Gene 588 (2016) 19–29

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Research paper

Integrative functional genomic analysis unveils the differing dysregulated metabolic processes across hepatocellular carcinoma stages

Vignesh Ramesh, Kumaresan Ganesan *

Cancer Genetics Laboratory, Department of Genetics, Centre for Excellence in Genomic Sciences, School of Biological Sciences, Madurai Kamaraj University, Madurai 625 021, India

ARTICLE INFO

Article history:
Received 1 November 2015
Received in revised form 17 April 2016
Accepted 18 April 2016
Available online 20 April 2016

Keywords:
Hepatocellular carcinoma
Metabolic process
Integrative genomics
Transcriptome
Inflammation
Lipid metabolism

ABSTRACT

Hepatocellular carcinoma (HCC) is a highly heterogeneous disease and the development of targeted therapeutics is still at an early stage. The 'omics' based genome-wide profiling comprising the transcriptome, miRNome and proteome are highly useful in identifying the deregulated molecular processes involved in hepatocarcinogenesis. One of the end products and processes of the central dogma being the metabolites and metabolic processes mediate the cellular functions. In recent years, metabolomics based investigations have revealed the major deregulated metabolic processes involved in carcinogenesis. However, the integrative analysis of the holistic metabolic processes with genomics is at an early stage. Since the gene-sets are highly useful in assessing the biological processes and pathways, we made an attempt to infer the deregulated cellular metabolic processes involved in HCC by employing metabolism associated gene-set enrichment analysis. Further, the metabolic process enrichment scores were integrated with the transcriptome profiles of HCC. Integrative analysis shows three distinct metabolic deregulations: i) hepatocyte function related molecular processes involving lipid/fatty acid/bile acid synthesis, ii) inflammatory processes with cytokine, sphingolipid & chondriotin sulphate metabolism and iii) enriched nucleotide metabolic process involving purine/pyrimidine & glucose mediated catabolic process, in hepatocarcinogenesis. The three distinct metabolic processes were found to occur both in tumor and liver cancer cell line profiles. Unsupervised hierarchical clustering of the metabolic processes along with clinical sample information has identified two major clusters based on AFP (alpha-fetoprotein) and metastasis. The study reveals the three major regulatory processes involved in HCC stages.

1. Introduction

Hepatocellular carcinoma (HCC) is the sixth most common cancer and second deadly cancer (Ferlay et al., 2015). HCC incidence is steadily rising and is endemic in Asia and Africa (Bosch et al., 2005; El-Serag, 2011). Only 40% of HCC patients are diagnosed at early stage and curative options such as liver transplantation or surgical resections are difficult for the patients diagnosed at late stage. This is due to the lack of proper screening platforms (Bruix and Llovet, 2002; Singal et al., 2013). Nevertheless, complete understanding and better characterization of the regulatory processes involved in liver tumorigenesis need to be investigated in unveiling the heterogeneous and complex nature of HCC (Lachenmayer et al., 2012; Ramakrishna et al., 2013). This warrants the need for a comprehensive cellular, molecular and functional genomic characterization of the processes in liver cancer for improved early diagnosis and to develop targeted therapeutics.

While the genomics approaches have advanced the understanding of the molecular processes in various cancers (Zender et al., 2006; Furge et al., 2007; Mattison et al., 2010), the unanswered questions in cancer biology can be answered by integrating the genomics and metabolomics as the metabolites are the end products and intermediates of cellular processes (Idle and Gonzalez, 2007). Thus, metabolomics in-hand with genomics might bridge the gap of genotype–phenotype relations in systems biology (Goodacre, 2005). Recent evidences show that metabolic pathways are good targets in cancer since the cancer cells have the altered metabolic flux and metabolism (Tennant et al., 2010). Glycolysis, pentose phosphate pathway, TCA and urea cycle are some of the pathways altered in colon and stomach cancers (Hirayama et al., 2009). Metabolomic signatures also have been derived from pancreatic and hepatocellular cancer cell lines for evaluating the anti-proliferative and anti-apoptotic effects of belinostat and bortezomib (Spratlin et al., 2011; Palmnas and Vogel, 2013). Thus, metabolomics are useful approaches in understanding the biological and clinical status and to use the information to develop screening platform for the diagnosis and aiding better therapeutic options for cancer patients. However, the

Abbreviations: AFP, alpha fetoprotein; ALP, alkaline phosphatase; BCLC, Barcelona-Clinic Liver Cancer; CLIP, Cancer of the Liver Italian Program; DEN, diethylnitrosamine; GEO, Gene Expression Omnibus; HCC, hepatocellular carcinoma; HCV, hepatitis C virus; TCA, tricarboxylic acid; TCGA, The Cancer Genome Atlas; TNM, tumor-node-metastasis; WGCNA, Weighted Gene Co-expression Network Analysis.

* Corresponding author.
E-mail address: kumar@oncocellomics.org (K. Ganesan).

http://dx.doi.org/10.1016/j.gene.2016.04.039
0378-1119/

studies on metabolomics are still in their infancy as the experimentation and the results are known to have high variation and also due to the conditions of metabolite extraction and their half-life (Koal and Deigner, 2010; Koek et al., 2011).

In recent years, gene-sets are highly useful in predicting the status of the signaling pathways and various biological, functional and regulatory processes in tumors (Ooi et al., 2009, 2011; Muthuswami et al., 2013; Tamilzhalagan et al., 2015). As an attempt to establish the approach bridging the gap between transcriptome & metabolome, the dysregulated metabolic processes were inferred from the mRNA profiles of liver tumors and liver cancer cell lines using gene-sets. The major deregulated metabolic processes in different sub-types of HCC were identified from this integrative functional genomic analysis of transcriptome and metabolic processes.

2. Materials and methods

2.1. Metabolic gene-sets and HCC profiles

In total, 445 metabolic process gene-sets related to metabolic, biochemical and enzymatic reactions were collected from Molecular Signature Database v4.0 (Subramanian et al., 2005). The collected metabolism associated gene-sets collectively represent different metabolisms of lipid, fatty acid, steroid, bile, carbohydrate, amine, amino acid, protein, acid, drug, hormone, nitrogen, nucleotide, cell cycle mediating, macromolecular/biopolymer, enzymatic activity, co-enzyme/co-factor, vitamin, alcohol, aromatic, organic, inorganic and acyl chain. The details of the collected gene-sets and their metabolic process category are provided in Supplementary Table 1. Gene expression profiles of HCC and non-tumors were obtained from Gene Expression Omnibus (GEO). From GSE14520 profile, 225 tumors and 220 non-tumor samples profiled with Affymetrix platform HT_HG-U133A alone were used for the analysis. The profile has the sample information for AFP and ALP levels, staging systems (CLIP, BCLC and TNM) and metastasis risk status. Gene expression profile of 20 liver cancer cell lines from GEO, GSE35818 profiled with Agilent Whole Human Genome 4x44 K platform was also used. The cell lines include well differentiated (HEPG2, HEP3B, PLC/PRF/5, HUH1, HUH7, JHH5 & JHH7) and poorly differentiated (JHH1, JHH2, JHH4, JHH6, HLE, HLF, SK HEP1, SNU-182, SNU-387, SNU-398, SNU-423, SNU-449 and SNU-475) cell lines. Gene expression profile of primary HCC profiled by RNA-Seq was obtained from TCGA (The Cancer Genome Atlas). The profile contains 371 HCC and 50 non-tumor tissues profiled with IlluminaHiSeq_RNASeqV2 platform.

2.2. Metabolic process enrichment analysis

The MAS5.0 intensity of the probes in the expression profile GSE14520 was averaged for the replicate genes. In the case of GSE35818, $\log_{10}$ normalized intensity profile was converted to numerical ratio value and averaged the replicate genes. Gene normalized count was used for TCGA profile samples. For the Z-score based enrichment analysis of HCC samples, non-tumor tissues were available to use as reference. However, the gene expression profile of 20 liver cancer cell lines (GSE35818) does not possess normal cell lines for the analysis. The microarray platform of the profile was Agilent dual channel array platform. Here, the reference was a mix of equal amounts of RNA from 19 liver cancer cell lines with the exception of JHH1. Since the reference was itself a mixture of equal amounts of RNA from 19 liver cancer cell lines, the median of the profile was considered as reference. The obtained series matrix file is ratio of Cy5/Cy3 representing test cell line/19 liver cancer cell line mix for all the genes. Therefore, the expression value (Cy5/Cy3) was considered as such for the Z-score calculation in enrichment analysis.

For each gene, fold expression of the tumor samples was calculated with respect to the average of non-tumor tissues as reference. Mean fold expression and standard deviation in fold expression of each sample in the whole gene expression profile were calculated. Similarly, mean fold expression value of the metabolic process gene-sets of each tumor and cell line samples was calculated by extracting the fold gene expression value of gene-sets from the respective profiles. The metabolic process enrichment score, Z-score was calculated by subtracting the mean fold expression of the whole expression profile from the mean fold expression of the gene-sets and divided by standard deviation in fold expression of the profile. Finally, the obtained value was multiplied with the square root of the number of genes in the metabolic process gene-set to attain the normalized Z-score value (Levine et al., 2006). Cumulative expression of mRNA modules was also calculated by the same approach in cell lines. The detailed method of deriving gene-set enrichment score (Z-score) is provided in Supplementary Method 1.

2.3. Unsupervised hierarchical clustering of Z-score based metabolic process enrichment

The derived 445 or selected metabolic processes related gene-set enrichment scores of HCC samples or cell lines were used as input in dChip software. Further, the clinical information of the samples such as AFP levels, metastasis risk, TNM, BCLC & CLIP staging information of tumors and differentiation status of cell lines was also uploaded along with the metabolic process related gene-set scores. Unsupervised hierarchical clustering analysis was performed with the Z-scores of metabolic processes and clinical information of the samples in dChip. The clustering parameters include the average linkage type, mean standardized with display range from -2 to $+2$ and 1-correlation as the distance metric. The sample clustering p-value was set <0.01. The unsupervised hierarchical clustering was visualized as heatmap with clusters of samples and metabolic process related gene-sets in dChip (Li and Wong, 2001).

2.4. Construction of the integrative network that connects mRNA modular expression and metabolic gene-set enrichment

A 'signed' co-expression based mRNA network was constructed with differentially expressed genes between tumor ($n = 225$) and non-tumor samples ($n = 220$) of the tumor profile GSE14520, profiled with HT_HG-U133A platform to understand the role of signaling pathways in HCC development (Vignesh and Kumaresan, unpublished). The network has identified 10 mRNA modules (co-expressed gene clusters). Functional, clinical and signaling pathway association of the mRNA modules has identified three categories of modules: i) hepatocyte (LD1), ii) inflammatory-stress (LD5–LD10) and iii) proliferative (LD2 & LD3) modules. Since modules consist of varying numbers of genes, the summarized expression of the genes in a module is specified as module eigengene value (Langfelder and Horvath, 2007). Correlations between eigengene expression of mRNA modules and the metabolic gene-set enrichment scores were derived using the inbuilt module-trait association matrix of WGCNA (Weighted Gene Co-expression Network Analysis) which provides Pearson's correlation value and significance values (Langfelder and Horvath, 2007). Significant pairs of mRNA modules and metabolic processes were considered based on the correlation between the modular eigengene values and metabolic gene-set enrichment scores. Interaction network was constructed for the modules with metabolic gene-sets based on the correlation values $r < -0.3$ & $r > +0.3$ with a p-value of <0.05. The correlation threshold of $r < -0.3$ & $r > +0.3$ value was set based on the previous study (Mukaka, 2012). The network was visualized using VisANT software (Hu et al., 2004).

2.5. Statistical analysis

The differentially activated metabolic gene-sets between the well and poorly differentiated cell lines were analyzed by employing two-tailed Student's t-test and significance was considered with a p-value <0.05.