nature medicine

A novel prognostic subtype of human hepatocellular carcinoma derived from hepatic progenitor cells

Ju-Seog Lee[1], Jeonghoon Heo[1], Louis Libbrecht[2], In-Sun Chu[3], Pal Kaposi-Novak[1], Diego F Calvisi[1], Arsen Mikaelyan[1], Lewis R Roberts[4], Anthony J Demetris[5], Zongtang Sun[6], Frederik Nevens[2], Tania Roskams[2] & Snorri S Thorgeirsson[1]

The variability in the prognosis of individuals with hepatocellular carcinoma (HCC) suggests that HCC may comprise several distinct biological phenotypes. These phenotypes may result from activation of different oncogenic pathways during tumorigenesis and/or from a different cell of origin. Here we address whether the transcriptional characteristics of HCC can provide insight into the cellular origin of the tumor. We integrated gene expression data from rat fetal hepatoblasts and adult hepatocytes with HCC from human and mouse models. Individuals with HCC who shared a gene expression pattern with fetal hepatoblasts had a poor prognosis. The gene expression program that distinguished this subtype from other types of HCC included markers of hepatic oval cells, suggesting that HCC of this subtype may arise from hepatic progenitor cells. Analyses of gene networks showed that activation of AP-1 transcription factors in this newly identified HCC subtype might have key roles in tumor development.

HCC is the fifth most common cancer in the world[1]. The incidence of HCC has doubled in the United States over the past 25 years and the incidence and mortality rates of HCC seem likely to double again over the next 10–20 years[2]. Individuals with HCC have a highly variable clinical course[3,4]. The prognostic variability of individuals with HCC supports the notion that HCC comprises several biologically distinct subgroups. This variability probably reflects a molecular heterogeneity that has not been appreciated from methods traditionally used to characterize HCC. Improving the classification of individuals with HCC would at minimum improve the application of currently available treatment modalities and at most offer new treatment strategies.

Cancer cells evolve from normal cells after accumulation of genetic and epigenetic alterations[5]. Although the gene expression patterns in cancer cells reflect these alterations, a considerable fraction of the gene expression program of the cancer cells is characteristic of the non-transformed cellular lineages from which the cancer originated[6]. Analysis of gene expression profiles of cancer cell lines indicated that neither physiological adaptation *in vivo* nor experimental adaptation *in vitro* were sufficient to overwrite the gene expression programs established during development[7]. These data suggest that the global gene expression profiles of tumors may provide crucial information on the cellular origin of the tumors. As it is established that HCC can have its origin in both adult hepatocytes and hepatic progenitor cells (for review see ref. 8), we decided to test whether global gene expression analysis of human HCC would identify subtypes of HCC derived from hepatocytes and hepatic progenitor cells. We adopted an experimental strategy involving the generation of gene expression data from multiple species suitable for integration and cross-comparison.

RESULTS

Cross-comparison of gene expression data

As an extension of our previous analysis strategy[9], we conducted a study to stratify individuals with HCC into homogeneous groups by integrating gene expression data from three different species. Using only orthologous genes, we integrated gene expression data from rat fetal hepatoblasts and adult hepatocytes and 61 cases of HCC from Chinese individuals. We sought to determine the fraction of human HCC that shares gene expression patterns with fetal hepatoblasts. Gene expression data from 39 cases of mouse HCC from five different mouse tumor models were also integrated as controls for hepatocyte-originated HCC because all of the transgenes in mouse models are under the control of the hepatocyte-specific *Alb1* promoter[10]. In hierarchical clustering analysis of the integrated data, the two previously recognized subgroups of rat hepatoblasts and hepatocytes and two subgroups of mouse HCC were still separated from each other (**Fig. 1a**)[9,11]. Notably, 14 human HCCs were tightly coclustered with rat fetal hepatoblasts from embryonic day (E)13 to E16 (hepatoblast

[1]Lab of Experimental Carcinogenesis, Center for Cancer Research, National Cancer Institute, National Institutes of Health, 37 Convent Drive, Room 4146, Bethesda, Maryland 20892, USA. [2]Departments of Morphology and Molecular Pathology, University of Leuven, Minderbroedersstraat 12, B-3000 Leuven, Belgium. [3]National Genome Information Center, Korea Research Institute of Bioscience and Biotechnology, 52 Eoeun-dong, Yuseong-gu, Taejun, 305-333, Korea. [4]Division of Gastroenterology and Hepatology, Mayo Clinic, 200 First Street SW, Rochester, Minnesota 55905, USA. [5]Thomas E. Starzl Transplant Institute, University of Pittsburgh Medical Center, UPMC Montefiore E-741, 200 Lothrop Street, Pittsburgh, Pennsylvania 15213, USA. [6]National Laboratory of Molecular Oncology, Cancer Institute, Chinese Academy of Medical Sciences, P.O. Box 2258, Beijing 100021, China. Correspondence should be addressed to S.S.T. (snorri_thorgeirsson@nih.gov).

Received 18 October 2005; accepted 3 February 2006; published online 12 March 2006; doi:10.1038/nm1377

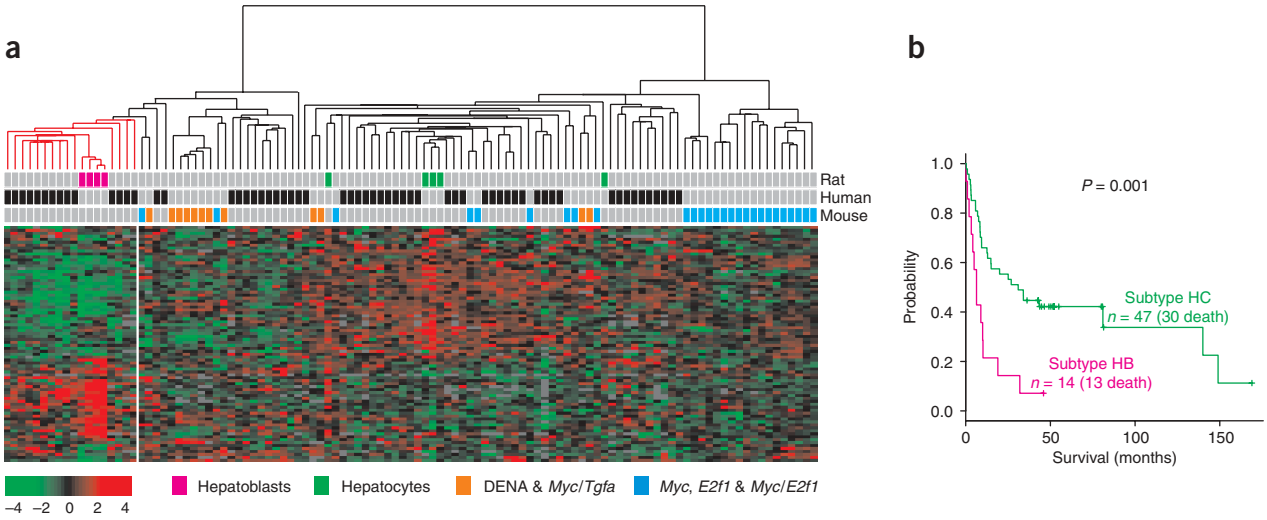

Figure 1 Hierarchical cluster analysis of integrated gene expression data from rat hepatoblasts and hepatocytes, and from mouse and human HCC. (**a**) A dendrogram and heat-map overview of the two-way hierarchical cluster analysis of gene expression data from 109 samples (9 rat samples from fetal hepatoblasts and adult hepatocytes, 39 mouse HCC from five different mouse models, and 61 HCCs from Chinese individuals), using 80 orthologous genes. Columns represent individual samples and rows represent each gene. Each cell in the matrix represents the expression level of a gene feature in an individual sample. Red and green in cells reflect high and low expression levels, respectively, as indicated in the scale bar ($\log_2$-transformed scale). Colored bars between dendrogram and heat-map represent rat, mouse and human samples, as indicated at the bottom of diagram. (**b**) Kaplan-Meier plots of overall survival of individuals with the HB and HC subtypes of HCC from hierarchical clustering analysis of integrated gene expression data. $P = 0.001$, log-rank test. +, censored data.

or HB subtype), whereas the rest of the human HCCs were either coclustered with rat hepatocytes or were excluded from the hepatoblast core cluster (hepatocyte or HC subtype). Members of this newly identified HB subtype of HCC also resided in a compact and easily separable three-dimensional space when viewed by a three-dimensional multidimensional scaling plot based on their overall similarity of expression patterns (data not shown), indicating a robust subtype. Coclustering of some human HCC with rat hepatoblasts suggested that HCC of the HB subtype might arise from bipotential hepatic progenitor cells that can differentiate into both hepatocytes and cholangiocytes[12,13], whereas the majority of human HCC might be derived from hepatocytes. Exclusion of mouse HCC from the hepatoblast core cluster also strongly supported this notion.

The two subtypes were similar with respect to the individuals' gender, age, serological level of α fetoprotein (AFP; >300 ng/ml), Edmonson grade and presence of cirrhosis in surrounding tissues (data not shown). Kaplan-Meier plots, however, indicated poorer survival in individuals with the HB subtype ($P < 0.001$, log-rank test) when compared with those who had the HC subtype (**Fig. 1b**). Thus, molecular differences that reflect the origin of tumor cells were associated with a marked difference in clinical outcome.

Construction and validation of prediction models

Having defined by unsupervised cluster analysis two distinctive subtypes of human HCC that reflect the origin of tumor cells, we next sought to validate these results using gene expression data from 78 white individuals with HCC. We applied five different statistical methods to determine whether gene expression patterns could be used to predict the likelihood of the HB or HC subtypes in the HCC validation cohort (**Fig. 2a**). Briefly, we identified the most differentially expressed genes between the two subtypes in the Chinese cohort (the training set). These genes were combined to form a series of classifiers that estimate the probability that a particular HCC belongs to subtype HB or HC. The number of genes in the classifiers was optimized to minimize misclassification errors during the leave-one-out cross-validation of the tumors in the training set. When applied to the white cohort (the test set), all five models produced consistent prediction patterns. All Kaplan-Meier plots in the test set showed significant differences between survival of individuals with subtype HB and HC that were independently predicted by the six different prediction models (**Fig. 2b**). These results showed not only strong association of gene expression patterns with the survival of the individuals, but also a robust reproducibility of these gene expression–based predictors.

Three distinct subgroups of HCC

Having identified two subtypes of HCC that may reflect the origin of tumor cells, we next sought to assess the relationship of the newly identified subtypes with two previously recognized subclasses of HCC (clusters A and B) that were classified by applying unsupervised cluster analysis of genome-wide gene expression patterns[14]. We first determined whether we could identify the two previously recognized subclasses of HCC in larger and more heterogeneous cohorts. We applied hierarchical cluster analysis to gene expression data from the pooled 139 human HCCs, and showed the presence of two previously recognized subclasses (**Fig. 3a**). Kaplan-Meier plots showed a significant difference between two subclasses (cluster A and B) in overall survival ($P < 0.001$; **Fig. 3b**) as well as between HB and HC subtypes ($P < 0.001$; **Fig. 3c**). Although the 22 HCCs in the HB subtype identified by both unsupervised and supervised methods (**Figs. 1** and **2**) were clustered together and shared gene expression patterns that are characteristic of fast-growing cells with the rest of the HCCs in cluster A (proliferation signature; **Fig. 3a**), HB and HC subtypes in cluster A were significantly different in overall survival ($P = 0.006$; **Fig. 3d**). Based on two independent signatures, hepatoblast gene expression signature and genome-wide global gene expression signature, we

further divided individuals with HCC into three subgroups and found a significant association of overall survival with the subgroups ($P < 0.001$; **Fig. 3e**). To further assess the clinical significance of the three subgroups, we compared the recurrence rate in the subgroups and found a significant association of recurrence rate with subgroups as well ($P < 0.001$; **Fig. 3f**). These findings support the hypothesis that the subgroups represent distinct categories of HCC that originate from malignant transformation of different cellular lineages in the liver and/or dissimilar mechanisms of transformation.

Biological insights into the HB subtype of HCC

As all HB subtype HCCs are part of cluster A and shared gene expression patterns with the rest of the HCCs in cluster A (**Fig. 3a**), we applied Venn diagram comparison of two gene lists to select gene expression patterns unique for the HB subtype and independent of the difference between clusters A and B. First, we generated two different gene lists by applying the two-sample *t*-test ($P < 0.001$; **Fig. 4a**). Gene list C represents the genes that were differentially expressed between clusters A and B, and gene list H represents the genes that were differentially expressed between the HB and HC subtypes. When we compared the two gene lists, we found three different patterns: C not H (2,836 genes), C and H (2,366 genes) and H not C (907 genes; **Fig. 4b**). Genes in the C not H category largely reflect the characteristics of cell proliferation and loss of liver functions in HCC (data not shown). Genes in the C and H category have both cluster-specific as well as subtype-specific expression patterns. Although expression patterns of many genes were similar in all HCCs of cluster A, we observed a more pronounced alteration of expression in some genes in HCCs of subtype HB. Notably, although expression of genes in the H not C category showed considerable differences between subtypes HB and HC in cluster A, we observed almost no differences between clusters A and B in subtype HC, signifying a unique gene expression signature that is only detectable in subtype HB (**Fig. 4b**).

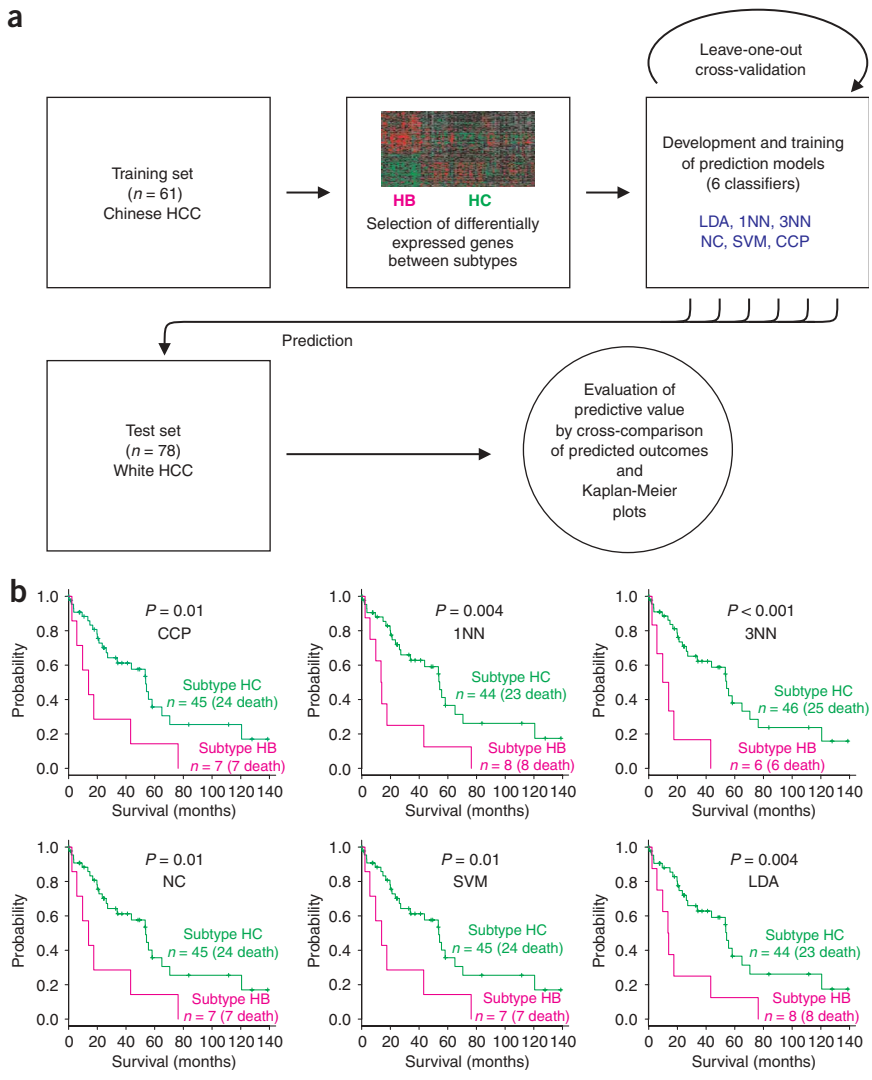

Figure 2 Construction of prediction models and evaluation of predicted outcome. (**a**) Schematic overview of the strategy used for the construction of prediction models and evaluation of predicted outcomes based on gene expression signatures. (**b**) Kaplan-Meier plots of overall survival of individuals with HCC in validation set predicted by compound covariate predictor (CCP), one nearest neighbor (1NN), three nearest neighbor (3NN), nearest centroid (NC), support vector machines (SVM) and linear discriminator analysis (LDA). The differences between groups were significant as indicated (log-rank test). +, censored data.

To identify the predominant signaling networks that might drive tumorigenesis in the HB subtype, we next carried out pathway analysis on the 907 genes unique to the HB subtype (**Fig. 4b**) using the Ingenuity Pathway Analysis. This analysis showed a series of putative networks of which the top 10 with high score (>10) are listed in **Supplementary Table 1** online. Functional connectivity of the top network (network 1) showed a strong over-representation of the AP-1 transcription factors FOS, FOSL2 and JUNB, suggesting that they might have important roles in tumorigenesis in HCC subtype HB. To further examine the involvement of AP-1 transcription factors, we used another pathway analysis tool, PathwayAssist. In addition to the 907 genes previously identified, we included 795 genes that showed significant differences between HB and HC subtypes in cluster A ($P < 0.001$, the two-sample *t*-test) to obtain a more comprehensive view of the gene networks in subtype HB (**Supplementary Fig. 1** online). To this end, we focused on genes that are connected to JUN-FOS heterodimers–AP-1 complex, in the networks, and found that many of the direct downstream targets of JUN and FOS are upregulated in the HB subtype (**Fig. 5** and **Supplementary Note** online). Although expression of FOS is higher in the HB subtype, expression of JUN is not altered in either HB or HC subtypes, indicating that the activation of AP-1 in the HB subtype is driven by JUN binding partners or by other mechanisms of JUN activation such as mutations or N-terminal phosphorylation.

The gene expression program that distinguishes the HB subtype from other types of HCC includes well-known markers of hepatic oval cells (adult hepatic progenitor cells).

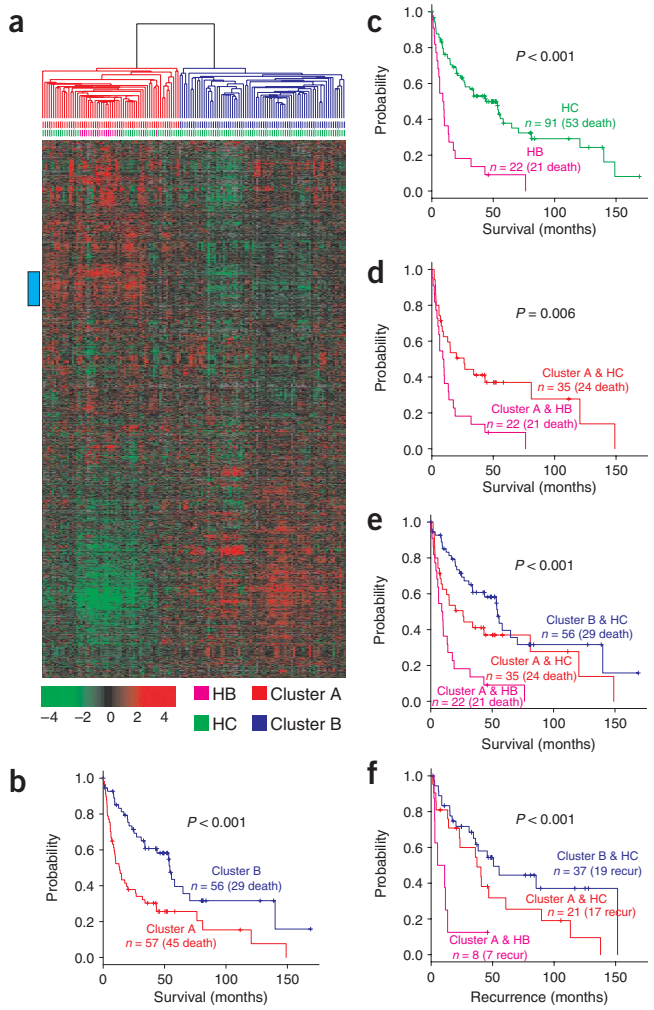

Figure 3 Three distinctive subtypes of HCC defined by gene expression patterns. (**a**) Hierarchical cluster analysis of 139 human HCCs. We selected genes with <30% of missing expression data across the tissues, as well as an expression ratio of at least twofold difference relative to reference in at least 10% (14 HCCs) of tissues for cluster analysis (3,992 genes). The light blue bar on the left indicates the cell proliferation gene expression signature. (**b**) Kaplan-Meier plots of overall survival of individuals with HCC grouped solely on the basis of gene expression profiling (clusters A and B). $P < 0.001$, log-rank test. (**c**) Kaplan-Meier plots of overall survival of individuals with HCC grouped on the basis of the similarity of gene expression patterns to rat hepatoblast or hepatocyte gene expression signatures (subtypes HB and HC). We identified 22 HB subtypes of HCC by either hierarchical cluster analysis of multispecies gene expression data set (**Fig. 1**) or prediction models (**Fig. 2**; $P < 0.001$). (**d**) Kaplan-Meier plots of overall survival of individuals with HB or HC subtype of HCC in cluster A ($P = 0.006$). (**e**) Kaplan-Meier plots of overall survival of three different subtypes of HCC identified by two independent analyses ($P < 0.001$). (**f**) Kaplan-Meier plots of HCC recurrence of three different subtypes of HCC identified by two independent analyses. Recurrence data were available from only 66 individuals ($P < 0.001$). +, censored data.

Expression of KRT7, KRT19 and VIM is significantly higher in the HB subtype of HCC in individuals of cluster A ($P < 0.001$, two-sample *t*-test), probably reflecting the derivation of these HCCs from hepatic progenitor cells (**Supplementary Fig. 2** online).

As our previous study showed a considerable correlation of prognosis with apoptosis and cell proliferation[14], we measured the proliferation and apoptosis rates of HCC belonging to subtypes HB and HC in individuals of cluster A. When the proliferation rates of tumor cells were assessed by immunostaining with antibodies to Ki-67, we found no difference between the HB and HC subtypes of cluster A HCCs, whereas proliferation rates were significantly higher in cluster A than cluster B regardless of subtype of HCCs in cluster A ($P < 0.001$, two-sample *t*-test; **Supplementary Fig. 3** online). Likewise, the apoptosis index showed similar differences (**Supplementary Fig. 3** online). Both results suggest that cell proliferation and apoptosis do not account for the difference in prognosis in the two subtypes of cluster A HCCs. We found, however, that five genes, *MMP1*, *PLAUR*, *TIMP1*, *CD44* and *VIL2* (which encodes EZRIN), all downstream targets of AP-1, are known to be involved in invasion and metastasis (**Fig. 5** and **Supplementary Note** online)[15–20]. EZRIN is a member of the ERM family that is crucial to maintaining cell adhesion, and recent studies identified it as a key component in metastasis of pediatric cancers[19–21]. EZRIN directly interacts with the cytoplasmic tail of CD44, which has also been suggested in many studies to be involved in metastasis[22]. This observation suggests that the HB subtype might have a more invasive phenotype, which may contribute to the poorer prognosis in this subtype.

Further biological insight into the HB subtype of HCC was obtained by correlating molecular subtype with clinicopathological features of HCCs (**Supplementary Table 2** online). This analysis showed that the HB subtype did not form a distinct morphological entity, even though HB subtype tended to be more invasive than the HC subtype.

Prognostic utility of hepatoblast signature in HCC

Previous and current studies generated two independent gene expression signatures in HCC[14]. The proliferation signature identified previously was the dominant characteristic that permitted the stratification of individuals into cluster A and B (**Fig. 3a**,**b**), whereas the hepatoblast stem cell signature identified here made it possible to stratify individuals into HB and HC subtypes (**Figs. 1** and **2**). To evaluate the prognostic power of the two independent signatures, we applied univariate and multivariate analysis of the signatures with known clinical and pathologic risk factors for progression of HCC. In agreement with previous reports[23,24], several tumor characteristics were associated with overall survival and recurrence in univariate Cox proportional hazards analysis (**Supplementary Table 3** online). Multivariate analyses that included all relevant pathological variables and the molecular subtype showed that only the HB subtype was independently associated with both recurrence and worse survival. Our findings suggest that the HB subtype retains its prognostic relevance even after the 'classical' pathological prognostic features have been taken into account. Moreover, hepatoblast stem cell gene expression signature was revealed independently of HCC gene expression patterns and clinicopathological features of the tumors, indicating that the potential clinical utility of the signature might come from better mechanistic understanding of HCC progression.

DISCUSSION

Here, we uncovered a subtype of HCC that has not been previously recognized by conventional diagnostic methods. Unequal distribution of expression patterns of genes enriched in fetal hepatoblasts and of direct downstream targets of AP-1 in subtypes with different survival suggest that distinct molecular features of HCC reflected in gene expression patterns govern the clinical phenotypes[9,14,25].

The variability in the prognosis of individuals with HCC probably reflects molecular heterogeneity. The heterogeneous features of HCC

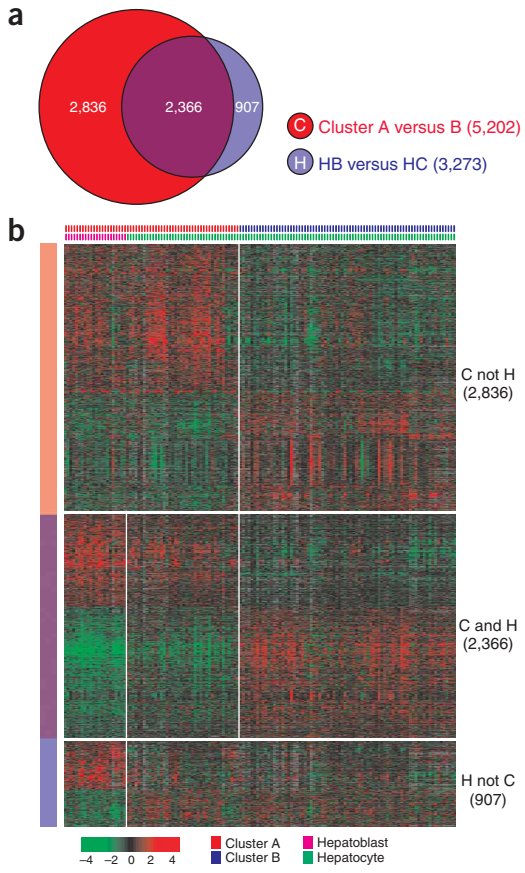

Figure 4 Cross-comparison of gene lists from two independent statistical tests. (**a**) Venn diagram of genes selected by univariate test (two-sample *t*-test) with multivariate permutations test (1,000 random permutations). The red circle (gene list C) represents genes differentially expressed between cluster A and B. The blue circle (gene list H) represents genes differentially expressed between subtypes HB and HC. Expression of 907 genes shows exclusive differences between HB and HC subtypes. We applied a cutoff *P* value of <0.001 to retain genes whose expression is significantly different between the two groups of tissues examined. In each comparison, the maximum allowed number of false-positive genes was 10 and the probability of getting the selected number of genes by chance if there are no real differences between the groups was 0. (**b**) Expression patterns of selected genes in the Venn diagram. Red and blue bars at the left side of the heat map represent genes in the Venn diagram. Colored bars at the top of the heat map represent tissues as indicated.

may also explain the high level of resistance against a range of therapeutic agents[26]. The high level of redundancy enables tumors to promote growth and survival in several ways. Therefore, it is difficult to estimate the efficacy of a treatment when it is applied to the entire heterogeneous group of individuals if the nature of heterogeneity is not delineated. By applying hierarchical clustering analysis of gene expression patterns from human HCC, mouse HCC and rat fetal hepatoblasts and adult hepatocytes, we identified a new prognostic subtype of HCC that shares gene expression patterns with fetal hepatoblasts. The HB subtype is distinguished from other types of HCC by the differential expression of hundreds of genes, and the robustness of this gene expression signature in the HB subtype was validated in an independent cohort of individuals with HCC.

Shared gene expression patterns of the HB subtype and fetal hepatoblasts suggest that this subtype of HCC may arise from adult hepatic progenitor cells, which have similar bipotential capacity as hepatoblasts to differentiate into mature hepatocytes or cholangiocytes. Further support for this idea is supplied by the finding that expression of well-known markers of hepatic oval cells, the early progenitors of adult liver stem cells, is found in the HB subtype of HCC. Lack of a hepatoblast-specific gene expression signature in mouse HCC is not surprising. All of the HCCs in our mouse models may arise from differentiated hepatocytes as a result of overexpression of oncogenes bay hepatocyte-specific minimum *Alb* promoter that is fully activated between 2 and 3 weeks after birth[27–29]. In addition, HCCs from mouse models are negative for hepatic oval cell markers (data not shown), suggesting that the contribution of oval cells to hepatocarcinogenesis in our mouse models is minimal or absent. These observations indicate that the gene expression signatures that distinguish the HB subtype from other types of HCC are unlikely to reflect the mature hepatocytic lineage.

Previous studies in diffuse large B-cell lymphoma and T-cell acute lymphoblastic leukemia indicated that the cellular origins of a tumor largely dictate the clinical outcome of patients[6,30], because mitogenic, motogenic and morphogenic responses as well as the propensity for apoptosis may vary at different stages of normal differentiation. In rodent liver injury models, oval cells emerge from the canals of Hering and later invade the entire lobular parenchyma[31]. Higher expression of genes involved in invasive phenotype (*MMP1*, *PLAUR*, *TIMP1*, *CD44* and *VIL2*) may reflect the cellular origin of these tumors, and account for the poor prognosis of individuals with the HB subtype.

Our network–based pathway analyses of gene expression have provided important insights into the pathogenesis of the HB subtype

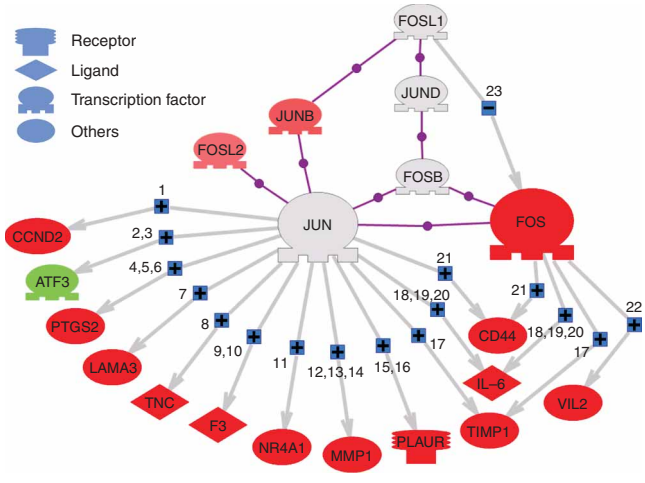

Figure 5 Gene networks of AP-1 transcription factors in the HB subtype of HCC. Genes interconnected with FOS and JUN (**Supplementary Fig. 1** online) were selected for further evaluation of the involvement of FOS and JUN in development of HCC of the HB subtype. Up- and downregulated genes in the HB subtype are indicated in red and green, respectively. Genes in gray are not on the list but are associated with the regulated genes. Gray lines and arrows represent the direction of transcriptional regulation, and + and – indicate positive and negative regulation of gene expression, respectively. Purple lines represent known physical interactions between genes connected. All references used to construct the networks are available in the **Supplementary Note** online and the numbers on the lines correspond to the references therein.

of HCC. Enrichment of predicted JUN and FOS activity in the HB subtype led us to hypothesize that the AP-1 complex might be the major driving force in tumorigenesis of the HB subtype. Previous studies showed that *Jun* is essential for normal hepatogenesis during embryonic development and it seems to be crucial for initiation of HCC development in mouse, as well[32,33]. Cell transformation frequently resurrects preexisting but dormant signaling pathways that were active during embryonic development[34]. Thus, our finding supports the growing understanding that signaling pathways that control embryonic development in vertebrates are also important in carcinogenesis in humans. Further studies, however, will be necessary to determine the roles of AP-1 in the newly identified HB subtype of HCC.

Although conventional application of morphological diagnosis supplemented with immunostaining of a few markers has been successful in broadly classifying HCC, it lacks the resolution needed to comprehensively identify prognostic subgroups of HCC. It also does not provide biological insight into the different subgroups, which is an essential requirement for the development of new therapeutic strategies. The identification of the HB subtype suggests that the use of interspecies comparison of gene expression patterns can improve the accuracy of the molecular classification of HCC[10,35].

Our observations in this study extend previous studies that postulated the presence of progenitor cell–derived HCC[8,36,37]. In fact, expression of two frequently used markers (*KRT7* and *KRT19*) for the progenitor cell origin of human HCC is significantly higher in the HB subtype. Our current data, however, do not rule out the possibility that the unique expression profile of HB subtype could be the result of malignant transformation of mature hepatocytes with concomitant dedifferentiation that results in the acquisition of progenitor-cell features. Regardless of how the HB subtype of HCC acquired the characteristic gene expression patterns, the unique gene set in the HB subtype will provide an opportunity to identify a set of markers that can accurately recognize by conventional immunostaining methods the poor prognosis of individuals with the HB subtype of HCC. In this context, it is of interest that we and other groups have shown that HCCs that immunostain positive for KRT19 have a worse prognosis than KRT19-negative HCCs[38–40]. Furthermore, it may be possible to use this set of markers to predict the clinical course of preneoplastic lesions (for example, dysplastic nodules) in the liver.

By applying two independent gene expression signatures, we were able to divide individuals with HCC into three subgroups characterized by statistically significant differences in clinical outcome. These findings support the notion that multiple molecular pathways dictate the development and different clinical outcomes of HCC. Our finding also indicates that the molecular features of HCC such as prognostic gene expression signatures are present at the time of diagnosis. Therefore, the use of gene expression profiling promises to improve molecular classification and prediction of outcomes in HCC. Furthermore, molecular stratification of individuals with HCC into homogeneous subgroups may provide opportunities for the development of new treatment modalities.

METHODS

Subjects and tissue samples. We obtained tumor tissues from 139 individuals undergoing surgical treatment for HCC from two ethnic groups (61 Chinese and 78 white). The median duration of follow up was 23.4 months; during this period, 74 individuals died. The median age of the individuals was 57, and 73.3% were male. Of the 78 white individuals, 17 underwent liver transplantation and 9 received palliative treatment (data from these 26 individuals were not included in the analysis of survival or tumor recurrence). Animal housing and care were in accordance with the guidelines from the Animal Care and Use Committee at the US National Cancer Institute. This study was approved by the Institutional Review Board of the Mayo Clinic, the Cancer Institute of the Chinese Academy of Medical Sciences, the University of Leuven and the US National Cancer Institute.

DNA microarrays. We obtained the Human Array-Ready Oligo Set (Version 2.0) containing 70-mer probes of 21,329 genes from Qiagen and oligo microarrays were produced at the Advanced Technology Center at the National Cancer Institute.

We isolated total RNAs from frozen liver tissue using CsCl density-gradient centrifugation. We pooled total RNAs from 19 normal livers for use as the reference for all microarray experiments. To obtain gene expression profile data from 139 human HCCs, we used 20 μg of total RNA from tissues to derive fluorescently (Cy-5 or Cy-3) labeled cDNA. We carried out at least two hybridizations for each tissue using a dye-swap strategy to eliminate dye-labeling bias as previously described[14].

We used previously published data for mouse HCC models and for rat hepatoblasts and hepatocytes[9,11]. We used average gene expression ratios from triplicate rat data sets.

Data analysis. We performed data transformation and normalization of gene expression as previously described[9,14]. As three different microarray platforms were used to generate gene expression profiles of rat hepatoblast and hepatocytes and of mouse and human HCCs, we selected orthologous genes that were present in all microarrays by using curated mammalian orthology from the Jackson Laboratory. A total of 80 orthologous genes were present in all microarrays. Because the rat gene expression data were generated using mouse cDNA microarrays[11], we pooled mouse and rat data together and considered them a single data set. Before integration of the two data sets, we standardized the expression of each gene to a mean of 0 and a standard deviation of 1 independently in both data sets. We applied hierarchical clustering analysis as previously described[9,14].

To avoid the idiosyncrasies of any particular prediction algorithm, we applied five different prediction methods to determine the validity of the HB and HC subtypes of HCC in Chinese individual: linear discriminator analysis, support vector machines, nearest centroid, nearest neighbor and compound covariate predictor. A gene expression data set from two predefined subclasses of human HCC from China was used to develop and train prediction methods. We identified the most differentially expressed genes between subtype HB ($n = 14$) and subtype HC ($n = 47$) in the human data set. We combined these genes (941 genes, $P < 1.0 \times 10^{-4}$, two-sample *t*-test) to form a series of classifiers that estimate the probability that a particular HCC tissue belongs to subtype HB or HC. The number of genes in the classifiers was optimized to minimize misclassification errors during the leave-one-out cross-validation of the data set from Chinese individuals.

To select genes that were differentially expressed between two groups of tissues, we used a class-comparison tool in BRB ArrayTools (v3.2) as a method for the two-sample *t*-test with the estimation of false discovery rate. We chose a cutoff ($P < 0.001$) to retain a maximum allowed number of false-positive genes to <10.

All statistical analyses were performed in R (version 2.0.1). We estimated the probabilities of overall survival according to the Kaplan-Meier method.

Pathway analysis. We used two different approaches to explore the functional relationships among the genes with altered expression in the HB subtype. The first approach, carried out with the Ingenuity Pathway Analysis tool, examined functional associations among genes and generated the gene networks with high significance on the basis that they had more of the interconnected genes present than would be expected by chance. The significance of each network was estimated by the scoring system provided by Ingenuity. The scores are determined by the number of differentially expressed genes within each of the networks and the strength of the associations among network members. Once over-represented genes that are functionally relevant in gene networks were identified, we validated their functional association by using the independent pathway analysis tool PathwayAssist (version 3.0, Ariadne Genomics).

Quantification of proliferation and apoptosis. We performed immunohistochemical staining and quantitative measurement of apoptosis as previously described[9].

Assessments of microvascular invasion. All assessments were performed on H&E-stained sections made from the paraffin-embedded biopsies with clear tumor edges by two investigators (L.L. and T.R.) using a multiheaded microscope. Microvascular invasion was defined as the unequivocal presence of tumor cells inside the lumen of a vessel which is recognized by the focal presence of endothelial lining.

URLs. The Jackson Laboratory: http://www.informatics.jax.org. BRB ArrayTools: http://linus.nci.nih.gov/BRB-ArrayTools.html.

Accession codes. Gene Expression Omnibus accession codes: human microarray platform, GPL1528; human HCC microarray data, GSE1898 and GSE4024; mouse microarray platform, GPL1529; mouse HCC microarray data, GSE1897.

Note: Supplementary information is available on the Nature Medicine website.

ACKNOWLEDGMENTS

This research was supported by the Intramural Research Program of the US National Institutes of Health, National Cancer Institute, Center for Cancer Research. L. Libbrecht is a "postdoctoraal onderzoeker" of the "F.W.O.-Vlaanderen." I.-S. Chu is supported by a grant from the Ministry of Science and Technology (21C Frontier Functional Human Genome Project), Korea.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Published online at http://www.nature.com/naturemedicine/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Parkin, D.M., Bray, F., Ferlay, J. & Pisani, P. Estimating the world cancer burden: Globocan 2000. *Int. J. Cancer* **94**, 153–156 (2001).
2. El Serag, H.B. & Mason, A.C. Rising incidence of hepatocellular carcinoma in the United States. *N. Engl. J. Med.* **340**, 745–750 (1999).
3. Bruix, J., Boix, L., Sala, M. & Llovet, J.M. Focus on hepatocellular carcinoma. *Cancer Cell* **5**, 215–219 (2004).
4. Llovet, J.M. *et al.* Natural history of untreated nonsurgical hepatocellular carcinoma: rationale for the design and evaluation of therapeutic trials. *Hepatology* **29**, 62–67 (1999).
5. Hanahan, D. & Weinberg, R.A. The hallmarks of cancer. *Cell* **100**, 57–70 (2000).
6. Alizadeh, A.A. *et al.* Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* **403**, 503–511 (2000).
7. Ross, D.T. *et al.* Systematic variation in gene expression patterns in human cancer cell lines. *Nat. Genet.* **24**, 227–235 (2000).
8. Roskams, T.A., Libbrecht, L. & Desmet, V.J. Progenitor cells in diseased human liver. *Semin. Liver Dis.* **23**, 385–396 (2003).
9. Lee, J.S. *et al.* Application of comparative functional genomics to identify best-fit mouse models to study human cancer. *Nat. Genet.* **36**, 1306–1311 (2004).
10. Lee, J.S., Grisham, J.W. & Thorgeirsson, S.S. Comparative functional genomics for identifying models of human cancer. *Carcinogenesis* **26**, 1013–1020 (2005).
11. Petkov, P.M. *et al.* Gene expression pattern in hepatic stem/progenitor cells during rat fetal development using complementary DNA microarrays. *Hepatology* **39**, 617–627 (2004).
12. Dabeva, M.D. & Shafritz, D.A. Hepatic stem cells and liver repopulation. *Semin. Liver Dis.* **23**, 349–362 (2003).
13. Thorgeirsson, S.S. & Grisham, J.W. Overview of recent experimental studies on liver stem cells. *Semin. Liver Dis.* **23**, 303–312 (2003).
14. Lee, J.S. *et al.* Classification and prediction of survival in hepatocellular carcinoma by gene expression profiling. *Hepatology* **40**, 667–676 (2004).
15. Chambers, A.F. & Matrisian, L.M. Changing views of the role of matrix metalloproteinases in metastasis. *J. Natl. Cancer Inst.* **89**, 1260–1270 (1997).
16. Andreasen, P.A., Egelund, R. & Petersen, H.H. The plasminogen activation system in tumor growth, invasion, and metastasis. *Cell. Mol. Life Sci.* **57**, 25–40 (2000).
17. Chang, C. & Werb, Z. The many faces of metalloproteases: cell growth, invasion, angiogenesis and metastasis. *Trends Cell Biol.* **11**, S37–S43 (2001).
18. Jothy, S. CD44 and its partners in metastasis. *Clin. Exp. Metastasis* **20**, 195–201 (2003).
19. Yu, Y. *et al.* Expression profiling identifies the cytoskeletal organizer ezrin and the developmental homeoprotein Six-1 as key metastatic regulators. *Nat. Med.* **10**, 175–181 (2004).
20. Khanna, C. *et al.* The membrane-cytoskeleton linker ezrin is necessary for osteosarcoma metastasis. *Nat. Med.* **10**, 182–186 (2004).
21. Curto, M. & McClatchey, A.I. Ezrin.a metastatic detERMinant? *Cancer Cell* **5**, 113–114 (2004).
22. Martin, T.A., Harrison, G., Mansel, R.E. & Jiang, W.G. The role of the CD44/ezrin complex in cancer metastasis. *Crit. Rev. Oncol. Hematol.* **46**, 165–186 (2003).
23. Kondo, K. *et al.* Risk factors for early death after liver resection in patients with solitary hepatocellular carcinoma. *J. Hepatobiliary Pancreat. Surg.* **12**, 399–404 (2005).
24. Gouillat, C., Manganas, D., Saguier, G., Duque-Campos, R. & Berard, P. Resection of hepatocellular carcinoma in cirrhotic patients: longterm results of a prospective study. *J. Am. Coll. Surg.* **189**, 282–290 (1999).
25. Lee, J.S. & Thorgeirsson, S.S. Genome-scale profiling of gene expression in hepatocellular carcinoma: classification, survival prediction, and identification of therapeutic targets. *Gastroenterology* **127**, S51–S55 (2004).
26. Kitano, H. Cancer robustness: tumour tactics. *Nature* **426**, 125 (2003).
27. Postic, C. & Magnuson, M.A. DNA excision in liver by an albumin-Cre transgene occurs progressively with age. *Genesis* **26**, 149–150 (2000).
28. Pinkert, C.A., Ornitz, D.M., Brinster, R.L. & Palmiter, R.D. An albumin enhancer located 10 kb upstream functions along with its promoter to direct efficient, liver-specific expression in transgenic mice. *Genes Dev.* **1**, 268–276 (1987).
29. Huh, C.G. *et al.* Hepatocyte growth factor/c-met signaling pathway is required for efficient liver regeneration and repair. *Proc. Natl. Acad. Sci. USA* **101**, 4477–4482 (2004).
30. Ferrando, A.A. *et al.* Gene expression signatures define novel oncogenic pathways in T cell acute lymphoblastic leukemia. *Cancer Cell* **1**, 75–87 (2002).
31. Sarraf, C. *et al.* Cell behavior in the acetylaminofluorene-treated regenerating rat liver. Light and electron microscopic observations. *Am. J. Pathol.* **145**, 1114–1126 (1994).
32. Eferl, R. *et al.* Liver tumor development. c-Jun antagonizes the proapoptotic activity of p53. *Cell* **112**, 181–192 (2003).
33. Hilberg, F., Aguzzi, A., Howells, N. & Wagner, E.F. c-jun is essential for normal mouse development and hepatogenesis. *Nature* **365**, 179–181 (1993).
34. Logan, C.Y. & Nusse, R. The Wnt signaling pathway in development and disease. *Annu. Rev. Cell Dev. Biol.* **20**, 781–810 (2004).
35. Lee, J.S. & Thorgeirsson, S.S. Genetic profiling of human hepatocellular carcinoma. *Semin. Liver Dis.* **25**, 125–132 (2005).
36. Robrechts, C. *et al.* Primary liver tumour of intermediate (hepatocyte-bile duct cell) phenotype: a progenitor cell tumour? *Liver* **18**, 288–293 (1998).
37. Libbrecht, L., Desmet, V., Van Damme, B. & Roskams, T. The immunohistochemical phenotype of dysplastic foci in human liver: correlation with putative progenitor cells. *J. Hepatol.* **33**, 76–84 (2000).
38. Wu, P.C. *et al.* Classification of hepatocellular carcinoma according to hepatocellular and biliary differentiation markers. Clinical and biological implications. *Am. J. Pathol.* **149**, 1167–1175 (1996).
39. Uenishi, T. *et al.* Cytokeratin 19 expression in hepatocellular carcinoma predicts early postoperative recurrence. *Cancer Sci.* **94**, 851–857 (2003).
40. Durnez, A. *et al.* The clinicopathological and prognostic relevance of cytokeratin 7 and 19 expression in hepatocellular carcinoma. *Hepatology* (in the press).