

Integrative analysis reveals novel driver genes and molecular subclasses of hepatocellular carcinoma

Liguang Yang^{1,2,*}, Zhengtao Zhang^{2,3,*}, Yidi Sun^{1,2}, Shichao Pang⁴, Qianlan Yao^{5,6}, Ping Lin^{1,2}, Jinming Cheng⁴, Jia Li¹, Guohui Ding^{1,7}, Lijian Hui³, Yixue Li^{1,2,4,8,9}, Hong Li^{1,2}

¹Bio-Med Big Data Center, CAS Key Laboratory of Computational Biology, CAS-MPG Partner Institute for Computational Biology, Shanghai Institute of Nutrition and Health, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China

²University of Chinese Academy of Sciences, Beijing 100049, China.

³State Key Laboratory of Cell Biology, CAS Center for Excellence in Molecular Cell Science, Shanghai Institute of Biochemistry and Cell Biology, Chinese Academy of Sciences, Shanghai 200031, China

⁴School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200240, China

⁵Department of Pathology, Fudan University Shanghai Cancer Center, Shanghai 200032, China

⁶Department of Oncology, Shanghai Medical College, Fudan University, Shanghai 200032, China

⁷Anhui Engineering Laboratory for Big Data of Precision Medicine, Anhui 234000, China

⁸Collaborative Innovation Center of Genetics and Development, Fudan University, Shanghai 200433, China

⁹Shanghai Center for Bioinformation Technology, Shanghai Academy of Science and Technology, Shanghai 201203, China

*Equal contribution

Correspondence to: Yixue Li, Hong Li; **email:** yxli@sibs.ac.cn, lihong01@sibs.ac.cn

Keywords: systematic integration, epigenome, stratifin, SPP1, liver cancer

Received: February 17, 2020

Accepted: August 25, 2020

Published: November 20, 2020

Copyright: © 2020 Yang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/3.0/) (CC BY 3.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

ABSTRACT

Hepatocellular carcinoma (HCC) is a heterogeneous disease with various genetic and epigenetic abnormalities. Previous studies of HCC driver genes were primarily based on frequency of mutations and copy number alterations. Here, we performed an integrative analysis of genomic and epigenomic data from 377 HCC patients to identify driver genes that regulate gene expression in HCC. This integrative approach has significant advantages over single-platform analyses for identifying cancer drivers. Using this approach, HCC tissues were divided into four subgroups, based on expression of the transcription factor E2F and the mutation status of TP53. HCC tissues with E2F overexpression and TP53 mutation had the highest cell cycle activity, indicating a synergistic effect of E2F and TP53. We found that overexpression of the identified driver genes, stratifin (SFN) and SPP1, correlates with tumor grade and poor survival in HCC and promotes HCC cell proliferation. These findings indicate SFN and SPP1 function as oncogenes in HCC and highlight the important role of enhancers in the regulation of gene expression in HCC.

INTRODUCTION

Hepatocellular carcinoma (HCC) is the second leading cause of cancer-related mortality worldwide [1]. In the United States, the incidence of HCC has been rising

faster than any other cancer type [2]. Common risk factors associated with HCC include hepatitis B or C infection, high alcohol consumption, autoimmune hepatitis, and metabolic diseases [3]. The intra-tumor and interpatient heterogeneities in HCC [4–6] make

early detection and effective therapies difficult, leading to the low five-year survival rates [7, 8]. Thus, it is crucial to understand the molecular mechanisms driving HCC.

Genome sequencing of HCC cohorts has revealed many genomic alterations, such as mutations of TP53, CTNNB1, and AXIN1, copy number alterations of CDKN2A/CDKN2B, and multiple alterations of TERT [9–14]. These genes are called drivers to distinguish them from randomly occurring passenger alterations. In addition to genetic drivers, epigenetic drivers may also play important roles in HCC [15–18]. While global DNA hypomethylation makes the genome unstable, promoter hypermethylation may silence tumor suppressor genes, such as SMPD3 and NEFH [19]. Since drivers contribute to tumorigenesis by different mechanisms, there is a need to identify the HCC drivers by integrative analysis of multiple omics data.

Current progress in computational algorithms demonstrates that integrative analysis of multi-omics data could increase our ability to identify likely drivers. For example, CONEXIC integrated matched copy number and gene expression data to identify the combination of modulators in melanoma [20]. Lemon-tree reconstructed module network and identified upstream regulators from multi-omics datasets [21]. Additionally, integrated analysis of multi-omics data may resolve cancer molecular classification with clinical relevance, and reveal previously unrecognized subgroups. The Cancer Genome Atlas (TCGA) study performed the first large scale multi-platform analysis of HCC [14]. They identified oncogenic drivers with significant mutations or copy number alterations, and found three subtypes by multi-platform integrative clustering. However, their integration approach did not consider the function effects of drivers. The shared data of TCGA project provide a valuable resource for us to improve the integration analysis. Since gene expression is an intermediate molecular phenotype that connects genome and phenotype, we hypothesize that alterations that regulate the expression of multiple genes are more likely to be the real drivers.

In this study, we performed an integrative analysis using five platform data from 377 HCC patients to identify driver genes that regulate gene expression in HCC. The integrative approach has obvious advantages over single-omics analysis. In addition, we utilized the driver events to classify HCC tissues into four subclasses having distinct prognostic implications and molecular characteristics. Furthermore, we used histone modification data to deepen our understanding of the regulation of gene expression in HCC.

RESULTS

Identification of HCC drivers by multi-omics integration

Multiple-omics data of 377 HCC tissues and 50 adjacent normal tissues were collected from TCGA project, including somatic mutations, copy number alterations, DNA methylation, and mRNA and miRNA expression. 9560 genes differentially expressed between cancer and normal tissues were divided into 241 co-expression modules by Gibbs sampling cluster algorithm. Potential upstream regulators were obtained from single-platform analysis or literature, including 69 mutated genes, 886 amplified and 1829 deleted genes, 570 differentially methylated genes, 196 differentially expressed microRNAs, 121 transcription factors, and 76 HCC associated genes (Supplementary Table 1, see Materials and Methods). These candidate regulators were assessed by an integrative multi-omics module network inference algorithm [21], which inferred the regulation relationship between co-expression modules and their associated regulators. High-scoring regulators were chosen as cancer drivers that might drive development of HCC. The final driver list was composed of 296 protein-coding genes and 88 microRNAs, of which 166 genes were not reported previously (Figure 1A, Supplementary Table 2, Supplementary Figure 1A).

To compare the differences between multi-omics and single-omics analyses, we utilized the same process, but only assigned candidate regulators from each single platform to identify the drivers. Only two significant drivers (TP53 and CTNNB1) regulating expression modules were identified from mutation data because of the low mutation frequency, but 172, 121, 55 and 46 drivers were identified from CNV, methylation, miRNA, and mRNA datasets, respectively. The multi-omics integrated analysis covered most of the drivers identified from single-omics analysis, but also identified new drivers (Figure 1B). Moreover, multi-omics integration identified drivers that simultaneously regulated multiple modules (Wilcoxon rank sum test $p = 0.0054$, Figure 1C). To investigate the association between driver genes and cancer hallmarks, we performed gene set enrichment analysis for drivers and found that the drivers generated from integrated analysis were significantly enriched in cancer hallmarks pathways (Chi-square test, $p = 1.194 \times 10^{-5}$ Figure 1D, Supplementary Figure 1B). These results indicate that the integrative approach can more accurately detect cancer drivers with functional implications.

Functions and characteristics of HCC drivers

Characteristics of the drivers generated from multi-omics integration are summarized in a circo plot

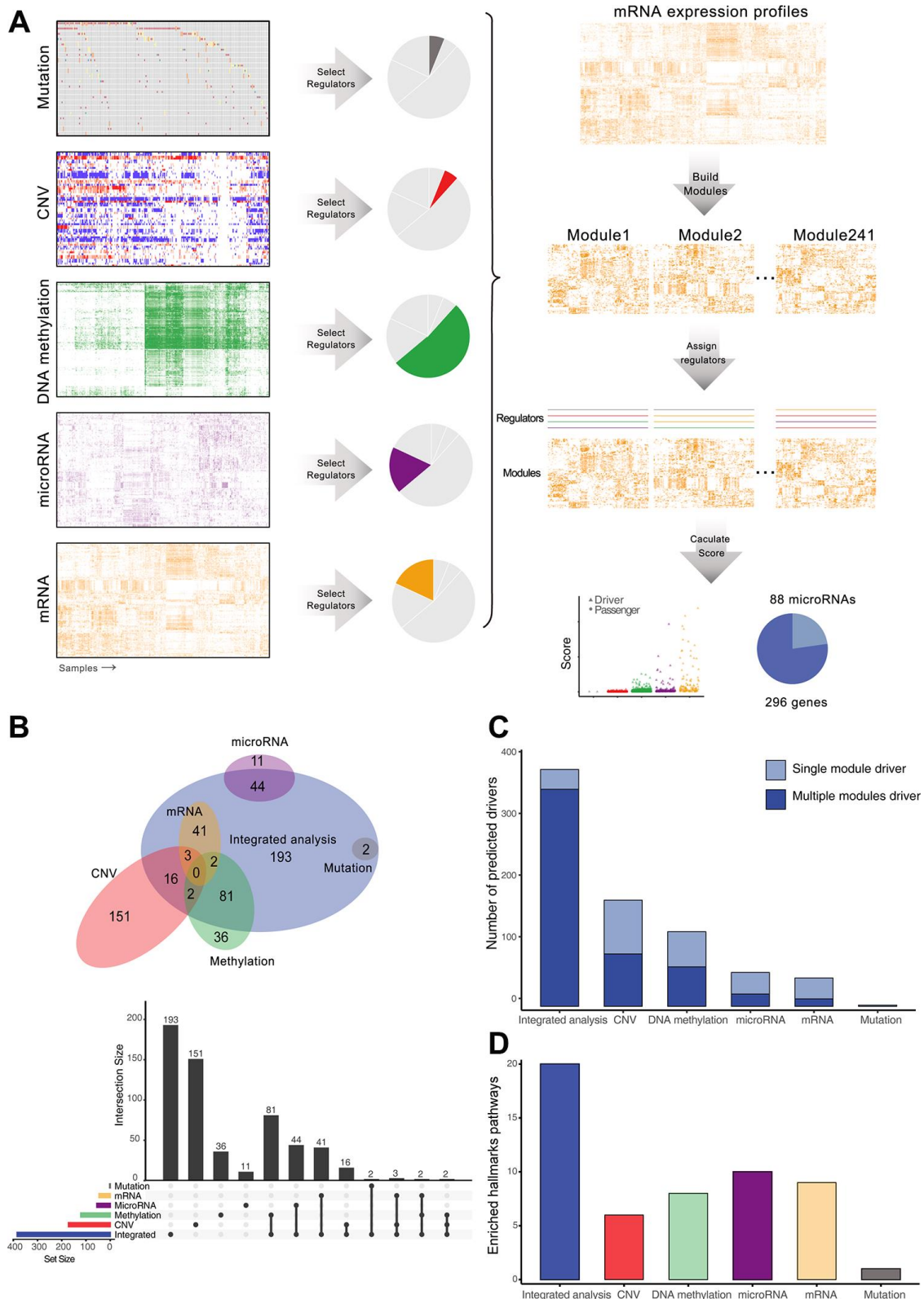


Figure 1. Identification and comparison of HCC drivers by integrative and single-platform approach. (A) Overview of the computational scheme for the systematic identification of drivers in HCC patients. Gray, red, green, purple, and orange represent mutations, copy number, DNA methylation, microRNA and mRNA, respectively. (B) Venn diagram and upset plot showing drivers recognized by different methods. (C) Regulation ability of drivers, measured by the number of modules. (D) Number of significantly enriched cancer hallmarks of driver genes.

(Figure 2A). Most of the drivers were down-regulated and hypermethylated in HCC tissues compared with normal tissues. Some drivers were enriched in copy number altered regions, such as chromosome 4q and 8q. We correctly identified known cancer genes, including 12 oncogenes, 42 tumor suppressor genes (TSGs), and 11 significantly mutated genes. We also found a number of novel driver genes that were not previously associated with HCC, including ANXA13, G6PC, STAT1, and JAK1. Some of them were significantly altered in at least two types of omics data, or were reported to participate in progression of other cancers [22–25].

Next, we performed a functional enrichment analysis to explore the biological function of these drivers. Tox analysis of these drivers showed that the most significantly associated toxicity phenotypes and clinical pathology endpoints included hepatocellular carcinoma, liver hyperplasia, and liver cirrhosis, corroborating our analysis (Supplementary Figure 2A). KEGG pathway analysis showed that the drivers were enriched not only in well-known pathways, such as cell cycle, WNT signaling, p53 signaling and HBV/HCV virus-related carcinogenesis, but also in several immune-related pathways, such as TGF-beta signaling, JAK-STAT signaling, and complement and coagulation cascades (Figure 2B). These pathways can be disrupted by different genes (Supplementary Figure 2B), and some genes have mutually exclusive mutation patterns (Supplementary Figure 2C, 2D); for example, CTNNB1, AXIN1 and APC in Wnt pathway, or RB1, CDKN1A and RBL1 in cell cycle pathway. Similar mutually exclusive patterns were observed in different alteration types of a single gene, such as a copy number deletion and mutation of AXIN1 (one-sided binomial tests $p = 0.001$). Such diverse alteration patterns may partly explain the heterogeneity of HCC.

Drivers have the potential to regulate expression of downstream genes. For example, genes in a 62-gene co-expression module were enriched in cell-cycle, and the predicted regulators were E2F1 expression, E2F2 expression, and TP53 mutation (Figure 2C). This is consistent with the existing knowledge that E2F transcription factor family and TP53 are crucial for cell cycle regulation. To explore the potential combination effect of E2F and TP53 on cell cycle, we classified HCC into four subgroups based on the expression of E2F1/E2F2 and the mutation status of TP53, and compared gene expression and cell cycle pathway scores among the four subgroups (Figure 2D–2E). Interestingly, HCC samples with E2F overexpression and TP53 mutation had the highest activity level of cell cycle pathway, indicating the synergistic effect of E2F and TP53.

Regulatory network of HCC drivers and experimental validation of typical driver genes

To understand connections between the identified drivers, we constructed literature-based regulatory networks using IPA. The functional categorization of the largest network revealed an important role in the regulation of migration and differentiation of tumor cells (Figure 3A). Some genes in this network have clear roles in HCC carcinogenesis, such as CTNNB1 and MYC. As for the rarely reported genes, we evaluated their expression changes in three independent datasets (GSE77314, GSE77509 and GSE97214), and analyzed their correlation with survival time and tumor grade. We found that the stratifin gene SFN was highly overexpressed in HCC (>8 fold in all datasets). Importantly, the increased expression of SFN and its decreased methylation correlated with the tumor grade (Figure 3B, 3C), and poor survival (Figure 3D).

SFN encodes 14-3-3 sigma protein, which regulates cell cycle and inhibits cysteine-type endopeptidase involved in apoptosis. To evaluate the role of SFN in HCC tumorigenesis, two independent sgRNAs were used to knockout SFN in two HCC cell lines (CLC7 and Huh1) by CRISPR/Cas9 system. Western blotting showed that SFN protein expression was efficiently suppressed in SFN knockout cell lines, indicating functional inactivation of SFN (Figure 3E). Compared to parental cell lines, the cell lines with suppressed SFN exhibited a markedly reduced proliferation (Figure 3F, 3G). Similar results can be observed with another highly expressed gene SPP1 (Supplementary Figure 3A–3D).

In order to further verify the reliability of the genes we obtained, we selected the top 50 genes with the highest integrative scores and upregulated in tumors for siRNA screening. Total 23 genes significantly inhibited cell proliferation after siRNA knockout, 17 of which were not reported in the literature (Supplementary Figure 3E). These experiments reveal the robustness of our results and overexpression of SFN and SPP1 in HCC promotes cancer cell proliferation.

Multiple-platform determinants of HCC subclasses

Previous cancer classification used highly variable genes to categorize patients into subclasses. The TCGA multi-platform classification used non-redundant copy number regions and the most variable CpG sites, mRNA, and miRNA for integrative clustering [14]. These approaches usually use a large number of genes, but the gene functions are not obvious. Since driver genes are master regulators of gene expression, we stratified HCC by using multi-platform data of driver genes. Samples were first clustered using each single.

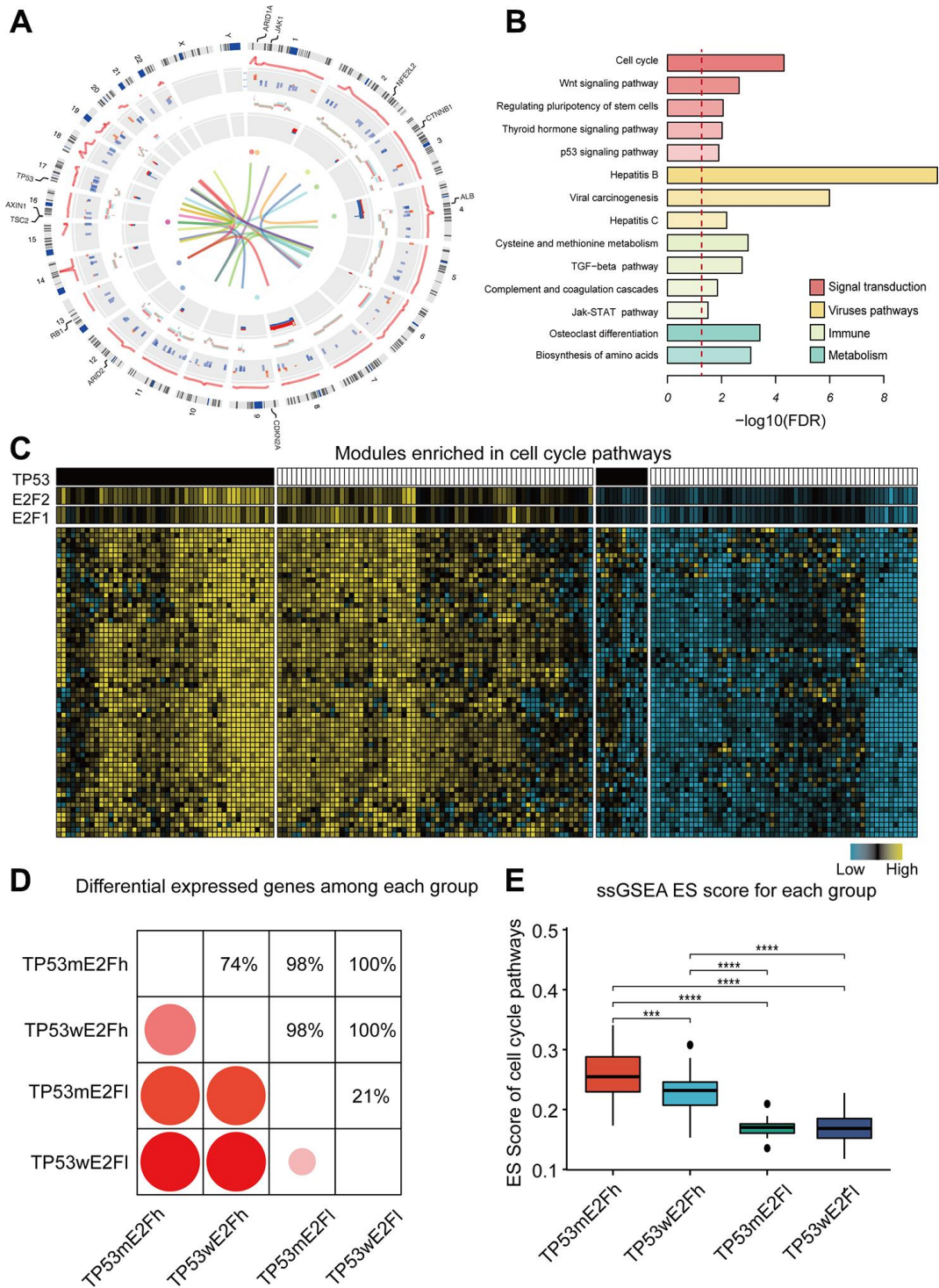


Figure 2. Functions and characteristics of HCC drivers. (A) Circos plot shows the alterations of 384 drivers. Circular tracks from outside to inside: regulation score, mRNA fold-change, copy number alteration, DNA methylation status, mutations, and protein interactions. The outermost labels indicate the significant mutated genes. Red represents high expression/methylation and CNV gain; blue represents low expression/methylation and CNV loss. Sex chromosomes are excluded. (B) Enriched KEGG pathways of the drivers; the dotted line indicates FDR=0.05. (C) Example of a co-expression module and its predicted regulators. Genes in this module are enriched in cell cycle pathway. HCC tissues were classified into four subgroups based on the combined alterations of E2F expression and TP53 mutation. (D) Comparison of the number of differentially expressed genes between any two subgroups. Only genes in the example module were used. (E) Comparison of the activity of cell cycle pathway among the four subgroups. ssGSEA was performed for each tumor using the genes in cell cycle pathway. Enrichment score (ES) represents the activity level of the pathway. TP53m and TP53w mean TP53 mutation and TP53 wildtype; E2Fh and E2Fl mean high and low expression of E2F. P value was determined by Wilcoxon rank sum test. ****P < 0.0001; ***P < 0.001; **P < 0.01; *P < 0.05.

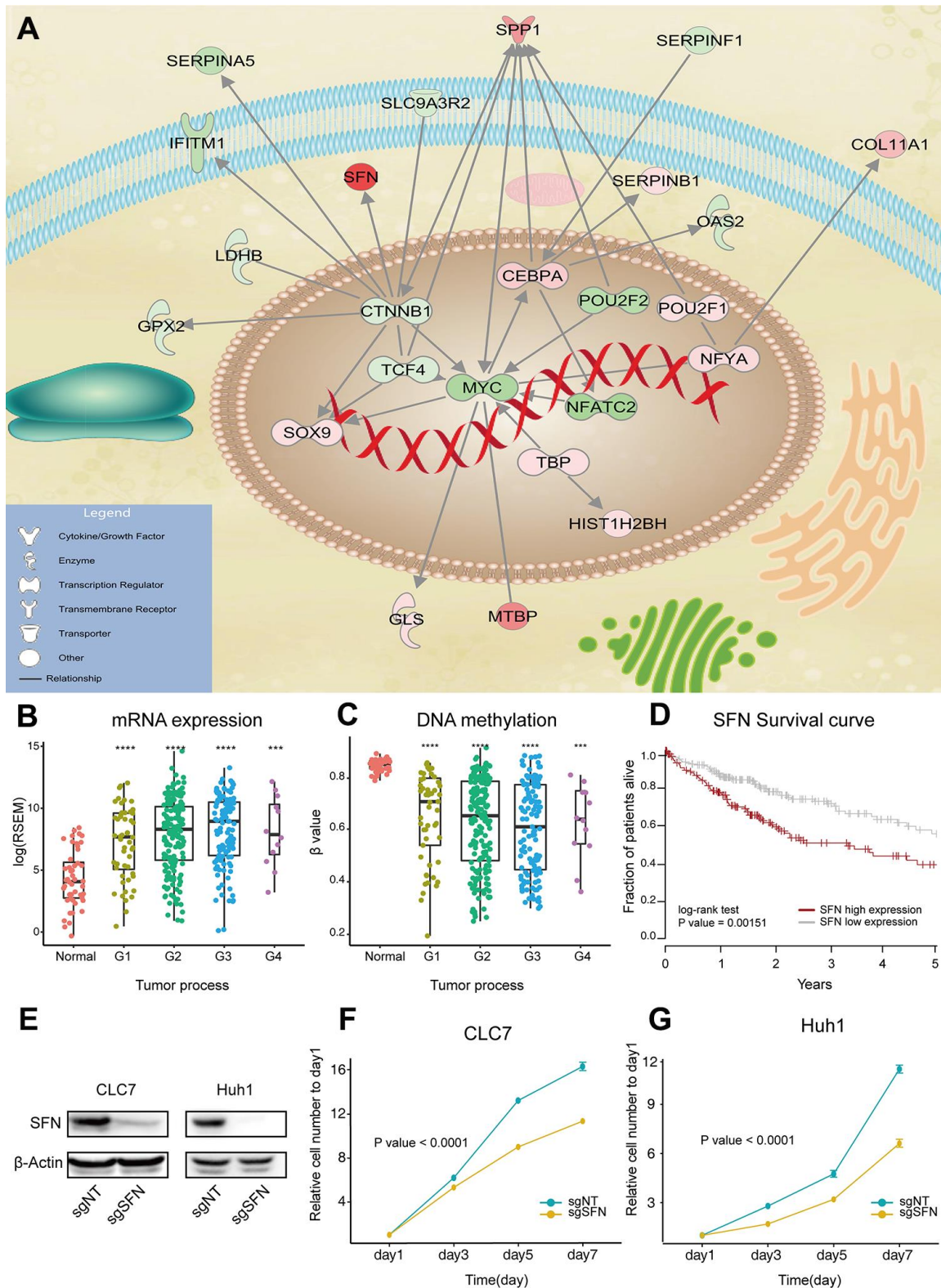


Figure 3. SFN functions as oncogene in HCC. (A) Max regulatory network of drivers, generated by IPA analysis. Edges represent direct regulatory relationships obtained from literature. Red and green nodes represent genes that are increased and decreased in HCC, respectively. (B, C) SFN expression and methylation correlate with HCC progression. P value was determined by Wilcoxon rank sum test: **** $P < 0.0001$; *** $P < 0.001$; ** $P < 0.01$; * $P < 0.05$. (D) Survival of HCC patients with high and low SFN expression. (E) SFN protein levels analyzed by western blotting in SFN knockout cells. (F, G) Proliferation of CLC7 and Huh1 cells transfected with sgRNAs targeting SFN ($n=3$, regression analysis).

platform, and then a cluster of cluster analysis (COCA) was performed to determine the integrative clusters [26]. This analysis revealed four robust HCC subclasses: C1 (n=128), C2 (n=22), C3 (n=72), and C4 (n=126) (Figure 4A). The COCA-subclasses displayed a higher similarity with mRNA subclasses compared with other platforms (Supplementary Figure 4A), similar to a previous study in oligodendroglial tumors²¹. To compare our results with the classic HCC classification, we applied Hoshida's [27] expression signatures to classify patients into three subclasses: S1 (n=112), S2 (n=74), and S3 (n=176). C1 predominantly consisted of Hoshida S3 tumors, whereas C4 predominantly consisted of Hoshida S1 tumors.

We observed significant survival differences among the subclasses identified by integrative analysis (Log-rank test $p = 0.03$). C2 tumors had significantly worse prognosis than other tumors, and C3 tumors had a better prognosis (Figure 4B). However, there was no

significant difference in overall survival using Hoshida's classification or single-platform classification in TCGA cohort (Supplementary Figure 4B–4D). To understand the characteristics of each subclass, we associated the subclasses with clinical, molecular, and signaling pathway features (Figure 4C). C1 tumors contained mutations in cell-cycle ($P = 0.0039$) and WNT ($P = 0.0031$) pathways, and in CTNNB1 gene ($P = 0.0053$). Differentially expressed genes between C1 and other tumors were enriched in DNA repair and viral carcinogenesis. C2 tumors were characterized by the mutations in NF- κ B pathway ($P=0.018$) and NBEA ($P=0.018$). NF- κ B links inflammation to cancer development and progression [28], and its over-expression in tumor tissues has been associated with a poor prognosis in different types of tumor [29]. Genes specifically expressed in C3 and C4 tumors were associated with immune response and T cell regulation. C4 tumors contained increased mutations in ARID1A ($P = 0.0035$), BAP1 ($P = 0.0017$) and IDH1 ($P = 0.05$)

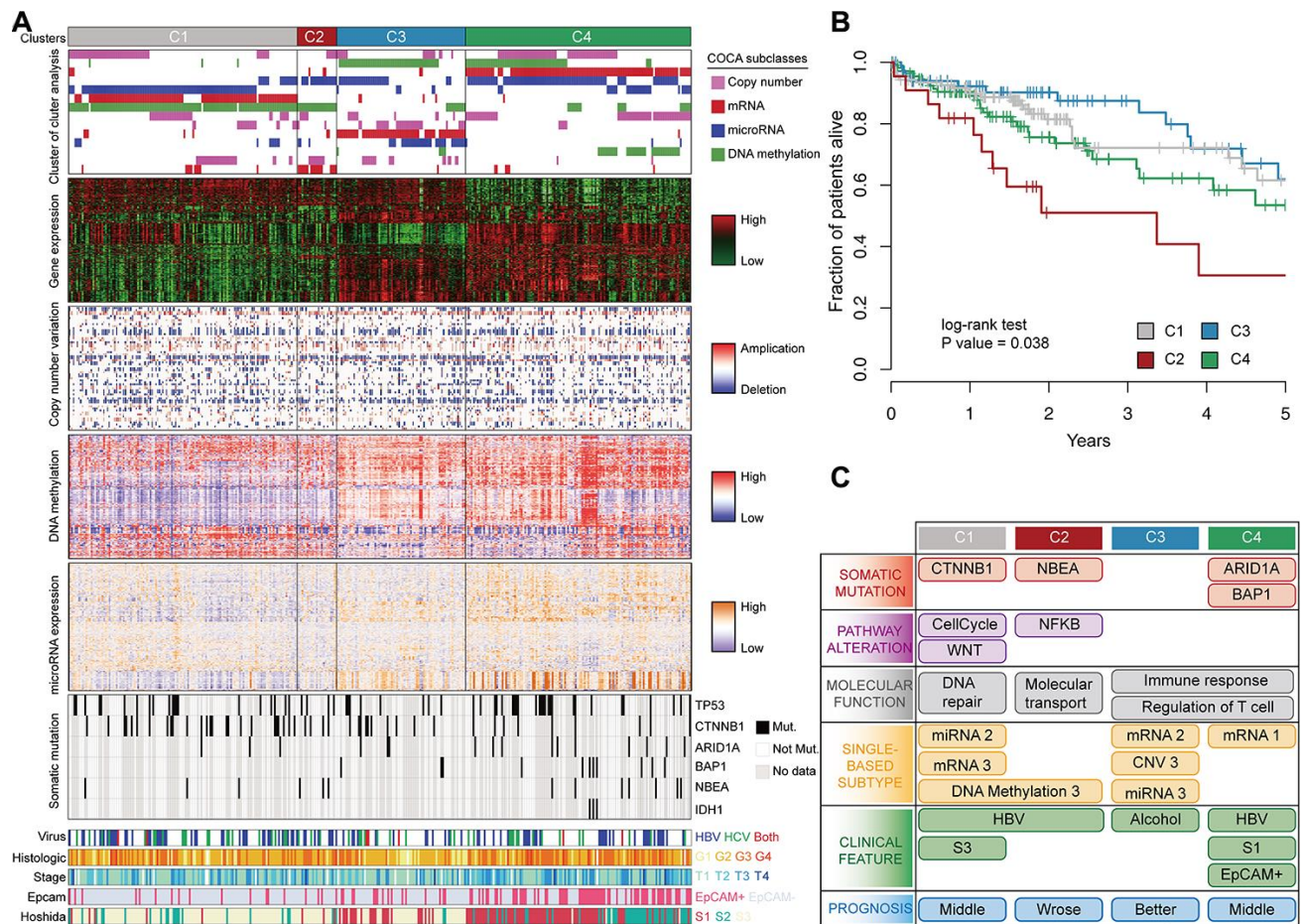


Figure 4. Multi-omics integration to identify HCC subclasses. (A) COCA subclasses of HCC identified by integrating multiple-platform data. Each column represents a patient; grey color in the spectrum of somatic mutation means patients without exome sequencing data. (B) Kaplan–Meier estimates of overall survival among different patient subclasses. (C) Schematic summary of molecular and clinical characteristics of the four HCC subclasses.

genes, were enriched in a signature of EpCAM positivity (Chi-square test $P = 1.9 \times 10^{-13}$, and overexpressed AFP ($P = 0.018$).

Roles of enhancers in HCC

Despite substantial advantages of multi-omics integration, expression changes of some driver genes cannot be explained by existing data. For example, copy number gains of AHR and PBX1 in tumors are significantly lower than in controls (Supplementary Figure 5). To investigate the epigenetic mechanisms of gene regulation, we defined chromatin states based on the combination of multiple epigenetic markers from the ENCODE project [30], and then defined enhancer regions by H3K4me1 and H3K27ac signals (see Materials and Methods).

Genes in actively transcribed regions exhibited increased expression compared to genes in inactive regions, and genes near strong enhancers exhibited increased expression compared to genes near weak enhancers (Supplementary Figure 6), reflecting the important regulatory role of chromatin states and enhancers [31]. Next, we compared the enhancer-related H3K4me1 signals between HCC and normal liver after excluding promoters, categorizing the genome into enhancer loss, enhancer stable, and enhancer gain based on the change in the H3K4me1 signal (Figure 5A, Supplementary Figure 7A). Loss of enhancer in HCC (40.8%) occurred more frequently than gain of enhancer (30.9%). This is consistent with a previous report demonstrating that the global enhancer expression is low in HCC [32]. We found that gene expression was significantly reduced in the regions with enhancer loss, especially those regulated by strong enhancers (Figure 5B). However, genes associated with stable enhancers and enhancer gain did not exhibit significant expression changes between HCC and normal tissues (Supplementary Figure 7B, 7C). We performed a functional enrichment analysis for genes associated with enhancer loss. In contrast to weak enhancers, inactivation of strong enhancers was associated with immunity and cancer pathways (Figure 5C), indicating that inactivation of strong enhancers plays an important role in HCC tumorigenesis.

Furthermore, we investigated the association between enhancer alterations and HCC driver genes. We first compared enhancers of driver genes with those of other genes. In normal liver, driver genes were located near strong enhancers compared to other genes (83.7% Vs. 72.9%, chi-square test, $p = 7.37 \times 10^{-6}$, Figure 5D). We also compared the enhancer alterations of driver genes between HCC and normal tissues, and found that the driver genes were associated with a loss of enhancers in HCC (45.6% loss vs. 16.9% gain, Figure 5E), similar

with the low global expression of enhancers in HCC genomes. Next, we compared the enhancer alterations of well-known oncogenes, TSGs, and other drivers. TSGs were more likely to locate near the regions with enhancer loss (59.1%, chi-square test, $p = 0.0793$, Figure 5F), resulting in their decreased expression. Finally, we explored the role of enhancers in the regulation of driver gene expression. Copy number gain and promoter hypermethylation lead to increased expression, while somatic mutation was associated with both decreased and increased expression. Integrative analysis of copy numbers, mutations, and DNA methylation could explain the expression changes of 61% drivers, while enhancers could explain the expression changes of 16% drivers (Figure 5G). For example, although there was an obvious copy number gain in AHR and PBX1 genes, the lost enhancer signals might have resulted in their decreased expression in HCC (Figure 5H).

DISCUSSION

The rapidly decreasing cost of omics experiments and increasing size of omics data have created an unprecedented opportunity to study cancer biology. However, multi-dimensional data pose a huge challenge for data analysis. In this study, we performed an integrative analysis of five platforms to identify driver genes and infer molecular classification of HCC. Compared to the original integrative analysis by TCGA network, our work is unique in several aspects: 1) Driver genes identified by TCGA are genes with significant somatic mutations or DNA copy number alterations; the TCGA approach does not consider the function effects of genomic alterations. Since we integrated several platforms and used a module network inference algorithm to evaluate the regulatory effects of genomic alterations on gene expression, the genes obtained by our approach are more likely to play important roles in HCC. 2) For multi-omics clustering, TCGA and our study used different genes and clustering algorithms. Since our study used driver genes instead of highly variable genes, the number of identified genes was smaller, and the subclasses had a better prognostic value than the TCGA subclasses. 3) Considering the alterations of histone methyltransferase [14], we analyzed histone modifications data and found the importance of gene enhancers for the regulation of gene expression.

Compared to single platform analysis, the integrative analysis of multi-platform data has obvious advantages. The integrative analysis can effectively remove random events from a single platform level and observe the true changes at multiple levels. The drivers identified from multi-omics analysis cover most of the drivers from

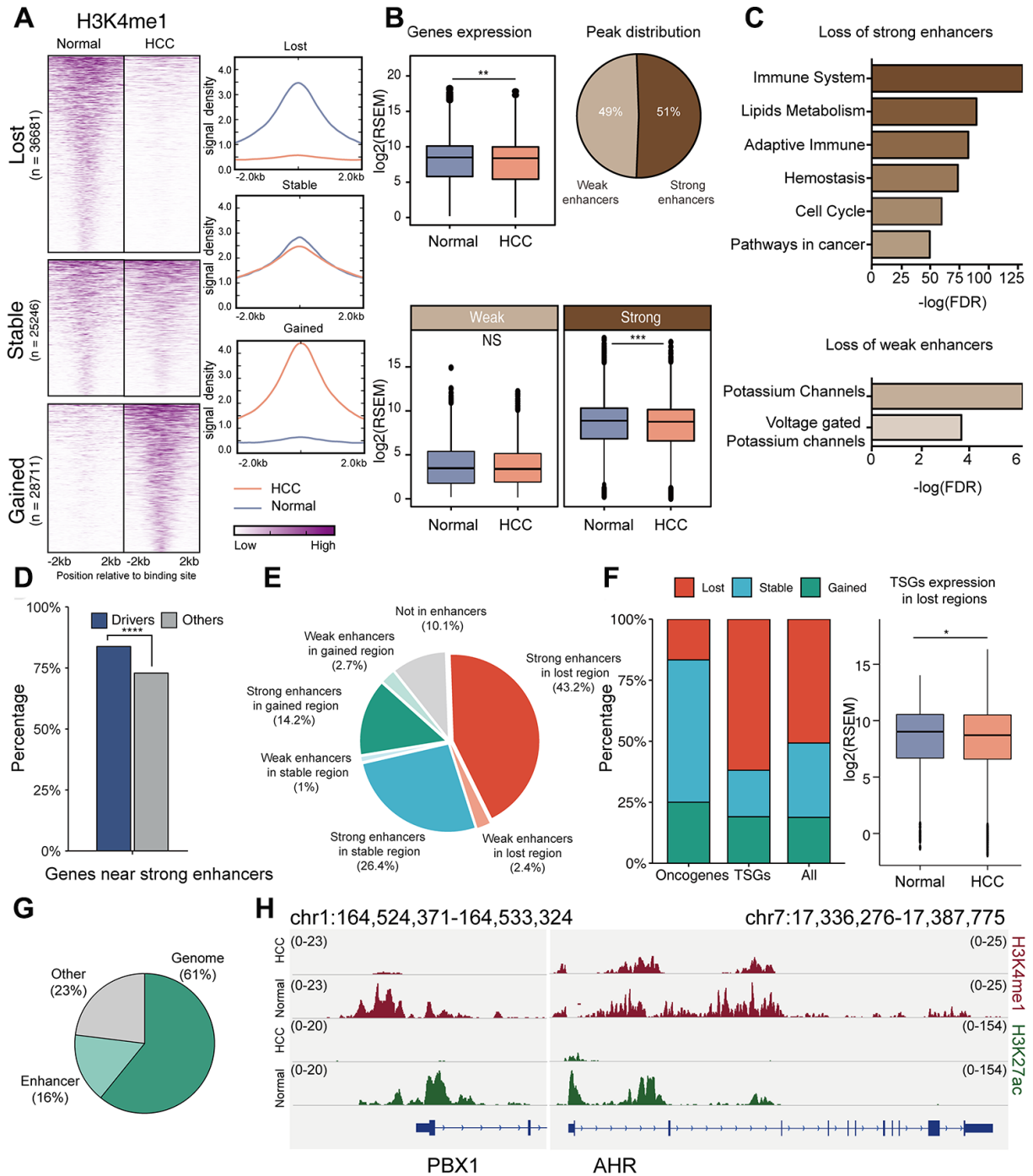


Figure 5. Role of enhancers in HCC. (A) Comparison of H3K4me1 signals in HCC cells (HepG2) and normal liver tissues. (B) Gene expression in enhancer loss regions in HCC and normal tissues. Enhancers were subdivided into strong and weak enhancers by H3K27ac signals. P value was determined by student's t test: ***P < 0.001; **P < 0.01; *P < 0.05; NS:P ≥ 0.05. (C) Functional enrichment for genes associated with the inactivation of strong/weak enhancers. (D) Percentage of driver genes located in strong enhancers. P value was determined by chi-square test: ****P < 0.0001. (E) Distribution of drivers in different regions; the regions were divided according to the types of enhancers and their expression in HCC and normal tissues. (F) Comparison of enhancer alterations of oncogenes, TSGs and other drivers. (G) Proportion of drivers that change expression. "Genome" means proportion of drivers whose down-regulation may be caused by mutation, hyper-methylation, or copy number deletion; up-regulation may be caused by mutations or copy number amplification. "Enhancer" means proportion of drivers whose down-regulation may be caused by enhancer loss, and up-regulation may be caused by enhancer gain. (H) ChIP-seq signal tracks for H3K4me1 and H3K27ac in the regions around PBX1 and AHR.

single-omics, act predominantly as hub genes regulating multiple modules, and correlate well with cancer hallmarks. Multi-omics integrative analysis can also reveal the combined action of different omics leading to disorders of the same genes or pathways. Subgroups derived from integrative analysis have more obvious survival differences than results from single-platform analysis. Overall, integration of multi-omics data is an effective strategy for identifying key genes involved in carcinogenesis.

Our study provides a comprehensive list of candidate driver genes in HCC. It will be important to conduct follow-up experimental studies to validate the roles of the identified novel drivers in HCC. For example, our computational analysis showed that the increased expression of SFN and SPP1 were associated with higher tumor grades and worse survival outcomes, indicating that SFN and SPP1 might act as oncogenes in HCC. Indeed, our knockdown experiments demonstrated that SFN and SPP1 promoted migration of HCC cells, validating the computational analysis. Although the exact oncogenic mechanism of SFN and SPP1 remains to be elucidated, our data demonstrate that integrative analysis can discover novel drivers with biological importance. Identification of the mechanisms of these candidate drivers will provide a more comprehensive and systematic understanding of HCC.

Additionally, our analysis provides a unique insight into the role of enhancers in HCC. Previous studies reported that histone methyltransferase EZH₂ was overexpressed in HCC, contributing to malignant transformation and poor prognosis [33, 34]. Due to the lack of histone modification data from a large HCC cohort, histone modifications cannot be directly integrated with other omics data. We used ChIP-seq data of H3K4me1 and H3K27ac obtained from HepG2 and normal liver tissues. We found that a loss of strong enhancers in HCC lead to a decreased expression of genes associated with carcinogenesis and immunity. Changes in the enhancer status could explain the expression changes of 16% drivers, which could not be explained by other types of omics data.

Together, our study shows that somatic mutations, copy number alterations, differential DNA methylation, and enhancer alterations coordinately regulate gene expression involved in hepatocarcinogenesis. A comprehensive driver list will provide a valuable resource for better understanding of the mechanisms responsible for HCC pathogenesis. The computational approach of integrating transcriptomic, genomic, and epigenetic data may be used also for other diseases.

MATERIALS AND METHODS

Data source

Multi-omics data of HCC were obtained from the Broad Institute TCGA Data Portal (<http://firebrowse.org>), including 377 HCC cases and 50 paired normal tissues on five platforms (RNA sequencing, DNA methylation arrays, miRNA sequencing, Affymetrix SNP arrays, and whole-exome sequencing). To minimize the possible batch effects, we applied the combat algorithm [35] to expression and methylation profiles.

Identification of cancer driver genes by multi-omics integration

Genomic, epigenetic and transcriptomic data were used to select the candidate gene regulators (Supplementary Table 1). 1) Significantly mutated genes were identified by the Mutsig2CV algorithm. 2) Copy number amplified or deleted regions were obtained from GISTIC algorithms. 3) Differentially methylated genes were identified with MethylMix [36], which uses a beta mixture model to find differential and transcriptionally predictive methylation states. 4) Experimentally validated miRNA-target interactions were collected from miRTarbase Database [37]. Spearman correlation coefficients between microRNAs and target genes were calculated; significantly negatively correlated miRNA-targets were selected by FDR <0.05. A list of transcription factors (TFs) [38, 39] was obtained from the UCSC Genome Browser, and HCC associated genes were obtained from literature [9–13, 40, 41].

Differentially expressed genes between tumor and normal tissues (student t test with FDR < 0.05) were regarded as downstream genes. An integrative multi-omics module network was reconstructed by Lemon-Tree to predict causal regulators [21]. It first infers co-expression modules from expression profiles of downstream genes; Gibbs sampling procedure was used to update the cluster assignment of each gene and sample. Then it employs regulation programs by fuzzy decision trees, which use the candidate regulators to predict the mean expression of genes in a module. Finally, a regulator-to-module score was computed by regulatory programs for that module. Top 1% of high-scoring regulators for each module, or top 10% of high-scoring regulators for sum scores of all modules were regarded as potential cancer drivers [21].

Functional enrichment analysis

Driver genes, genes in co-expression modules, and differentially expressed genes in HCC subclasses were subjected to functional enrichment analysis. Cancer

hallmarks were downloaded from MSigDB [42]. Enriched biological pathways and disease processes were found by DAVID [43] and GSEA (FDR<0.05). Potential toxic effects and molecular networks were generated by Ingenuity Pathways Analysis software (IPA) [44].

Cell culture

HCC cell line Huh1 was purchased from Japanese Collection of Research Bio-resources (JCRB, <http://cellbank.nibiohn.go.jp/english/>). HCC cell line CLC1 and CLC7 was established from Chinese HCC patients by Dr. Lijian Hui' lab. CLC1 and CLC7 cells were cultured in RPMI 1640 (HyClone) supplemented with 10% FBS (HyClone), 1 × ITS (Insulin, Transferrin, Selenium Solution), 40 ng/mL EGF (epithelial growth factor), 100 U/ml penicillin, and 100 µg/ml streptomycin. Huh1 cells were cultured in DMEM (Gibco) supplemented with 10% FBS, 100 U/ml penicillin, and 100 µg/ml streptomycin. Both cell lines were maintained in a humidified incubator at 37° C with 5% CO₂, and passaged every 3 days.

SFN and SPP1 knockout

To knockout SFN and SPP1, we used CRISPR/Cas9 system. 20-nucleotide single-guide RNA (sgRNA) sequences targeting SFN (5' AGCCTTCATGAAAG GCGCCG) and SPP1 (5' TGTACTCACTGGTAT GGCAC) were designed using the CRISPR design tool at <http://crispr.mit.edu/>. The sgRNA was cloned into the lentiCRISPR V2 vector (Addgene plasmid # 52961) and co-transfected with psPAX2 (Addgene plasmid # 12260) and pMD2.G (Addgene plasmid # 12259) packaging plasmids into 293FT cells. Virus containing supernatants were collected 48 h after transfection, filtered to remove cells using 0.45 µm filter, and then used to infect HCC cell lines CLC1, CLC7 and Huh1 in the presence of 8 µg/ml polybrene. 16 h post-transfection, cells were selected with puromycin (1 µg/ml; 72 h); SFN silencing was confirmed by Sanger sequencing and western blotting.

Cell proliferation assay

Cell proliferation was determined by CellTiter-Glo reagent (Promega). 50 µL of CellTiter-Glo reagent was added to each well of 96-well plates, and after a 10 min incubation at room temperature, luminescence was measured by EnVision Multilabel Reader (PerkinElmer), and normalized to day 1.

SiRNA screening

A panel of 4X siRNAs for 50 genes (4 siRNAs/gene) were picked up from Human Genome siRNA library

(Dharmacon) and dispensed 10 µl siRNA/well in 384-well plate by Biomek FX (Beckman Coulter). Three siRNAs were used as nontargeting control. 0.1 µl RNAimax (Invitrogen) in 10 µl serum-free opti-MEM medium (Thermo Fisher Scientific) was added by Multidrop Combi Reagent Dispenser (Thermo Fisher Scientific). Following a 20 min incubation, CLC7 cell line was seeded at a density of 3000 cells/well in 30 µl medium. The plate was transferred to the incubator for 96 hr. At the end point of treatment, each well was added 25 µl CellTiter-Glo reagent (Promega), and after 10 min incubation in room temperature, the luminescent signals were measured by EnVision Multilabel Reader (PerkinElmer) to determine the cell viabilities. The siRNA screening was performed twice.

Multi-platform based molecular subclasses

Cluster of Cluster analysis (COCA) [26] was used to classify HCC into subclasses. First, using drivers as features, data from each molecular platform (copy number alteration, DNA methylation, mRNA, and microRNA expression) were clustered as described [26]. The robustness of the clusters was estimated according to the factorization rank and cophenetic correlation coefficient. Then, a second-level clustering was derived from the class assignments of individual molecular platforms, and the optimal combination of clusters was determined by maximizing the consensus cumulative distribution function. The final COCA subclasses represent a common cluster among the single platforms.

Using differentially expressed genes in each subclass, pathway enrichment analysis was done to assess the potential functions of the subclasses. Frequently mutated pathways in each subclass were evaluated by the proportion of tumors with any gene alteration. Survival curves were estimated and plotted using the Kaplan-Meier method. Log-rank tests were used to compare the survival curves among COCA subgroups. To associate molecular subclasses with clinical variables, Fisher's exact test and two-way analysis of variance (ANOVA) were used to examine the difference in categorical and continuous variables among COCA subclasses, respectively. All statistical analyses were performed in R.

ChIP-Seq analysis

ChIP-seq data of H3K4me1 and H3K27ac for HCC cell line (HepG2) and liver tissues were downloaded from the ENCODE database. ChIP-seq signal tracks were quantile-normalized using Haystack [45] with 50-bp windows. The blacklisted regions were downloaded from ENCODE and excluded in subsequent analysis. Promoter regions were

identified by HOMER v4.10 [46]. The remaining differential peaks between HCC and normal tissues were determined by 2x fold-change. Heatmap and profile plot were generated using DeepTools v2.1.0 [47]. Genes associated with ChIP-seq peaks were identified via GREAT [48] with the basal plus extension (20kb) association rule. Genomic contexts were visualized by Integrative Genomics Viewer (IGV) software [49].

Chromatin state and enhancer

To delineate the chromatin contexts, 18 chromatin states (E1~E18) of normal liver tissues and HCC cells (HepG2) were downloaded from <http://compbio.mit.edu/roadmap> [30]. Genome was subdivided into active regions (E1~E12) and inactive regions (E13~E18) by the chromatin state composition. H3K4me1 and H3K27ac are histone markers for active enhancers. Regions absent promoters with both H3K4me1 and H3K27ac marks were called strong enhancers, while regions with only H3K4me1 marks were called weak enhancers. Wilcoxon rank sum test was used to compare gene expression levels in different genomic regions. Chi-square test was used to compare the occurrence of super-enhancers in HCC and normal samples.

AUTHOR CONTRIBUTIONS

LY, HL and YL conceived of the project. LY designed the computational analysis. ZZ designed and performed experiments. LY generated figures and tables. HL and YL supervised the research. LY, ZZ and HL wrote the manuscript. YS, SP, QY, PL, JC, JL, GD and JH participated in discussions and interpretations of the data and results.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

FUNDING

This work was supported by the National Natural Science Foundation of China (31771472), National Key Research and Development Project (2016YFC0901704, 2017YFC1201200, 2019YFC1315804), CAS Youth Innovation Promotion Association, SA-SIBS Scholarship Program, Chinese Academy of Sciences (KFJ-STQ-QYZD-126, ZDBS-SSW-DQC-02), Shanghai Municipal Science and Technology Major Project (No.2018SHZDZX01), LCNBI and ZJLab.

REFERENCES

1. Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray F. Cancer

incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer*. 2015; 136:E359–86.

<https://doi.org/10.1002/ijc.29210> PMID:25220842

2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin*. 2019; 69:7–34. <https://doi.org/10.3322/caac.21551> PMID:30620402
3. El-Serag HB, Rudolph KL. Hepatocellular carcinoma: epidemiology and molecular carcinogenesis. *Gastroenterology*. 2007; 132:2557–76. <https://doi.org/10.1053/j.gastro.2007.04.061> PMID:17570226
4. Jeng KS, Chang CF, Jeng WJ, Sheen IS, Jeng CJ. Heterogeneity of hepatocellular carcinoma contributes to cancer progression. *Crit Rev Oncol Hematol*. 2015; 94:337–47. <https://doi.org/10.1016/j.critrevonc.2015.01.009> PMID:25680939
5. Friemel J, Rechsteiner M, Frick L, Böhm F, Struckmann K, Egger M, Moch H, Heikenwalder M, Weber A. Intratumor heterogeneity in hepatocellular carcinoma. *Clin Cancer Res*. 2015; 21:1951–61. <https://doi.org/10.1158/1078-0432.CCR-14-0122> PMID:25248380
6. Buczak K, Ori A, Kirkpatrick JM, Holzer K, Dauch D, Roessler S, Endris V, Lasitschka F, Parca L, Schmidt A, Zender L, Schirmacher P, Krijgsveld J, et al. Spatial tissue proteomics quantifies inter- and intratumor heterogeneity in hepatocellular carcinoma (HCC). *Mol Cell Proteomics*. 2018; 17:810–25. <https://doi.org/10.1074/mcp.RA117.000189> PMID:29363612
7. Befeler AS, Di Bisceglie AM. Hepatocellular carcinoma: diagnosis and treatment. *Gastroenterology*. 2002; 122:1609–19. <https://doi.org/10.1053/gast.2002.33411> PMID:12016426
8. Liu J, Liu JF, Wang K, Yan ZL, Wan XY, Huang AM, Wang YZ, Li J, Xia Y, Shi LH, Jiao BH, Zhang Y, Shen F. Loss of function of Notch1 identifies a poor prognosis group of early stage hepatocellular carcinoma following hepatectomy. *Oncol Rep*. 2015; 34:3174–86. <https://doi.org/10.3892/or.2015.4300> PMID:26398566
9. Fujimoto A, Furuta M, Totoki Y, Tsunoda T, Kato M, Shiraishi Y, Tanaka H, Taniguchi H, Kawakami Y, Ueno M, Gotoh K, Ariizumi S, Wardell CP, et al. Erratum: whole-genome mutational landscape and characterization of noncoding and structural mutations in liver cancer. *Nat Genet*. 2016; 48:700. <https://doi.org/10.1038/ng0616-700a> PMID:27230686
10. Fujimoto A, Totoki Y, Abe T, Boroevich KA, Hosoda F, Nguyen HH, Aoki M, Hosono N, Kubo M, Miya F, Arai Y,

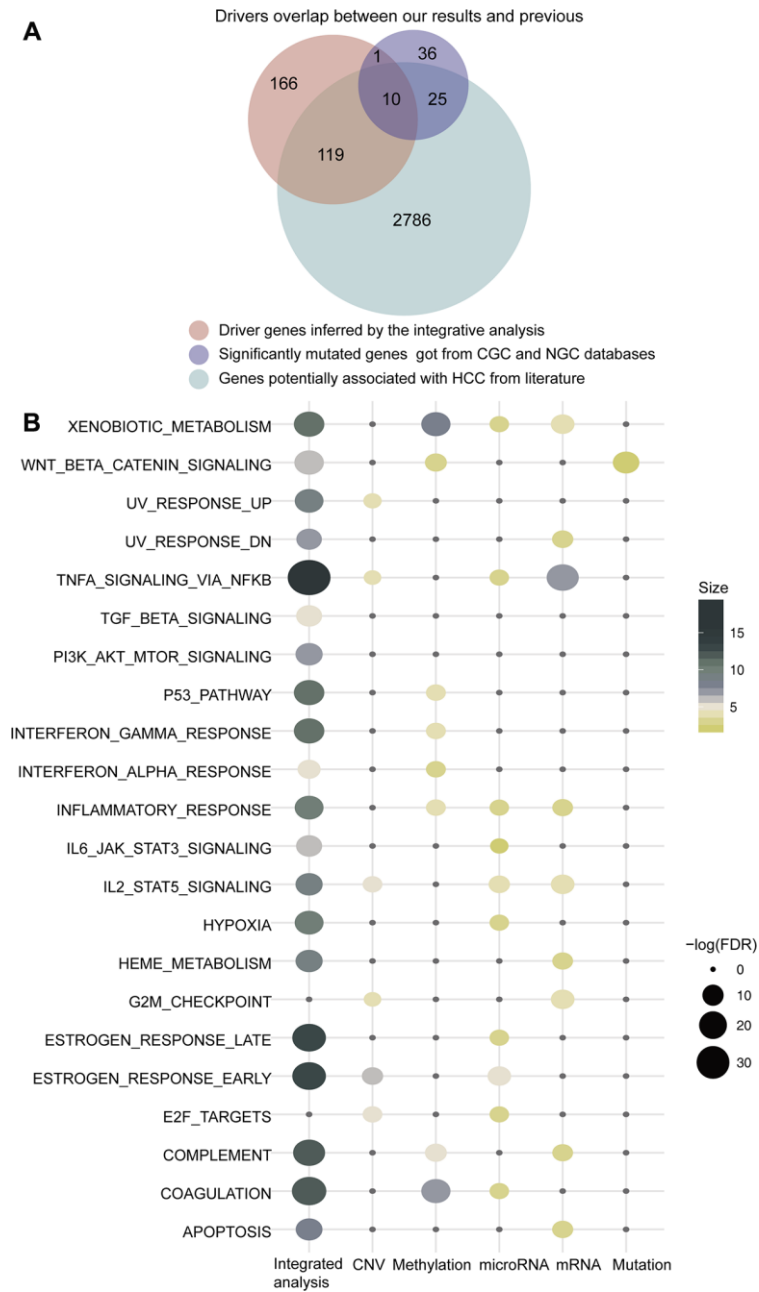
- Takahashi H, Shirakihara T, et al. Whole-genome sequencing of liver cancers identifies etiological influences on mutation patterns and recurrent mutations in chromatin regulators. *Nat Genet.* 2012; 44:760–64.
<https://doi.org/10.1038/ng.2291> PMID:22634756
11. Schulze K, Imbeaud S, Letouzé E, Alexandrov LB, Calderaro J, Rebouissou S, Couchy G, Meiller C, Shinde J, Soysouvanh F, Calatayud AL, Pinyol R, Pelletier L, et al. Exome sequencing of hepatocellular carcinomas identifies new mutational signatures and potential therapeutic targets. *Nat Genet.* 2015; 47:505–11.
<https://doi.org/10.1038/ng.3252> PMID:25822088
 12. Sung WK, Zheng H, Li S, Chen R, Liu X, Li Y, Lee NP, Lee WH, Ariyaratne PN, Tennakoon C, Mulawadi FH, Wong KF, Liu AM, et al. Genome-wide survey of recurrent HBV integration in hepatocellular carcinoma. *Nat Genet.* 2012; 44:765–69.
<https://doi.org/10.1038/ng.2295> PMID:22634754
 13. Totoki Y, Tatsuno K, Covington KR, Ueda H, Creighton CJ, Kato M, Tsuji S, Donehower LA, Slagle BL, Nakamura H, Yamamoto S, Shinbrot E, Hama N, et al. Trans-ancestry mutational landscape of hepatocellular carcinoma genomes. *Nat Genet.* 2014; 46:1267–73.
<https://doi.org/10.1038/ng.3126> PMID:25362482
 14. Cancer Genome Atlas Research Network. Electronic address: wheeler@bcm.edu; Cancer Genome Atlas Research Network. Comprehensive and Integrative Genomic Characterization of Hepatocellular Carcinoma. *Cell.* 2017; 169:1327–1341.e23.
<https://doi.org/10.1016/j.cell.2017.05.046> PMID:28622513
 15. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA Jr, Kinzler KW. Cancer genome landscapes. *Science.* 2013; 339:1546–58.
<https://doi.org/10.1126/science.1235122> PMID:23539594
 16. Flavahan WA, Gaskell E, Bernstein BE. Epigenetic plasticity and the hallmarks of cancer. *Science.* 2017; 357:eaal2380.
<https://doi.org/10.1126/science.aal2380> PMID:28729483
 17. Xu P, Wang J, Sun B, Xiao Z. Integrated analysis of miRNA and mRNA expression data identifies multiple miRNAs regulatory networks for the tumorigenesis of colorectal cancer. *Gene.* 2018; 659:44–51.
<https://doi.org/10.1016/j.gene.2018.03.050> PMID:29555201
 18. Zhang Y, Huang J, Li Q, Chen K, Liang Y, Zhan Z, Ye F, Ni W, Chen L, Ding Y. Histone methyltransferase SETDB1 promotes cells proliferation and migration by interacting with Tiam1 in hepatocellular carcinoma. *BMC Cancer.* 2018; 18:539.
<https://doi.org/10.1186/s12885-018-4464-9> PMID:29739365
 19. Revill K, Wang T, Lachenmayer A, Kojima K, Harrington A, Li J, Hoshida Y, Llovet JM, Powers S. Genome-wide methylation analysis and epigenetic unmasking identify tumor suppressor genes in hepatocellular carcinoma. *Gastroenterology.* 2013; 145:1424–35.e1.
<https://doi.org/10.1053/j.gastro.2013.08.055> PMID:24012984
 20. Akavia UD, Litvin O, Kim J, Sanchez-Garcia F, Kotliar D, Causton HC, Pochanard P, Mozes E, Garraway LA, Pe'er D. An integrated approach to uncover drivers of cancer. *Cell.* 2010; 143:1005–17.
<https://doi.org/10.1016/j.cell.2010.11.013> PMID:21129771
 21. Bonnet E, Calzone L, Michoel T. Integrative multi-omics module network inference with lemon-tree. *PLoS Comput Biol.* 2015; 11:e1003983.
<https://doi.org/10.1371/journal.pcbi.1003983> PMID:25679508
 22. Guo T, Chen T, Gu C, Li B, Xu C. Genetic and molecular analyses reveal G6PC as a key element connecting glucose metabolism and cell cycle control in ovarian cancer. *Tumour Biol.* 2015; 36:7649–58.
<https://doi.org/10.1007/s13277-015-3463-6> PMID:25926381
 23. Jiang G, Wang P, Wang W, Li W, Dai L, Chen K. Annexin A13 promotes tumor cell invasion in vitro and is associated with metastasis in human colorectal cancer. *Oncotarget.* 2017; 8:21663–73.
<https://doi.org/10.18632/oncotarget.15523> PMID:28423508
 24. Cerezo M, Guemiri R, Druillenec S, Girault I, Malka-Mahieu H, Shen S, Allard D, Martineau S, Welsch C, Agoussi S, Estrada C, Adam J, Libenciuc C, et al. Translational control of tumor immune escape via the eIF4F-STAT1-PD-L1 axis in melanoma. *Nat Med.* 2018; 24:1877–86.
<https://doi.org/10.1038/s41591-018-0217-1> PMID:30374200
 25. Albacker LA, Wu J, Smith P, Warmuth M, Stephens PJ, Zhu P, Yu L, Chmielecki J. Loss of function JAK1 mutations occur at high frequency in cancers with microsatellite instability and are suggestive of immune evasion. *PLoS One.* 2017; 12:e0176181.
<https://doi.org/10.1371/journal.pone.0176181> PMID:29121062
 26. Hoadley KA, Yau C, Wolf DM, Cherniack AD, Tamborero D, Ng S, Leiserson MD, Niu B, McLellan MD, Uzunangelov V, Zhang J, Kandoth C, Akbani R, et al, and

- Cancer Genome Atlas Research Network. Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. *Cell*. 2014; 158:929–44.
<https://doi.org/10.1016/j.cell.2014.06.049>
PMID:25109877
27. Hoshida Y, Nijman SM, Kobayashi M, Chan JA, Brunet JP, Chiang DY, Villanueva A, Newell P, Ikeda K, Hashimoto M, Watanabe G, Gabriel S, Friedman SL, et al. Integrative transcriptome analysis reveals common molecular subclasses of human hepatocellular carcinoma. *Cancer Res*. 2009; 69:7385–92.
<https://doi.org/10.1158/0008-5472.CAN-09-1089>
PMID:19723656
28. Song R, Song H, Liang Y, Yin D, Zhang H, Zheng T, Wang J, Lu Z, Song X, Pei T, Qin Y, Li Y, Xie C, et al. Reciprocal activation between ATPase inhibitory factor 1 and NF- κ B drives hepatocellular carcinoma angiogenesis and metastasis. *Hepatology*. 2014; 60:1659–73.
<https://doi.org/10.1002/hep.27312> PMID:25042864
29. Wu D, Wu P, Zhao L, Huang L, Zhang Z, Zhao S, Huang J. NF- κ B expression and outcomes in solid tumors: a systematic review and meta-analysis. *Medicine (Baltimore)*. 2015; 94:e1687.
<https://doi.org/10.1097/MD.0000000000001687>
PMID:26448015
30. Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, Ziller MJ, Amin V, Whitaker JW, Schultz MD, et al, and Roadmap Epigenomics Consortium. Integrative analysis of 111 reference human epigenomes. *Nature*. 2015; 518:317–30.
<https://doi.org/10.1038/nature14248> PMID:25693563
31. Tsang FH, Law CT, Tang TC, Cheng CL, Chin DW, Tam WV, Wei L, Wong CC, Ng IO, Wong CM. Aberrant super-enhancer landscape in human hepatocellular carcinoma. *Hepatology*. 2019; 69:2502–17.
<https://doi.org/10.1002/hep.30544>
PMID:30723918
32. Chen H, Li C, Peng X, Zhou Z, Weinstein JN, Liang H, and Cancer Genome Atlas Research Network. A pan-cancer analysis of enhancer expression in nearly 9000 patient samples. *Cell*. 2018; 173:386–99.e12.
<https://doi.org/10.1016/j.cell.2018.03.027>
PMID:29625054
33. Chen Y, Lin MC, Yao H, Wang H, Zhang AQ, Yu J, Hui CK, Lau GK, He ML, Sung J, Kung HF. Lentivirus-mediated RNA interference targeting enhancer of zeste homolog 2 inhibits hepatocellular carcinoma growth through down-regulation of stathmin. *Hepatology*. 2007; 46:200–08.
<https://doi.org/10.1002/hep.21668>
PMID:17596871
34. Sudo T, Utsunomiya T, Mimori K, Nagahara H, Ogawa K, Inoue H, Wakiyama S, Fujita H, Shirouzu K, Mori M. Clinicopathological significance of EZH2 mRNA expression in patients with hepatocellular carcinoma. *Br J Cancer*. 2005; 92:1754–58.
<https://doi.org/10.1038/sj.bjc.6602531>
PMID:15856046
35. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*. 2007; 8:118–27.
<https://doi.org/10.1093/biostatistics/kxj037>
PMID:16632515
36. Gevaert O. MethylMix: an R package for identifying DNA methylation-driven genes. *Bioinformatics*. 2015; 31:1839–41.
<https://doi.org/10.1093/bioinformatics/btv020>
PMID:25609794
37. Chou CH, Chang NW, Shrestha S, Hsu SD, Lin YL, Lee WH, Yang CD, Hong HC, Wei TY, Tu SJ, Tsai TR, Ho SY, Jian TY, et al. miRTarBase 2016: updates to the experimentally validated miRNA-target interactions database. *Nucleic Acids Res*. 2016; 44:D239–47.
<https://doi.org/10.1093/nar/gkv1258>
PMID:26590260
38. Yan Z, Shah PK, Amin SB, Samur MK, Huang N, Wang X, Misra V, Ji H, Gabuzda D, Li C. Integrative analysis of gene and miRNA expression profiles with transcription factor-miRNA feed-forward loops identifies regulators in human cancers. *Nucleic Acids Res*. 2012; 40:e135.
<https://doi.org/10.1093/nar/gks395> PMID:22645320
39. Wang X, Yan Z, Fulciniti M, Li Y, Gkatzamanidou M, Amin SB, Shah PK, Zhang Y, Munshi NC, Li C. Transcription factor-pathway coexpression analysis reveals cooperation between SP1 and ESR1 on dysregulating cell cycle arrest in non-hyperdiploid multiple myeloma. *Leukemia*. 2014; 28:894–903.
<https://doi.org/10.1038/leu.2013.233> PMID:23925045
40. Guichard C, Amaddeo G, Imbeaud S, Ladeiro Y, Pelletier L, Maad IB, Calderaro J, Bioulac-Sage P, Letexier M, Degos F, Clément B, Balabaud C, Chevet E, et al. Integrated analysis of somatic mutations and focal copy-number changes identifies key genes and pathways in hepatocellular carcinoma. *Nat Genet*. 2012; 44:694–98.
<https://doi.org/10.1038/ng.2256>
PMID:22561517
41. Kan Z, Zheng H, Liu X, Li S, Barber TD, Gong Z, Gao H, Hao K, Willard MD, Xu J, Hauptschein R, Rejto PA, Fernandez J, et al. Whole-genome sequencing identifies recurrent mutations in hepatocellular carcinoma. *Genome Res*. 2013; 23:1422–33.
<https://doi.org/10.1101/gr.154492.113>
PMID:23788652

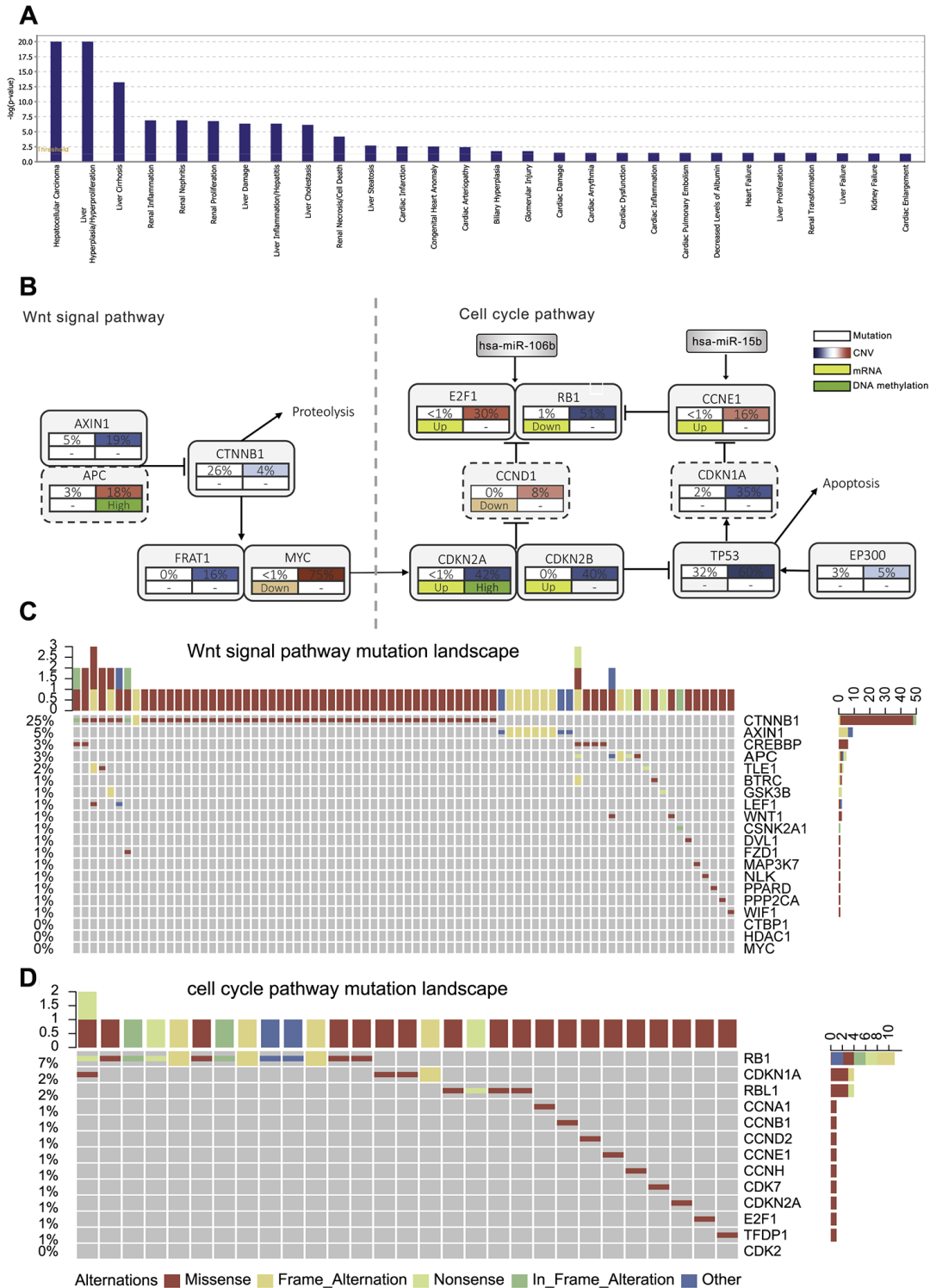
42. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA*. 2005; 102:15545–50.
<https://doi.org/10.1073/pnas.0506580102>
PMID:[16199517](https://pubmed.ncbi.nlm.nih.gov/16199517/)
43. Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc*. 2009; 4:44–57.
<https://doi.org/10.1038/nprot.2008.211>
PMID:[19131956](https://pubmed.ncbi.nlm.nih.gov/19131956/)
44. Krämer A, Green J, Pollard J Jr, Tugendreich S. Causal analysis approaches in ingenuity pathway analysis. *Bioinformatics*. 2014; 30:523–30.
<https://doi.org/10.1093/bioinformatics/btt703>
PMID:[24336805](https://pubmed.ncbi.nlm.nih.gov/24336805/)
45. Pinello L, Xu J, Orkin SH, Yuan GC. Analysis of chromatin-state plasticity identifies cell-type-specific regulators of H3K27me3 patterns. *Proc Natl Acad Sci USA*. 2014; 111:E344–53.
<https://doi.org/10.1073/pnas.1322570111>
PMID:[24395799](https://pubmed.ncbi.nlm.nih.gov/24395799/)
46. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*. 2010; 38:576–89.
<https://doi.org/10.1016/j.molcel.2010.05.004>
PMID:[20513432](https://pubmed.ncbi.nlm.nih.gov/20513432/)
47. Ramírez F, Dündar F, Diehl S, Grüning BA, Manke T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res*. 2014; 42:W187–91.
<https://doi.org/10.1093/nar/gku365> PMID:[24799436](https://pubmed.ncbi.nlm.nih.gov/24799436/)
48. McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM, Bejerano G. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol*. 2010; 28:495–501.
<https://doi.org/10.1038/nbt.1630>
PMID:[20436461](https://pubmed.ncbi.nlm.nih.gov/20436461/)
49. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. Integrative genomics viewer. *Nat Biotechnol*. 2011; 29:24–26.
<https://doi.org/10.1038/nbt.1754>
PMID:[21221095](https://pubmed.ncbi.nlm.nih.gov/21221095/)

SUPPLEMENTARY MATERIALS

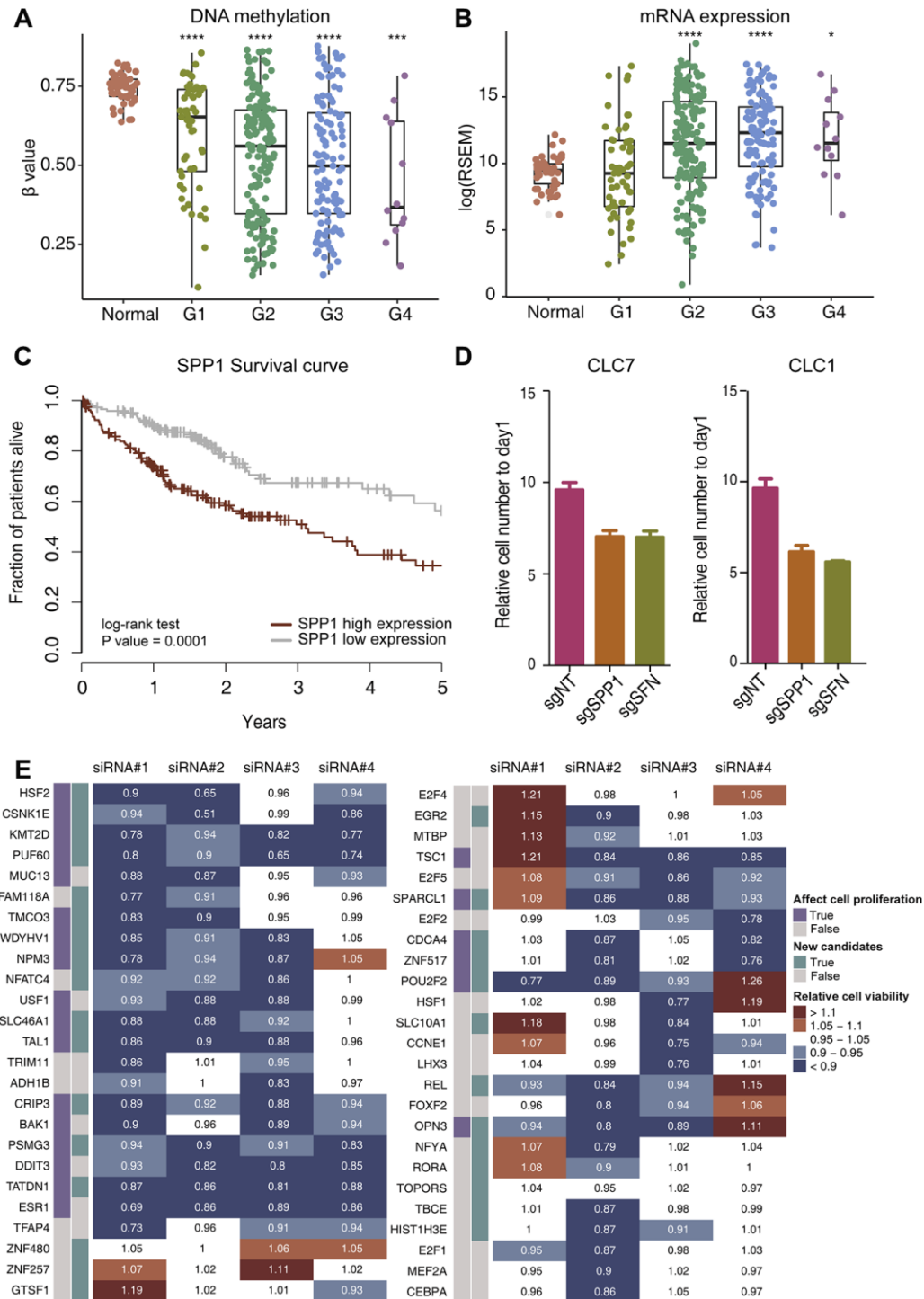
Supplementary Figures



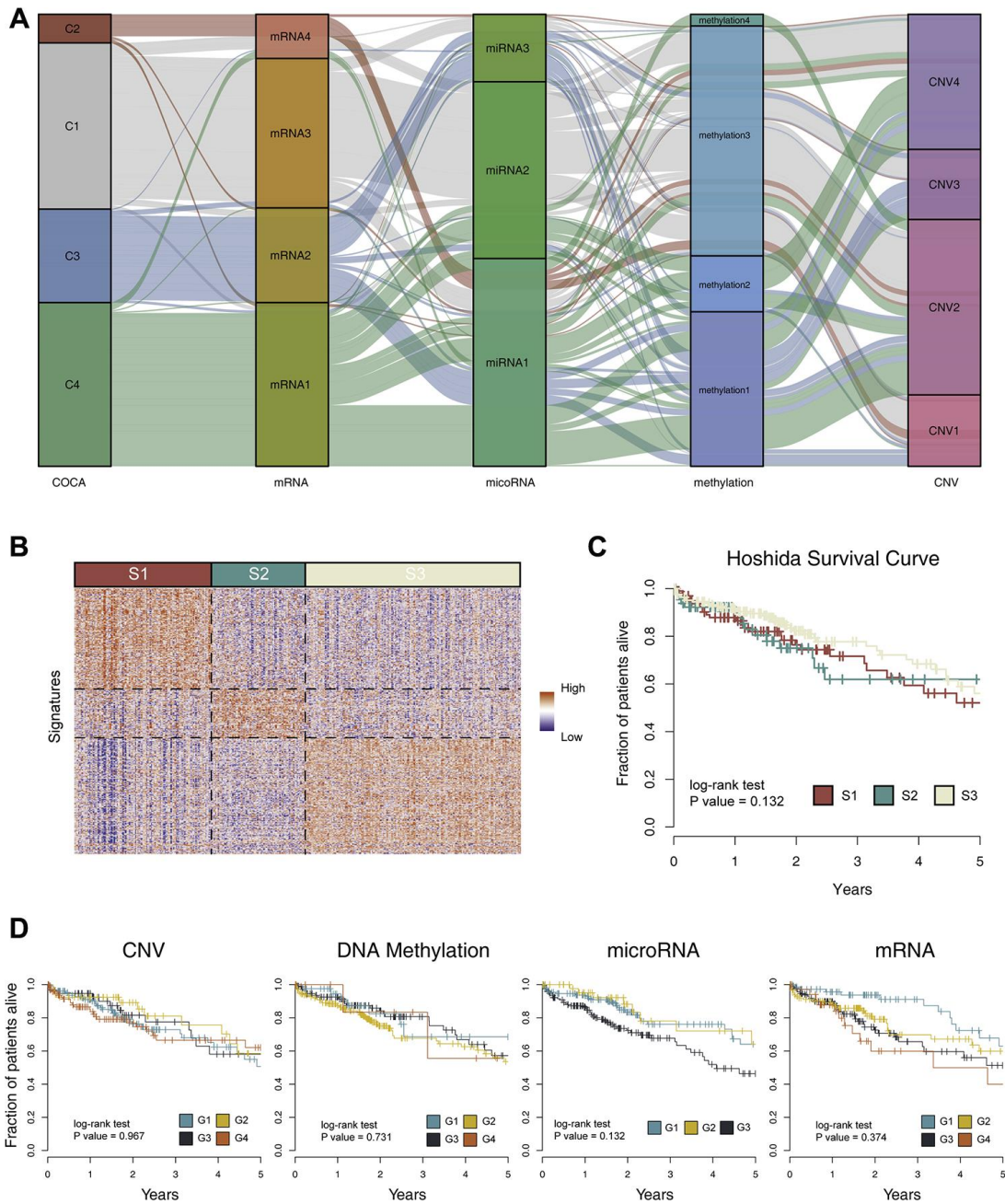
Supplementary Figure 1. Comparison between our results and previously published methods. (A) Comparison of the driver genes identified by our integrative analysis, and mutated genes and genes potentially associated with HCC identified from literature. Previously reported significantly mutated HCC-associated genes were obtained from Cancer Gene Census (CGC) and Network of Cancer Genes (NGC) databases. Genes potentially associated with HCC were obtained from PubMed by searching “hepatocellular carcinoma” and “gene” in the abstracts. (B) Cancer hallmarks enrichment for drivers identified by different omics data. Drivers from the integrated analysis were significantly enriched in the cancer hallmarks related terms, compared to independent analyses from each omics. Cancer hallmark terms were downloaded from MsigDB datasets. The size of the circle indicates the significance of the enrichment and the color indicates the number of drivers.



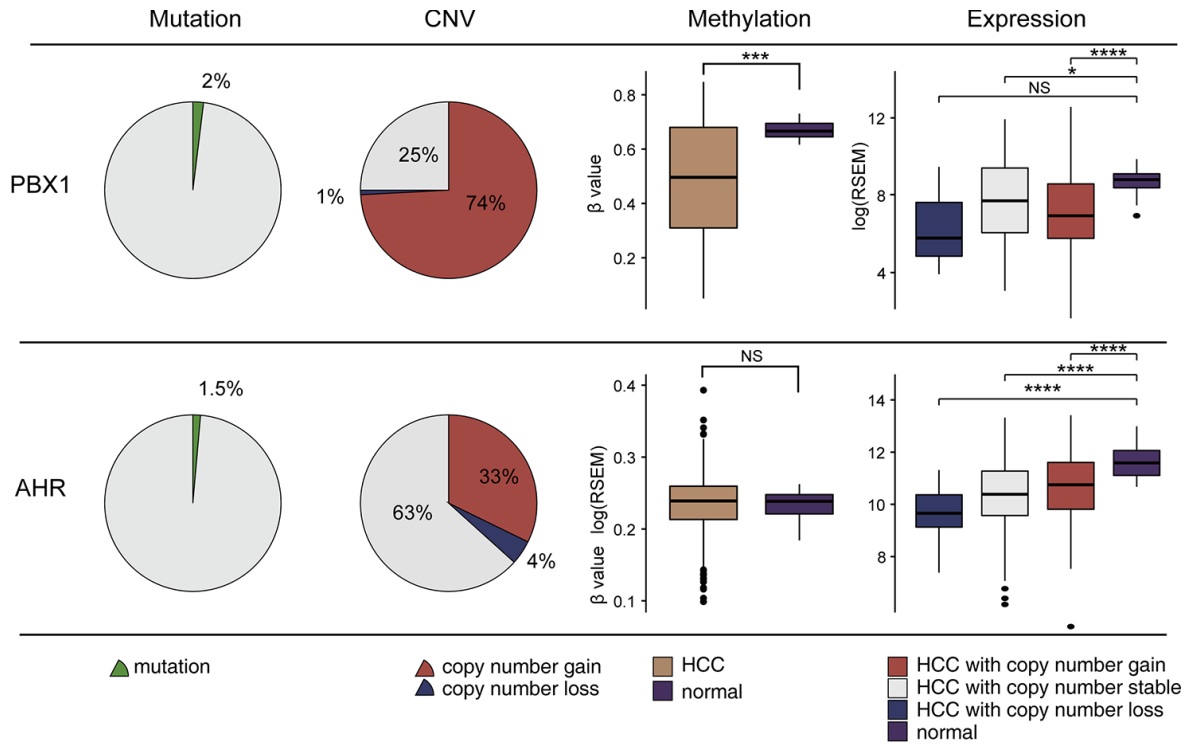
Supplementary Figure 2. Multi-omics landscape of Wnt and cell cycle pathways associated with HCC driver genes. (A) Tox analysis of drivers; hepatocellular carcinoma, liver hyperplasia/hyperproliferation and liver cirrhosis were significantly enriched, indicating correlation with liver cancer. **(B)** Overview of alterations in Wnt and cell cycle pathways. The solid rectangles represent drivers and dashed rectangles represent necessary linking genes in pathways that are not drivers. Colors represent the percentage of patients with mutations, copy number alterations, and dysregulation of methylation and expression. **(C, D)** Mutation landscape of Wnt and cell cycle pathway.



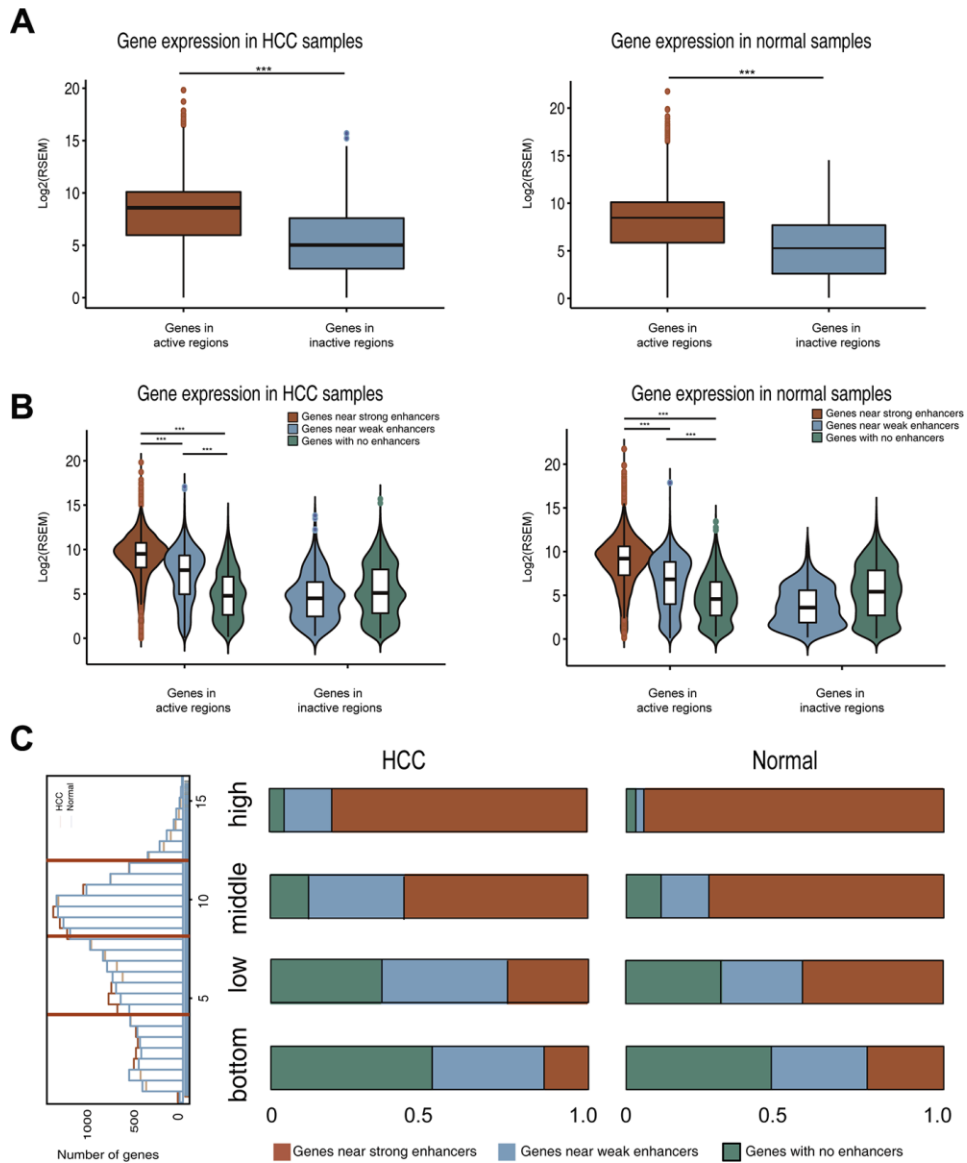
Supplementary Figure 3. Experimental validation of typical driver genes. (A, B) SPP1 methylation and expression correlate with HCC progression. P value was determined by Wilcoxon rank sum test: **** $p < 0.0001$; *** $p < 0.001$; ** $p < 0.01$; * $p < 0.05$. (C) Survival of HCC patients with high and low SPP1 expression. (D) Proliferation of CLC7 and CLC1 cells transfected with sgRNAs targeting SPP1 and SFN ($n=3$, regression analysis) after 96 hours. The relative cell number is normalized by the first day. (E) siRNA screening for top 50 genes with the highest integrative scores and upregulated in tumors. The genes that affect cell proliferation are determined by a 10% decrease or a 10% increase in cell proliferation rate in at least half of the repeated experiments after siRNA knockout.



Supplementary Figure 4. Comparison of integrative classification and other single-platform methods. (A) The consistency of integrative clustering and single-platform clustering. (B) The heat-map depicts gene expression profiles of 619 genes in Hoshida’s subtypes. (C, D). Kaplan–Meier estimates of the overall survival. Patients were classified according to Hoshida’s signature genes or single-platform drivers.

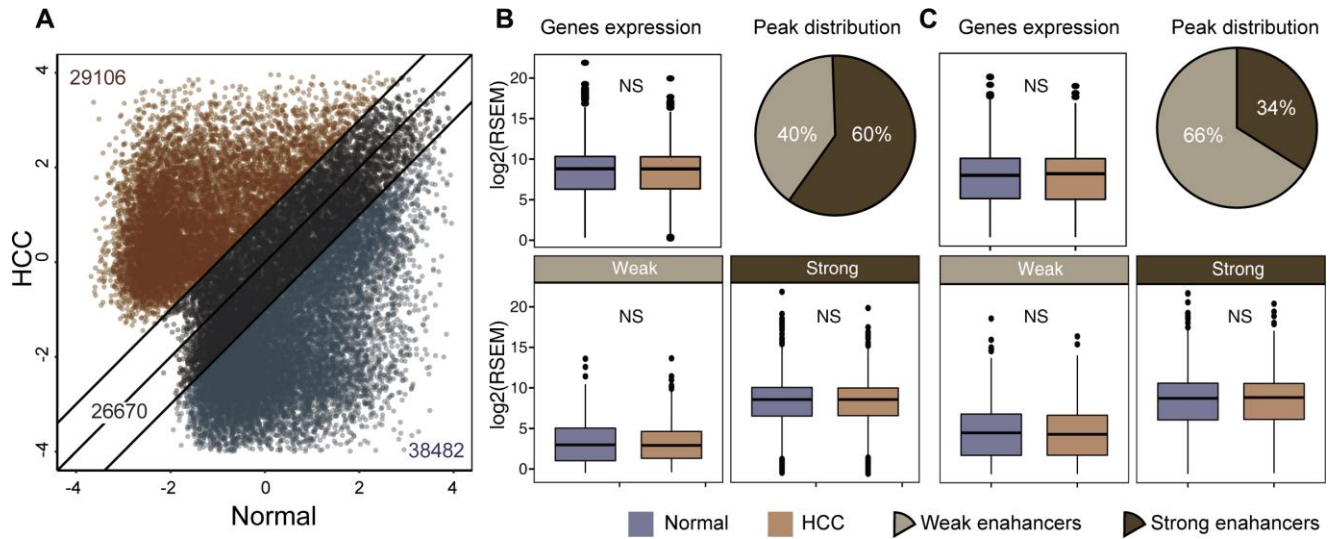


Supplementary Figure 5. Genomic alteration and expression of PBX1 and AHR. P value was determined by Wilcoxon rank sum test: ***P<0.001; **P<0.01; *P<0.05; NS:P≥0.05.



Supplementary Figure 6. Relationship of chromatin states and enhancers with gene expression in HCC and normal samples.

(A) Genes in activation regions (chromatin states E1-E12) show higher levels of gene expression than in inactivated regions (chromatin states E13-E18). (B) Enhancers significantly affect gene expression in activation regions. Strong enhancers are regions marked with both H3K4me1 and H3K27ac, and weak enhancers are regions only marked with H3K4me1, excluding promoters. (C) High-expression genes show greater association with strong enhancers. Genes associated with genome regions are assigned by 20kb from the TSS. P value was determined by Wilcoxon rank sum test: *** $P < 0.001$; ** $P < 0.01$; * $P < 0.05$.



Supplementary Figure 7. Enhancer subgroups and gene expression in stable and gain regions. (A) Scatter plot of H3K4me1 signal in HCC cell line (HepG2) and normal liver tissue. Lost, stable, and gained peaks were labelled by blue, black, and red, respectively. (B, C) Gene expression in enhancer stable and gained regions. P value was determined by student's t test: ***P<0.001; **P<0.01; *P<0.05; NS P≥0.05.

Supplementary Tables

Please browse Full Text version to see the data of Supplementary Tables 1 and 2.

Supplementary Table 1. List of candidate gene regulators selected by comparing the genomic, epigenetic and transcriptomic data between HCC and normal tissues.

Supplementary Table 2. List of HCC driver genes identified by integrative analysis of five platforms.