Published in final edited form as:
Cancer Res. 2010 April 15; 70(8): 3034–3041. doi:10.1158/0008-5472.CAN-09-2823.

Identification of a cholangiocarcinoma-like gene expression trait in hepatocellular carcinoma

Hyun Goo Woo[4,5], **Jeong-Hoon Lee**[1], **Jung-Hwan Yoon**[1], **Chung Yong Kim**[1], **Hyo-Suk Lee**[1], **Ja June Jang**[2], **Nam-Joon Yi**[3], **Kyung-Suk Suh**[3], **Kuhn Uk Lee**[3], **Eun Sung Park**[6], **Snorri S. Thorgeirsson**[4], and **Yoon Jun Kim**[1]

[1]Department of Internal Medicine and Liver Research Institute, Seoul National University College of Medicine, Seoul 110-744, Korea

[2]Department of Pathology, Seoul National University College of Medicine, Seoul 110-744, Korea

[3]Department of Surgery, Seoul National University College of Medicine, Seoul 110-744, Korea

[4]Laboratory of Experimental Carcinogenesis, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892

[5]Department of Physiology, Ajou University School of Medicine, Suwon 443-749, Korea

[6]Department of Systems Biology, Division of Cancer Medicine, University of Texas MD Anderson Cancer Center, Houston, TX 77030

Abstract

Hepatocellular carcinoma (HCC) and cholangiocarcinoma (CC) are the major adult liver cancers. The existence of combined hepatocellular-cholangiocarcinoma (CHC), a histopathological intermediate form between HCC and CC, suggests phenotypic overlap between these tumors. Here, we applied an integrative oncogenomic approach to address the clinical and functional implications of the overlapped phenotype between these tumors. By performing gene expression profiling of human HCC, CHC, and CC, we identified a novel HCC subtype, namely, CC-like HCC (CLHCC), which expressed CC-like traits (CC signature). As like CC and CHC, CLHCC showed aggressive phenotype with shorter recurrence-free and overall survival. In addition, we found that CLHCC coexpressed embryonic stem cell-like expression traits (ES signature) suggesting its derivation from bipotent hepatic progenitor cells. By comparing the expression of CC signature with previous ES-like, hepatoblast-like, or proliferation-related traits, we observed that that the prognostic value of the CC signatures is independent of the expression of those signatures. In conclusion, we suggest that the acquisition of CC like-expression traits play a critical role in the heterogeneous progression of HCC.

Keywords

microarray; stem cell; prognosis; cholangiocarcinoma-like HCC

Correspondence: Yoon Jun Kim, M.D., Seoul National University College of Medicine, 28 Yongon-dong, Chongno-gu, Seoul, 110-744, Korea, yoonjun@snu.ac.kr, Tel: 82-2-2072-3081, Fax: 82-2-743-6701.

Data deposition footnote: The raw data of the microarray experiments are available in GEO database (http://www.ncbi.nlm.nih.gov/geo) with an accession number GSE15765.

The authors declared no potential conflict of interest.

Introduction

Hepatocellular carcinoma (HCC) and cholangiocarcinoma (CC) are the major primary liver cancers in adults. HCC is enormously heterogeneous with dismal clinical outcome (1), while CC is more difficult to diagnose and to treat compared with HCC showing worse prognosis (2). In addition, a rare form of combined hepatocellular-cholangiocarcinoma (CHC) has been reported to have intermediate characteristics between HCC and CC (3). Most HCC and CC are believed to be derived from hepatocytes and cholangiocytes, respectively. Meanwhile, CHC have been suggested to be derived from bipotential liver stem cells which can differentiate into either hepatic or biliary progenitor cells (4). In addition, several studies have suggested the stem/progenitor cell origin of liver cancers including HCC (5, 6), CC (7), and cholangiolocellular carcinoma (8). Such heterogeneous differentiation status of cellular origin suggests the phenotypic overlap between HCC, CHC, and CC.

Liver stem cells have bipotential to differentiate into either hepatic or biliary progenitor cells (9, 10). However, the cancer phenotypes derived from the biliary committed cells during developmental hierarchy were not investigated thoroughly. In this study, to address the biological and clinical implications of the overlapped phenotype between biliary and hepatic cell-traits in liver cancer, we applied an integrative genomic approach by comparing the expression profiles of HCC, CHC, and CC. By calculating the expression levels of CC-like traits (CC signature), we identified a novel subtype of HCC, namely CC-like HCC (CLHCC) which might be derived from biliary lineage cells. Further evaluation of the CC-like trait with the previous stem cell-derived traits such as embryonic stem (ES) cell-like (11) and hepatoblast (HB)-like (12) signatures was helpful to elucidate the heterogeneous progression of liver cancers implying their cellular origins from different developmental stages.

Materials and Methods

Patients and diagnosis of tumor types

We prospectively collected intrahepatic 70 HCC, 7 CHC, and 13 CC specimens from the patients who had surgical treatment for tumor at Seoul National University (SNU) Hospital. Intraoperative ultrasonography confirmed that no distant metastases or space-occupying lesions existed in the non-resected remnant liver of any of the individuals in this study. Patients with extrahepatic tumor were excluded. All the patients were determined to have received curative resection by examining the presence of residual tumors at surgical margin. The study protocol was approved by the institutional review board at the Seoul National University Hospital. Histological diagnosis using H&E-stained slides were reviewed by two experienced pathologists, independently. Validation of tumor types was performed by using the immune-reactivity of anti-hepatocyte antibody and CK7/CK19, respectively (*i.e.*, HCC, +/−; CHC, +/+; CC, −/+). All CHC specimens were the mixed type C according to Allen and Lisa classification (3).

Microarray experiments

Microarray experiments were performed using Affymetrix HGU-133A2 GeneChip as described previously (13). Raw data of 90 tumors including our previous data (13) were normalized by RMA method (14), and further set the mean values of each gene and each sample to zero. For multiple tagged gene features for same Entrez Gene identifier, the gene feature with the highest magnitude (*i.e.* sum of square of the expression levels) was used as a representative gene feature.

Construction of gene expression compendium of HCC (HCCcomp)

The gene expression compendium of HCC (HCCcomp) was constructed by collecting six different data sets including data from LEC (*i.e.*, data from Laboratory of Experimental Carcinogenesis, National Cancer Institute, GSE1898 and GSE4024), GSE5975 (15), E-TABM-36 (16), GSE9843 (17), SNU (data from Seoul National University, GSE15765), and a new platform data for formalin-fixed, paraffin-embedded (FFPE) tumors (GSE10186) (18). Publicly available data were downloaded from NCBI GEO or ArrayExpress databases. Non-HCC samples in the data sets were excluded, and each dataset was normalized as described above. Finally, 712 HCC gene expression profiles were compiled.

The signatures for stem cell-like expression traits and canonical pathways

A collection of ES-related signatures (*i.e.*,ES1, ES2, sox2, oct4, myc1, and myc2), polycomb-target gene signatures (*i.e.*, Suz2, Eed, H3K27 bound, and PRC2), and a proliferation-related gene signature (prol) were obtained from the previous study (11). The rat hepatoblast (HB) gene signature expressed in early fetal liver development were also obtained from the publication site (12). The homologous human and rat genes were linked via the NCBI HomoloGene database. For the enrichment test of canonical functions, the gene sets in each of biological process Gene Ontology (GO) terms with at least 3 genes were used.

Estimation of the enrichment of gene expression signatures

The enrichment of a gene set in individual tumors was determined by Kolmogorov-Smirnov (KS-) test. Briefly, for each individual gene expression profile, two *P*-values for the estimates D+ and D− were calculated by KS-test, which determine the significance of the directional (positive or negative) enrichment of distributions of the signature. The enrichment scores S_{D+} and S_{D-} for a given signature was calculated as $-\log_{10}$ (*P*-value) from D+ and D−, respectively. The enrichment score S was defined as S_{D+} if $S_{D+} > S_{D-}$ and $-S_{D-}$ if $S_{D+} < S_{D-}$. The samples with $|S|>2$ ($P<0.01$) were regarded to be significantly enriched. The enrichment patterns across sample groups were determined by calculating fraction of the significantly enriched samples in a group, and the significance of the group enrichment was calculated by hypergeometric test as described previously (11).

Gene network analysis

Genetic network was constructed using PathwayStudio (Ariadne Genomics version 6.2). All the direct interactions among an interested gene set were identified from the curated database provided by the software. Of the subnetworks constructed from the gene set, the largest sub-network with significant enrichment ($P<0.05$, calculated from the software) was identified as a key regulatory network.

Results

Clinical and transcriptomic characteristics of CC, CHC, and HCC

The clinical features of HCC, CHC, and CC were summarized in Supplementary Table S1. Compared with the HCC patients, the CC patients were much older (>55) and had lower serum levels of AFP and platelets. Histologically, all the CC tumors were multiple nodular types implying their aggressive behavior. Kaplan-Meir plot analyses revealed that the CC patients had poorer recurrence-free survival (RFS) and overall survival (OS) compared with those of HCC patients (Supplementary Fig. S1A, B). The CHC patients showed poor prognosis similar to CC patients suggesting the closer likeliness of CC and CHC tumors in agreement with previous study (19).

First, we sought to evaluate whether the expression profiles are ready to classify the tumor types. Unsupervised clustering analysis showed that the HCC and CC were well stratified indicating the marked difference between HCC and CC (Supplementary Fig. S1C). 5 out of 7 CHCs were clustered together with the CC tumors supporting the prognostic similarity between CHC and CC. No significant batch effect was found between the newly added HCC samples and the previous ones (Supplementary Fig. S1C). Next, to verify the reflection of the conserved tumor type characteristics in the gene expression profiles, the functional characteristics of each tumor type was evaluated by calculating the signature enrichment in the GO hierarchy. Considering the heterogeneity of the expression patterns in individual tumors, the signature enrichment analysis was applied to each tumor (for details see Materials and Methods). The HCC specimens were enriched with the metabolism- and immune-related functions, while the CC specimens were enriched with the development/differentiation- or metastasis/adhesion-related functions (Supplementary Fig. S2). The CHC showed similar pattern with the CC. These data indicate that the distinct prognostic values of each tumor class are well reflected in the gene expression patterns. Therefore, we suggest that the prognostic differences among the tumor types are *bona fide* tumor characteristics rather than the casual associations with other clinical factors such as operability of the tumors, time to diagnosis, or patient's health condition.

Identification of a novel subtype cholangiocarcinoma-like HCC (CLHCC)

Next, to determine the expression of CC-like traits in HCC, we identified differentially expressed genes between HCC and CC (n=2,188) by using 10,000 permuted two-sample *T*-test ($P<0.001$) and fold difference greater than 2 (false discovery rate > 0.0023). These genes included many putative candidates for the prognostic biomarkers as well as biomarkers for the differential diagnosis of CC. Notably, well-known biomarkers for CC or hepatic progenitor cells (HPC) such as *KRT19* (CK19, 6.0 fold), *TACSTD1* (EpCam, 4.1 fold), and *PROM1* (CD133, 3.0 fold) were identified. Also, known CC biomarkers such as *CEACAM6* (20), *MUC1* (21), and *CLDN4* (22) were identified indicating the usefulness of the CC signature as novel differential biomarkers for CC. Supervised clustering of the tumors with these genes revealed that a fraction of HCC (14 out of 70) was clustered together with CC samples. These tumors, referred to CC-like HCC (CLHCC), showed shorter recurrence-free survival (RFS) and overall survival (OS) compared with other HCC ($P = 5.21 \times 10^{-6}$, 0.004, respectively, log-rank test, Fig. 1A, B). No significant association with clinical features was found (Supplementary Table S2).

Independent validation of the prognostic value of the CLHCC

To validate and characterize the CLHCC in independent data sets with larger sample size, we constructed a gene expression compendium of HCC (HCCcomp, n=712) by concatenating 6 independent HCC data sets. We defined the most differentially expressed genes between CLHCC and other HCC as classifiers for CLHCC, which we referred to CC-like expression trait in HCC (*i.e.*, CC signature, n=625, false discovery rate > 0.0073, Supplementary Table 3). Then, the HCC tumors were classified based on the expression status of the CC signature. Since the clustering-based classification shown in Fig. 1 can be influenced by sample composition in the data set to be tested, we calculated the individual enrichment scores of CC signature based on KS-test. The tumors expressing both up- (CC_UP, $P<0.01$) and down- (CC_DOWN, $P<0.01$) regulated genes were classified as C1 (representing CLHCC, n=190), while the other tumors were classified as C2 (n=522). Notably, each subtype in the C1-C2 classification showed homogeneous expression pattern independent of microarray platforms and patient cohorts, which may support the consistency and robustness of our classification (Fig. 2A).

For independent validation, the prognostic values between C1 and C2 tumors were evaluated in two independent cohorts of Chinese (n=61) and Caucasian (n=78) patients in the LEC data set which had been described in a previous study (5). The C1 tumors showed shorter RFS and OS compared with the C2 tumors in both cohorts (Fig. 2B, C). In addition, univariate and multivariate analyses in the validation data set (n=139, Table 1) and test data set (n=70, Supplementary Table S4) also demonstrated that the C1-C2 classification is significant in predicting RFS as well as OS. Taken together, we suggest that the CC signature is a strong predictor for poor prognosis.

As like CC, GO analysis showed that the C1 tumors were enriched with the proliferation, metastasis/adhesion, and development-related functions reflecting their aggressive phenotype (Fig. 3A,B). By contrast, the C2 tumors were enriched with metabolism-related genes which might be due to the high metabolic rate of well-differentiated hepatocyte-derived HCC. Next, we further sought to identify key regulators of the CC signature. Of the genes directly interacted with the CC_UP genes (n=251), the *TP53* expression subnetwork was identified as the most prominent subnetwork with significant enrichment (P=0.036) indicating the finding is not likely to be observed by chance (Fig. 3C and Supplementary Table S5). This result may support the regulatory role of *TP53* for the aggressive phenotype of CLHCC.

Comparison of the expression of ES and CC signatures

Considering the hypothesis that the CHC is originated from bipotential liver stem/progenitor cells (4), we examined whether the CLHCC (C1) express stem cell-like traits. We evaluated multiple ES-related signatures (*i.e.*, ES1, ES2, and the target genes for Nanog, Oct4, Sox2, and Myc) and the polycomb group target signatures (*i.e.,* Suz2, Eed, H3K27 bound, and PRC2 target genes) which were previously known to play important roles in maintaining the undifferentiating status of ES cells (11, 23-27). Strikingly, the C1 tumors showed significant enrichment of the ES signatures and combined repression of polycomb target genes (Fig. 4A). Another stemness trait, a hepatoblast–derived signature (HB signature) was also evaluated, which showed similar expression patterns with the ES signatures. In addition, to exclude the possible influence of the proliferation-related genes in the ES signature, we subtracted them from the CC, HB, and ES signatures (noprol), but no significant influence was found (Fig. 4A). When we compared the genes in these signatures, only the 36 HB genes (5.8 %) and the 26 ES1 genes (4.2 %) were overlapped with the CC signature genes (Supplementary Fig. S3A), implying that the prognostic difference between C1-C2 classes is not likely to be confounded by the co- enrichment of the ES or HB signatures. As expected, no significant influence was found on the CC signature enrichment by subtracting the HB signature and/or ES signature genes (*i.e.*, ES1) from the CC signature genes (*i.e.*, CC_noHB, CC_noES, and CC_noHBnoES) (Fig. 4B). Taking these results together, we suggest that the prognostic value of the CC signature is independent of the coexpression of the HB, ES, or proliferation-related genes.

For independent validation, we next evaluated the relationship between CC and ES signatures in each data set which were compiled in HCCcomp. Remarkably, all the 6 data sets showed significant enrichment of ES signatures in C1 tumors compared with C2 tumors. These findings strongly indicate the robustness and consistency of the coexpression of CC and ES signatures regardless of microarray platforms and patient cohorts (Fig. 4C).

Next, when we further examined the individual tumors, not all the C1 tumors coexpress the ES signature (Supplementary Fig. S3B). This may imply the distinct expression of CC and ES signatures to some extent, which might be derived from intermediate cellular origins during the sequential development stages from the primitive liver stem cells to biliary committed cells. With this concern, we investigated whether the combined expression status

of ES and CC signature can reflect distinct prognostic phenotypes. The HB signature was not considered in the analysis because of its intermediate property between ES and CC signature. When we re-classified the 209 tumors (LEC and SNU) which survival data are available into four classes based on the expression of CC and ES signatures (*i.e.*, $CC^{+}ES^{+}$, $CC^{+}ES^{-}$, $CC^{-}ES^{+}$, and $CC^{-}ES^{-}$), the $CC^{+}ES^{+}$ tumors showed the worst prognosis for both RFS (hazard ratio, HR=2.84; 95% CI =1.51−5.34; P=7.42 x 10^{-4}) and OS (HR=2.98; 95% CI=1.79−4.98; P=1.2 × 10^{-5}) compared with the $CC^{-}ES^{-}$ showing the best prognosis (Fig. 4D). The $CC^{+}ES^{-}$ (n=21) and $CC^{-}ES^{+}$ (n=26) tumors showed intermediate prognostic values indicating the correlation between prognostic outcomes and the expression status of the CC and ES signatures. Although further validation might be required, this finding may support the idea that the cellular origin at different developmental stage plays a pivotal role in the heterogeneous clinical outcome of HCC.

Discussion

In this study, we addressed the heterogeneity of HCC by identifying a fraction of HCC expressing the CC-like traits. Although we profiled relatively small samples of CC, the huge expression difference between CC and HCC (as like “apple and orange”) allowed us to identify the robust CC signature. The functional and clinical relevance of the CC signature could be further validated by independent data sets (Fig. 2B, C).

The CC signature was concomitantly expressed with the ES or HB signatures suggesting the stem-like features of CLHCC, which could be validated by all the 6 independent HCC data sets. In fact, biliary markers such as CK-7 and CK-19 are frequently used as HPC markers (10, 28, 29). Therefore, the expression of stem-like traits or biliary traits alone may not be specific in discriminating the stem cell origin tumors from the biliary cell origin tumors. Further evaluation using HCCcomp revealed the presence of $CC^{+}ES^{-}$ or $CC^{-}ES^{+}$ tumors, indicating the existence of intermediate transition of the expression of those signatures. The ES signature is presumed to be derived from more primitive and pluripotent stem cells. The HB signature may express in the tumors derived from hepatoblast cells, whereas the CC signature may express in the tumors derived from the biliary lineage cells including premature and mature cholangiocytes. Based on this concept, the cellular origin of HCC from the sequential development stages can be postulated by the expression status of the ES, HB, and CC signatures as illustrated in Supplementary Fig. S4. This suggests the different cellular origin of HCC during transition from pluripotent to differentiated CC or HCC may play a critical role in the heterogeneous progression of HCC. However, we cannot exclude the possibility that trans-differentiation and de-differentiation of hepatocytes or cholangiocytes during cancer development can contribute to the acquisition of these signatures.

Our analysis dissecting the heterogeneous HCC based on the expression of CC signature could classify the tumors into homogeneous and distinct prognostic phenotypes. Targeting molecular pathways specific to such subpopulations would be more effective for the development of personalized clinical strategy (30). We suggested *TP53* pathway may play a pivotal role in the development of the CC signature-expressing HCC (CLHCC). The association of *TP53* mutation with poor prognosis is well known in many cancer types (reviewed in 31). Moreover, recent studies have demonstrated the critical role of *TP53* in the control of neural and glioma stem/progenitor cell renewal and differentiation (32, 33). These findings consistently support the pivotal role of *TP53* in the aggressive progression of the HCC harboring CC-like or stem-like traits.

In conclusion, our findings provide novel biological and clinical insights into the CC-like traits in HCC, emphasizing the critical role of the developmental stage of the cell of origin in HCC pathogenesis.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This project was supported by the Korea Research Foundation Grant (grant no. KRF-2005-041-E00160), a grant of the National R&D Program for Cancer Control, Ministry for Health, Welfare & Family Affairs, Republic of Korea (0920280), and the Basic Research Laboratory (BRL) program through the National Research Foundation of Korea funded by the Ministry of Education, Science and Technology (2009-0087452).

References

1. Khan SA, Thomas HC, Davidson BR, Taylor-Robinson SD. Cholangiocarcinoma. Lancet. 2005; 366:1303–14. [PubMed: 16214602]
2. Blechacz B, Gores GJ. Cholangiocarcinoma: advances in pathogenesis, diagnosis, and treatment. Hepatology. 2008; 48:308–21. [PubMed: 18536057]
3. Allen RA, Lisa JR. Combined liver cell and bile duct carcinoma. Am J Pathol. 1949; 25:647–55. [PubMed: 18152860]
4. Zhang F, Chen XP, Zhang W, et al. Combined hepatocellular cholangiocarcinoma originating from hepatic progenitor cells: immunohistochemical and double-fluorescence immunostaining evidence. Histopathology. 2008; 52:224–32. [PubMed: 18184271]
5. Lee JS, Heo J, Libbrecht L, et al. A novel prognostic subtype of human hepatocellular carcinoma derived from hepatic progenitor cells. Nat Med. 2006; 12:410–6. [PubMed: 16532004]
6. Yamashita T, Ji J, Budhu A, et al. EpCAM-positive hepatocellular carcinoma cells are tumor-initiating cells with stem/progenitor cell features. Gastroenterology. 2009; 136:1012–24. [PubMed: 19150350]
7. Nomoto K, Tsuneyama K, Cheng C, et al. Intrahepatic cholangiocarcinoma arising in cirrhotic liver frequently expressed p63-positive basal/stem-cell phenotype. Pathol Res Pract. 2006; 202:71–6. [PubMed: 16377099]
8. Komuta M, Spee B, Vander Borght S, et al. Clinicopathological study on cholangiolocellular carcinoma suggesting hepatic progenitor cell origin. Hepatology. 2008; 47:1544–56. [PubMed: 18393293]
9. Sell S, Leffert HL. Liver cancer stem cells. J Clin Oncol. 2008; 26:2800–5. [PubMed: 18539957]
10. Mishra L, Banker T, Murray J, et al. Liver stem cells and hepatocellular carcinoma. Hepatology. 2009; 49:318–29. [PubMed: 19111019]
11. Ben-Porath I, Thomson MW, Carey VJ, et al. An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. Nat Genet. 2008; 40:499–507. [PubMed: 18443585]
12. Petkov PM, Zavadil J, Goetz D, et al. Gene expression pattern in hepatic stem/progenitor cells during rat fetal development using complementary DNA microarrays. Hepatology. 2004; 39:617–27. [PubMed: 14999680]
13. Woo HG, Park ES, Cheon JH, et al. Gene expression-based recurrence prediction of hepatitis B virus-related human hepatocellular carcinoma. Clin Cancer Res. 2008; 14:2056–64. [PubMed: 18381945]
14. Irizarry RA, Hobbs B, Collin F, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics. 2003; 4:249–64. [PubMed: 12925520]
15. Jia HL, Ye QH, Qin LX, et al. Gene expression profiling reveals potential biomarkers of human hepatocellular carcinoma. Clin Cancer Res. 2007; 13:1133–9. [PubMed: 17317821]

16. Boyault S, Rickman DS, de Reynies A, et al. Transcriptome classification of HCC is related to gene alterations and to new therapeutic targets. Hepatology. 2007; 45:42–52. [PubMed: 17187432]
17. Villanueva A, Chiang DY, Newell P, et al. Pivotal role of mTOR signaling in hepatocellular carcinoma. Gastroenterology. 2008; 135:1972–83. 83 e1–11. [PubMed: 18929564]
18. Hoshida Y, Nijman SM, Kobayashi M, et al. Integrative transcriptome analysis reveals common molecular subclasses of human hepatocellular carcinoma. Cancer Res. 2009; 69:7385–92. [PubMed: 19723656]
19. Cazals-Hatem D, Rebouissou S, Bioulac-Sage P, et al. Clinical and molecular analysis of combined hepatocellular-cholangiocarcinomas. J Hepatol. 2004; 41:292–8. [PubMed: 15288479]
20. Ieta K, Tanaka F, Utsunomiya T, Kuwano H, Mori M. CEACAM6 gene expression in intrahepatic cholangiocarcinoma. Br J Cancer. 2006; 95:532–40. [PubMed: 16868542]
21. Higashi M, Yonezawa S, Ho JJ, et al. Expression of MUC1 and MUC2 mucin antigens in intrahepatic bile duct tumors: its relationship with a new morphological classification of cholangiocarcinoma. Hepatology. 1999; 30:1347–55. [PubMed: 10573510]
22. Nishino R, Honda M, Yamashita T, et al. Identification of novel candidate tumour marker genes for intrahepatic cholangiocarcinoma. J Hepatol. 2008; 49:207–16. [PubMed: 18490072]
23. Lee TI, Jenner RG, Boyer LA, et al. Control of developmental regulators by Polycomb in human embryonic stem cells. Cell. 2006; 125:301–13. [PubMed: 16630818]
24. Assou S, Le Carrour T, Tondeur S, et al. A meta-analysis of human embryonic stem cells transcriptome integrated into a web-based expression atlas. Stem Cells. 2007; 25:961–73. [PubMed: 17204602]
25. Li Z, Van Calcar S, Qu C, Cavenee WK, Zhang MQ, Ren B. A global transcriptional regulatory role for c-Myc in Burkitt's lymphoma cells. Proc Natl Acad Sci U S A. 2003; 100:8164–9. [PubMed: 12808131]
26. Fernandez PC, Frank SR, Wang L, et al. Genomic targets of the human c-Myc protein. Genes Dev. 2003; 17:1115–29. [PubMed: 12695333]
27. Boyer LA, Lee TI, Cole MF, et al. Core transcriptional regulatory circuitry in human embryonic stem cells. Cell. 2005; 122:947–56. [PubMed: 16153702]
28. Durnez A, Verslype C, Nevens F, et al. The clinicopathological and prognostic relevance of cytokeratin 7 and 19 expression in hepatocellular carcinoma. A possible progenitor cell origin. Histopathology. 2006; 49:138–51. [PubMed: 16879391]
29. Wu PC, Fang JW, Lau VK, Lai CL, Lo CK, Lau JY. Classification of hepatocellular carcinoma according to hepatocellular and biliary differentiation markers. Clinical and biological implications. Am J Pathol. 1996; 149:1167–75. [PubMed: 8863666]
30. Knight B, Tirnitz-Parker JE, Olynyk JK. C-kit inhibition by imatinib mesylate attenuates progenitor cell expansion and inhibits liver tumor formation in mice. Gastroenterology. 2008; 135:969–79. 79 e1. [PubMed: 18602920]
31. Petitjean A, Achatz MI, Borresen-Dale AL, Hainaut P, Olivier M. TP53 mutations in human cancers: functional selection and impact on cancer prognosis and outcomes. Oncogene. 2007; 26:2157–65. [PubMed: 17401424]
32. Wang Y, Yang J, Zheng H, et al. Expression of mutant p53 proteins implicates a lineage relationship between neural stem cells and malignant astrocytic glioma in a murine model. Cancer Cell. 2009; 15:514–26. [PubMed: 19477430]
33. Zheng H, Ying H, Yan H, et al. p53 and Pten control neural and glioma stem/progenitor cell renewal and differentiation. Nature. 2008; 455:1129–33. [PubMed: 18948956]

Abbreviations

CC	cholangiocarcinoma
HCC	hepatocellular carcinoma
CHC	combined hepatocellular-cholangiocarcinoma

ES	embryonic stem
HPC	hepatic progenitor cell
CLHCC	cholangiocarcinoma-like hepatocellular carcinoma
HB	hepatoblast
HCCcomp	gene expression compendium of HCC

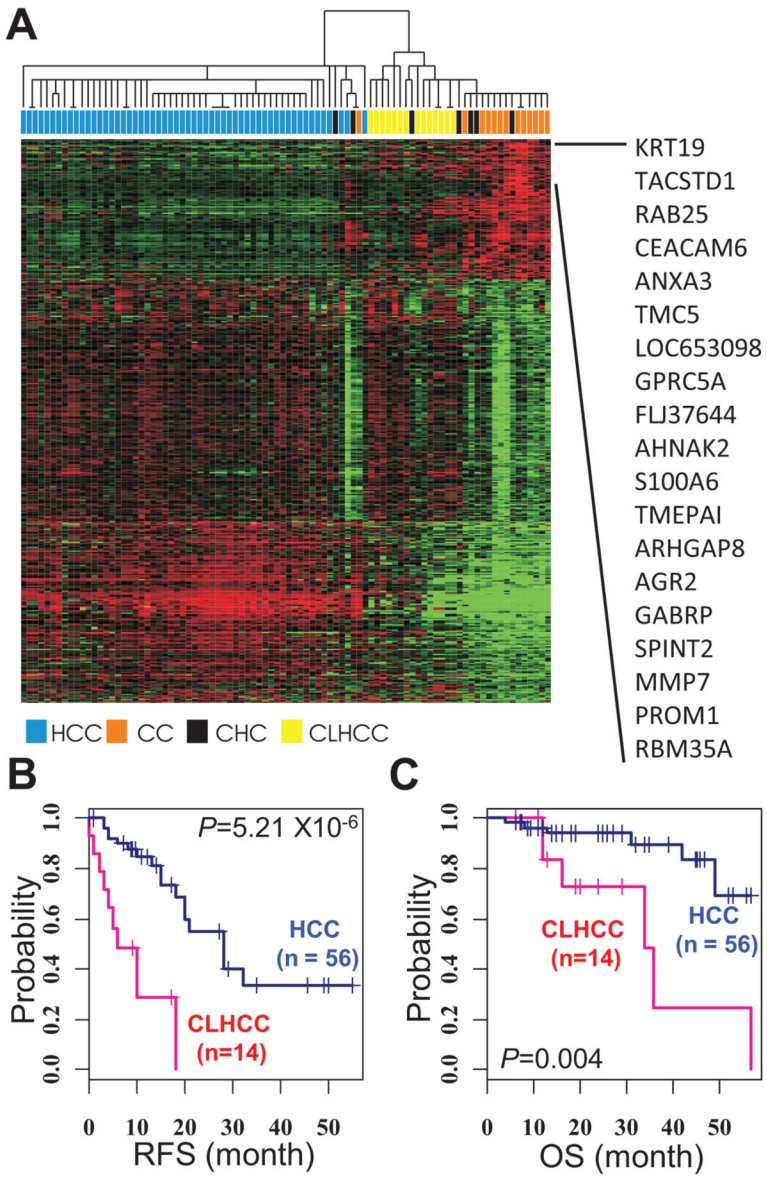

Fig. 1. Identification of cholangiocarcinoma-like HCC (CLHCC)

A. Supervised clustering of CC, CHC, and HCC based on the expression of the CC signature (*left*). The top 20 most differential expressed genes between CC and HCC are indicated (*right*). **B**. Kaplan-Meir plot analyses for recurrence-free survival (RFS) and overall survival (OS) between HCC and CLHCC.

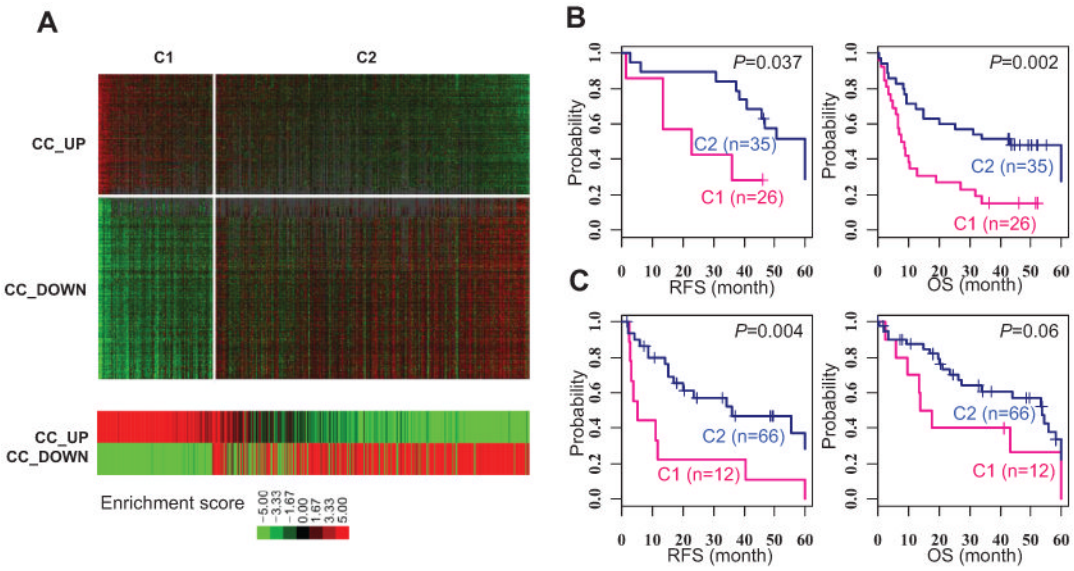

Fig. 2. Validation of the CC signature in HCCcomp

A. The gene expression profiles of HCCcomp were classified into 2 groups based on the expression of the CC signature (*top*). The enrichment scores for CC_UP and CC_DOWN signatures are shown (*bottom*). **B, C.** Kaplan-Meir plots for recurrence-free survival (RFS, *left*) and overall survival (OS, *right*) in independent cohorts of Chinese (n=61, **B**) and Caucasian (n=78, **C**), respectively. The follow up time for recurrence-free survival (RFS) and overall survival (OS) are truncated to five years.

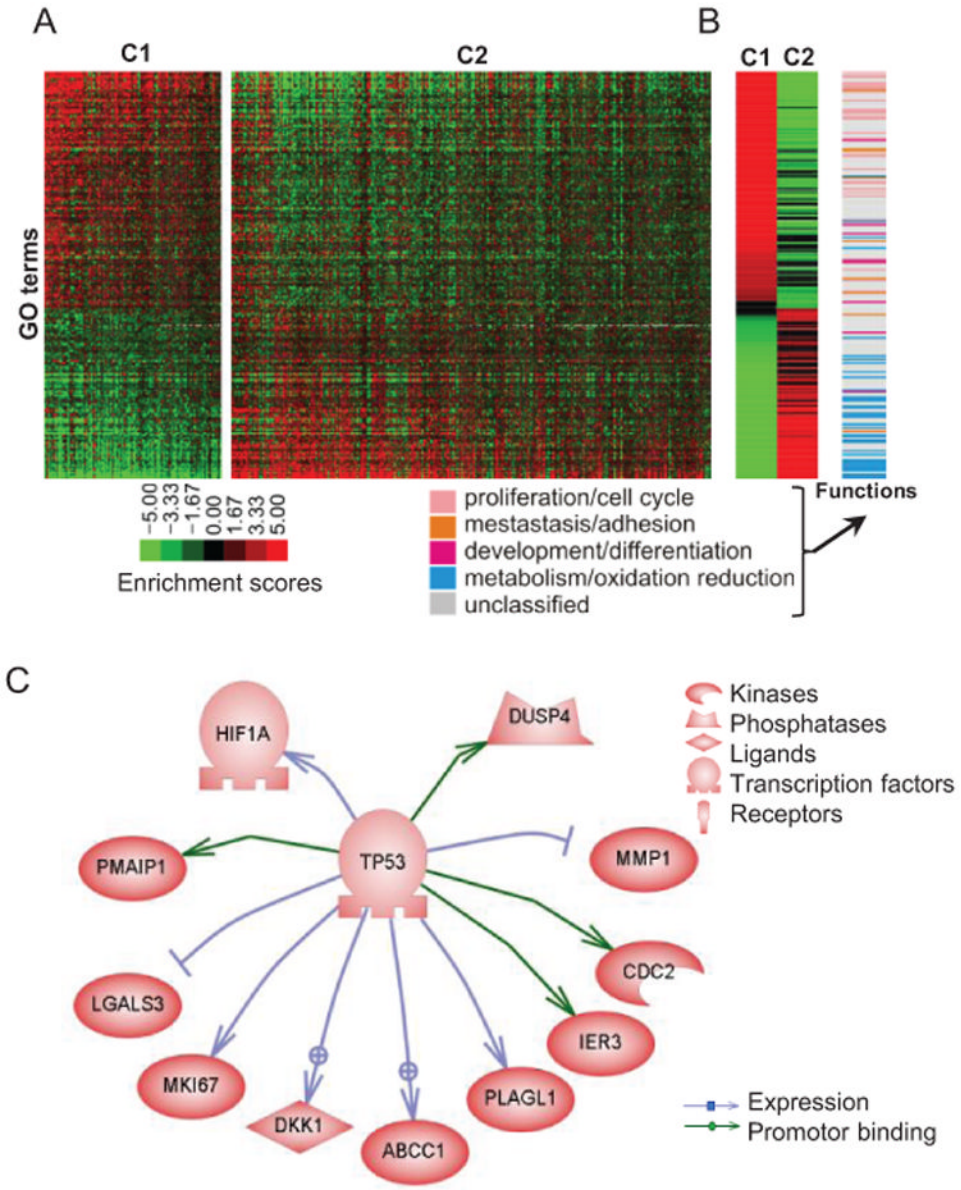

Fig. 3. Functional characteristics of the CC signature

A. The functional enrichment scores in the Gene Ontology hierarchy for each group C1 and C2. A total of 196 GO terms are significantly enriched at least in a group (P<0.01, hypergeometric test). **B**. The enrichment scores in each group are shown by bar-views (*right bar*). The enriched patterns of the GO terms for the 4 functional categories are indicated by different colors, respectively (*the rightmost bar*). The 4 functional categories are chosen which show prominent enrichment of the enrichment scores in each group. Other GO terms which are not prominent in the functional grouping are indicated as “unclassified”. **C**. Network analysis indicates the *TP53* as a most prominent regulator of the CC_UP signature with significant enrichment (P<0.036).

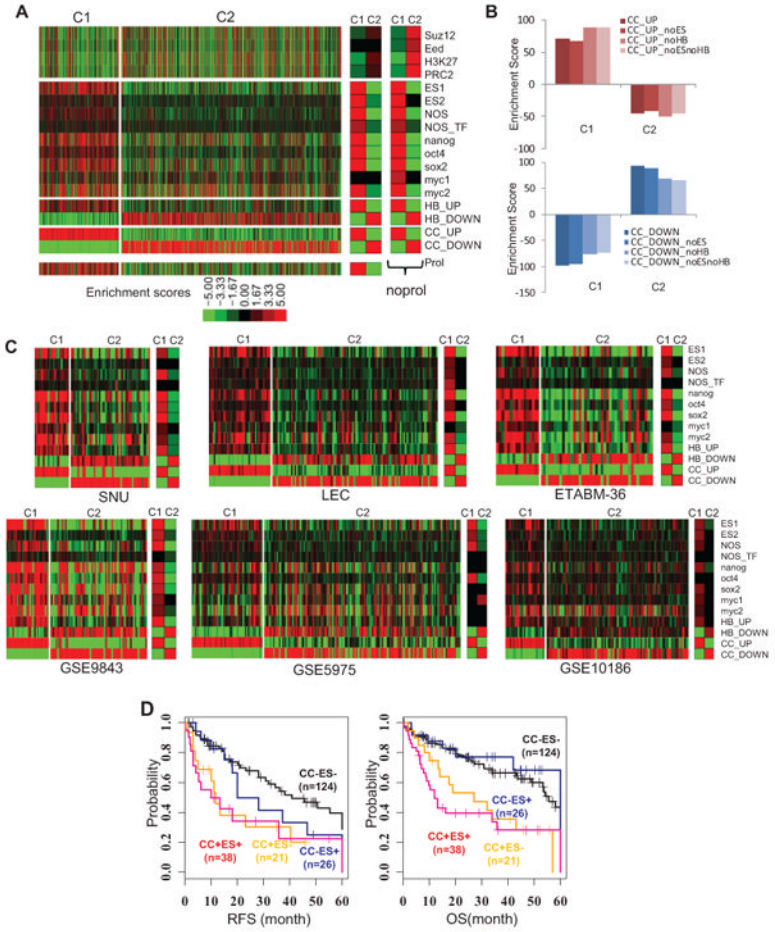

Fig. 4. Comparison of the CC signature with ES signatures

A. The enrichment of ES signatures and polycomb target gene sets in C1 and C2 classes in HCCcomp are shown. The enrichment scores of ES signatures without proliferation signature (no_prol) are shown (*right bar*). **B.** Bar plots for the enrichment scores of the CC_UP (*top*) and CC_DOWN (*bottom*) signatures and the CC_UP and CC_DOWN signatures subtracted by ES (noES), HB (noHB), or ES and HB (noESnoHB) are shown. **C.** The enrichment ES signatures in 6 independent data sets are shown. For each data set, the group enrichment in C1 and C2 tumors are indicated in the right bars (P<0.05). **D.** Kaplan-Meir plots analyses for recurrence-free survival (RFS, *left*) and overall survival (OS, *right*) based on the expression status of CC and ES signatures in the integrated LEC and SNU data sets (n=209). The CC^+ represents C1 tumors, and ES^+ represents the tumors express ES1 signature.

Table 1

Univariate and multivariate analyses for the selected clinical factors in validation data set

Clinico-pathological features	Number of patients	Hazard Ratio	95% Confidence Interval	P-value (recurrence-free survival)	Hazard Ratio	95% Confidence Interval	P-value (overall survival)
Univariate analysis							
Sex (Male/Female)	102/37	1.30	0.68-2.47	0.423	1.38	0.83-2.30	0.213
Age (55≤/>55, year)	77/62	0.84	0.46-1.54	0.574	1.29	0.82-2.04	0.271
Size (5>/≤5, cm)	66/43	1.11	0.51-2.42	0.786	1.38	0.76-2.51	0.290
Grade (III,IV/I,II)	80/59	1.38	0.76-2.51	0.291	1.60	1.00-2.54	0.047
AFP (>300/≤300,ng/ml)	50/54	1.63	0.71-3.72	0.245	2.29	1.20-4.36	0.0097
CC signature (C1/C2)	39/100	2.86	1.48-5.52	0.001	2.67	1.66-4.29	0.00002
Multivariate analysis							
Sex (Male/Female)		1.28	0.47-3.52	0.63	1.01	0.44-2.32	0.99
Age (55≤/>55, year)		1.47	0.60-3.65	0.40	1.74	0.88-3.45	0.11
Size (5>/≤5, cm)		1.12	0.46-2.71	0.81	1.02	0.50-2.05	0.96
Grade (III,IV/I,II)		2.00	0.85-4.71	0.11	1.45	0.72-2.92	0.29
AFP (>300/≤300,ng/ml)		1.28	0.52-3.13	0.59	2.14	1.11-4.12	0.024
CC signature (C1/C2)		3.70	1.23-11.14	0.02	3.54	1.84-6.82	0.00015