sets. DEG identification criteria were adj. P < 0.05 and |log FC (fold change)| ≥1.

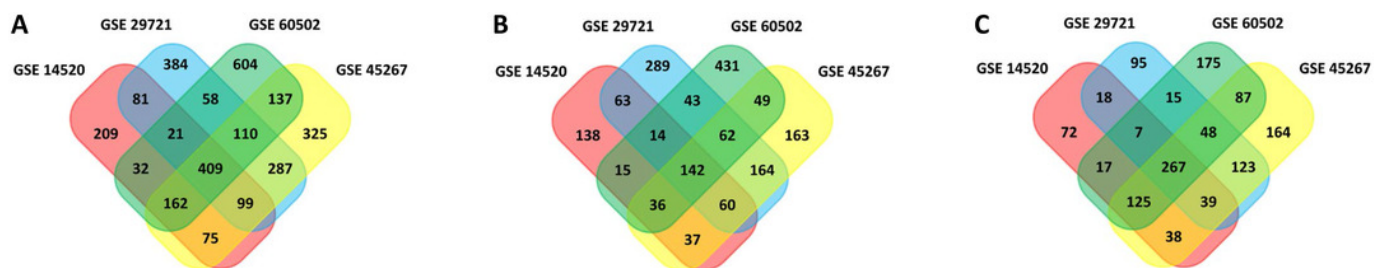



Figure 2

DEG PPI network and modular analysis.

(A) STRING was employed to create a PPI network of 403 nodes and 3502 edges, visualized using Cytoscape software. Genes that are upregulated are shown by red nodes, while those that are downregulated are blue. The MCODE plug-in was used to analyze highlighted regions. (B) Clustering module 1 scored 58.492 with 62 nodes and 1784 edges. (C) Clustering module 2 scored 11.529 with 18 nodes and 98 edges. (D) Clustering module 3 scored 10.917 with 25 nodes and 131 edges.

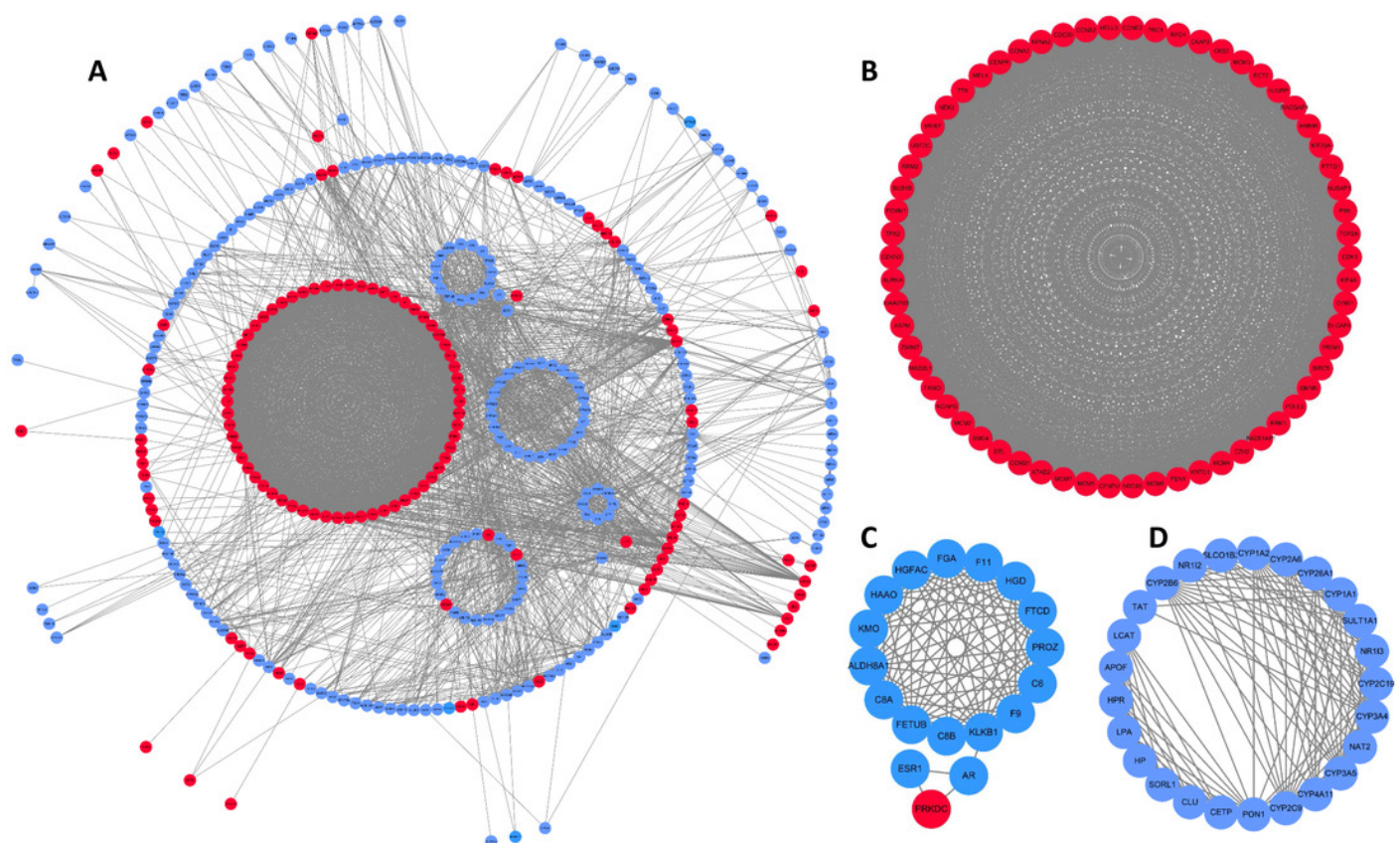



Figure 3

Identification of nine core HCC-associated genes.

(A) Nine core genes in clustering module 1 screened by 12 algorithms using the cytoHubba plug-in intersected with Hub genes. (B) Core gene expression profiles in 421 TCGA liver cancer tissues, including 50 solid normal tissues and 371 primary tumors, **P < 0.01. (C) The UCSC databased was used for core gene hierarchical clustering.

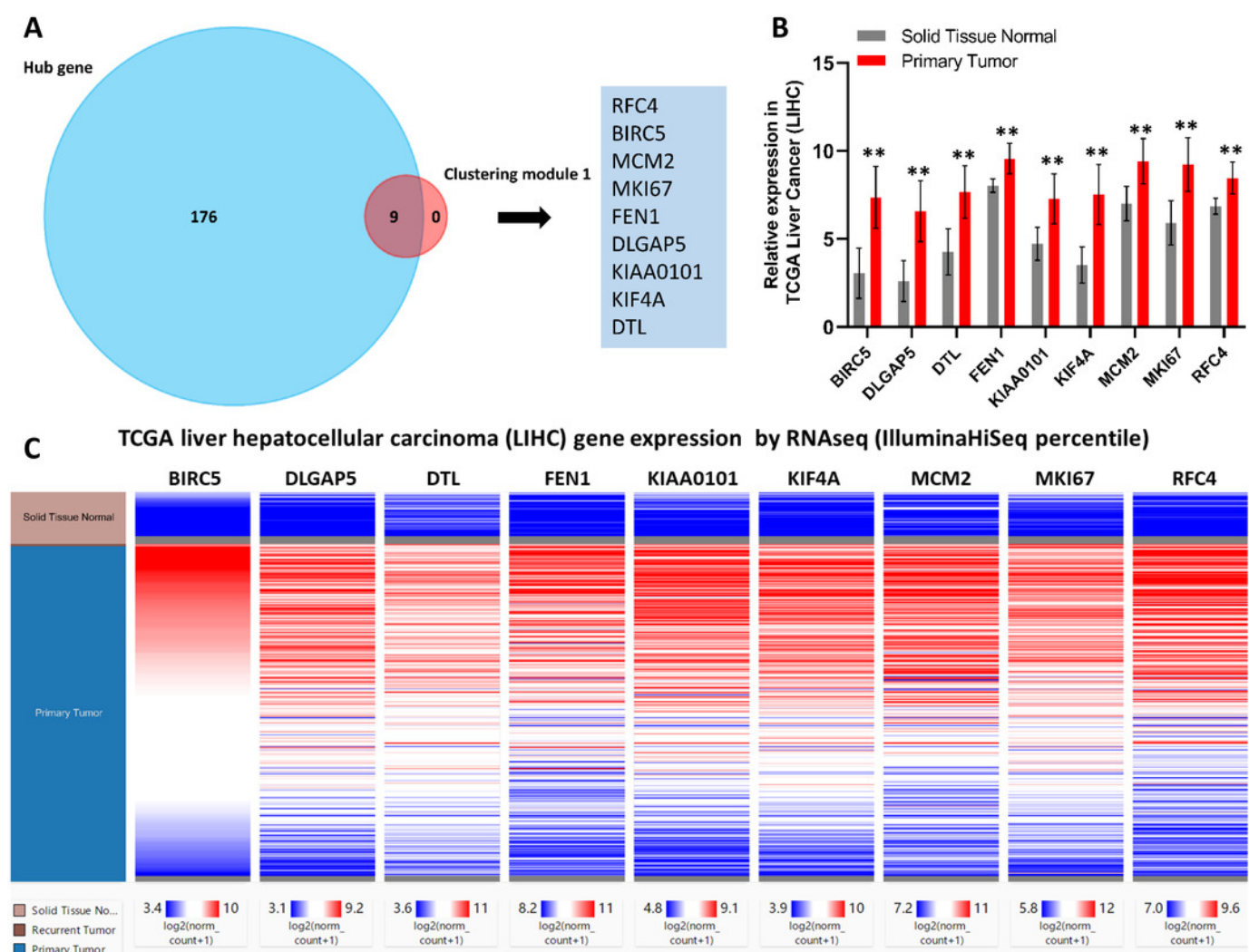



Figure 4

Association between core genes and overall survival.

Association between core genes and overall survival (A, BIRC5; B, DLGAP5; C, DTL; D, FEN1; E, KIAA0101; F, KIF4A; G, MCM2; H, MKI67 and I, RFC4) in those with HCC. CI, confidence interval; HR, hazard ratio. High- and low-risk groups are shown in red and blue, respectively. P < 0.05 was the significance threshold.

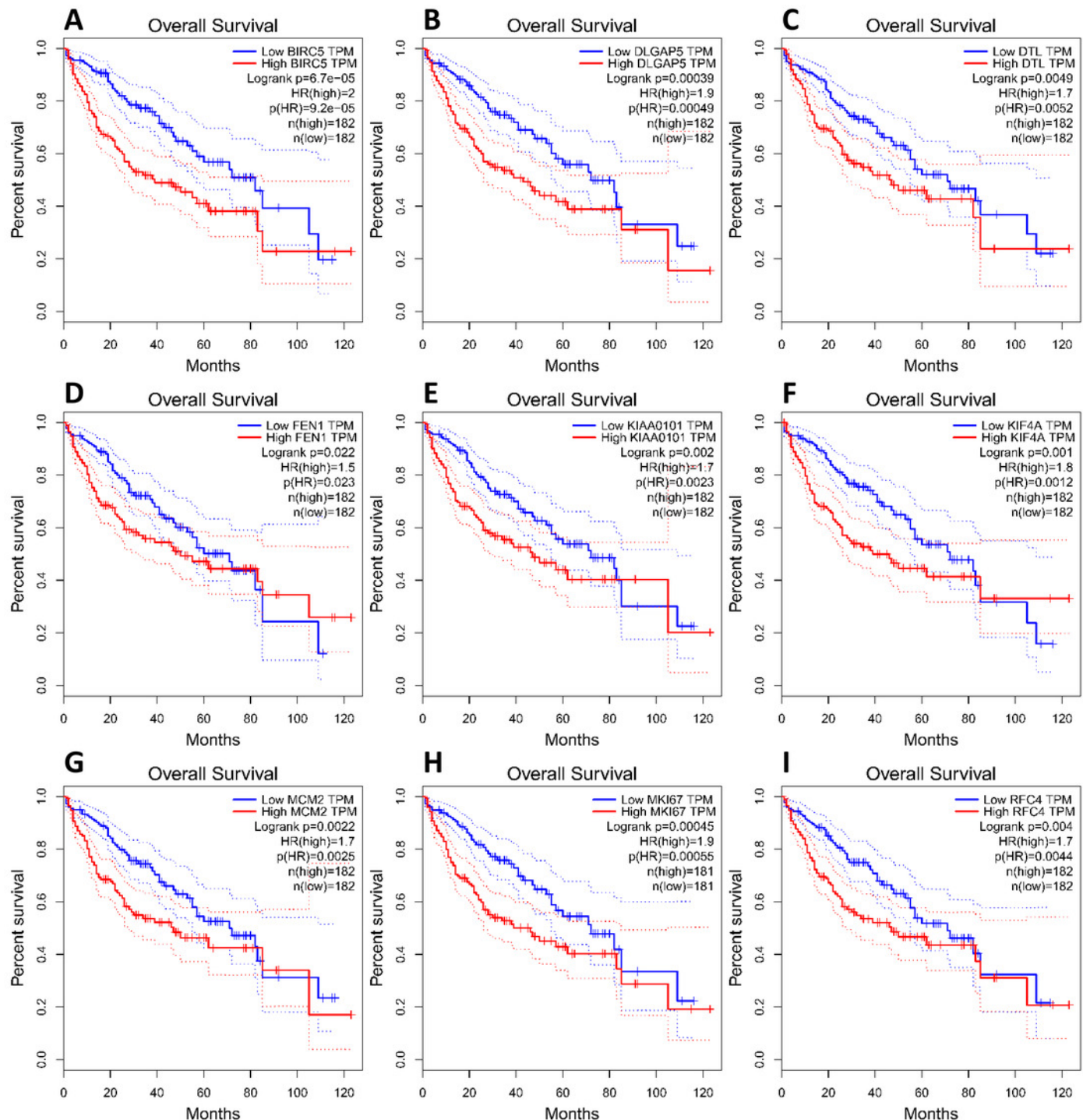



Figure 5

Association between core genes and disease-free survival.

Association between core genes and disease-free survival (A, BIRC5; B, DLGAP5; C, DTL; D, FEN1; E, KIAA0101; F, KIF4A; G, MCM2; H, MKI67 and I, RFC4) in those with HCC. CI, confidence interval; HR, hazard ratio. High- and low-risk groups are shown in red and blue, respectively. P < 0.05 was the significance threshold.

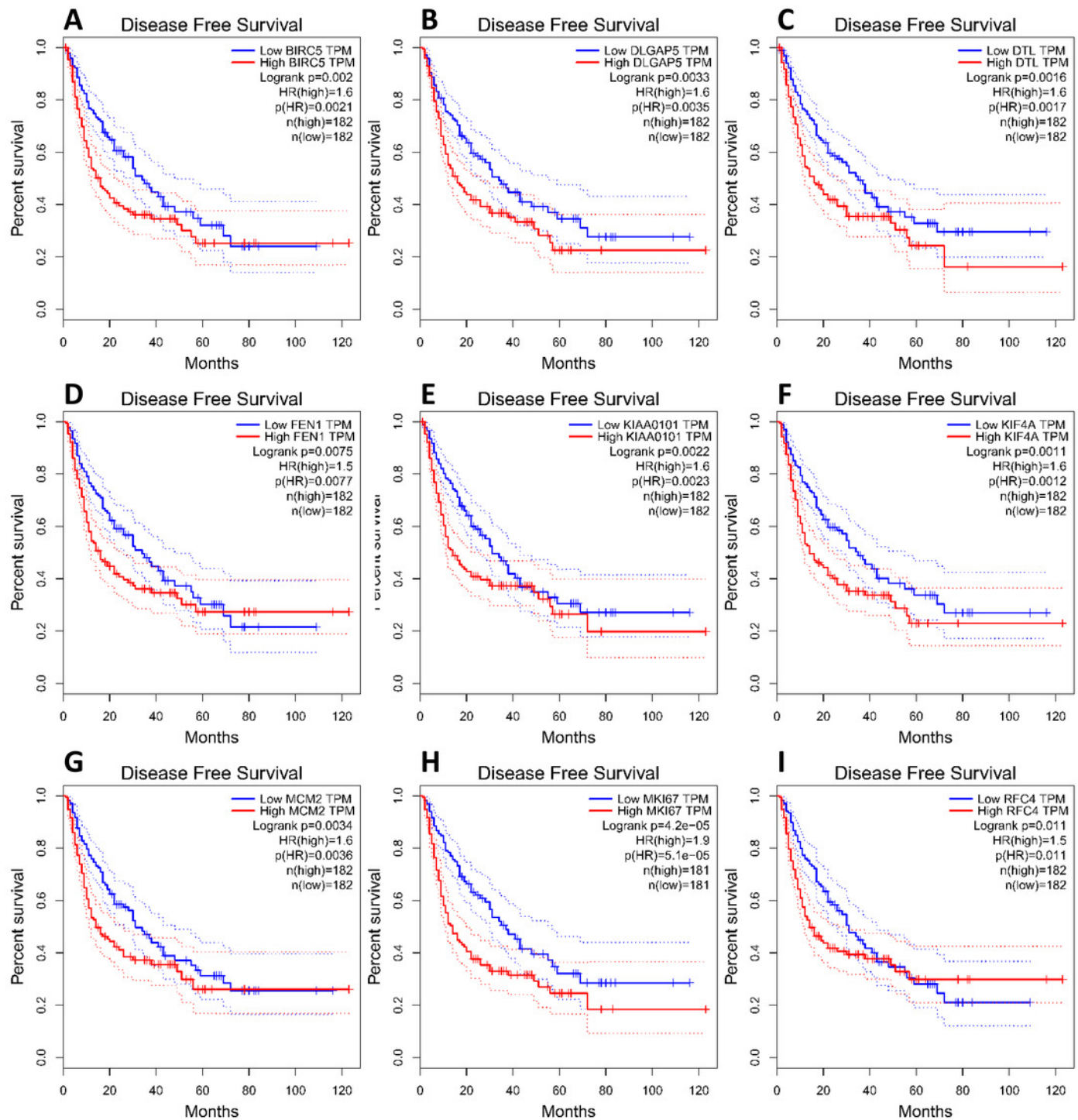



Figure 6

Correlations between the expression of FEN1 and MCM2, BIRC5, and RFC4.

(A) A PPI network for the nine core genes generated with the Cbioportal database; (B) The correlation between FEN1 and MCM2 in 421 TCGA liver cancer tissues including 50 solid normal tissues and 371 primary tumors; (C) The correlation between FEN1 and BIRC5 in 421 TCGA liver cancer tissues including 50 solid normal tissues and 371 primary tumors; (D) The correlation between FEN1 and RFC4 in 421 TCGA liver cancer tissues including 50 solid normal tissues and 371 primary tumors. P < 0.05 was the significance threshold.

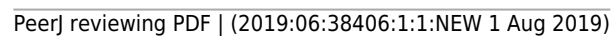



Figure 7

The expression of FEN1 in cancers including HCC.

The expression of FEN1 in various tumor tissue types was analyzed in the ONCOMINE (A) and FIREBROWSE databases (B); (C) The expression of FEN1 in three different HCC-related chip data sets was analyzed in the ONCOMINE database.

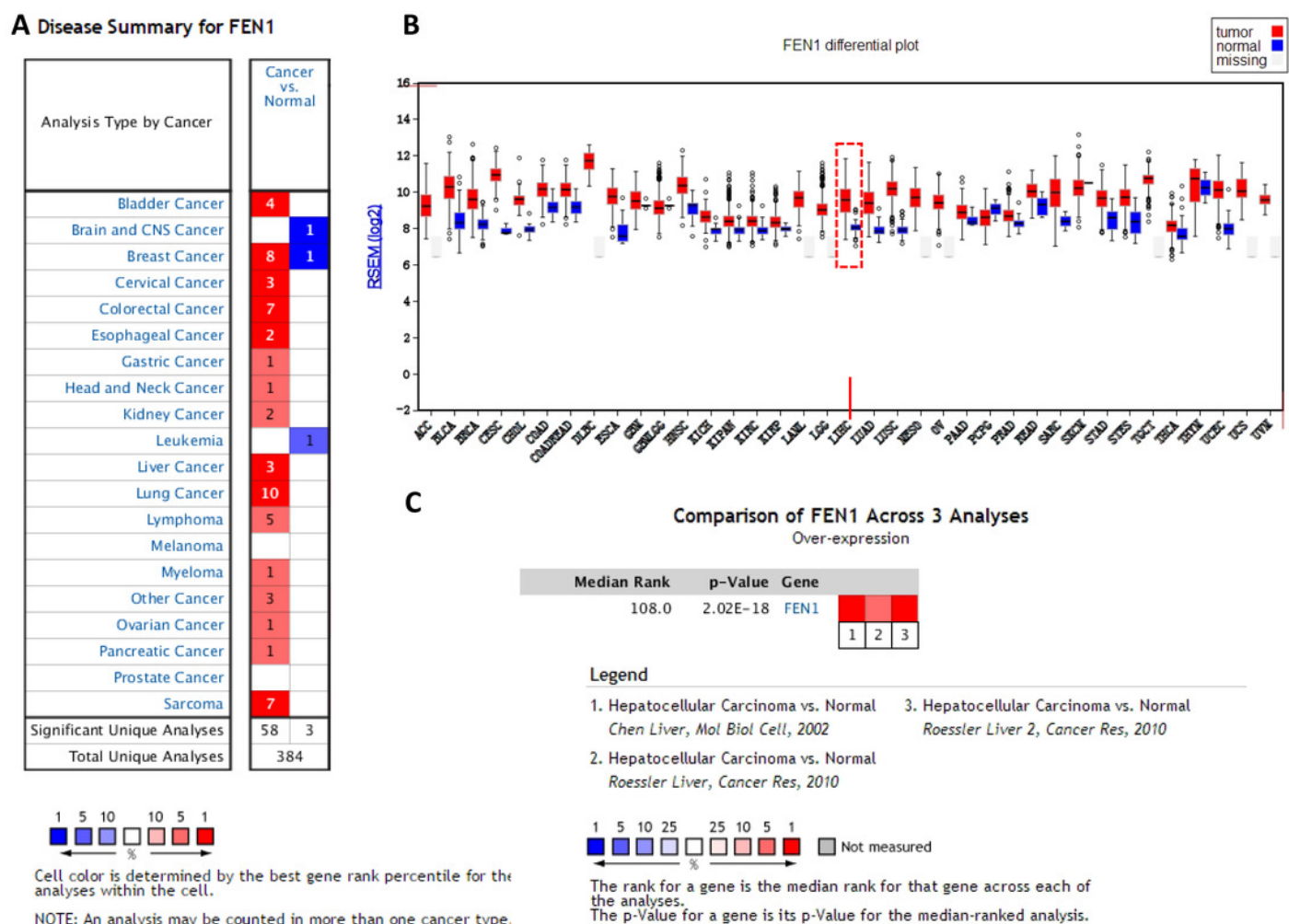
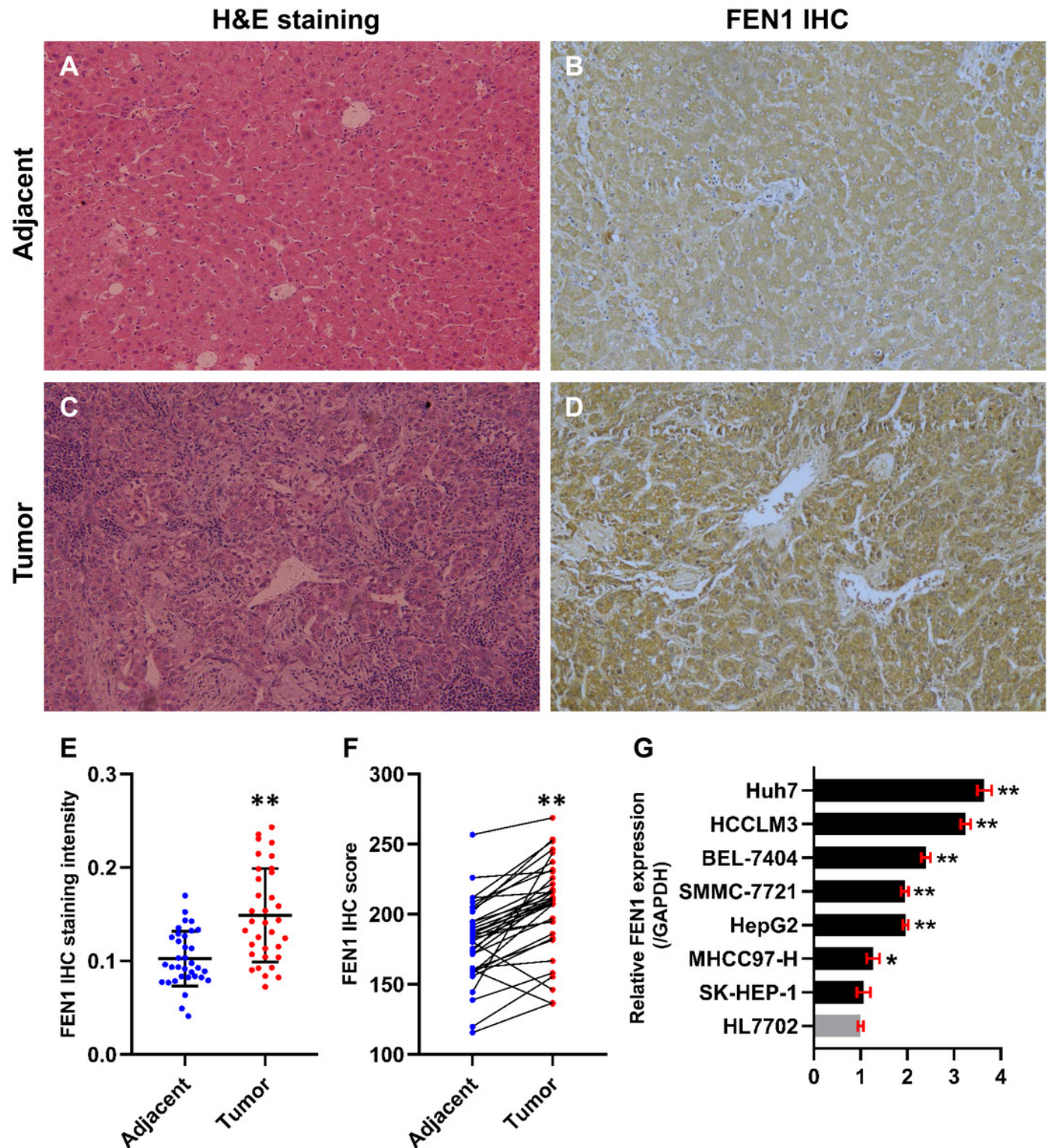



Figure 8

FEN1 was up-regulated in HCC tissues and hepatoma cell lines.

H&E staining of adjacent tissue (A) and HCC tissue (C). (B) FEN1 IHC in adjacent tissue. (D) FEN1 IHC in HCC tissue. (E) FEN1 IHC staining quantification; n=34/group. (F) FEN1 IHC scores; n=34/group. (G) FEN1 expression in control HL7702 liver cells and in hepatoma cell lines (as indicated); *P < 0.05, **P < 0.01.

H&E staining   **FEN1 IHC**

Detailed information of the GEO datasets in this study.

1   **Table.1** Detailed information of the GEO datasets in this study.

Series accession	Species	Type			Platform
GSE14520	Homo sapiens	Expression array	profiling	by	GPL3921 Affymetrix HT Human Genome U133A Array
GSE29721	Homo sapiens	Expression array	profiling	by	GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array
GSE45267	Homo sapiens	Expression array	profiling	by	GPL570 [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array
GSE60502	Homo sapiens	Expression array	profiling	by	GPL96 [HG-U133A] Affymetrix Human Genome U133A Array

2   GEO, Gene Expression Omnibus.

Correlation between FEN1 expression and clinicopathological features in 34 paired HCC patients

2 **Table 2** Correlation between FEN1 expression and clinicopathological features

3 in 34 paired HCC patients

Clinicopathological features	Cases (n=34)	FEN1 expression		P value
		High (%)	Low (%)	
All case (n=34)	34	25	9	
Gender				1.000
Male	29	21(61.8%)	8(23.5%)	
Female	5	4(11.8%)	1(2.9%)	
Age(y)[a]				0.697
<52	15	12(35.3%)	3(8.8%)	
≥52	19	13(38.2%)	6(17.6%)	
Tumor size				0.047*
<5cm	19	11(32.4%)	8(23.5%)	
≥5cm	15	14(41.2%)	1(2.9%)	
Tumor multiplicity				0.348
Single	27	21(61.8%)	6(17.6%)	
Multiple	7	4(11.8%)	3(8.8%)	
TNM stage				1.000
I ~ II	23	17(50.0%)	6(17.6%)	
III ~ IV	11	8(23.5%)	3(8.8%)	
Pathological grade				1.000
Well/moderate	27	20(58.8%)	7(20.6%)	
Poor	7	5(14.7%)	2(5.9%)	
Metastasis				0.013*
With	12	12(35.3%)	0(0.0%)	
Without	22	13(38.2%)	9(26.5%)	
HBsAg				1.000
Positive	30	22(64.7%)	8(23.5%)	
Negative	4	3(8.8%)	1(2.9%)	
Liver cirrhosis				0.687
With	12	8(23.5%)	4(11.8%)	
Without	22	17(50.0%)	5(14.7%)	
Serum AFP				0.725
＞ 400 ng/ml	18	12(35.3%)	6(17.6%)	
≤ 400 ng/ml	16	9(26.5%)	7(20.6%)	
Vascular invasion				0.024*
Present	14	13(38.2%)	1(2.9%)	
Absent	20	11(32.4%)	9(26.5%)	

4   a: patients were divided according to the median age; AFP: alpha-fetoprotein; HBsAg: hepatitis B surface

5 antigen; *: $P < 0.05$.