

Comparison of gene expression in hepatocellular carcinoma, liver development, and liver regeneration

Tingting Li · Bingbing Wan · Jian Huang ·
Xuegong Zhang

Received: 25 November 2009 / Accepted: 4 March 2010 / Published online: 1 April 2010
© Springer-Verlag 2010

Abstract Proliferation of liver cells can be observed in hepatocarcinogenesis, at different stages of liver development, and during liver regeneration after an injury. Does it imply that they share similar molecular mechanisms? Here, the transcriptional profiles of hepatocellular carcinoma (HCC), liver development, and liver regeneration were systematically compared as a preliminary attempt to answer this question. From the comparison, we found that advanced HCC mimics early development in terms of deprived normal liver functions and activated cellular proliferation, but advanced HCC and early development differ in expressions of cancer-related genes and their transcriptional controls. HCC and liver regeneration demonstrate

different expression patterns as a whole, but regeneration is similar to dysplasia (pre-stage of HCC) in terms of their proximity to the normal state. In summary, of these three important processes, the carcinogenic progress carries the highest variance in expression; HCC pre-stage shares some resemblance with liver regeneration; and advanced HCC stage displays similarity with early development.

Keywords Hepatocellular carcinoma · Liver development · Liver regeneration · Microarray · Functional preference · Transcriptional regulation

Communicated by S. Hohmann.

Electronic supplementary material The online version of this article (doi:10.1007/s00438-010-0530-y) contains supplementary material, which is available to authorized users.

T. Li (✉)
Department of Biomedical Informatics,
Peking University Health Science Center,
38 Xueyuan Rd, Beijing 100191, China
e-mail: litt@hsc.pku.edu.cn

T. Li · X. Zhang
MOE Key Laboratory of Bioinformatics and Bioinformatics
Division, TNLIST/Department of Automation,
Tsinghua University, Beijing 100084, China

B. Wan · J. Huang
Chinese National Human Genome Center at Shanghai,
351 Guo shou-jing Road, Shanghai 201203, China

J. Huang
National Engineering Center for Biochip at Shanghai,
151 Libin Road, Zhangjiang Hi-Tech Park,
Pudong, Shanghai 201203, China

Introduction

The liver is one of the largest glands in the human body. It carries out a number of complex functionalities, such as detoxification, glycogen storage, and secretion of digestive fluid, which are all essential for life. Since there is no replacement or compensation for liver function, liver diseases are considered as a grave threat to human lives. Hepatocellular carcinoma (HCC) is one of the most common tumors in East Asia. Despite people's familiarity with the etiology of HCC, such as hepatitis virus infection, alcohol abuse, and aflatoxin B1 intoxication (Thorgeirsson and Grisham 2002), the outcome of its treatment remains limited. Thus, it is imperative for biologists to understand the molecular mechanisms of HCC for better diagnosis and therapy. High-throughput microarray technologies have provided new avenues to explore expression patterns of genes related to liver dysfunctions, including their transcriptional controls from a genomic perspective. For example, by utilizing high-density oligonucleotide microarrays, Wurmbach et al. (2007) identified a set of gene markers whose expression patterns can be used to track the progression

of Hepatitis C virus (HCV)-induced liver carcinogenic process or distinguish cancerous from normal liver cells.

In general, HCC and other cancers develop when cells proliferate at an unusually high rate. This high rate of cell proliferation can also be observed during the processes of liver development and regeneration. Therefore, comparison of gene expression profiles between these two processes and HCC progression may shed light on the molecular nature of liver cancer. We and others found that transcriptional reprogramming induced in HCC mimics that of the developing liver cells (Coulouarn et al. 2005; Li et al. 2009). The global similarity in transcriptional profiles between developmental and cancerous cells has also been observed in a variety of tissues (Naxerova et al. 2008). Since liver development and regeneration both require simultaneous proliferation and acquisition of highly specialized cellular functions (Kelley-Loughnane et al. 2002), it is intriguing to find how much they share in common at the transcriptome level. A microarray study of Otu et al. (2007), however, revealed that liver restoration after hepatectomy and liver development differed dramatically in the gene expression profiles. But similarities and differences between liver regeneration and HCC still remain to be explored.

Motivated by the above considerations, gene expression patterns in stepwise liver carcinogenic processes were compared with liver development and regeneration. Our comparative analyses elucidated their relationships regarding functional groups and transcriptional control, and identified several genes that may be related to HCC progression.

Materials and methods

Microarray data

In our previous study, gene expression profiles of mouse liver development were generated by Affymetrix Mouse 430 2.0 chips (Li et al. 2009). They can be accessed in the GEO database under accession number GSE13149. Gene expressions were measured at 14 different stages during the C57/B6 mouse liver development: E11.5 (embryonic day 11.5), E12.5, E13.5, E14.5, E15.5, E16.5, E17.5, E18.5, Day 0 (day of birth), Day 3, Day 7, Day 14, Day 21, and normal adult liver (control sample).

Gene expression data of mouse liver regeneration (GSE6998) and human HCC (GSE6764) were downloaded from the GEO database. The mice undergoing partial hepatectomy in GSE6998 dataset were female CD-1 mice at 10 weeks of age (Otu et al. 2007). During their operation, the three anterior lobes of the liver were tied first and then resected. After the 2/3 hepatectomy, the liver was returned to the abdominal cavity and the abdomen closed in two

layers. The regenerated liver tissues were harvested at 0, 1, 2, 6, 12, 18, 24, 30, 48, and 72 h after 2/3 hepatectomy. Gene expression data were generated by Affymetrix Mouse 430 2.0 chips. Normal liver tissue (0 h) was taken as reference point.

Human HCC data were produced by Wurmbach et al. (2007) using HG U133 Plus2 chips. The dataset was comprised of 65 disease samples of four disease groups representing a stepwise carcinogenic process from preneoplastic lesions (cirrhosis and dysplasia) to HCC, which included 13 samples from cirrhotic tissue, 17 dysplastic nodules, 18 early HCCs, and 17 advanced HCCs. This dataset also contained ten healthy liver tissue samples. The mean expression levels of healthy liver tissues were used as reference values.

All of the original.CEL files were annotated with CDF files generated by Dai et al. (2005) (Version 11.0.1, Entrez gene) using the software RMA (Irizarry et al. 2003). From the updated genome assembly, it was found that the definitions of many genes/transcripts on the microarrays are out-of-date. Dai et al. (2005) aligned the probes to different sources of genomic data to filter out problematic probes. Using their definition, this study mapped probe sets of 17,726 Entrez genes for human liver disease data, and 16,331 Entrez genes for mouse liver development and regeneration data.

Cross-platform comparison and homology mapping

After probe-set mapping, corresponding mouse orthologs of human genes were retrieved from Ensembl's Biomart system, and only one-to-one ortholog pairs (12,480 pairs) were kept for analysis. To compare data from different platforms, we defined the expression level for each gene as its ratio relative to the mean value of control samples (Table S1). Hierarchical clustering based on Uncentered Correlation Coefficient was carried out using Cluster 3.0 (Eisen et al. 1998) and visualized in TreeView 1.60. Expression levels of each gene were adjusted by log-transforming, mean-centering, and normalizing to unit variance across different samples before clustering.

Detecting differentially expressed genes

In identifying the subset of human genes potentially responsible for HCC, we gathered stepwise carcinogenic processes ranging from preneoplastic lesions (cirrhosis and dysplasia) to HCC as a time series, and then used EDGE to identify the differentially expressed genes during this time course. EDGE is a significance method for identifying differentially expressed genes in a time course study (Storey et al. 2005). We chose the q value cutoff 1×10^{-4} , and sifted 3,556 genes as HCC progressed. For the 3,556 genes, 2,788

one-to-one human–mouse ortholog pairs and corresponding expression profiles were obtained (Table S2).

KEGG pathway analysis

The manually curated pathway file of human (*hsa_pathway.list*, version 2009/01/05) and the map title file (*map_title.tab*, version 2009/01/04) were downloaded from the KEGG Website (<http://www.genome.ad.jp/kegg/pathway.html>). GeneMerge (Castillo-Davis and Hartl 2003) was used to identify pathways whose component genes occur more often in the group of our interest than what is expected by chance. GeneMerge produced a *p* value from hypergeometric test to rank over-represented pathways. To avoid inflated false positives as a consequence of multiple testing, the Bonferroni corrected *p* value of 0.01 was used as cutoff.

Identifying enriched motifs

Motifclass in CREAD package (Smith et al. 2006) was employed to search for over-represented motifs in the promoter regions of selected genes relative to the background promoter set. The background set was constructed by 1,000 genes randomly selected from the human genome. We defined promoter regions as 1,000 bp upstream to 200 bp downstream from annotated TSS sites, and retrieved their sequences from the UCSC Genome Browser. Known sequence motifs were obtained from the vertebrate subset of TRANSFAC (Matys et al. 2003) and JASPAR (Sandelin et al. 2004), and were represented by the position weight matrix. Then, all of the motifs were ranked based on relative error rates by Motifclass. In Motifclass, the relative error rate is equal to $1 - (\text{sensitivity} + \text{specificity})/2$. Here, sensitivity is the proportion of study set whose promoters included the motif; specificity is proportion of background promoter sets which do not include the motif. We required the relative error rates for selected motifs to be significantly lower than random promoters (*p* value < 0.01); and only reported top ten ranked motifs.

Results

Comparison of gene expression across liver development, liver regeneration, and HCC

Our analysis is based on a large-scale comparison of gene expression across liver development, liver regeneration, and HCC. After processing and homology mapping, expressions of 12,480 human–mouse one-to-one orthologous Entrez genes were retained. To ensure that the data from different platforms and species are comparable to each

other, we applied a previously proposed referencing strategy (Kaiser et al. 2007) that defines the expression level for each gene in each sample as its ratio relative to the mean level of corresponding normal control samples. The relative expression profiles of all 12,480 genes are listed in supplementary Table S1.

To investigate the global expression patterns of different biological processes, expression profiles from different samples were grouped by hierarchical clustering. As shown in Fig. 1a, human cirrhotic and dysplasia samples were first clustered with mouse regeneration samples and then with mouse development samples; next, human early HCC and advanced HCC samples were clustered together in another clade. It has been reported that in many tissues, tumorigenesis and developmental processes have similar expression patterns (Kaiser et al. 2007; Naxerova et al. 2008), which were not observed in Fig. 1a. However, after exclusion of late development samples, early development (E11.5 to E17.5, as defined in our previous work; Li et al. 2009) and cancer were placed under the same hierarchy (Fig. 1b), suggesting a closer relationship between early development and HCC.

Comparing Fig. 1a and b, we can find that the inclusion of late developmental stages influenced the final clustering results dramatically. In Fig. 1a, the early and late developmental stages are more similar to each other and are clustered together firstly. All these developmental stages as a whole are not so similar to HCC. In hierarchical clustering, the clades which are grouped together in the previous step can no longer be re-evaluated; so all of these developmental stages as a whole are separated with HCC (Fig. 1a). But after we excluded the late developmental stages, the similarity between early development and HCC was finally revealed (Fig. 1b). Besides the relations between development and HCC, in Fig. 1a and b, we also noticed that the regeneration samples demonstrate similar expression pattern with cirrhotic and dysplasia, especially with the dysplasia.

The similarity and difference between HCC and liver development

The above-mentioned hierarchical clusterings suggest partial similarity between development and HCC, so we asked what gene groups contribute to the similarities and what groups can explain the differences. Here, two-dimensional clustering was performed on 2,788 genes whose mRNA levels were altered during human carcinogenesis (Table S2). The resulting heatmap clearly demonstrates four gene groups (Fig. 2), from which we can see that Groups A and C are similar, while Groups B and D are different, in relation to liver development and HCC.

The expression patterns of advanced HCC and early development are similar in Groups A and C. In Group A,

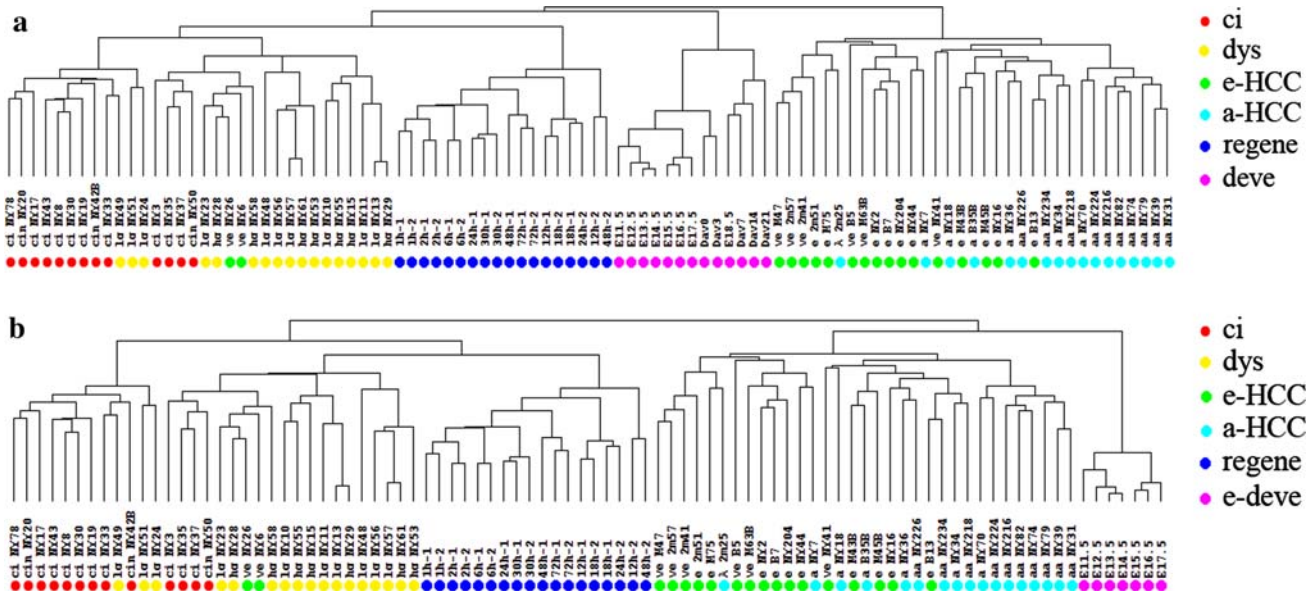


Fig. 1 **a** Clustering dendrogram for the development, regeneration, and different stages of HCC. Here, *ci* stands for cirrhotic tissue, *dys* stands for dysplastic nodules, *e-HCC* stands for early HCC, *a-HCC* stands for advanced HCC, *regene* stands for regeneration, and *deve* stands for development. **b** Clustering dendrogram for the early

development, regeneration, and different stages of HCC. Human HCC and mouse early development samples were placed side-by-side in contrast to **a**. Here, *e-deve* stands for early development stages which include E11.5 to E17.5

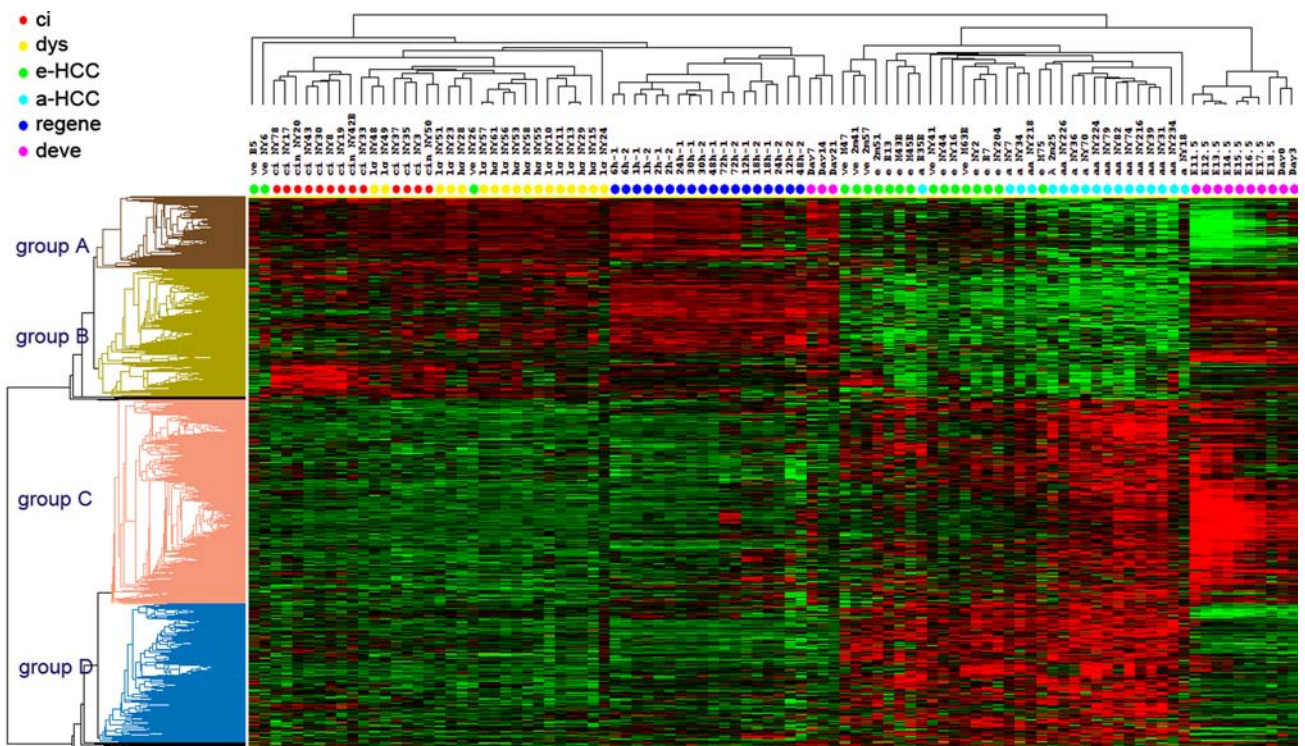


Fig. 2 Heatmap of 2,788 genes whose expression levels are significantly changed during human carcinogenesis. *Columns* represent samples from different processes, and *rows* represent genes. It is clear that

four groups of genes with distinct expression patterns emerged. The definition for each label is the same as those in Fig. 1

the genes are suppressed in both advanced HCC and early development; in Group C, the genes are both elevated in these two processes. Functional analysis indicates that

genes in Group A mainly execute normal liver functions such as “fatty acid metabolism” (Table 1). Motifclass was further used to sieve enriched cis-regulatory elements from

Table 1 Over-represented KEGG pathways (adjusted p value < 0.01) for different gene groups in Fig. 2

	Enriched KEGG pathways
Group A	Valine, leucine and isoleucine degradation; propanoate metabolism; complement and coagulation cascades; fatty acid metabolism; tryptophan metabolism; PPAR signaling pathway; beta-alanine metabolism; butanoate metabolism; glycine, serine and threonine metabolism; bile acid biosynthesis; lysine degradation; pyruvate metabolism; benzoate degradation via CoA ligation; glycolysis/gluconeogenesis; retinol metabolism
Group B	TGF-beta signaling pathway; cytokine–cytokine receptor interaction
Group C	Purine metabolism; homologous recombination; mismatch repair; aminoacyl-tRNA biosynthesis; DNA replication; nucleotide excision repair; pyrimidine metabolism; cell cycle; base excision repair

Table 2 Top ten enriched motifs for different gene groups in Fig. 2

Group A			Group B			Group C			Group D		
Matrix AC ^a	TF	p value	Matrix AC	TF	p value	Matrix AC	TF	p value	Matrix AC	TF	p value
MA0046	HNF-1	0	M00071	E47	0	M00940	E2F-1	0	M00803	E2F	0
M00762	DR1	0	M00176	AP-4	0	MA0028	Elk-1	0	M00695	ETF	0
M01105	ZBRK1	0	M00238	BARBIE	0.006	M00695	ETF	0	M00322	c-Myc:Max	0
M01020	TBX5	0	M00467	Roaz	0.001	MA0004	ARNT	0	MA0080	SPI-1	0
M00071	E47	0	MA0090	TEF-1	0.003	M00778	AhR	0	M01072	HIC1	0
M00117	C/EBPbeta	0.001	M00657	PTF1-beta	0.005	M00652	Nrf-1	0	M00807	EGR	0
M00974	SMAD	0.004	M00446	Spz1	0.005	M00651	NF-muE1	0	M00652	Nrf-1	0
M00262	Staf	0.003	M01081	Zec	0.008	M00801	CREB	0	M00986	Churchill	0
M00418	TGIF	0.006	M00159	C/EBP	0.006	M01035	YY1	0	M00797	HIF-1	0
M00159	C/EBP	0.008	MA0098	c-ETS	0.001	M00273	R	0	M00008	Sp1	0

^a Matrix accession number in TRANSFAC or JASPAR database

promoter regions of genes in Group A. It showed that the motifs ranking on top are corresponding to liver-enriched transcriptional factors HNF1 and HNF4 (Table 2). From the above analysis we can find that genes in Group A should play important roles in normal liver functions. Suppressed expression of these liver function-related genes in advanced HCC and early development reveals the immaturity of liver-related functions in early development, and the loss of these functions in cancer, respectively.

Group C genes have high expression patterns in both advanced HCC and early development. Functional analyses showed that genes in this group preferentially participate in pathways such as “cell cycle” and “DNA replication” (Table 1). Binding sites of the key cell cycle regulator E2F1 were ranked at the top of the enriched motifs for Group C genes. These observations suggest that liver development and HCC share similar pathways that promote cellular growth. When looking more details, gene expressions of Group C showed slightly negative correlation between HCC and development: gene expression levels rise from early HCC to advanced HCC; meanwhile, expression levels decline across liver development. This indicates that those promoting cellular growth pathways are strengthened from early HCC to advanced HCC and weakened during liver development.

Despite similar expression profiles between HCC and development in Groups A and C, their differences are apparent in Groups B and D. Group B genes distinguish HCC from all other processes by the lower-than-average expressions during cancerous stages, while Group D exhibits the opposite pattern (Fig. 2). Functional analyses reveal that “TGF-beta signaling pathway” and “cytokine–cytokine receptor interaction” are enriched in Group B (Table 1). The tumor suppressor genes such as DCN (Salomaki et al. 2008; Seidler et al. 2006), BMPR2 (Pouliot et al. 2003), and ID4 (Noetzel et al. 2008) in TGF-beta signaling pathway were included in Group B. Furthermore, Group B also included three TGF-beta signaling pathway genes which play direct roles in early stages of HCC and HCC differentiation (ID1, ID2, and ID3; Damdinsuren et al. 2005). For the cytokine–cytokine receptor interaction pathway, we found that cancer-related genes such as NGFR (Jin et al. 2007), CXCL14 (Pelicano et al. 2009; Wentz et al. 2008), and CCL2 (Lu et al. 2006) were also included in Group B. The opposite expression pattern between HCC and liver development of these cancer-related genes in Group B may contribute to the difference between HCC and development.

Different expression profiles between HCC and development can also be tracked by the difference of cis-regulatory

elements. Motif analysis by Motifclass revealed that promoter regions of genes in Groups C and D share enriched motifs corresponding to transcription factors E2F and ETF. But the enriched motifs of transcription factors MYC and EGR are specific to Group D (Table 2). From Fig. 2, we can see that genes in Group C are elevated in both HCC and development, while genes in Group D are elevated in HCC but inhibited in development. MYC is a well-known oncogene. In our data, it is down-regulated in the early cancerous stages (Fig. 3a). EGR is up-regulated in both regeneration and development, but repressed in cirrhotic, dysplasia, and cancer (especially in its late stages) (Fig. 3b). This expression pattern is consistent with the previous report that EGR is a cancer suppressor gene (Krones-Herzig et al. 2003, 2005). The specific enrichment of these cancer-related motifs in the promoter regions of Group D should contribute to the difference between HCC and development. Taken together, differences between HCC and development should be a result of concerted activities of cancer-related genes and their transcriptional controls.

The relation between liver regeneration and dysplasia

Generally, all of the above-mentioned clustering results demonstrate that HCC and liver regeneration are quite different from each other. But in Figs. 1 and 2, we can see that regeneration and cirrhosis/dysplasia are clustered together, which reveals the similarity at the mRNA level between regeneration and cirrhosis/dysplasia, especially between regeneration and dysplasia (Fig. 1b). This similarity between regeneration and dysplasia required our further attention. Considering all of the 12,480 genes, we then tried to identify which among them contributed to common

expression patterns of regeneration and dysplasia. Genes whose expression levels changed at least twofold in more than 50% of the samples in regeneration relative to the normal control samples were identified. The same filtering was applied on the dysplasia data. Finally, 143 genes and 75 genes were obtained for dysplasia and regeneration, respectively. Surprisingly, the two gene sets share only seven members; among the seven genes, only one has the same direction. Although dysplasia and regeneration share similar expression pattern as a whole, their few overlap here reveals apparent difference of these two processes.

To explain the inconsistency that dysplasia and regeneration show both global similarity and apparent difference at a finer scale, we employed PCA analysis. All expression profiles from different samples were projected onto the first two principle component axes (Fig. 4). The *x* and *y* axes explained 23 and 17% of the total variance, respectively. Since expression level for each gene was its ratio relative to normal control samples and was log-transformed before PCA analysis, the origin of coordinates corresponded to the normal state. The PCA result demonstrated that human and mouse samples form two clumps extending at different directions, with the cirrhotic, dysplasia and regeneration samples gather closer to each other and are all located near normal state (Fig. 4). The similarity between regeneration and dysplasia does not result from the same group of co-regulated genes, but is caused by their close resemblance to normal tissues.

Regeneration leads to the restoration of normal tissue, whereas dysplasia heralds the progression of cancer. Given that both have similar expression patterns to normal state, their different destinies might be subject to the activities of a number of genes that “switch on” the tumor progression. We found five genes SULT1E1, PDK4, MT1G, FOS, and

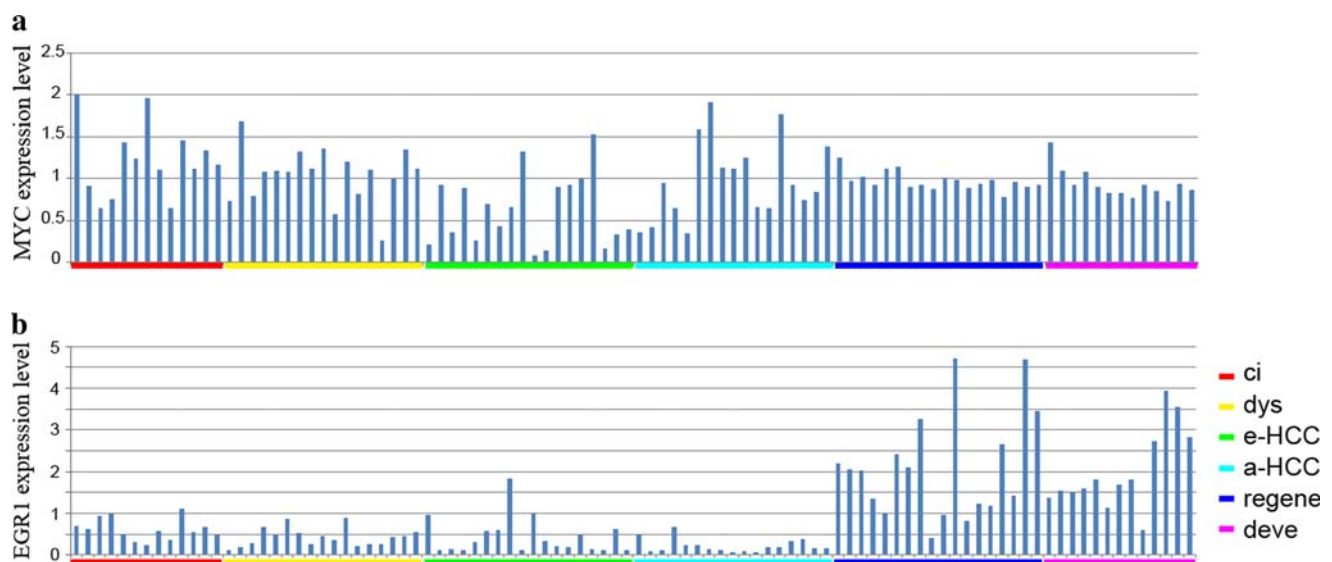


Fig. 3 Expression profiles of MYC (a) and EGR1 (b) during different time points in cirrhotic tissue, dysplastic nodules, early HCCs, advanced HCCs, regeneration, and development. The definition for each label is the same as those in Fig. 1

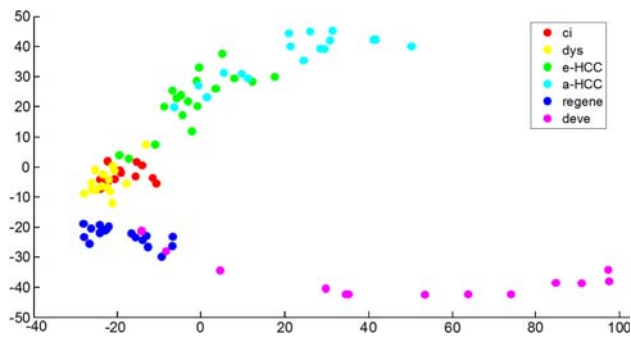


Fig. 4 Projection of all sample points to the plane spanned by the first two principle component axes. The origin of coordinate represents the normal state. Although mouse regeneration and development points set aside from human data points, extending to two different directions, regeneration, cirrhosis, and dysplasia gather close to each other near the origin of coordinate. The definition for each label is the same as those in Fig. 1

S100A8 whose mRNA level changed more than tenfold between regeneration and dysplasia and at least twofold relative to control in each process. It happened that all of them were up-regulated in regeneration, but suppressed in dysplasia (supplementary Fig. S1). Among them, FOS (part of AP-1 transcription factor complex) and S100A8 (dimerizing with S100A9) reportedly play roles in cancer progression. The FOS protein family dimerized with Jun proteins to form an AP-1 transcription factor complex. They play a central role in the proliferation and differentiation of normal tissue, as well as in oncogenic transformation and tumor progression (Mahner et al. 2008). Consistent with this, the expression level of Jun changed more than fivefold between regeneration and dysplasia in the data of this study (supplementary Fig. S1). S100A8 is a Ca^{2+} -binding protein of the S100 family. It naturally forms stable heterodimeric complex with S100A9 (Arai et al. 2008; Leukert et al. 2006). The expression profile of S100A9 was not included in our data since it was filtered by the one-to-one ortholog pairs' criterion. We checked manually and found that the expression level of S100A9 is very similar with S100A8 and is above by ninefold in regeneration as compared to dysplasia. S100A8 and S100A9 induced inflammatory activation of the endothelial cells (Roth et al. 2003); they are strong markers for various cancers and play important roles in the development and progression of carcinomas (Arai et al. 2008; Hermani et al. 2005; Moon et al. 2008; Turovskaya et al. 2008). Based on the above analyses and reports, we hypothesized that three remaining genes, SULT1E1, PDK4, and MT1G, may also be related to cancer progression.

Discussion

In this study, the transcriptional profiles of HCC, liver development, and liver regeneration based on high-density

microarray data were systematically compared. With these data, a comprehensive analysis on gene expression patterns, functional preferences, and transcriptional regulations for different gene groups was obtained.

Similarities between liver cancer and development had already been observed in previous studies. Here, we elucidated that the similarity is partial. Advanced HCC mimics prenatal development in terms of deprived normal liver functions and activated cellular proliferation, but they differ in activities of cancer-related genes and their transcriptional controls. The analyses revealed that genes with similar expression patterns between HCC and development prefer to be involved in pathways such as “fatty acid metabolism” and “glycolysis/gluconeogenesis” which are related to normal liver functions, and “cell cycle” and “DNA replication” pathways which are related to cellular proliferation. While genes with opposite expression patterns between HCC and development prefer to be involved in the “TGF-beta signaling pathway” and “cytokine–cytokine receptor interaction” pathways. It is well known that the “TGF-beta signaling pathway” is one of the most important pathways in cancer. Many genes in “cytokine–cytokine receptor interaction” pathway such as NGFR, CXCL14, and CCL2 were also reported to be cancer-related. Besides these cancer-related pathways and genes, we also found the specific enrichment of cancer-related motifs in the promoter regions of the genes which have opposite expression patterns between HCC and development. Further study on these genes and their transcriptional factors may provide new clues to cancer genesis.

Liver has its unique capability to regenerate. Although its molecular mechanism is poorly understood, similarities between liver regeneration and dysplasia were found in this study. Further analysis revealed that this similarity does not result from the same set of co-regulated genes, but due to their proximity to the normal state. This observation suggests that: firstly, the mechanisms of regeneration and dysplasia are significantly different from each other; secondly, the mechanisms of liver regeneration and dysplasia may not be as complicated as HCC since they are much closer to normal liver cells. Getting the similar expression patterns of pre-cancer stage and liver regeneration to normal state, we proposed that the regulation on a relatively few number of key genes might trigger the different progressions of cancer and regeneration. Our analyses found that SULT1E1, PDK4, MT1G, FOS, and S100A8 may be some candidates for key genes. Among them, FOS and S100A8 have been reported to be important players in cancer progression. Though we have not found many reports on SULT1E1, PDK4, and MT1G, we hypothesized that these three genes may likely contribute to HCC progression.

Acknowledgments We thank Xueya Zhou for helpful discussions. This work is partially supported by NSFC (30625012, 60721003, and 60905014), Tsinghua University SIST Basic Research Fund, Chinese National Key Program on Basic Research (2006CB910402), Shanghai Commission for Science and Technology (08JC14164), China Postdoctoral Science Foundation (20070420097), and Shanghai Postdoctoral Science Foundation (07R214138).

References

- Arai K, Takano S, Teratani T, Ito Y, Yamada T, Nozawa R (2008) S100A8 and S100A9 overexpression is associated with poor pathological parameters in invasive ductal carcinoma of the breast. *Curr Cancer Drug Targets* 8:243–252
- Castillo-Davis CI, Hartl DL (2003) GeneMerge—post-genomic analysis, data mining, and hypothesis testing. *Bioinformatics* 19:891–892
- Coulouarn C, Derambure C, Lefebvre G, Daveau R, Hiron M, Scotte M, Francois A, Daveau M, Salier JP (2005) Global gene repression in hepatocellular carcinoma and fetal liver, and suppression of dudulin-2 mRNA as a possible marker for the cirrhosis-to-tumor transition. *J Hepatol* 42:860–869
- Dai M, Wang P, Boyd AD et al (2005) Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. *Nucleic Acids Res* 33:e175
- Damdinsuren B, Nagano H, Kondo M et al (2005) Expression of Id proteins in human hepatocellular carcinoma: relevance to tumor dedifferentiation. *Int J Oncol* 26:319–327
- Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* 95:14863–14868
- Hermani A, Hess J, De Servi B, Medunjanin S, Grobholz R, Trojan L, Angel P, Mayer D (2005) Calcium-binding proteins S100A8 and S100A9 as novel diagnostic markers in human prostate cancer. *Clin Cancer Res* 11:5146–5152
- Irizarry RA, Hobbs B, Collin F, Beazer-Barclay YD, Antonellis KJ, Scherf U, Speed TP (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* 4:249–264
- Jin H, Pan Y, He L et al (2007) p75 neurotrophin receptor inhibits invasion and metastasis of gastric cancer. *Mol Cancer Res* 5:423–433
- Kaiser S, Park YK, Franklin JL et al (2007) Transcriptional recapitulation and subversion of embryonic colon development by mouse colon tumor models and human colon cancer. *Genome Biol* 8:R131
- Kelley-Loughnane N, Sabla GE, Ley-Ebert C, Aronow BJ, Bezerra JA (2002) Independent and overlapping transcriptional activation during liver development and regeneration in mice. *Hepatology* 35:525–534
- Krones-Herzig A, Adamson E, Mercola D (2003) Early growth response 1 protein, an upstream gatekeeper of the p53 tumor suppressor, controls replicative senescence. *Proc Natl Acad Sci USA* 100:3233–3238
- Krones-Herzig A, Mittal S, Yule K, Liang H, English C, Urcis R, Soni T, Adamson ED, Mercola D (2005) Early growth response 1 acts as a tumor suppressor in vivo and in vitro via regulation of p53. *Cancer Res* 65:5133–5143
- Leukert N, Vogl T, Strupat K, Reichelt R, Sorg C, Roth J (2006) Calcium-dependent tetramer formation of S100A8 and S100A9 is essential for biological activity. *J Mol Biol* 359:961–972
- Li T, Huang J, Jiang Y, Zeng Y, He F, Zhang MQ, Han Z, Zhang X (2009) Multi-stage analysis of gene expression and transcription regulation in C57/B6 mouse liver development. *Genomics* 93:235–242
- Lu Y, Cai Z, Galson DL, Xiao G, Liu Y, George DE, Melhem MF, Yao Z, Zhang J (2006) Monocyte chemotactic protein-1 (MCP-1) acts as a paracrine and autocrine factor for prostate cancer growth and invasion. *Prostate* 66:1311–1318
- Mahner S, Baasch C, Schwarz J, Hein S, Wolber L, Janicke F, Milde-Langosch K (2008) C-Fos expression is a molecular predictor of progression and survival in epithelial ovarian carcinoma. *Br J Cancer* 99:1269–1275
- Matys V, Fricke E, Geffers R et al (2003) TRANSFAC: transcriptional regulation, from patterns to profiles. *Nucleic Acids Res* 31:374–378
- Moon A, Yong HY, Song JI et al (2008) Global gene expression profiling unveils S100A8/A9 as candidate markers in H-ras-mediated human breast epithelial cell invasion. *Mol Cancer Res* 6:1544–1553
- Naxerova K, Bult CJ, Peaston A, Fancher K, Knowles BB, Kasif S, Kohane IS (2008) Analysis of gene expression in a developmental context emphasizes distinct biological leitmotifs in human cancers. *Genome Biol* 9:R108
- Noetzel E, Veeck J, Niederacher D, Galm O, Horn F, Hartmann A, Knuchel R, Dahl E (2008) Promoter methylation-associated loss of ID4 expression is a marker of tumour recurrence in human breast cancer. *BMC Cancer* 8:154
- Otu HH, Naxerova K, Ho K, Can H, Nesbitt N, Libermann TA, Karp SJ (2007) Restoration of liver mass after injury requires proliferative and not embryonic transcriptional patterns. *J Biol Chem* 282:11197–11204
- Pelicano H, Lu W, Zhou Y, Zhang W, Chen Z, Hu Y, Huang P (2009) Mitochondrial dysfunction and reactive oxygen species imbalance promote breast cancer cell motility through a CXCL14-mediated mechanism. *Cancer Res* 69:2375–2383
- Pouliot F, Blais A, Labrie C (2003) Overexpression of a dominant negative type II bone morphogenetic protein receptor inhibits the growth of human breast cancer cells. *Cancer Res* 63:277–281
- Roth J, Vogl T, Sorg C, Sunderkotter C (2003) Phagocyte-specific S100 proteins: a novel group of proinflammatory molecules. *Trends Immunol* 24:155–158
- Salomaki HH, Sainio AO, Soderstrom M, Pakkanen S, Laine J, Jarvelainen HT (2008) Differential expression of decorin by human malignant and benign vascular tumors. *J Histochem Cytochem* 56:639–646
- Sandelin A, Alkema W, Engstrom P, Wasserman WW, Lenhard B (2004) JASPAR: an open-access database for eukaryotic transcription factor binding profiles. *Nucleic Acids Res* 32:D91–D94
- Seidler DG, Goldoni S, Agnew C, Cardi C, Thakur ML, Owens RT, McQuillan DJ, Iozzo RV (2006) Decorin protein core inhibits in vivo cancer growth and metabolism by hindering epidermal growth factor receptor function and triggering apoptosis via caspase-3 activation. *J Biol Chem* 281:26408–26418
- Smith AD, Sumazin P, Xuan Z, Zhang MQ (2006) DNA motifs in human and mouse proximal promoters predict tissue-specific expression. *Proc Natl Acad Sci USA* 103:6275–6280
- Storey JD, Xiao W, Leek JT, Tompkins RG, Davis RW (2005) Significance analysis of time course microarray experiments. *Proc Natl Acad Sci USA* 102:12837–12842
- Thorgeirsson SS, Grisham JW (2002) Molecular pathogenesis of human hepatocellular carcinoma. *Nat Genet* 31:339–346
- Turovskaya O, Foell D, Sinha P et al (2008) RAGE, carboxylated glycans and S100A8/A9 play essential roles in colitis-associated carcinogenesis. *Carcinogenesis* 29:2035–2043
- Wente MN, Mayer C, Gaida MM, Michalski CW, Giese T, Bergmann F, Giese NA, Buchler MW, Friess H (2008) CXCL14 expression and potential function in pancreatic cancer. *Cancer Lett* 259:209–217
- Wurmbach E, Chen YB, Khitrov G et al (2007) Genome-wide molecular profiles of HCV-induced dysplasia and hepatocellular carcinoma. *Hepatology* 45:938–947