Reviewer’s report

Title: Integrative network analysis identifies key genes and pathways in the progression of Hepatitis C Virus induced Hepatocellular Carcinoma

Version: 1 Date: 29 June 2011

Reviewer: Pierre-Olivier Vidalain

Reviewer’s report:

Siyuan Zheng and collaborators have developed an integrative network approach and applied it to study deregulated events in HCV-induced hepatocellular carcinoma (HCC). They combined human protein-protein interaction data with gene expression profiles encompassing normal liver and four consecutive pathological stages from cirrhosis to HCC. This was performed using a module searching strategy that was developed and described by the same group in a recent publication (Jia P. & al. Bioinformatics, 2011). They identified four interaction modules enriched for differentially expressed genes corresponding to diseases progression. The functional annotation of each network showed a high degree of consistency with current knowledge of HCC. Interestingly, authors clearly identified CDC2 as a master gene deregulated (together with binding partners) during hepatocarcinogenesis. This is an elegant, well-written and well-conducted study that clearly highlights cellular modules deregulated as a consequence of HCV-induced HCC. In addition, this provides an interesting list of potential drug targets for the development of therapeutic approaches.

Minor Essential Revisions:

It is surprising that there was no attempt to compare interaction modules identified in this report with available datasets of HCV-host protein interactions. HCV proteins have been reported to bind numerous cellular proteins. It would be interesting to determine if some of these cellular proteins that are directly targeted by HCV components also belong to the deregulated interaction modules described in this manuscript.

In the “methods” section, authors “ranked the genes by their weights and selected the top 500 genes”. Thus, a list of 500 deregulated genes was extracted from each node-weighted protein interaction network corresponding to each stage of the disease. I understand that 4 lists of 500 genes were established (normal-to-cirrhosis; cirrhosis-to-dysplasia; dysplasia-to-early HCC; early-to-advanced HCC). Is that correct? This should be clarified in the text.

From each list of 500 seed genes, an interaction module was extracted using the algorithm described in Jia P. & al. Identified interaction modules are composed of only 40-60 genes (Table I). Thus, a list of 500 genes is shrunk to a highly-significant sub-network that is about ten times smaller. But the algorithm developed by Jia P. & al is used to expand a network from seed genes. Even after reading the manuscript by Jia P. & al., it is unclear to me how their algorithm
can reduce instead of increasing the size of the network. This should be better explained.

Discretionary Revisions:

Table I. Horizontal lines should be added to clearly identify hub genes associated to each network.

Figure 4 and supplementary figure 2 are almost identical (Figure 4 is showing only major core components). Supplementary figure 2 should be presented as Figure 1, and Figure 4 should be eliminated.

Level of interest: An article of importance in its field

Quality of written English: Acceptable

Statistical review: No, the manuscript does not need to be seen by a statistician.

Declaration of competing interests:

I declare that I have no competing interests