Pathol. Oncol. Res. (2013) 19:251–258
DOI 10.1007/s12253-012-9576-y

RESEARCH

Bioinformatics Analysis Reveals Potential Candidate Drugs for HCC

Xiao-Bing Huang · Jing Li · Lu Zheng · Guo-Hua Zuo · Ke-Qiang Han · Hong-Yan Li · Ping Liang

Received: 13 June 2012 / Accepted: 24 September 2012 / Published online: 23 January 2013

Abstract In our study, we used the GSE17967 series to identify differentially expressed genes between cirrhosis and hepatocellular carcinoma, aiming to analyse the mechanism of the progression of cirrhosis to hepatocellular carcinoma and identify the sub-pathways closely related to this progression, and find the small molecule drugs to interfere this progression. From the result of our study, we find that many small molecule drugs closely related with carcinoma have been linked by our method. We also find some new small molecule drugs related to this progression. It is demonstrated that bioinformatics analysis is useful in identification of the candidate drugs in hepatocellular carcinoma.

Keywords Hepatocellular carcinoma · Cirrhosis · Sub-pathways · Small molecule drugs

Introduction

Hepatocellular carcinoma is the fifth most common cancer worldwide and the most common form of liver cancer, being responsible for 80 % of the primary malignant liver tumors in adults [1, 2]. The 5-year relative survival rate is about 7 % and causes more than 6 million deaths annually worldwide [1]. The disease is most prevalent in Eastern and Southeastern Asia, and Middle Africa, with more than half of the patients being reported from China [2, 3].

Much is known about the development and causes of HCC. Patients with cirrhosis of the liver have been identified as being at risk for hepatocellular carcinoma, and hepatocellular carcinoma is the principal cause of death in patients with cirrhosis. It is unclear why this tumor frequently accompanies cirrhosis [5–7]. Hepatocellular carcinoma may be either the inevitable consequence of longstanding hepatic disease or an independent response to a hepatic insult common to hepatocellular carcinoma and cirrhosis [4]. Although the importance of the association of hepatocellular carcinoma with cirrhosis is still obscure, such an association provides a means to identify patients at high risk for hepatocellular carcinoma [5].

Even in developed countries, potentially curative therapies are offered to only one in every four patients coming to highly committed centers [10, 11]. Curative therapies, such as resection, liver transplantation, or percutaneous treatments, benefit only 25 % of patients and are the only chance to improve life expectancy. Despite the implementation of surveillance programmes for early hepatocellular carcinoma, most tumors are diagnosed at advanced stages, for which no standard therapy has been established [11–14].

Extensive epidemiological studies over the years have identified major risk factors of HCC and many advances have been made to understand the pathogenesis of HCC. However, little is known about molecular mechanisms that lead to carcinogenesis. Abrupt changes that occur in liver tissues due either to viral infection or exposure to hepatotoxic agents cause significant changes in the cellular signaling pathways and alter gene expression resulting in tumor formation [6]. The most important mechanism of liver cancer progression is cell proliferation. There is no dominant

Electronic supplementary material The online version of this article (doi:10.1007/s12253-012-9576-y) contains supplementary material, which is available to authorized users.

X.-B. Huang · J. Li · L. Zheng · G.-H. Zuo · K.-Q. Han · H.-Y. Li · P. Liang (✉)
Department of Hepatobiliary Surgery, Xinqiao Hospital, Third Military Medical University,
400037, NO.2 Xin Qiao Street, Sha Ping Ba District,
Chongqing, People's Republic of China
e-mail: PiangLiang2012@hotmail.com

pathway responsible for it in HCC, but gene expression studies have resulted in major molecular classes of HCC, according to gene sets responsible for cell proliferation and survival [16–18].

DNA microarray analysis as a global approach is applied to investigate physiological mechanisms in health and disease [7]. High-throughput technologies such as single-nucleotide polymorphism array, complementary DNA microarray, and protein mass spectrometry have changed the molecular approach to liver cancer [8]. Genomic expression profiling evolves as a useful tool to identify novel pathomechanisms in human cancer.

The purposes of this paper are to analysis the mechanism of the progression of cirrhosis to hepatocellular carcinoma, identify biological pathways in this progression, thus we can identify tumor markers that are useful for early detection of tumors, to predict prognosis, or to find new therapeutic targets with their underlying molecular mechanism of action.

Methods

Data Source

Affymetrix Microarray Data

One transcription profile of GSE17967 [9] was obtained from a public functional genomics data repository GEO (http://www.ncbi.nlm.nih.gov/geo/) which are based on the Affymetrix GPL571 platform data (Affymetrix Human Genome U133A 2.0 Array). Only 47 hepatocelluar carcinoma chips and 16 cirrhosis control chips are available.

The Connectivity Map Data

The connectivity map resource can be used to find connections among small molecules sharing a mechanism of action, chemicals and physiological processes, and diseases and drugs. It stores transcription expression profiles of human genome that interfered by active micromolecule, including 6,100 groups of micromolecule interfering experiments, and total 7,056 profiles. We downloaded all the profile data to analyse.

Pathway Data

KEGG (Kyoto Encyclopedia of Genes and Genomes) is a collection of online databases dealing with genomes, enzymatic pathways and biological chemicals [10]. The PATHWAY database records networks of molecular interactions in the cells, and variants of them specific to particular organisms (http://www.genome.jp/kegg/). Total 300 pathways were collected from KEGG. All the pathways were translated into "KO-KO" digraph using SubpathwayMiner, and sub-pathways were identified by "K-Clique" method.

Protein Interaction Data

Human Protein Reference Database (HPRD; http://www.hprd.org/) is a resource for experimentally derived information about the human proteome including protein–protein interactions, post-translational modifications (PTMs) and tissue expression. It contains 37041protein-protein interactions which involves 9,518 proteins.

Methods

Differentially Expressed Genes (DEGs) Analysis

For the GSE17967 dataset, the limma method [11] was used to identify DEGs. The original expression datasets from all conditions were processed into expression estimates using the RMA method with the default settings implemented in Bioconductor, and then constructed the linear model. To circumvent the multi-test problem which might induce too much false positive results, the Benjamini & Hochberg method [12] was used to adjust the raw *P*-values into false discovery rate (FDR). The DEGs only with the fold change value larger than 1.5 and p-value less than 0.05 were selected.

Significant Analysis of Pathway

We adopted an impact analysis that includes the statistical significance of the set of pathway genes, and also considered other crucial factors such as the magnitude of each gene's expression change, the topology of the signaling pathway, their interactions, etc. In this model, the Impact Factor (IF) of a pathway Pi is calculated as the sum of two terms:

$$IF(Pi) = \log\left(\frac{1}{pi}\right) + \frac{\sum_{g \in Pi} |PF(g)|}{|\triangle E| \cdot N_{de}(Pi)}$$

The first term is a probabilistic term that captures the significance of the given pathway P_i from the perspective of the set of genes contained in it.

It is obtained by using the hyper geometric model in which p_i is the probability of obtaining at least the observed number of differentially expressed genes, N_{de}, just by chance.

The second term is a functional term that depends on the identity of the specific genes that are differentially

expressed as well as on the interactions described by the pathway (i.e., its topology).

The second term sums up the absolute values of the perturbation factors (PFs) for all genes g on the given pathway P_i.

The PF of a gene g is calculated as follows:

$$PF(g) = \triangle E(g) + \sum_{u \in USg} \beta_{ug} \cdot \frac{PF(u)}{Nds(u)}$$

In this equation, the first term $\Delta E\ (g)$ captures the quantitative information measured in the gene expression experiment. The factor $\Delta E\ (g)$ represents the normalized measured expression change of the gene g. The first term $\Delta E\ (g)$ in the above equation is a sum of all PFs of the genes u directly upstream of the target gene g, normalized by the number of downstream genes of each such gene N_{ds} (u), and weighted by a factor β_{ug}, which reflects the type of interaction: $\beta_{ug} = 1$ for induction, $\beta_{ug} = -1$ for repression (KEGG supply this information about the type of interaction of two genes in the description of the pathway topology). US_g is the set of all such genes upstream of g. We need to normalize with respect to the size of the pathway by dividing the total perturbation by the number of differentially expressed genes on the given pathway, N_{de} (P_i). In order to make the *IF*s as independent as possible from the technology, and also comparable between problems, we also divided the second term in equation 1 by the mean absolute fold change ΔE, and calculated across all differentially expressed genes. The results of the significance analysis of pathway were shown in Table 1.

Results

Differentially Expressed Genes Between Cirrhosis and Hepatocellular Carcinoma

To get differentially expressed genes between cirrhosis and hepatocellular carcinoma, we obtained publicly available microarray data sets GSE17967 from GEO. After microarray analysis, the differentially expressed genes with the fold change value larger than 1.5 of GSE17967 and p-value less than 0.05 were selected. The p-value less than 0.05 were chosen as the threshold. Finally, we got 888 genes which expressed differently between the samples of cirrhosis and hepatocellular carcinoma. Among the 888 genes, 76 % (671) of them shows down-regulation, while 24 % shows up-regulation. We hypothesize that the 888 DEGs closely correlated with the progression of cirrhosis to hepatocellular carcinoma, and these genes play important roles in the progression.

Significant Protein Interactions in Hepatocellular Carcinoma

To analyse the mechanism in the progression of cirrhosis to hepatocellular carcinoma on function level, we performed network analysis of the cancer-related genes at first. We mapped all the 888 genes to protein interaction network (PPI), and extracted the maximum link component as the cirrhosis canceration related network, that is PCHN (Progression of Cirrhosis to HCC Network). This network contains 241 nodes (44 up-regulated genes and 197 down-regulated genes), 248 sides (Fig. 1). Besides, we analysed the network topological properties of the 214 canceration related genes that mapped to PPI and all genes in PPI (Fig. 2). It is suggested that genes related to carcinogenesis

Table 1 Pathway significant analysis

pathwayID	pathwayName	p-value	FDR
path:04070_1	Phosphatidylinositol signaling system	4.27E-05	0.039204
path:04650_10	Natural killer cell mediated cytotoxicity	5.20E-05	0.039204
path:04270_14	Vascular smooth muscle contraction	0.000102	0.039204
path:04810_32	Regulation of actin cytoskeleton	0.000129	0.039204
path:04070_7	Phosphatidylinositol signaling system	0.000142	0.039204
path:04666_10	Fc gamma R-mediated phagocytosis	0.000163	0.039204
path:05215_10	Prostate cancer	0.000163	0.039204
path:04510_10	Focal adhesion	0.000209	0.039204
path:04510_8	Focal adhesion	0.000232	0.039204
path:04510_19	Focal adhesion	0.000232	0.039204
path:04210_22	Apoptosis	0.000237	0.039204
path:04512_1	ECM-receptor interaction	0.000237	0.039204
path:05200_3	Pathways in cancer	0.000251	0.039204
path:05222_4	Small cell lung cancer	0.000251	0.039204

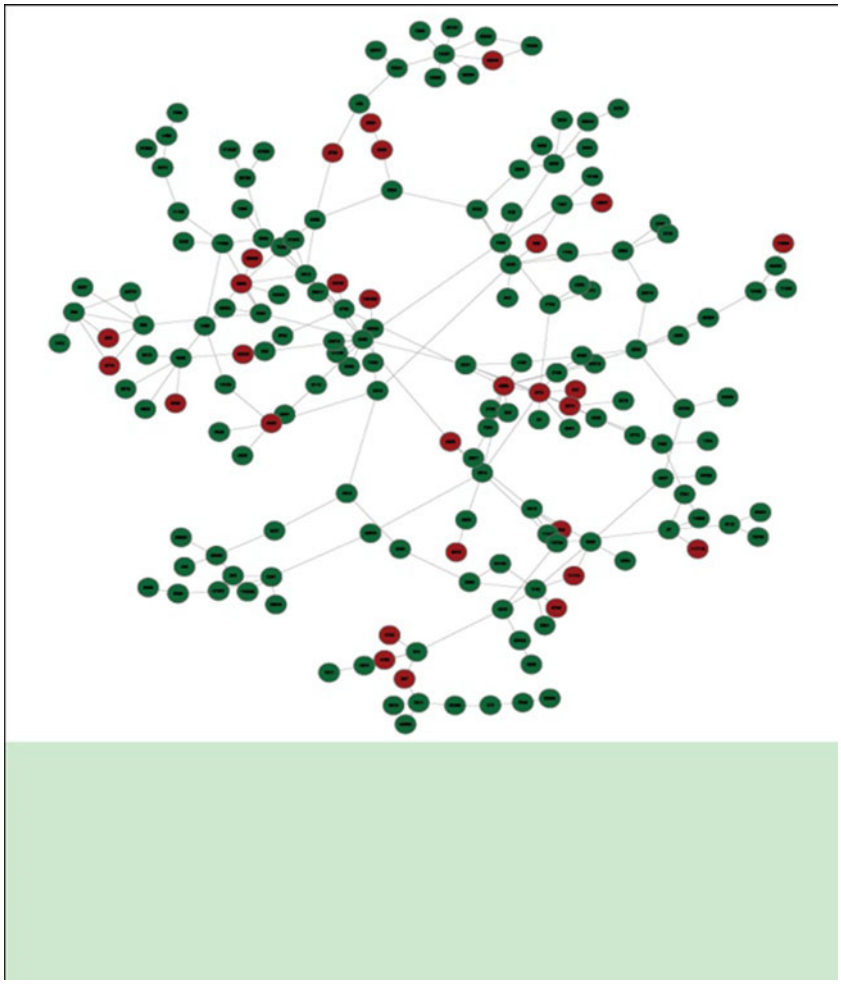

Fig. 1 Progression of cirrhosis to hepatocellular carcinoma network. The red nodes refer to up-regulated genes in the progression of cirrhosis to hepatocellular carcinoma and green nodes refer to down-regulated genes

were significantly higher than the average level of all genes in the protein interaction network on the four topological properties: degree, betweenness, closeness and cluster coefficient.

Significant Pathway in Hepatocellular Carcinoma

To identify the relevant pathways changed in progression of cirrhosis to hepatocellular carcinoma, we used a statistical approach on pathway level. Significance analysis at single gene level may suffer from the limited number of samples and experimental noise that can severely limit the power of the chosen statistical test. Pathway can provide an alternative way to relax the significance threshold applied to single genes and may lead to a better biological interpretation. So, we adopted a pathway based impact analysis method that contained many factors including the statistical significance of the set of differentially expressed genes in the pathway, the magnitude of each gene's expression change, the topology of the signaling pathway, and their interactions and so on. Here, we chose FDR >0.5 as the significant threshold, and obtained 11 pathways that closely related to the progression of cirrhosis to hepatocellular carcinoma. The impact analysis method yields many significant pathways contained Phosphatidylinositol signaling system, Natural killer cell mediated cytotoxicity, Vascular smooth muscle contraction and so on (Table 1).

Significant Sub-pathway That Interfered by Small Molecule

According to the records that 1,309 small molecules interference on cancer cells in CMap database, we preformed differentially expressed genes analysis of profile data (with and without medicine), and obtained gene collection of each small molecule and its DEGs. 1,221 small molecules have DEGs. Each small molecule corresponding to a collection of DEGs, we preformed KEGG sub-pathway enrichment analysis of the gene collection and got all the sub-pathways that significant influenced by the small molecule (FDR >0.5 as the significant threshold). There are 192 small molecules that have significant enrichment sub-pathways, which

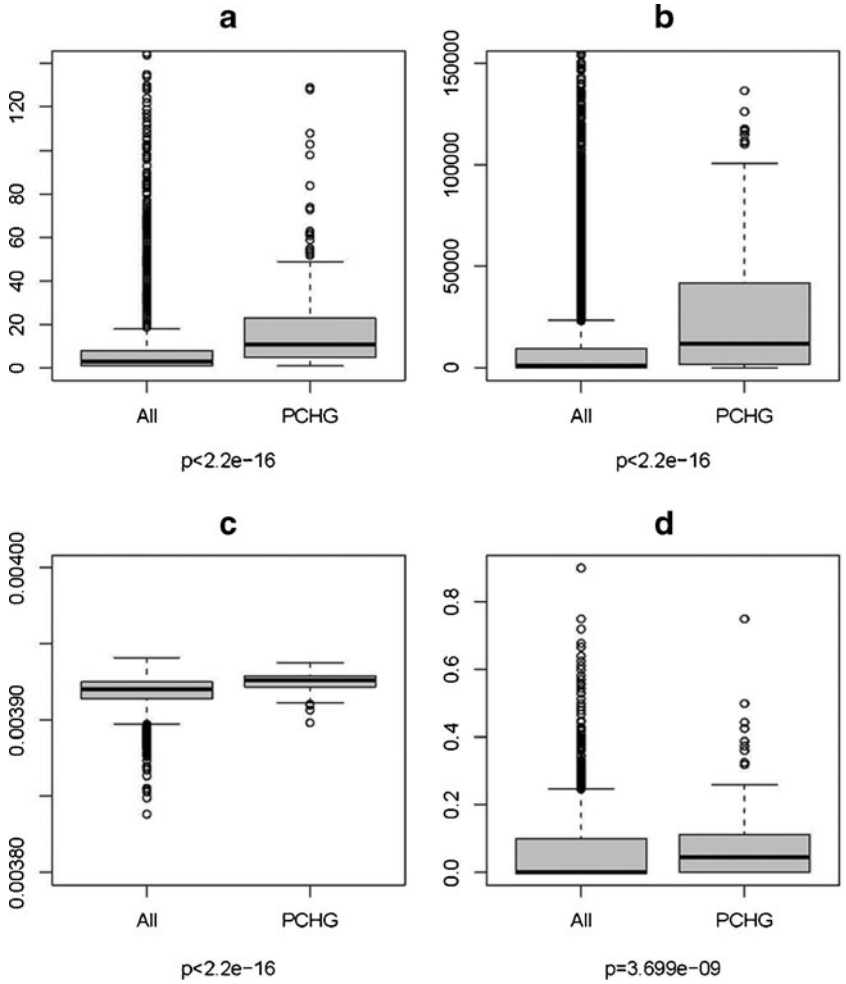

Fig. 2 Topological properties analysis of genes related to canceration in protein network. Subfigure **a** is degree, subfigure **b** is betweenness, subfigure **c** is closeness and subfigure **d** is cluster coefficient. Compared to all the genes in PPI, the four topological properties of genes related to canceration were significant higher

involve 698 sub-pathways. (Detailed results are shown in Supplement Table 1).

Identification of Small Molecules That Target to the Pathway Closely Related to the Progression of Cirrhosis to Hepatocellular Carcinoma

Combined the two results above, and picked out the common sub-pathways, and then we can find out the small molecules that target to the pathway closely related to the progression of cirrhosis to hepatocellular carcinoma. Then, we can hypothesis that the small molecule drugs could perturb the progression of cirrhosis to hepatocellular carcinoma. The results are shown in Table 2.

By integrating the relationships above, a network of interference relationship was built between small molecules and sub-pathways that interferes the canceration (Fig. 3). In this network, cephaeline and thapsigargin can perturb one or several sub-pathways by themselves, while other small molecules perturb pathways by cooperating with other small molecules.

Table 2 The intersection of molecules and 10 sub-pathways in canceration

molecule	p-value	CommonSub-pathways
Fludrocortisone	3.47E-06	3
Trichostatin A	1.52E-05	7
Fenoprofen	1.71E-05	2
Vincamine	0.000169527	2
Tretinoin	0.000317576	3
Metronidazole	0.000347426	3
Diphenylpyraline	0.000447698	3
Latamoxef	0.000700323	3
Camptothecin	0.002570411	4
Adiphenine	0.025472494	1
Cephaeline	0.033099908	2
Thapsigargin	0.082583795	1
Vorinostat	0.087518059	2
Withaferin A	0.102215947	1
Anisomycin	0.198897429	2
Lycorine	0.284518179	1

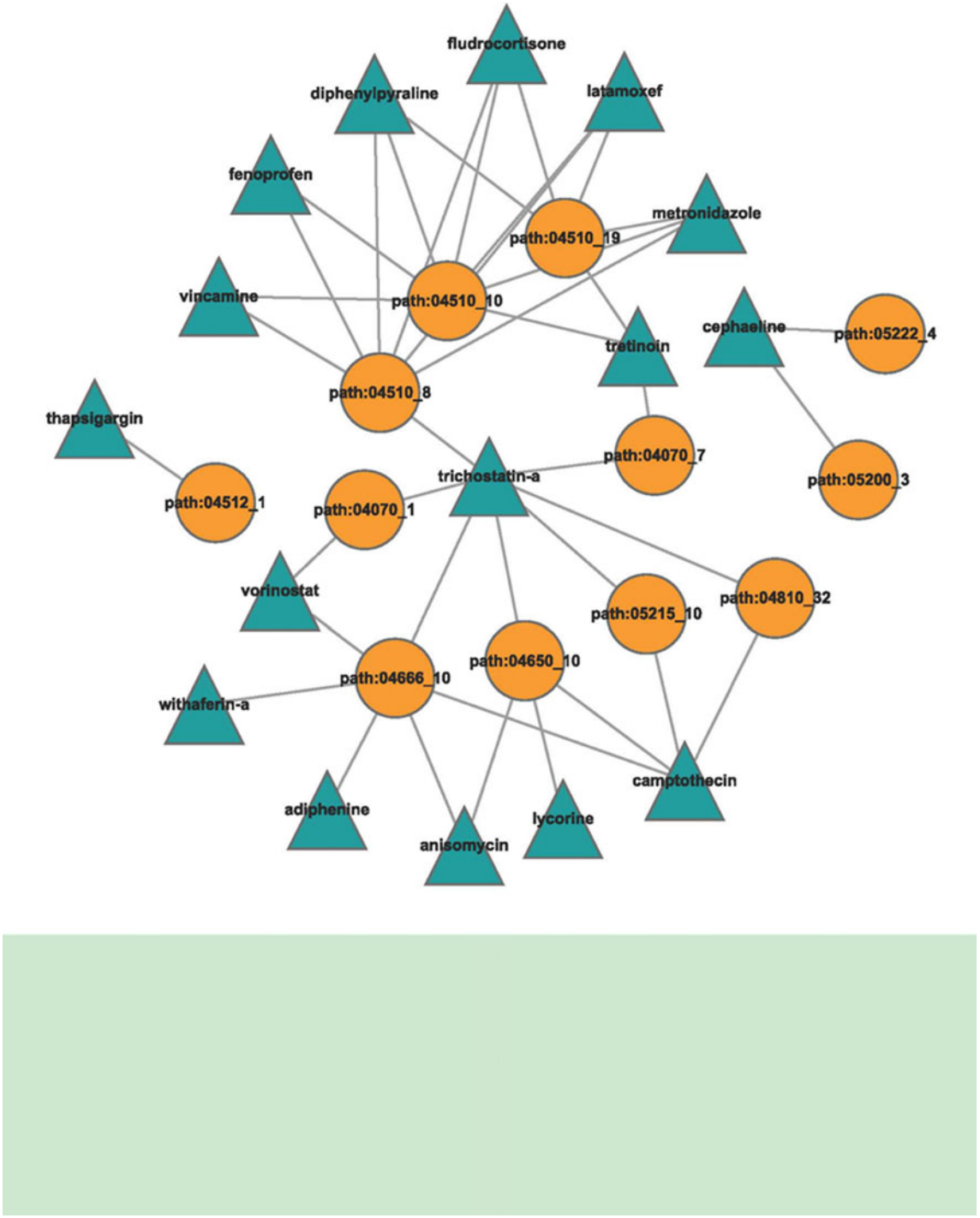

Fig. 3 Interference network between small molecule and sub-pathways that related to canceration. Blue triangle represents small molecule, orange circle represents sub-pathways that related to canceration

Discussion

Currently, hepatocellular carcinoma (HCC) has poor prognosis, because it is often diagnosed at an advanced stage. Heterogeneous phenotypic and genetic traits of affected individuals and a wide range of risk factors have classified it a complex disease [6]. HCC normally develops as a consequence of underlying liver disease and is most often associated with cirrhosis [13]. HCC is not amenable to standard chemotherapy and is resistant to radiotherapy. In most cases, surgical resection and liver transplantation remain the only curative treatment options. Therefore, development of novel, effective drugs or drug combinations is of prime importance. Extensive research over the past decade

has identified a number of molecular biomarkers as well as cellular networks and signaling pathways affected in liver cancer. Recent studies using a combination of "omics" technologies, microRNA studies, combinatorial chemistry, and bioinformatics are providing new insights into the gene expression and protein profiles during various stages of the disease [6].

In the past decade, many signaling pathways are identified associated with cancer development, such as JAK/STAT, MAPK/ERK, PI3K/AKT, NF-kB, Wnt, TGF-β, etc. [14]. Identically, these pathways play a significant role in hepatocellular carcinoma. From the result of significant pathway in hepatocellular carcinoma, we could find that many biological pathways closely related with carcinoma have been linked by our method. The significant enrichment pathways involve ten pathways: Phosphatidylinositol signaling system, Natural killer cell mediated cytotoxicity,Vascular smooth muscle contraction,Regulation of actin cytoskeleton, Fc gamma R-mediated phagocytosis, Prostate cancer, Focal adhesion, Apoptosis, ECM-receptor interaction, Pathways in cancer and Small cell lung cancer. The 10 sub-pathways showed close interaction in their corresponding pathways. In the progression of cirrhosis to hepatocellular carcinoma, these sub-pathways showed a significant imbalance. Therefore, further analysis about these sub-pathways would play a guiding role in the progression, and drug screening for these significant imbalanced sub-pathways will have important clinical significance.

The result of Interference network construction between small molecule and sub-pathways that related to canceration revealed many small molecule drugs that maybe effective in interfering carcinogenesis, for example, fludrocortisones, trichostatin a, fenoprofen, vincamine, tretinoin, metronidazole, diphenylpyraline, latamoxef, camptothecin, adiphenine, cephaeline, thapsigargin, vorinostat, withaferin a, anisomycin, lycorine and so on. Some of them, such as trichostatin A, camptothecin and vorinostat were proved to have anti-cancer effects by previous study.

Trichostatin A (TSA) inhibits the eukaryotic cell cycle during the beginning of the growth stage. It is a member of a larger class of histone deacetylase inhibitors (HDIs or HDACIs) that have a broad spectrum of epigenetic activities. Thus, TSA has some potential as an anti-cancer drug [15]. One suggested mechanism is that TSA promotes the expression of apoptosis-related genes, leading to cancerous cells surviving at lower rates, thus slowing the progression of cancer [16]. TSA was found to have potent antiproliferative activity in eight breast cancer cell lines using the sulforhodamine B assay.

Camptothecin (CPT), a plant alkaloid with antitumor activity, has been shown to be a potent inhibitor of nucleic acid synthesis and a strong inducer of DNA strand breaks in mammalian cells. Camptothecin class of compounds has been demonstrated to be effective against a broad spectrum of tumors. Their molecular target has been firmly established to be human DNA topoisomerase I (topo I). CPT inhibits topo I by blocking the rejoining step of the cleavage/religation reaction of topo-I, resulting in accumulation of a covalent reaction intermediate, the cleavable complex.

Vorinostat is a member of a larger class of compounds that inhibit histone deacetylases. It is the first drug approved for the treatment of cutaneous manifestations in patients with cutaneous T cell lymphoma (CTCL). It causes the accumulation of acetylated histones and induces cell cycle arrest and/or apoptosis of some transformed cells [17]. A recent study suggested that vorinostat also possesses some activity against recurrent glioblastoma multiforme, resulting in a median overall survival of 5.7 months (compared to 4–4.4 months in earlier studies). Further brain tumor trials are planned in which vorinostat will be combined with other drugs. In clinical trials, Vorinostat has shown significant anticancer activity against both hematologic and solid tumors at doses well tolerated by patients [18].

Other small molecule drugs, such as fludrocortisones, fenoprofen, vincamine, tretinoin etc., also observed in our interference network, and they maybe play an important role in interference of the progression of cirrhosis to hepatocellular carcinoma. A deeper understanding about the interference network remains an area of intense research activity in future.

Summery

Our interference network is useful in investigating the complex interacting mechanisms of the progression of cirrhosis to hepatocellular carcinoma. However, further experiments are still needed to confirm the conclusion.

References

1. El-Serag HB, Rudolph KL (2007) Hepatocellular carcinoma: epidemiology and molecular carcinogenesis. Gastroenterology 132 (7):2557–2576. doi:10.1053/j.gastro.2007.04.061
2. Avila MA, Berasain C, Sangro B, Prieto J (2006) New therapies for hepatocellular carcinoma. Oncogene 25(27):3866–3884. doi:10.1038/sj.onc.1209550
3. Yuan R, Jiang C, Hong K, Yu X, Wu L, Liu T, Liu X, Tang X, Cai H, Shao J (2011) Genetic variation in the Fat10 gene is associated with risk of hepatocellular carcinoma in a Chinese population. Asian Pac J Cancer Prev 12(8):2117–2122
4. Johnson PJ, Williams R (1987) Cirrhosis and the aetiology of hepatocellular carcinoma. J Hepatol 4(1):140–147
5. Colombo M, de Franchis R, Del Ninno E, Sangiovanni A, De Fazio C, Tommasini M, Donato MF, Piva A, Di Carlo V, Dioguardi N (1991) Hepatocellular carcinoma in Italian patients with

cirrhosis. N Engl J Med 325(10):675–680. doi:10.1056/NEJM199109053251002
6. Aravalli RN, Steer CJ, Cressman EN (2008) Molecular mechanisms of hepatocellular carcinoma. Hepatology 48(6):2047–2063. doi:10.1002/hep. 22580
7. Spies M, Dasu MR, Svrakic N, Nesic O, Barrow RE, Perez-Polo JR, Herndon DN (2002) Gene expression analysis in burn wounds of rats. Am J Physiol Regul Integr Comp Physiol 283(4):R918–R930. doi:10.1152/ajpregu.00170.2002
8. Llovet JM, Bruix J (2008) Molecular targeted therapies in hepatocellular carcinoma. Hepatology 48(4):1312–1327. doi:10.1002/hep. 22506
9. Su H, Hu N, Yang HH, Wang C, Takikita M, Wang QH, Giffen C, Clifford R, Hewitt SM, Shou JZ, Goldstein AM, Lee MP, Taylor PR (2011) Global gene expression profiling and validation in esophageal squamous cell carcinoma and its association with clinical phenotypes. Clin Cancer Res 17(9):2955–2966. doi:10.1158/1078-0432.CCR-10-2724
10. Kanehisa M (2002) The KEGG database. Novartis Found Symp 247:91–101; discussion 101–103, 119–128, 244–152
11. Smyth GK (2004) Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol 3:Article3. doi:10.2202/1544-6115.1027
12. Benjamini Y, Drai D, Elmer G, Kafkafi N, Golani I (2001) Controlling the false discovery rate in behavior genetics research. Behav Brain Res 125(1–2):279–284
13. Zhou L, Liu C, Meng FD, Qu K, Tian F, Tai MH, Wei JC, Wang RT (2012) Long-term prognosis in hepatocellular carcinoma patients after hepatectomy. Asian Pac J Cancer Prev 13(2):483–486
14. Dreesen O, Brivanlou AH (2007) Signaling pathways in cancer and embryonic stem cells. Stem Cell Rev 3(1):7–17
15. Drummond DC, Noble CO, Kirpotin DB, Guo Z, Scott GK, Benz CC (2005) Clinical development of histone deacetylase inhibitors as anticancer agents. Annu Rev Pharmacol Toxicol 45:495–528. doi:10.1146/annurev.pharmtox.45.120403.095825
16. Shankar S, Srivastava RK (2008) Histone deacetylase inhibitors: mechanisms and clinical significance in cancer: HDAC inhibitor-induced apoptosis. Adv Exp Med Biol 615:261–298. doi:10.1007/978-1-4020-6554-5_13
17. Minucci S, Pelicci PG (2006) Histone deacetylase inhibitors and the promise of epigenetic (and more) treatments for cancer. Nat Rev Cancer 6(1):38–51. doi:10.1038/nrc1779
18. Marks PA, Breslow R (2007) Dimethyl sulfoxide to vorinostat: development of this histone deacetylase inhibitor as an anticancer drug. Nat Biotechnol 25(1):84–90. doi:10.1038/nbt1272