

HCCNet: an integrated network database of hepatocellular carcinoma

Cell Research (2010) 20:732-734. doi:10.1038/cr.2010.67; published online 18 May 2010

Dear Editor,

As a complex disease, the development and progression of hepatocellular carcinoma (HCC) involves the interactions of multiple proteins, genes and miRNAs in various biological pathways, and it has been extensively studied with different high-throughput techniques. However, efforts to integrate multiple data sources at different levels, especially with regard to biological pathways and interaction networks, are still negligible in the HCC research field. We have built a database of the HCC network (HCCNet) by integrating interactions of multiple proteins, genes and miRNAs in biological pathways, and manually collecting all of the HCC-related genes and miRNAs from the literature in combination with a bioinformatic analysis of the collected HCC expression data (Supplementary information, Data S1). Currently, there are 37 811 experimentally confirmed protein-protein interactions (PPIs), 9 148 experimentally confirmed transcriptional regulatory interactions (TRIs), 114 miRNA-target gene interactions, 2 234 high-confidence HCC-related genes and 160 HCC-related miRNAs available in the database. The database also provides an online graphic analysis tool to view the interactions among HCC-related proteins, genes and miRNAs. HCCNet is a helpful platform to explore the molecular mechanisms that underlie human HCC. The database can be accessed at <http://www.megabionet.org/hcc>.

We collected data from experimentally confirmed PPIs from the Human Protein Reference Database (HPRD) [1]. PPIs that were accompanied by at least one literature citation (PubMed ID) and confirmed by at least one experimental method were thought to be of high quality. Examples of experimentally confirmed TRIs were collected from either published transcriptional regulation studies or the TRANSFAC 7.0 public database. TRIs that were supported by at least one independent wet experimental study were thought to be of high quality. These high-quality PPIs and TRIs were selected for further biological network construction in our HCCNet system. Moreover, we used “miRNA” and “hepatocellular

carcinoma” as key words to search for literature in the PubMed database, and added miRNAs that were reported to be differentially expressed in human HCC, including their target genes from these published papers to our system. Even though some of the collected target genes for miRNAs were merely predicted and not proven, we believed that the results were useful, as they were provided by authors and subjected to the peer review process.

High-confidence HCC-related genes were obtained through microarray analysis and literature mining. We first used “hepatocellular carcinoma” as key words to search the GEO database for published HCC microarray data. We then selected those microarray data generated from human HCC patients. After microarray analysis, we selected those DEGs that were up- or down-regulated in more than half of the samples in at least one independent microarray study, and that were supported by at least two independent human HCC microarray studies as high-confidence HCC-related genes (Supplementary information, Data S1). We used “gene” and “hepatocellular carcinoma” as key words to search for published papers from the PubMed database. We then selected the HCC-related genes that were supported by at least two published manuscripts and labeled these as high-confidence HCC-related genes. With the criteria mentioned above, 2 234 high-confidence HCC-related genes were inputted into the HCCNet.

miRNAs are not as well studied as genes are in the field of HCC. To include as much data and information related to miRNAs as possible, we did not select high-confidence HCC-related miRNA as stringently as we did for HCC-related genes. Instead, HCC-related miRNAs that were supported by at least one literature source were selected as high-confidence HCC-related miRNAs. A total of 160 miRNAs fit into this category and are regarded as high-confidence HCC-related miRNAs.

Currently, there are two databases, OncoDB.HCC [2] and EHCO [3], which have collected thousands of genes that have been shown to be differentially expressed in some microarray studies related to liver cancer. The EHCO has also done some network analysis among the

hundreds of genes based on predicted PPI data [3]. However, efforts to integrate multiple data sources at different levels, especially with regard to biological pathways and interaction networks, are still exiguous in the field of HCC. Both OncoDB.HCC and EHCO are a collection of single HCC genes and do not address more complex interactions.

As an attempt to integrate multiple data sources at different levels, especially with regard to biological pathways and interaction networks, we built the HCCNet database by integrating experimentally confirmed PPIs, TRIs and miRNA-target gene interactions into a biological network (Supplementary information, Data S1). Gene expression patterns in HCC, KEGG pathways, Gene Ontology and subcellular localization were used to annotate nodes on the network. We also selected high-confidence HCC-related genes and miRNAs, and mapped them to the biological network (Supplementary information, Data S1). We made a comparison of the HCC-related genes in HCCNet with the genes in other HCC databases (OncoDB.HCC and EHCO). The results show that about half of the genes in our database are also collected in the other two databases (Supplementary information, Figure S1), indicating a greater number of HCC genes in the HCCNet. OncoDB.HCC did not include miRNAs and EHCO only offered a downloadable file that contained about 70 miRNAs in HCC. Compared to these sources, HCCNet offers a higher quantity and quality of HCC-related miRNAs. Moreover, HCCNet is not a simple collection of HCC-related genes and miRNAs, as it organizes and displays HCC factors in an interactive way. An online graphic tool was developed to display and analyze the HCC-related biological network in HCCNet. As a result, HCCNet is a platform to explore the molecular mechanism that underlies the development and progression of HCC.

In the current release version, there are 37 811 PPIs, 9 148 TRIs, 114 miRNA-target gene interactions, 2 234 high confident HCC-related genes and 160 HCC-related miRNAs available in the HCCNet database.

The development and progression of HCC cannot be determined by a single factor or a simple collection of factors, but rather by interactions of multiple proteins, genes and miRNAs in various biological pathways. Previous studies on protein–protein interaction networks in liver cancer have found many interactions among proliferation and apoptosis-related proteins and differential glycoproteins; thus, it is proposed that instead of analyzing a single or a few proteins, a “molecule groups” concept should be introduced into the diagnosis and metastasis prediction of HCC [4]. Hence, instead of analyzing a single or a small subset of proteins, the

“molecule groups” concept has been introduced into our HCCNet system. In this way, a HCC-related gene/miRNA that interacts with others could be used as a network entry point to draw a HCC-related biological network and discover potential HCC-related functional modules. HCCNet offers a variety of ways (Figure 1) for users to gain network entry. The first and most important way is to search for high-confidence HCC-related genes in HCCNet by symbol, alias or full name of the gene of interest. For users who do not have a particular gene in mind, alternative approaches are offered in HCCNet: one could select high-confidence HCC-related genes by pathway and subcellular localization on the query page; one could also select high-confidence HCC-related genes by chromosome on the chromosome viewer page; we have also provided a high-confidence HCC-related gene list sorted alphabetically by gene symbol; and lastly, one could search for HCC-related miRNAs by symbol or select one in the HCC-related miRNA list. The target genes of a miRNA are provided in both the list and the search result. If the high-confidence HCC-related gene/miRNA is a network entry, one can find the hyperlink “show network” on the page. This link leads the user to the display center, in which they can view the displayed biological network expanded from the network entry node with our graphic tool.

All the annotations, such as expression patterns in HCC, biological pathways, subcellular localizations, functions and gene ontology annotations, can be found in detailed information pages of the gene entry. Furthermore, if the queried gene is expressed in normal liver, related hyperlinks can be used to view the detailed transcriptomics and proteomics information about this gene in the ATP-HL system.

HCCNet not only lists the high-confidence HCC-related genes that are mapped to the KEGG pathways, but also maps these genes to the pathway in a graphic way (Supplementary information, Figure S2). In this way, it will be easier for users to study how these genes cooperate or how this pathway is activated in HCC.

With HCCNet, when the user studies the molecular network environments of a high-confidence HCC-related gene or miRNA, and checks if any kind of functional module exists when the gene or miRNA is activated in HCC, one should gain network entry by searching or selecting a gene or miRNA from the gene or miRNA list first. Then one can view the biological network expanded from the entry node with our graphic tool. Alternatively, one can expand the network from a group of entries selected by biological pathway and subcellular localization. If the user wants to expand the biological network in this way, they should select a pathway and a subcellular lo-

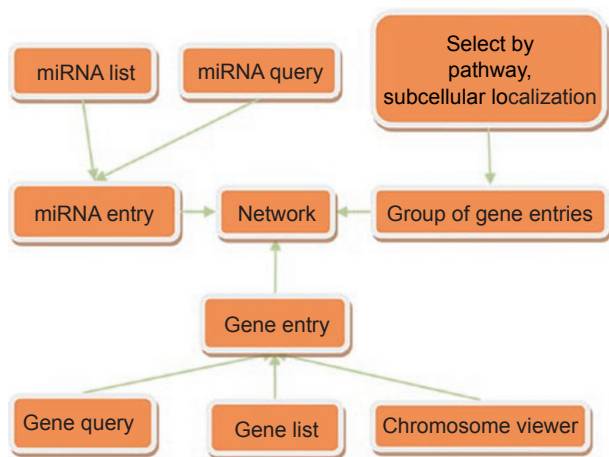


Figure 1 Information flowchart of HCCNet.

calization on the query page, and then choose the button “view network”. With our graphic tool (Supplementary information, Figure S3), the network can be expanded from the node of network entry or nodes of entries by edges. With the optimization of algorithms, the components of the displayed network can be expanded to 600 nodes or until no new node can be added by the edges. In this way, users can analyze and visualize the HCC-related genes and miRNAs in a global view. In addition, users can filter nodes by pathway, subcellular localization, function, degree or other options that are listed to the right of the display. Furthermore, we also provide a function that users can expand the network from the selected limbic node. One can check the selection function on the right menu.

In conclusion, HCCNet is a useful platform when exploring the molecular mechanisms that underlie HCC (Supplementary information, Data S1). In the current release version of HCCNet, the interactions of miRNA-target genes are not as reliable as PPIs and TRIs (Supplementary information, Data S1). However, the selection criterion will become stricter with the accumulation of more related experimentally validated data. The collection of HCC-related genes and miRNAs will continue in the future. To better improve the system, HCCNet will keep pace with the novel discoveries and progress in the liver cancer research field. One of our goals is to identify those signal transduction pathways that have significant changes and explore the relationships between those

pathways during the pathological processes. At the same time, we will also catalog the downstream genes for each altered signaling pathway in live cancer and ultimately fulfill the goal of identifying new potentially relevant liver cancer genes and new mechanisms.

Acknowledgments

We are very thankful to Dr Dušan M Jeftinija (Department of Neuroscience & Anatomy, University of Louisville, Kentucky, USA) for his help during the manuscript preparation. This work was supported by grants from the State Key Program of Basic Research of China (2007CB108800, 2009CB918400, 2010CB912102), the Hi-Tech Research and Development Program of China (2006AA02Z313), National Natural Science Foundation of China (30870575) and Science and Technology Commission of Shanghai Municipality (06DZ22923).

Bing He¹, Xiaojie Qiu¹, Peng Li^{2,3}, Lishan Wang¹, Qi Lv¹, Tieliu Shi^{1,4}

¹Center for Bioinformatics and Computational Biology, and The Institute of Biomedical Sciences, School of Life Sciences, East China Normal University, 500 Dongchuan Road, Shanghai 200241, China; ²College of Life Sciences, The Northeast Forestry University, Harbin, Heilongjiang 150040, China; ³Daqing Institute of Biotechnology, Northeast Forestry University, Daqing, Heilongjiang 163316, China; ⁴Shanghai Information Center for Life Sciences, Chinese Academy of Science, 319 Yueyang Road, Shanghai 200031, China

Correspondence: Tieliu Shi
E-mail: tlshi@sibs.ac.cn

References

1. Keshava Prasad TS, Goel R, Kandasamy K, *et al.* Human Protein Reference Database – 2009 Update. *Nucleic Acids Res* 2008; **37**(Database issue):D767-772.
2. Su WH, Chao CC, Yeh SH, Chen DS, Chen PJ, Jou YS. OncoDB.HCC: an integrated oncogenomic database of hepatocellular carcinoma revealed aberrant cancer target genes and loci. *Nucleic Acids Res* 2007; **35**:D727-731.
3. Hsu CN, Lai JM, Liu CH, *et al.* Detection of the inferred interaction network in hepatocellular carcinoma from EHCO (Encyclopedia of Hepatocellular Carcinoma genes Online). *BMC Bioinformatics* 2007; **8**:66.
4. Zhou HJ, Liu YK, Li Z, Yun D, Shun QL, Guo K. Analysing protein-protein interaction networks of human liver cancer cell lines with diverse metastasis potential. *J Cancer Res Clin Oncol* 2007; **133**:663-672.

(Supplementary information is linked to the online version of the paper on the *Cell Research* website.)