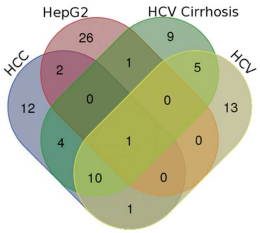
Molecular BioSystems

The gene expression profiling for hepatocellular carcinoma by network analysis approach shows a dominance of Intrinsically Disordered Proteins (IDPs) between HUB nodes

Journal:	*Molecular BioSystems*
Manuscript ID:	MB-ART-06-2015-000434.R1
Article Type:	Paper
Date Submitted by the Author:	23-Jul-2015
Complete List of Authors:	Singh, Sakshi; SUN, Colonna, Giovanni; SUN, Di Bernardo, Giovanni; SUN, Bergantino, Francesca; INT PAscale, Cammarota, Marcella; SUN, Costantini, Susan; INT Pascale,



The results evidence a strong presence of intrinsically disordered proteins in key roles as HUBs in the HCC sub-networks.

Journal Name

ARTICLE

The gene expression profiling for hepatocellular carcinoma by network analysis approach shows a dominance of Intrinsically Disordered Proteins (IDPs) between HUB nodes

Sakshi Singh[1], Giovanni Colonna[2], Giovanni Di Bernardo[3], Francesca Bergantino[4], Marcella Cammarota[3], Giuseppe Castello[4], Susan Costantini[4]

We have analyzed by means of a networking analysis the transcriptomic data from patients with hepatocellular carcinoma (HCC) after viral HCV infection at the various stages of the disease by the publicly available E-MTAB-950 dataset and compared with those obtained in our group from HepG2 cells, a cancer cell line that lacks the viral infection. By a sequential pruning of data, taking also into account the data from cells of healthy patients as blank, we were able to have a distribution of HUB genes for the various stages that characterize the disease and finally we isolated a metabolic sub-net specific of HCC alone. The general picture is that the basic organization to sustain energetically and metabolically the cells both normal and diseased is the same but a complex cluster of sub-networks controlled by HUB genes drives with high metabolic flexibility and plasticity the HCC progression. In particular, we have extracted from HepG2 cells a sub-net of genes strictly correlated to other HUB genes of the network but specific for the HCC and mainly devoted to: i) control at chromatin levels of cell division; ii) control of ergastoplasmatic stress through protein degradation and misfolding; iii) control of the immune response also through an increase of mature T-cells in thymus. This sub-set is characterized by 26 HUB genes coding for intrinsically disordered proteins with high ability to interact with numerous molecular partners. Moreover, we have also noted that periphery molecules, that is, with one or very few interactions (e.g., cytokines or post-translational enzymes), which does not have a central role in the clusters that make up the global metabolic network; essentially have roles of information transporters. The results evidence a strong presence of intrinsically disordered proteins in key roles as HUBs in the sub-networks that characterize the various stages of the disease, conferring a structural plasticity to the net nodes but an inherent functional versatility to the whole metabolic net. Thus, our present article provide a novelty in targeting the intrinsic disorder in HCC networks to dampen the cancer effects and providing new insight into the potential mechanisms of HCC. Taken together, the present findings suggest novel targets to design strategies for drug design and may support a rationale intervention in the pharmacotherapy of HCC and other associated diseases.

[1]*Dottorato in Biologia Computazionale, Dipartimento di Biochimica, Biofisica e Patologia generale, Seconda Università degli Studi di Napoli, Napoli, Italy*

[2] *Servizio di Informatica Medica, Azienda Ospedaliera Universitaria, Seconda Università di Napoli, Napoli, Italy*

[3] *Dipartimento di Medicina Sperimentale, Seconda Università degli Studi di Napoli, Napoli, Italia*

[4]*CROM, Istituto Nazionale Tumori "Fondazione G. Pascale" - IRCCS, Napoli, Italia*

***Corresponding Author:**

E-Mail: susan.costantini@unina2.it; s.costantini@istitutotumori.na.it (SC)

Introduction

Hepatocellular carcinoma (HCC) is a primary malignancy of the liver and the second most malignancy related deaths worldwide [1]. In particular, HCC is the fifth and ninth common cancer in male and female respectively [1]. The chronic viral infection of hepatitis B and C virus (HBV and HCV), the consumption of alcohol and smoking are the main factors that trigger liver diseases and HCC [2]. The obesity and type 2 diabetes are also known to be causative agents for HCC [34] through non-alcoholic fatty liver or fatty liver disease [5] as well as the exposure to vinyl chloride or polyvinyl chloride makes people more susceptible for this type of cancer [2]. However, in literature it is also reported that the iron load and estrogen-progesterone combined oral contraceptives (OC) induce an increase of the HCC risk [2]. Its diffusion changes on geographical regions, ethnic groups, sex group and environmental conditions [2]. Despite recent advances in diagnosis and management, the median survival of HCC patients is less than 8 months [6], and the surgical resection, the liver transplantation, and the local ablation remain the only curative modalities of HCC [7,8] even if its recurrence occurs in up to 70% of patients within 5 years after resection [9,10].

The microarray approach has been already adopted by many researchers using different samples to find driver genes and potential molecular markers able to improve its early

detection and prognosis [11–14] In details, Lau et al. [15] were the firsts to use microarray technology to compare gene expression profiles of HCC and non-HCC liver tissues [16]. Since then, multiple comparative studies have been published and allowed the identification of a number of potential genetic pathways, deregulated in the context of liver carcinogenesis [17]. Among these, Wnt-signaling pathway, p53-signaling pathway, TGF-β, MAPK, IGF-2 and the Jak/Stat pathway were demonstrated to be differentially regulated by means of microarray experiments [18–22]. However, some of the most differentially expressed categories of genes between HCC and non-tumor liver tissues are related to cell cycle progression, RNA splicing, protein degradation, cell adhesion, metabolic enzymes, detoxification, immune response, extracellular matrix and cytoskeleton, DNA damage repair system, and apoptosis, and also cytokines, growth factors, oncogenes, tumor suppressors, and GTP-binding proteins[13,18,19,23,24].

Also massively parallel sequencing approaches have been recently used [25] to characterize individual cancer patients to identify somatic and germ line alterations but they agree in suggesting that several mutations, the knowledge of which has already given rise to specific therapies, were also present in healthy tissue and therefore not due to cancer. This is an inherent problem that is due to the multifactorial origin of HCC and complicates the understanding of the molecular mechanisms and makes uncertain the identification of genes that should guide the progression of cancer.

In addition, networks studies can provide useful insights on highly connected genes and informational flow in networks. In fact, the centrality indices of networks such as degree distribution, betweenness, centrality measures, and HUBness, permit to identify HUB genes that are those most correlated and thus driving genes. In this way it is possible to extract from the set of total genes fewer genes, which are more specifically related to particular physiopathological situations.

In this paper we have focused our attention on detecting HUB-genes key players in HCC+HCV, in an attempt to learn more about this terrible disease because the Southern Italy shows a high mortality trend for liver cancer just in HCV patients [26]. We have collected microarray experiments from 225 liver tissues comprising samples from normal healthy individuals, and from patients with only HCV, with HCV-related cirrhosis and with HCC from HCV-related cirrhosis, evaluating the differentially expressed genes in the different disease stages through a network analysis. Then, we have pruned these data by means of those obtained from our microarray analysis on HepG2 cell line, a model for HCC cancer without viral infection, compared to normal hepatocytes, in order to identify HUB genes, which were specific for HCC in absence of the metabolic effects due to viral progression. The results evidence a strong presence of intrinsically disordered proteins in key roles as HUBs of the sub-networks that characterize the various stages of the disease conferring a structural plasticity to the net nodes and functional versatility to the whole metabolic net. Moreover, we have for the first time isolated a sub-net specifically related to HCC control which has been entirely found made of IDPs.

Methods

Data samples and differentially expressed gene analysis

We extracted the gene expression data obtained from 40 normal liver tissues and from liver tissues of 61 HCV, 17 HCV-related cirrhosis and 107 HCC with HCV-related cirrhosis patients by the publicly available E-MTAB-950 dataset (www.ebi.ac.uk/arrayexpress) obtained using the Affymetrix GeneChip Human Genome U133 Plus 2.0 [27]. We used Robust Multi-array Average or Robust Multi-chip Average (RMA) for the normalization and selected the up and down expressed genes concerning a fold change value above 2 and below -2, respectively, comparing the gene expression data in HCV or HCV-related cirrhosis or HCC with HCV-related cirrhosis

patients in respect to healthy controls. In details, the RMA normalization begins with a computing background corrected perfect match intensities for each perfect match cell on every gene chip.

Then, we re-analyzed the microarray data obtained recently in our laboratory on normal hepatocytes and hepatoma cell line (HepG2) using DESeq tool [28] in R package and a similar fold change value above 2 and below -2, to select the up- and down-expressed genes respectively in HepG2 cells compared to normal hepatocytes. In this case, we used DESeq because RMA is the common approach adopted for analyzing the affymetrix microarray whereas DESeq with Illumina data with low samples.

Network analysis

We analyzed all the differentially expressed genes by the network analysis using Cytoscape [29] against known network of human interactome compiled from Pathway Commons [30], Biological General Repository for interaction Datasets (BioGRID) [31], Human Protein Reference Database (HPRD) [32], ConsensusPathDB [33], Database of Interacting Proteins (DIP) [34], Breast Cancer Information Core (BIC), Michigan Molecular Interactions (MiMI) [35]. Hence, we extracted from the human interactome the differently expressed genes obtained for HCV, HCV-related cirrhosis and HCC with HCV-related cirrhosis in comparison to normal tissues and for HepG2 cells compared to normal hepatocytes. In particular, we considered only the connected component of these seed networks for statistical and functional analysis by using different tools, Netanalyzer [29], DAVID [36] and BiNGO [37], in Cytoscape package, and performed some statistical analysis and three measures of centrality: degree, betweeness, closeness centrality [38]. The degree of a node explains the number of interactions of a particular node with other nodes in the network and the distribution probability of these degrees over the whole network known as degree distribution. However, concerning that the power law is a functional relationship between two quantities, where one quantity varies as a power of another, the power law degree distribution implies the scale-free property (”Rich gets richer effect”) of the network [39]. This property helps to predict the HUB nodes that play important role in the network [40]. Also the betweenness of a node indicates the importance of this node in the network and its involvement in different pathways and how other interactions in the network are controlled by this node [41]. The closeness centrality of a node, ranging from 0 to 1 [42], is defined as the reciprocal of the average shortest path length, and measures the speed of information flow through this node to reachable nodes in the network [42]. Concerning the other statistical analysis, the average clustering coefficient of a network, ranging from 0 to 1 [43], helps to predict the modularity in networks [43], the network density is defined as proportion of all potential connections in a network with actual connections [44], the centralization of a graph explains the overall integration of a network [45] whereas the heterogeneity shows how the network is heterogeneous, and its values is very low in PPI networks and ranges from 0.218±0.129 [46].

Other Analyses

Functional and pathway analysis was performed by BINGO and DAVID tool[47].The prediction of miRNAs having the differentially expressed genes as putative target genes was done by the miRWalk algorithm with eight established miRNA-target prediction programs, i.e., DIANA-microT, miRanda, miRDB, PicTar, PITA, RNA22, RNAhybrid and TargetScan [48].

Intrinsic Disorder - The related protein sequences corresponding to the differentially expressed genes were extracted from UniProt database. To assess the percentage of residues involved into intrinsic disorder we have used the DisProt tool subdividing the sequences in three major groups extracted on the basis of similar contents of disorder (10-

15%, 15-50% and over 50%). However, we used a window value equal to 11[49].

Gene Paralog search - GeneDecks Partner Hunter is an analysis tool which provides a similarity metric by extracting shared descriptors among genes, based on the rich annotation within the GeneCards compendium of human genes in the GeneCards Platform [50]. This analysis is also able to extract putative functional paralogs, namely genes that are similar to the query gene based on combinatorial similarity of attribute annotations. For the sequence paralogy attribute, if a partner candidate is also identified as a sequence paralog (SP), then it is assigned a value of 1 for this attribute and 0 otherwise.

Results

Differentially expressed gene analysis in the data sets

We selected the number of differentially expressed genes in HCV, HCV-related cirrhosis and HCC with HCV-related cirrhosis tissues compared to normal tissues and in HepG2 cells compared to normal hepatocytes (**Table 1)** using the procedure reported in Methods. Then, we performed the seed network analysis on the differentially expressed genes using as background the entire human interactome composed by 15,352 nodes and 281,797 interactions [51]. In details, the seed network approach has been used to extract our network from the human proteome by means of some basic nodes that in our case are the differentially expressed genes taken from experiments. However, the seed network obtained on differentially expressed genes in HCV tissues compared to healthy liver tissues, is composed of 1,708 nodes and 11,452 interactions (S1-A **Fig.)**, that in HCV-related cirrhosis tissues of 1,419 nodes with 8,259 interactions (S2-A **Fig.)**, that in HCC with HCV-related cirrhosis tissues of 1756 nodes and 15420 interactions (S3-A **Fig.)**, and that in HepG2 of 250 nodes with 754 interactions (S4-A **Fig**.).

Table 1 the number of differentially expressed genes in HCV, HCV-related cirrhosis and HCC with HCV-related cirrhosis tissues compared to normal tissues and in HepG2 cells compared to normal hepatocytes.

Conditions	Up-regulated genes	Down regulated genes
HCV	2460	288
HCV-related cirrhosis	2155	32
HCC with HCV-related cirrhosis	2311	160
HepG2 cells	371	280

We compared the number of the exclusive and common genes present in the three different cases (HCV, HCV-related cirrhosis and HCC with HCV-related cirrhosis) as shown in **Fig. 1**. In details, we can underline that: i) the 1201 genes present only in the case of HCV network are mainly involved in oxidation reduction, oxidative phosphorylation and cellular respiration, ii) the 589 genes present only in HCV-related cirrhosis network are involved in cell adhesion, cell motion and response to organic substances, and iii) the 955 genes present only in HCC with HCV-related cirrhosis play an important role in cell cycle, chromatin modification and mitosis. In addition, the 555 genes present in HCV, HCV-related cirrhosis and HCC with HCV-related cirrhosis are involved in the establishment of the protein localization, macromolecule catabolic processes and intracellular transport. The 1104 common genes in HCV and HCV-related cirrhosis network play a role in immune response, defense response and regulation of cell proliferation; whereas the 1012 common genes in HCV-related cirrhosis and HCC with HCV-related cirrhosis network mainly participate in regulation of cellular component size, vascular development

and blood vessel development. The groups of 974 genes that are common in HCV and HCC with HCV-cirrhosis networks play an important role in RNA processing, protein localization and proteolysis involved in cellular protein catabolic process.

As shown in **Fig. 1**, we have 25 genes involved in all the cases of HCV, HCV-related cirrhosis, HCC with HCV-related cirrhosis and HepG2 cell-line. This group of gene is mainly involved in positive regulation of B cell apoptosis and DNA metabolic process. Moreover, the 431 genes present only in HepG2 network are mainly involved in acute inflammatory response, M phase of mitotic cell cycle and response to wounding while the 139 common genes in HCC with HCV-cirrhosis and HepG2 network (85 + 20 + 25 + 9) play an important role in M phase of mitotic cell cycle and microtubule cytoskeleton organization. The 78 common genes in HCV and HepG2 networks (25 + 24 + 20 + 9) are involved in the regulation of the hydrolase activity, response to metal ions and response to wounding. The 102 genes present in HCV cirrhosis and HepG2 networks (25 + 20 + 37 + 20) are mainly involved in actin cytoskeleton organization. The genes commonly expressed in case of HCV-related cirrhosis, HCC with HCV-related cirrhosis and HepG2 networks play also a role in ectoderm development and reproductive developmental process. The greater number of the genes, among those shared between HCC with HCV-related cirrhosis and HepG2 cell-line, are mainly found to be playing important role in M phase (24 genes), cell cycle phase (25 genes) and organelle fission (17 genes) with significant p-values. These genes are also found to be involved in other specific functions like humoral immune response mediated by circulating immunoglobulin, complement activation, chromosome segregation, and activation of plasma proteins involved in acute inflammatory response.

The common nodes between HCC with HCV-related cirrhosis and HepG2 cell-line comprise also high degree nodes like PCNA, AURKA, HNRNPA1, H2AFX, MCM6, HLA-B, KPNA2 and ILF3. The ILF3 gene correspond to the NF90 protein which is involved in mitotic controls and post transcriptional phenomenon, but also in the expression of the gene itself and this property is exploited during the viral multiplication in cells [52]. This gene has also been found to participate in HCC [53] thus, because of its high degree of distribution, it is an important node in the whole network.

Comparing the expression values for the following genes, AXIN2, TOP2A, ILF3, CDC20, PEG10 and DKK1, in HCC with HCV-related cirrhosis and in HepG2 cell line we found that these genes had expression values of fold change much higher in HepG2 cell line suggesting the probable involvement of these genes in HCC than in the viral infection. Moreover, SPINK1 gene resulted up-expressed in HCC with HCV-related cirrhosis, but in HepG2 cell-line this is highly down expressed. In the literature it is reported that the loss of SPINK1 function is found in urothelial carcinoma [54] and its up expression in pancreatitis [55] and in prostate cancer [56].

Network analysis on differentially expressed genes in HCV tissues

The network obtained on differentially expressed genes in HCV tissues compared to healthy liver tissues (S1-A **Fig.)** presents a density equal to 0.008 with a characteristic path length of 3.296 implying the shortest travel between any two nodes, a clustering coefficient of 0.256 index of the network modularity, a network centralization of 0.132, a network heterogeneity of 1.441 and an average number of neighbors of 13.41 (**Table 2)**. The plot of the node degree distribution showed a decreasing trend demonstrating that this network had scale free property and presented robustness against random failures (S1-B **Fig**.). The analysis of the putatively important nodes in this network, detected on the basis of betweenness centrality, degree, average shorted path length and closeness centrality is presented in S1 **Table**. The nodes having the highest distribution degrees were CUL3, FN1,

KIAA0101 and EEF1A1 with 238, 195, 175 and 169 degrees, respectively.

Table 2 Statistical analysis for the all diseased case networks.

Statistical analysis of seed network	HCV network	HCV related Cirrhosis network	HCC with HCV etiology network	HepG2 network	HepG2 network (1st order)
Nodes	1708	1419	1756	250	6509
Interactions	11452	8259	15420	754	220381
Network centralization	0.132	0.129	0.230	0.157	0.745
Average neighbours	13.41	11.641	17.563	6.03	67.715
Network heterogeneity	1.441	1.467	1.482	1.119	1.581
Characteristic path length	3.296	3.265	3.011	3.642	2.336
Clustering coefficient	0.256	0.234	0.269	0.278	0.331
Network Density	0.008	0.008	0.010	0.02	0.01

Moreover, we evaluated the betweenness centrality that provides inferences on the importance of genes on the basis of load placed on the given node in the network, and, hence, information about the core skeleton of the network. Betweenness centrality demonstrated an increasing trend (S1-C **Fig.)** with maximum load placed on: FN1, CUL3, FBXO6, EGFR and KIAA0101 (S1 **Table)**. In details, between the top 30 nodes (**Table 3)**, three down-expressed genes resulted mainly involved in molecular function of cell proliferation, cell adhesion and migration processes whereas twenty-seven up-expressed in functions like nucleotide excision repair, localization of cell and cell death and most importantly in cell cycle by DAVID and BiNGO tool. Moreover, from the pathway analysis by DAVID tool, for example, EGFR, CUL2, JUN, RAC1, SMAD4, SMAD2, STAT1, STAT3, and FN1 genes were found to be involved in cancer pathway (with p-value = 4.59E-04), whereas RPA1, RPA2, PCNA, and RPA3 (with p value = 3.03E-04) in mismatch repair (S2 **Table)**.

Then, we focused our attention on nodes showing HUB–HUB interactions (**Fig. 2)** and we verified that CUL3, FN1, EEF1A1, COPS5 and KIAA0101 were highly interacting in a sub network. These nodes are involved in functions like RNA polymerase activity, MAP/ERK kinase kinase activity, and ribonucleoside binding (S2 Table).

Network analysis on differentially expressed genes in HCV-related cirrhosis tissues

The seed network on differentially expressed genes in HCV-related cirrhosis tissues compared to healthy liver tissues (S2-A **Fig.)** had a density equal to 0.008, a clustering coefficient of 0.234, a network heterogeneity of 1.467, the characteristic path length of 3.265 and the average number of neighbors of 11.6 (**Table 2)**. As in the case of HCV network, the plot of the node degree distribution showed a decreasing trend demonstrating that also this network has scale free property (S2-B **Fig. 2)**. This network had some very high degree nodes like FN1, YWHAZ, MDM2, COPS5 and ACTB with 195, 141, 135, 127 and 126 degrees, respectively (**Table 3 and S3 Table)**. Moreover, we evaluated the betweenness centrality that showed an increasing trend (S2-C **Fig.)** with maximum load placed on: FN1, MDM2, FBXO6, COPS5 and MYC. On the basis of betweenness centrality, degree, average shorted path length and closeness centrality values we found the top degree genes. They were all up-expressed in HCV-related cirrhosis tissues compared to normal liver samples (**Table 3 and S3** Table), and involved in some pathways like cell cycle, pathogenic infection, adherens junction and pathways in cancer to indicate the most significance (S4 **Table)**. In **Fig. 3** one can see the organization of the most significant sub networks with their HUB nodes and relative interactions playing important functional roles in the whole network of HCV-related cirrhosis network. The presence of genes involved into viral infection progression is clearly shown as Pathogenic infection (p = 1.10E-06) as well as genes involved in cancer and leukocyte migration (S4 **Table)**.

Network analysis on differentially expressed genes in HCC with HCV-related cirrhosis tissues

The seed network on differentially expressed genes in HCC with HCV-related cirrhosis tissues compared to healthy liver tissues (S3-A **Fig**.). The statistical analysis evidences that in this network the clustering coefficient is of 0.269, the

Table 3 Top 30 degree nodes in the four seed networks obtained for HCV, HCV-related cirrhosis, HCC with HCV-related cirrhosis, and HepG2 and for the first order network of HepG2. We report the list of the genes from highest and lowest degree value.

HCV	HCV-related cirrhosis	HCC with HCV-related cirrhosis	HepG2	HepG2 first order
CUL3	FN1	SUMO2	AURKA	UBC
FN1	YWHAZ	SUMO1	PCNA	SUMO2
KIAA010	MDM2	FN1	ACTB	NRF1
EEF1A1	COPS5	KIAA0101	AURKB	APP
COPS5	ACTB	COPS5	CDC20	CUL3
CAND1	FBXO6	MDM2	CSNK2A1	ELAVL1
YWHAZ	MYC	YWHAZ	UBD	SUMO1
FBXO6	ITGA4	CAND1	ZWINT	TP53
EGFR	HSP90AB1	FBXO6	CCNB1	HSP90AA1
RPA2	YWHAB	RPA1	MCM6	FN1
PCNA	CAND1	HSP90AB1	H2AFX	CDK2
ITGA4	YWHAQ	PCNA	SRC	EEF1A1
RPA1	TUBA1A	CDK1	MCM3	ESR1
CUL2	VCAM1	HNRNPA1	CHEK1	YWHAZ
YWHAE	YWHAE	HSPD1	UBE2C	KIAA0101
HNRNPC	FYN	CUL2	MAD2L1	CUL1
ICT1	CUL2	PRKDC	CENPA	MDM2
POLR2E	GAPDH	RAN	SFN	GRB2
GAPDH	RPA1	AURKA	RFC4	COPS5
VCAM1	RAC1	YWHAB	CCNA2	CSNK2A1
YWHAB	PCNA	CTNNB1	HNRNPA1	UBD
RAN	XPO1	UBR5	CKAP5	HSPA5
RPA3	RAN	XPO1	BIRC5	HSPA8
EED	DDX3X	PABPC1	SERPING1	NEDD8
YWHAQ	JUN	NPM1	CDC25C	SRC
DNAJA1	LCK	HNRNPC	HLA-B	ACTB
CUL5	HSPB1	CSNK1A1	KIF2C	CAND1
POLR2I	SMAD2	YWHAQ	MCM4	MYC
FYN	CSNK1A1	CCT3	SPC24	TUBB
ATP5B	VIM	RANBP2	INCENP	SIRT7

network centralization of 0.23, the network heterogeneity of 1.482, the characteristic path length of 3.01, and the average number of neighbors of 17.56 (**Table 2)**. Moreover, the density is equal to 0.01 which is a value greater than those obtained for HCV and HCV related cirrhosis networks because this network have more edges per nodes. It is important to underline that HCC with HCV-related cirrhosis network is more centralized and more clustered, compared to HCV and HCV-related cirrhosis networks as well as the characteristic path length is lower, suggesting an easy travel from one node to another. As in the case of HCV and HCV-related cirrhosis networks, the plot of the node degree distribution showed a decreasing trend demonstrating that also this network had scale free property (S3-B **Fig**.). This network is composed of very high degree nodes like SUMO1, SUMO2, FN1 and KIAA0101 with 421, 269, 209 and 205 degrees, respectively. Moreover, we evaluated the betweenness centrality that showed an increasing trend (S3-C **Fig.)** with maximum load placed on: SUMO2, SUMO1, FN1, COPS5 and MDM2.

On the basis of betweenness centrality, degree, average shorted path length and closeness centrality values we found top 30 degree genes that were all up-expressed in HCC with HCV-related cirrhosis tissues compared to normal cases (**Tables 3 and S5 Table)**. As one can see in S6 **Table** the distribution of their functions and pathways is very large with high p-values (5.75E-16 for pathways and 3.34E-07 for functions). This suggests that the concomitant effect of cancer and viral infection has a strong impact on the metabolism with a strong increase of functional activities. The overall picture that we can observe is functionally very complex with strong activity at nuclear level and of post-translational modifications, due to the progression of the chronic inflammation started from the viral infection. However, S6 **Table** shows also an evident methylation activity, suggesting epigenetic modifications, as well as those of various kinases. In few words we have a greater number of metabolic clusters strictly connected with short path length between them and much more edges for node. Functionally speaking the cancer activates new metabolic pathways and this certainly leads to more global metabolic energy expenditure for the organism.

Also in this case, we focused on HUB–HUB interactions and verified that two HUB nodes such as SUMO2 and SUMO1 **Fig. 4**, are highly connected with each other and these genes play an important role in SUMoylation.[57] Moreover, the other interconnected HUB nodes like HSP90AB1, SMAD2, YWHAZ etc. are related to DNA replication origin binding, single–stranded RNA binding, ligase activity, p53 binding and RNA binding functions (S6 **Table)**.

HepG2 Network

We performed the network analysis on down- and up-expressed genes resulted common by our previous study and DESeq analysis [28] (S4-A F**ig**.). This network presents a density of 0.024, the clustering coefficient of 0.278, the heterogeneity network of 1.12, the centralization of network of 0.158, the characteristic path length of 3.642, the average number of neighbors of 6.03 (**Table 2)**. The degree distribution and betweenness distribution of nodes follow power low explaining the scale free property of the network (S4-B and C **Figs**.). The highest betweenness exhibiting nodes in the network are ACTB, PCNA, UBD, AURKA and CSNK2A1. These nodes play a bridge role with the rest of the network.

The highest 30 degree nodes ranged from 45 to 13 (**Tables 3 and S7 Table)** and many of these genes are also involved in HCV, HCV-related cirrhosis and HCC with HCV–related cirrhosis. The functional and pathway analysis evidenced that some HUB genes are involved in ATP, nucleotide and kinase activities and in the following pathways like cell cycle, DNA replication, and p53 signaling pathways (S8 **Table)**. In the sub

network of HepG2 HUB nodes like AURKA, AURKB, PCNA, ACTB (**Fig. 5)** and other nodes are highly connected and hence the sub network is playing an important role in outlay of whole network.

Moreover, to check the independence of HepG2 differentially expressed genes in human proteome, we analyzed the first order network composed from 6,509 nodes and 220,381 interactions (S5-A **Fig**.). In details, its density is equal to 0.01 with the heterogeneity of 1.58, and the average number of neighbors was equal to 67.7 (**Table 2)**. Moreover, the characteristic path length of HepG2 first order network was equal to 2.336; this value is lesser than that in other networks and is index of the fast flow of information in this network (**Table 2)**. However it is highly centralized (with 0.745) with a clustering coefficient of 0.331 that is higher compared to all the other networks showing the importance of seed nodes in the network (**Table 2)**. The degree and betweenness distribution plots of first order of HepG2 network follow the power law showing the scale free behavior of the network (S5-B, C **Figs**.). The high betweeness nodes in the first order network are UBC, APP, NRF1, SUMO2 and ELAVL1.

In the first order of HepG2 (**Tables 3 and S9 Table)** we found the highest degree nodes of the seed network PCNA, ACTB, UBD and CSNK2A1 implying their importance in diseased condition. While UBC, SUMO2, NRF1, SUMO1, TP53, HSP90AA1, FN1, CUL1, MDM2, GRB2, COPS5, HSPA5, SRC, CAND1, MYC, TUBB and SIRT7 are among most interacting nodes with a high degree of network. The HUB nodes like CDK2, TP53 and MDM2 with other high degree node are involved in cell cycle, prostate cancer and pathways in cancer, NEDD8, CDK2 and CUL1 in regulation of p27 phosphorylation during cell cycle progression and CSNK2A1, TP53, MYC and CUL1 in Wnt signaling pathway. The high degree nodes of first order of HepG2 network are also found to be involved in molecular functions of protein modification by small conjugation, regulation of apoptosis and mitotic cell cycle. They show specific considerable metabolic complexity related essentially to boost cell proliferation through the control of the cell cycle. An interesting observation is related to the functions expressed by these genes primarily aimed at controlling the formation of complexes, also ATP dependent, and phosphorylation.

Comparison between HUB nodes

We compared the presence and the role of HUB nodes in four networks of HCV, HCV-related cirrhosis, HCC with HCV-related cirrhosis and HepG2 cell line (**Fig. 6 and S10 and S11 Tables)**. Since our aim was to discriminate between HCC in presence or in absence of viral infection we focused mainly on the comparison between the HUB nodes common between HCC with HCV-related cirrhosis and HepG2 cell line or resulted specific for two situations.

In particular, PCNA was a common HUB node in all four networks. It is involved in cell cycle and in cancer pathways, was found implicated in the liver related infection and in HCC [58] and its dysregulation determines both tumor progression as well as the outcome of anticancer treatment [59,60]. AURKA and HNRNPA1 were two HUB nodes present in the two HCC with HCV-related cirrhosis and HepG2 networks. Both AURKA, a centrosome-associated serine/threonine kinase, and HNRNPA1, belonging to the A/B subfamily of ubiquitously expressed heterogeneous nuclear ribonucleoproteins, resulted up-expressed frequently in HCC, and to correlate with high grade and high stage, indicating that their role in the development and progression of HCC [61,62].

However, twelve genes resulted HUB nodes only in the HCC with HCV-related cirrhosis network.

Among these genes, CTNNB1 is involved in Wnt/beta-catenin pathway and in cellular survival [63], HSPD1 in stress response in the mitochondria, CDK1 in cell division cycle, UBR5 in mitotic non-disjunction and chromosome instability [64],

PABPC1 in cytoplasmic regulatory processes of mRNA metabolism, HNRNPC belonging to heterogeneous nuclear ribonucleoproteins in pre-mRNA processing, NPM1 in regulation of ARF/p53 pathways [65] and SUMO1, SUMO2 and RANBP2 in the sumoylation and nuclear export pathway. Moreover, since CCT3 is a molecular chaperone, and PRKDC a serine/threonine-protein kinase, they develop their function by binding different other proteins and forming the protein complexes.

Finally twenty six genes resulted HUB nodes only in HepG2 network. In details, CCNA2, CCNB1, MCM 3, 4 and 6 are involved in cell cycle checkpoint signaling pathways [66,67], SPC24 and ZWINT in kinetochore and participate in anaphase of cell cycle [68], CENPA and INCENP in centromere function, RFC4 in elongation of multiprimed DNA template [69], CDC20 and BIRC5 in apoptosis [70], CDC25C in G1/S and G2/M checkpoints [71], H2AFX in the nucleosome [72], KIF2C and MAD2L1 in cell division, CKAP5 in centrosomal microtubule assembly [73], and UBD and UBE2C in mitotic non-disjunction and chromosome instability[64]. Moreover, CSNK2A1 is a casein kinase, CHEK1 a serine/threonine kinase and SRC a proto-oncogene tyrosine-protein kinase that phosphorylate a large number of substrates, SFN is implicated in Akt/mTOR pathway and in p53 activation [74], and HLA-B plays a critical role in the immune system and SERPING1 is a serpin peptidase inhibitor found involved in hepatocellular carcinoma [75,76].

Discussion

The HCC presentation has significantly changed over the past years. Although its etio-pathogenesis is still not clearly elucidated, it is now manifest that the disease is clearly multifactorial[77] and often develops in patients with underlying cirrhotic liver disease of various etiologies. The cirrhotic liver is characterized by fibrosis, inflammation, necrosis, and ongoing regeneration, which support the carcinoma progression, including the modification of numerous biochemical pathways[78]. The biochemical transformations might be also induced by external and environmental factors thus the origin of the genetic changes that, for example, increase cellular proliferation, are not easily discernible. Further, whereas we can exclude the familial and hereditary cancers clearly demonstrated of genetic origin, for the remaining [79] we need to consider the well-documented influence of external molecules as well as dietary components on cancer initiation and progression trough epigenetic modifications [80]. As a logical consequence, this leads to think a sequential interaction between: stressors, epigenetic, metabolic network, and cancer.

Aim of our study is to extract information by means of a less common approach, i.e., by analyzing the inter-genic relationships through networks [39]. In this article we compare differentially the relationships existing between two extreme situations, i.e., pure cancer cells without any presence of the virus infection (HepG2 cells) and cells of tissues that characterize the various stages of the disease progression from infection, HCV-related cirrhosis and HCC with HCV-related cirrhosis. Our attempt is to isolate, as much as possible, the contributions of genes and their related proteins that operate together to specifically sustain the HCC proliferation. Since no gene works by itself but it is part of the whole metabolic network together with its product (the protein) so that, in principle, any biologically altered function of the network that we suspect associated to pathologically different situations, can be evidenced through those genes that change their expression. In our case, we have identified the genes that play a key role as HUB nodes, because they, having extensive interrelationships between genes, effectively control or mediate the biological activities in the network.

The overall picture that emerges from our experiments is the presence of mutual interactions involved in metabolic networks typically linked to biochemical pathways, grouped in metabolic modules. The most noticeable consequence of

this property is the presence of few highly connected HUBs that hold connected the whole metabolic network. These HUB genes control modules that link inter-connected biochemical processes. The validity of our network clustering analysis can only be assessed in terms of biological relevance. In S2, S4, S6 and S8 **Tables** we report the biological functions related to these genes, some of which have been found also by other researchers, for instance, the genes involved in oocyte meiosis and progesterone mediated oocyte maturation [81], and the general frame work of the common genes (Fig. 1). Since we are working with a regulatory network of coding HUB genes, topological clusters should correspond to groups of proteins involved in processes convergent on similar biological functions, and probably located in the same subcellular compartment. The analyzed networks show an equivalent high degree of clustering, which implies the existence of topological modules that represent highly interlinked local regions involved in the same disease with increased tendency to interact with each other, therefore, all the extracted genes show important functional activities as proteins supporting the cancer cell proliferation and controlling at chromatin level the cell division. We note also strong HUB-HUB interactions that increase the functional connectivity between network and sub-network regions sharing and strengthening functional properties in topologically defined metabolic sub-areas. From the Venn diagram in F**ig. 2,** we note that 26 genes are specifically related to HepG2, but since all code for real proteins, they deserve a more detailed analysis. In S8 **Table** we report that analysis. In general, their functional properties can be roughly summarized in three major groups: i) control at chromatin level of cell division; ii) control of ergastoplasmatic stress through protein degradation and misfolding; iii) control of the immune response also through an increase of mature T-cells in thymus. We would like also to highlight that the complex tridimensional and hierarchical structure of the chromatin organization certainly requires a large number of proteins that are involved in the same process. These complex sets of interactions seems to represent the mechanistic foundation for much of the cell function trough complex networks of protein-protein interactions [82]. Among these there are also some kinases which act mainly as a regulatory nuclear nodes which integrate and coordinate numerous signals leading to appropriate cellular response (CSNK2A1) [83] or as molecular sensors for DNA damage (PRKDC) [84]. Our observations show that the nodes of the first order HepG2 sub-network, as proteins, utilize a large number of physical interactions, corresponding to a large number of different molecular partners.

Our present knowledge of cell physiology suggests that only the intrinsic disorder, exerted by the IDPs, can provide molecules so flexible to be able to interact with a wide range of partners. In fact, IDPs have been found to be involved in a number of human diseases, including cancer as well many studies have also been performed on individual IDPs, where they were always found involved in important and different metabolic roles often related to cancer [85]. However, their ability to recognize and interact with multiple partners can be fully explained only if we demonstrate that all our nodes, when translated into proteins, contain, at varying degrees, structural regions rich in intrinsically disordered segments [86]. Our bioinformatics analysis shows that this is exactly the case (S12 Table). An important recent observation that support our findings is that the transcription, which controls the complex and crucial biological functions predominantly localized inside the cell nucleus, is a functional process intimately related to the IDPs [87]. Also, an analysis (not shown) of the other proteins involved in the progression of HCC, even in the presence of the viral infection, does conclude that also the majority of them is IDP. In few words, our results suggest that the HUB proteins control the key nodes but probably also the general metabolic organization supporting the HCC proliferation. Recent studies suggest that

signaling and regulatory roles carried out by IDPs require them to be tightly regulated, and that altered IDP abundance may lead to disease [88]. The wide presences of IDPs in cells are generating a mounting interest in understanding the structure of these proteins but this aspect is still an open challenge. In fact, although IDPs, or their regions (IDRs), hold many biological functions, their molecular mechanisms, which are often elusive to the experimental characterization, revolve around their ability to act as centers for many protein-protein interactions, visibly diverging from those of the classic globular proteins.

In principle, their function may be controlled by post-translational modifications that lead to structural changes during the interaction with the target [89]. However a multisite phosphorylation can give rise to a wider range of functional responses, allowing the same protein to bind many different molecular targets with various functional consequences. Undoubtedly, the combinatorial post-translational modifications with a great number of isoforms [90] add complexity to regulatory networks and provide a mechanism for tissue-specific signaling. However, we must not forget that IDP functionalities are also context dependent [89], therefore, an open question is to know on a global metabolic scale how the connectivity between metabolic nodes translates into the interactions between HCV and HCC [91]. A final issue is that the shortest path length observed between nodes of the first order reflects an overall network connectivity, where hub nodes are central to network topology, but the finding that these nodes have an average of about 220 interacting partners suggests that they might be considered mainly as date hubs and not party hubs [92]. Thus, the hub topology seems favorable to drug design by means of well-tuned attacks against target HUB nodes, so generating the most massive functional effects [92].

From our analysis it seems that HCC should not be understood through a vision connected only with the genetic mutations, rather, it should be considered as a novel tissue, in which the cancer cells interact with the surrounding metabolic environment, communicant traits that promote their own survival. Further, the ability of the cancer cell population to regenerate and reprogram themselves in response to hostile microenvironment, and ultimately persisting in their proliferative state, is controlled by intrinsically disordered proteins. We have found that the nodes of these metabolic networks are occupied by proteins that are structurally rather flexible and extended, with physico-chemical properties diametrically opposed to those of the classic globular proteins, which allow them to easily switch from one partner to another [93] simply by means of post-translational modifications. Moreover, they easily buffer numerous mutations, but can change their function (i.e. the partner) mainly by mutations in Ser, Thr or Tyr, the targets of post-translational modifications [90]. Thus, independently from the mutation site, it is amazing the infinite number of possible functional variations that a sub-network containing them, may implement as a consequence of the kinase action. From this point of view the peripheral molecules are important to shift the HUB nodes of metabolic sub-nets to different functions. In practice, the presence of these functionally multiform nodes suggests that the flow of information, for instance such as that transported by cytokinome or by kinome, is likely one of the critical points in the control of the metabolic routing between stages of the disease. In fact, the joint effects between IDPs, informative molecules, and post-translational modifications offer to networks a high metabolic plasticity, and proliferative flexibility with no or very few changes in genome.

Another noticeable observation come from the recent finding that a long non-coding RNA (HULC), highly up regulated in liver cancer, plays an important role in HCC, and that the PKA (protein kinase A) pathway may be involved in its up-regulation. Some Authors demonstrated that HULC may act as an endogenous 'sponge', which down-regulates a series of miRNAs activities, including miR-372 [94]. Where, the

related inhibition of miR-372 can lead to the expression increase of its target genes. Therefore, we have evaluated the presence of targets for miR-372 in our lists of differentially expressed genes as well among the HUB genes in the networks related to HCC with HCV-related cirrhosis and HepG2 cells. We performed a prediction study of the putative genes that can be the target of miR-372 by miRWalk algorithm [48] and, considering a consensus of five out eight tools, we selected the genes among those found differentially expressed according to our data. In this way, we obtained a total of 2344 targets for miR-372, among which two groups of 348 and 44 genes were common to the lists of up-expressed genes in HCC with HCV-related cirrhosis and HepG2 cells, respectively. In details, 1/348 (UBR5) and 3/44 (MCM3, MCM4 and MCM6) were HUB genes in the related networks S13 **Table**. Moreover, recently it has been published that HULC [95] contributes to the perturbations in circadian rhythm of the hepatoma cells inducing an hepato-carcinogenesis promotion. This is also confirmed from our analysis which has evidenced that among the possible targets for miR-372 in our lists of up-expressed genes in HCC with HCV-related cirrhosis and HepG2 cells S13 **Table**, have been found 15 genes involved into the circadian rhythm. Moreover, to support this result, we have selected also the best 100 out of 679 inferred partners found for HULC that share the same disorder (i.e., HCC). As we can see on S14 **Table**, we have found three genes like MDM2 [96], a nuclear-localized E3 ubiquitin ligase that can promote tumor formation by targeting tumor suppressor proteins, such as p53, for proteosomal degradation; MYC [97], a transcription factor that regulates cell cycle progression, apoptosis and cellular transformation as transcription of specific target genes; and CDK1 [98], a catalytic subunit of the highly conserved protein kinase complex known as M-phase promoting factor (MPF), which is essential for G1/S and G2/M phase transitions of eukaryotic cell cycle, functionally correlated to HULC in HCC among the top first order HUB nodes specific for HepG2. However, also another gene BIRC5 [99], (a multitasking protein that has dual roles in promoting cell proliferation and preventing apoptosis) has been found among the top 30 degree nodes of HepG2 genes. These findings seem to correlate the origin of the HCC to possible alterations through HULC. However, this result falls within the recent observation indicating that RNA editing modification may play an important role in the development of HCC [100]. Interestingly, also this observation seems to support the sequence: stress, epigenetic event, metabolic network changes, and cancer development.

Actually, our results offer a reasonable conceptual foundation for understanding how metabolic alterations may contribute to cancer if treated organically. In fact, it is now well known that chemical modifications of histones and DNA control the epigenetic gene regulation, and that the malignancy is pervasive in tissues also through the disruption of the epigenetic control [101] inducing changes which allow cancer cells to progress [102] with molecular mechanisms similar to those determined by instability and mutations of the genome. At the same time, many of the enzymes involved into such chemical modifications are sensitive to metabolic changes also due to diet [103]. However, these aspects deserve deeper investigation to shed light on the origin of HCC when not connected to the viral infection.

The general consideration that can be done in the presence of a so widespread participation of IDPs at metabolic key points of the cells affected by cancer, but also by the viral infection, is that we are in the presence of molecular mechanisms supported by nodes that have a inherently infinite molecular adaptation that allows them to coordinate, with continuity, new metabolic changes, also unfavorable. However, any network of genes coding for real proteins, in normal physiological conditions has a possible pattern of interactions dictated by their concentration and by the time of their physical presence. Both these parameters are tightly regulated by the genome but the genes of which we are

speaking about are over expressed. This means that their local concentration in cell rises and thus favors the protein-protein interactions, which are strongly concentration dependent. In this case we have a noise affecting both time and mass. But, because the HCC sub-network is made of IDPs, their over expression opens to many new potential interactions. If we consider that the same is also valid for the numerous kinases that show disordered regions, we can reasonably hypothesize a new balance of the interaction pattern, which is able to metabolically switch to a new phenotype. All this dramatically changes the present vision on various sequential mutations for the cancerous progression. Probably we can also consider a genomic variation following the initial insult, but nothing excludes that it can also be of epigenetic nature rather than a true gene mutation, or even due to stressors of still un-known nature that, for instance, modify some critical step of the circadian system. Therefore, any curative approach acting only on a local metabolic pathway or aimed to hit a local signaling system may be frustrated because it underestimates the enveloping role of the global metabolic network in cellular functioning. It is evident from our data that HCC involves abnormal metabolic states that change the normal tissue physiology and lead to tissue dysregulation, which today we know that it can be done through many concomitant metabolic options [104], and this is not surprising because in an non-deterministic complex system, as the cellular metabolism is, many different space-time solutions are possible for a single stress. A second consideration comes from the observation that the main HUB nodes in HepG2 have a field of action essentially nuclear and most of them are transcription factors. In few words the origin of each variation resides mainly on the control of the chromatin function exerted by genes coding for IDPs. All this, with due caution, seems a sophisticated form of metabolic parasitism, exerted through the control of the global network. The main question remains: where does the primary insult?

The last consideration is on the general metabolic role exerted by the molecules that are allocated to the periphery of the various modules of the network, that is, molecules such as cytokines, many kinases, and those enzymes devoted to post-translational modifications. They have not key positions, because they play an informative work, namely to carry the information needed to determine adequate flows of mass and energetic needs between functional sub-nets, under the changing metabolic pressure exerted by the HUB networks. This is a view at present not easy to pursue because of the pervasive presence at key nodes of the IDPs. In fact, about these proteins, we have some knowledge only from a functional point of view but still little about their structural behavior in solution and of the molecular mechanisms in which they are involved.

Acknowledgements

We thank Marilina Russo for her help in the figure preparation.

References

1 J. Ferlay, I. Soerjomataram I, R. Dikshit, S. Eser, C. Mathers, M. Rebelo, D. M. Parkin, D. Forman D and F. Bray, *Int. J. Cancer*, 2014, **00**.

2 S.-C. Chuang, C. La Vecchia and P. Boffetta, *Cancer Lett.*, 2009, **286**, 9–14.

3 S. C. Larsson and a Wolk, *Br. J. Cancer*, 2007, **97**, 1005–8.

4 Y. Qian and F. Jian-Gao, .

5 H. B. El-Serag and K. L. Rudolph, *Gastroenterology*, 2007, **132**, 2557–76.

6 H. B. El-Serag, *Gastroenterology*, 2004, **127**, 27–34.

7 J. Bruix and M. Sherman, *Hepatology*, 2005, **42**, 1208–1236.

8 J. Bruix and J. M. Llovet, *Hepatology*, 2002, **35**, 519–524.

9 J. M. Llovet, J. Fuster and J. Bruix, *Hepatology*, 1999, **30**, 1434–1440.

10 P. M. Loria, A. Duke, J. B. Rand and O. Hobert, *elsevier*, 2003, **38**, 200–207.

11 T. Maass, I. Sfakianakis, F. Staib, M. Krupp, P. R. Galle and A. Teufel, *Curr. Genomics*, 2010, **11**, 261–8.

12 H. Dong, X. Ge, Y. Shen, L. Chen, Y. Kong, H. Zhang, X. Man, L. Tang, H. Yuan, H. Wang, G. Zhao and W. Jin, *BMC Med. Genomics*, 2009, **2**, 5.

13 W. Wang, J. X. Peng, J. Q. Yang and L. Y. Yang, *Dig. Dis. Sci.*, 2009, **54**, 2729–35.

14 B. Skawran, D. Steinemann, A. Weigmann, P. Flemming, T. Becker, J. Flik, H. Kreipe, B. Schlegelberger and L. Wilkens, *Mod. Pathol.*, 2008, **21**, 505–16.

15 W. Y. Lau, P. B. Lai, M. F. Leung, B. C. Leung, N. Wong, G. Chen, T. W. Leung and C. T. Liew, *Oncol. Res.*, 2000, **12**, 59–69.

16 W. Y. Lau, P. B. Lai, M. F. Leung, B. C. Leung, N. Wong, G. Chen, T. W. Leung and C. T. Liew, *Oncol. Res.*, 2000, **12**, 59–69.

17 J. Liang, J. Lv and Z. Liu, *Exp. Mol. Pathol.*, 2015, **98**, 540–545.

18 O. Delpuech, J.-B. Trabut, F. Carnot, J. Feuillard, C. Brechot and D. Kremsdorf, *Oncogene*, 2002, **21**, 2926–2937.

19 H. Okabe, S. Satoh, T. Kato and T. Progression, *Cancer Res.*, 2001, **61**, 2129–2137.

20 N. Iizuka, M. Oka, H. Yamada-okabe, N. Mori, T. Tamesa and T. Okada, 2002, 3939–3944.

21 E. Wurmbach, Y. B. Chen, G. Khitrov, W. Zhang, S. Roayaie, M. Schwartz, I. Fiel, S. Thung, V. Mazzaferro, J. Bruix, E. Bottinger, S. Friedman, S. Waxman and J. M. Llovet, *Hepatology*, 2007, **45**, 938–947.

22 S. Boyault, D. S. Rickman, A. De Reyniès, C. Balabaud, S. Rebouissou, E. Jeannot, A. Hérault, J. Saric, J. Belghiti, D. Franco, P. Bioulac-Sage, P. Laurent-Puig and J. Zucman-Rossi, *Hepatology*, 2007, **45**, 42–52.

23 and P. O. B. Xin Chen, Siu Tim Cheung, Samuel So, Sheung Tat Fan, Christopher Barry, John Higgins, Kin-Man Lai, Jiafu Ji, Sandrine Dudoit, Irene O.L. Ng, Matt van de Rijn, David Botstein, *Mol. Biol. Cell*, 2002, **13**, 1929–1939.

24 P. Zindy, L. Andrieux, D. Bonnier, O. Musso, S. Langouët, J. P. Campion, B. Turlin, B. Clément and N. Théret, *FEBS Lett.*, 2005, **579**, 95–99.

25 S. Jones, V. Anagnostou, K. Lytle, S. Parpart-li, M. Nesselbush, D. R. Riley, M. Shukla, B. Chesnick, M. Kadan, E. Papp, K. G. Galens, D. Murphy, T. Zhang, L. Kann, M. Sausen, S. V Angiuoli, L. A. D. Jr and V. E. Velculescu, *Sci. Transl. Med.*, 2015, **7**, 283ra53.

26 M. Montella, M. Malvezzi, M. Grimaldi, F. Nocerino, F. Frigeri, A. Pinto, A. Giudice and A. Crispo, *Hepat. Mon.*, 2013, **13**, e12725.

27 G. Rustici, N. Kolesnikov, M. Brandizi, T. Burdett, M. Dylag, I. Emam, A. Farne, E. Hastings, J. Ison, M. Keays, N. Kurbatova, J. Malone, R. Mani, A. Mupo, R. Pedro Pereira, E. Pilicheva, J. Rung, A. Sharma, Y. A. Tang, T. Ternent, A. Tikhonov, D. Welter, E. Williams, A. Brazma, H. Parkinson and U. Sarkans, *Nucleic Acids Res.*, 2013, **41**, D987–90.

28 S. Anders and W. Huber, *Genome Biol.*, 2010, **11**, R106.

29 M. S. Cline, M. Smoot, E. Cerami, A. Kuchinsky, N. Landys, C. Workman, R. Christmas, I. Avila-Campilo, M. Creech, B. Gross, K. Hanspers, R. Isserlin, R. Kelley, S. Killcoyne, S. Lotia, S. Maere, J. Morris, K. Ono, V. Pavlovic, A. R. Pico, A. Vailaya, P.-L. Wang, A. Adler, B. R. Conklin, L. Hood, M. Kuiper, C. Sander, I. Schmulevich, B. Schwikowski, G. J. Warner, T. Ideker and G. D. Bader, *Nat. Protoc.*, 2007, **2**, 2366–2382.

30 E. G. Cerami, B. E. Gross, E. Demir, I. Rodchenkov, O. Babur, N. Anwar, N. Schultz, G. D. Bader and C. Sander, *Nucleic Acids Res.*, 2011, **39**, D685–90.

31 C. Stark, B.-J. Breitkreutz, T. Reguly, L. Boucher, A. Breitkreutz and M. Tyers, *Nucleic Acids Res.*, 2006, **34**, D535–9.

32 S. Peri, J. D. Navarro, R. Amanchy, T. Z. Kristiansen, C. K. Jonnalagadda, V. Surendranath, V. Niranjan, B. Muthusamy, T. K. B. Gandhi, M. Gronborg, N. Ibarrola, N. Deshpande, K. Shanker, H. N. Shivashankar, B. P. Rashmi, M. a Ramya, Z. Zhao, K. N. Chandrika, N. Padma, H. C. Harsha, a J. Yatish, M. P. Kavitha, M. Menezes, D. R. Choudhury, S. Suresh, N. Ghosh, R. Saravana, S. Chandran, S. Krishna, M. Joy, S. K. Anand, V. Madavan, A. Joseph, G. W. Wong, W. P. Schiemann, S. N. Constantinescu, L. Huang, R. Khosravi-Far, H. Steen, M. Tewari, S. Ghaffari, G. C. Blobe, C. V Dang, J. G. N. Garcia, J. Pevsner, O. N. Jensen, P. Roepstorff, K. S. Deshpande, A. M. Chinnaiyan, A. Hamosh, A. Chakravarti and A. Pandey, *Genome Res.*, 2003, **13**, 2363–71.

33 A. Kamburov, C. Wierling, H. Lehrach and R. Herwig, *Nucleic Acids Res.*, 2009, **37**, D623–8.

34 I. Xenarios, L. Salwinski and X. Duan, *Nucleic acids …*, 2002, **30**, 303–305.

35 M. Jayapandian, A. Chapman, V. G. Tarcea, C. Yu, A. Elkiss, A. Ianni, B. Liu, A. Nandi, C. Santos, P. Andrews, B. Athey, D. States and H. V Jagadish, *Nucleic Acids Res.*, 2007, **35**, D566–71.

36 D. W. Huang, B. T. Sherman and R. a Lempicki, *Nat. Protoc.*, 2009, **4**, 44–57.

37 S. Maere, K. Heymans and M. Kuiper, *Bioinformatics*, 2005, **21**, 3448–3449.

38 L. Freeman, *Sociometry*, 1977.

39 A.-L. Barabási, N. Gulbahce and J. Loscalzo, *Nat. Rev. Genet.*, 2011, **12**, 56–68.

40 J. Wu, Y. Tan, H. Deng and D. Zhu, 1–8.

41 J. Yoon, A. Blumer and K. Lee, *Bioinformatics*, 2006, **22**, 3106–8.

42 M. Newman, *Soc. Networks*, 2005, **27**, 39–54.

43 E. Ravasz, A. Somera and D. Mongru, *Science (80-.).*, 2002, **297**, 1551–1555.

44 J. Dong and S. Horvath, *BMC Syst. Biol.*, 2007, **1**, 24.

45 L. C. Freeman, *Soc. Networks*, 1978, **1**, 215–239.

46 E. Estrada, *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.*, 2010, **82**, 1–8.

47 S. Maere, K. Heymans and M. Kuiper, *Bioinformatics*, 2005, **21**, 3448–9.

48 H. Dweep, C. Sticht, P. Pandey and N. Gretz, *J. Biomed. Inform.*, 2011, **44**, 839–47.

49 K. Peng, S. Vucetic, P. Radivojac, C.J. Brown, A.K. Dunker and Z. Obradovic, *Journal of Bioinformatics and Computational Biology*, 2005, **3 (1)**, 35-60.

50 G. Stelzer, I. Dalah, T. I. Stein, Y. Satanower, N. Rosen, N. Nativ, D. Oz-Levi, T. Olender, F. Belinky, I. Bahir, H. Krug, P. Perco, B. Mayer, E. Kolker, M. Safran and D. Lancet, *Hum. Genomics*, 2011, **5**, 709–717.

51 M.-S. Kim, S. M. Pinto, D. Getnet, R. S. Nirujogi, S. S. Manda, R. Chaerkady, A. K. Madugundu, D. S. Kelkar, R. Isserlin, S. Jain, J. K. Thomas, B. Muthusamy, P. Leal-Rojas, P. Kumar, N. a Sahasrabuddhe, L. Balakrishnan, J. Advani, B. George, S. Renuse, L. D. N. Selvan, A. H. Patil, V. Nanjappa, A. Radhakrishnan, S. Prasad, T. Subbannayya, R. Raju, M. Kumar, S. K. Sreenivasamurthy, A. Marimuthu, G. J. Sathe, S. Chavan, K. K. Datta, Y. Subbannayya, A. Sahu, S. D. Yelamanchi, S. Jayaram, P. Rajagopalan, J. Sharma, K. R. Murthy, N. Syed, R. Goel, A. a Khan, S. Ahmad, G. Dey, K. Mudgal, A. Chatterjee, T.-C. Huang, J. Zhong, X. Wu, P. G. Shaw, D. Freed, M. S. Zahari, K. K. Mukherjee, S. Shankar, A. Mahadevan, H. Lam, C. J. Mitchell, S. K. Shankar, P. Satishchandra, J. T. Schroeder, R. Sirdeshmukh, A. Maitra, S. D. Leach, C. G. Drake, M. K. Halushka, T. S. K. Prasad, R. H. Hruban, C. L. Kerr, G. D. Bader, C. a Iacobuzio-Donahue, H. Gowda and A. Pandey, *Nature*, 2014, **509**, 575–81.

52 D. Guan, N. Altan-Bonnet, A. M. Parrott, C. J. Arrigo, Q. Li, M. Khaleduzzaman, H. Li, C.-G. Lee, T. Pe'ery and M. B. Mathews, *Mol. Cell. Biol.*, 2008, **28**, 4629–4641.

53 H.-W. Wang, T.-H. Hsieh, S.-Y. Huang, G.-Y. Chau, C.-Y. Tung, C.-W. Su and J.-C. Wu, *BMC Genomics*, 2013, **14**, 736.

54 M. Rink, K. Park, B. G. Volkmer, E. Xylinas, J. Hansen, E. K. Cha, B. D. Robinson, R. Hautmann, R. Küfer, O. Engel, F. K. Chun, R. Dahlem, M. a. Rubin, S. F. Shariat and J. M. Mosquera, *Urol. Oncol. Semin. Orig. Investig.*, 2013, **31**, 1716–1724.

55 O. Itkonen and U. H. Stenman, *Clin. Chim. Acta*, 2014, **431**, 260–269.

56 N. Nicolaiew, V. Basset, F. Semprez, F. Vacherot, G. Ploussard, P. Soyeux, P. Maill, A. De Taille and Y. Allory, 2015, 1–9.

57 C. F. de la Cruz-Herrera, M. Campagna, V. Lang, J. Del Carmen González-Santamaría, L. Marcos-Villar, M. S. Rodríguez, a Vidal, M. Collado and C. Rivas, *Oncogene*, 2014, 1–9.

58 C. Li, X. Wu, H. Zhang, G. Yang, M. Hao, S. Sheng, Y. Sun, J. Long, C. Hu, X. Sun, L. Li and J. Zheng, *elsevier*, 2015, 1–6.

59 I. Stoimenov and T. Helleday, *Biochem. Soc. Trans.*, 2009, **37**, 605–613.

60 S. Roa, E. Avdievich, J. U. Peled, T. Maccarthy, U. Werling, F. L. Kuang, R. Kan, C. Zhao, A. Bergman, P. E. Cohen, W. Edelmann and M. D. Scharff, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 16248–16253.

61 Y.-M. Jeng, *Clin. Cancer Res.*, 2004, **10**, 2065–2071.

62 Z.-J. Zhou, Z. Dai, S.-L. Zhou, X.-T. Fu, Y.-M. Zhao, Y.-H. Shi, J. Zhou and J. Fan, *Int. J. Cancer*, 2013, **132**, 1080–9.

63 R. Yuan, K. Wang, J. Hu, C. Yan, M. Li, X. Yu, X. Liu, J. Lei, W. Guo, L. Wu, K. Hong and J. Shao, *Cancer Res.*, 2014, **74**, 5287–300.

64 E. E. Bates, O. Ravel, M. C. Dieu, S. Ho, C. Guret, J. M. Bridon, S. Ait-Yahia, F. Brière, C. Caux, J. Banchereau and S. Lebecque, *Eur. J. Immunol.*, 1997, **27**, 2471–7.

65 T. Velusamy, M. J. Kiel, A. A. Sahasrabuddhe, D. Rolland, C. A. Dixon, N. G. Bailey, B. L. Betz, N. A. Brown, A. C. Hristov, R. A. Wilcox, R. N. Miranda, L. J. Medeiros, Y. K. Jeon, K. V Inamdar, M. S. Lim and K. S. J. Elenitoba-Johnson, *Blood*, 2014, **124**, 3768–71.

66 D. Cortez, G. Glick and S. J. Elledge, *Proc. Natl. Acad. Sci.*, 2004, **101**, 10078–10083.

67 J. E. Olson, X. Wang, V. S. Pankratz, Z. S. Fredericksen, C. M. Vachon, R. A. Vierkant, J. R. Cerhan and F. J. Couch, *Breast Cancer Res. Treat.*, 2011, **125**, 221–8.

68 D. a Starr, R. Saffery, Z. Li, a E. Simpson, K. H. Choo, T. J. Yen and M. L. Goldberg, *J. Cell Sci.*, 2000, **113 (Pt 1**, 1939–1950.

69 M. Chen, Z. Q. Pan and J. Hurwitz, *Proc. Natl. Acad. Sci. U. S. A.*, 1992, **89**, 5211–5215.

70 N. Bah, L. Maillet, J. Ryan, S. Dubreil, F. Gautier, A. Letai, P. Juin and S. Barillé-Nion, *Cell Death Dis.*, 2014, **5**, e1291.

71 S. M. Hamid, S. Cicek, S. Karamil, M. B. Ozturk, B. Debelec-Butuner, B. Erbaykent-Tepedelen, L. Varisli, C. Gonen-Korkmaz, K. Yorukoglu and K. S. Korkmaz, *Mol. Cell. Endocrinol.*, 2014, **383**, 38–47.

72 M. Evert, M. Frau, M. L. Tomasi, G. Latte, M. M. Simile, M. A. Seddaiu, A. Zimmermann, S. Ladu, T. Staniscia, S. Brozzetti, G. Solinas, F. Dombrowski, F. Feo, R. M. Pascale and D. F. Calvisi, *Br. J. Cancer*, 2013, **109**, 2654–64.

73 L. Cassimeris, B. Becker and B. Carney, *Cell Motil. Cytoskeleton*, 2009, **66**, 535–45.

74 G. Imreh, H. V. Norberg, S. Imreh and B. Zhivotovsky, *J. Cell Sci.*, 2011, **124**, 2951–63.

75 S. Manekeller, T. Sauerbruch, H.-P. Fischer, P. Propping and A. Hirner, *Z. Gastroenterol.*, 2010, **48**, 1211–4.

76 S. Ezzikouri, A. E. El Feydi, L. El Kihal, R. Afifi, M. Benazzouz, M. Hassar, A. Chafik, P. Pineau and S. Benjelloun, *Arch. Med. Res.*, 2008, **39**, 236–41.

77 S. Coulon, V. Legry, F. Heindryckx, C. Van Steenkiste, C. Casteleyn, K. Olievier, L. Libbrecht, P. Carmeliet, B. Jonckx, J.-M. Stassen, H. Van Vlierberghe, I. Leclercq, I. Colle and A. Geerts, *Hepatology*, 2013, **57**, 1793–805.

78 M. J. Alter, *World J. Gastroenterol.*, 2007, **13**, 2436–41.

79 P. Anand, A. B. Kunnumakkara, A. B. Kunnumakara, C. Sundaram, K. B. Harikumar, S. T. Tharakan, O. S. Lai, B. Sung and B. B. Aggarwal, *Pharm. Res.*, 2008, **25**, 2097–116.

80 K. S. Bishop and L. R. Ferguson, *Nutrients*, 2015, **7**, 922–47.

81 B. Jin, W. Wang, G. Du, G.-Z. Huang, L.-T. Han, Z.-Y. Tang, D.-G. Fan, J. Li and S.-Z. Zhang, *Eur. Rev. Med. Pharmacol. Sci.*, 2015, **19**, 592–601.

82 A. K. Dunker, M. S. Cortese, P. Romero, L. M. Iakoucheva and V. N. Uversky, *FEBS J.*, 2005, **272**, 5129–48.

83 M. Sayed, S. Pelech, C. Wong, A. Marotta and B. Salh, *Oncogene*, 2001, **20**, 6994–7005.

84 T. Reitsema, D. Klokov, J. P. Banáth and P. L. Olive, *DNA Repair (Amst).*, 2005, **4**, 1172–81.

85 J. Y. Yang, A. Dunker, J. S. Liu, X. Qin, H. R. Arabnia, W. Yang, A. Niemierko, Z. Chen, Z. Luo, L. Wang, Y. Liu, D. Xu, Y. Deng, W. Tong and M. Yang, *BMC Bioinformatics*, 2014, **15 Suppl 1**, I1.

86 M. Sickmeier, J. A. Hamilton, T. LeGall, V. Vacic, M. S. Cortese, A. Tantos, B. Szabo, P. Tompa, J. Chen, V. N. Uversky, Z. Obradovic and A. K. Dunker, *Nucleic Acids Res.*, 2007, **35**, D786–93.

87 T. Frege and V. N. Uversky, *Biochem. Biophys. Reports*, 2015, **1**, 33–51.

88 V. Vacic and L. M. Iakoucheva, *Mol. Biosyst.*, 2012, **8**, 27–32.

89 K. J. Niklas, S. E. Bondos, A. K. Dunker and S. A. Newman, *Front. cell Dev. Biol.*, 2015, **3**, 8.

90 S. Costantini, A. Sharma, R. Raucci, M. Costantini, I. Autiero and G. Colonna, *BMC Evol. Biol.*, 2013, **13**, 60.

91 K.-I. Goh and I.-G. Choi, *Brief. Funct. Genomics*, 2012, **11**, 533–42.

92 J.-D. J. Han, N. Bertin, T. Hao, D. S. Goldberg, G. F. Berriz, L. V Zhang, D. Dupuy, A. J. M. Walhout, M. E. Cusick, F. P. Roth and M. Vidal, *Nature*, 2004, **430**, 88–93.

93 V. N. Uversky, C. J. Oldfield and A. K. Dunker, *Annu. Rev. Biophys.*, 2008, **37**, 215–46.

94 J. Wang, X. Liu, H. Wu, P. Ni, Z. Gu, Y. Qiao, N. Chen, F. Sun and Q. Fan, *Nucleic Acids Res.*, 2010, **38**, 5366–83.

95 M. Cui, M. Zheng, B. Sun, Y. Wang, L. Ye and X. Zhang, *Neoplasia*, 2015, **17**, 79–88.

96 J. D. Oliner, K. W. Kinzler, P. S. Meltzer, D. L. George and B. Vogelstein, *Nature*, 1992, **358**, 80–3.

97 J. Gearhart, E. E. Pashos and M. K. Prasad, *N. Engl. J. Med.*, 2007, **357**, 1469–1472.

98 M. E. Crosby, *Yale J. Biol. Med.*, 2007, 80, 141.

99 N. K. Sah, Z. Khan, G. J. Khan and P. S. Bisen, *Cancer Lett.*, 2006, **244**, 164–71.

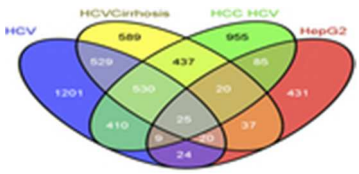
100 L. Kang, X. Liu, Z. Gong, H. Zheng, J. Wang, Y. Li, H. Yang, J. Hardwick, H. Dai, R. T. P. Poon, N. P. Lee, M. Mao, Z. Peng and R. Chen, *Genomics*, 2015, **105**, 76–82.

101 M. D. Schultz, Y. He, J. W. Whitaker, M. Hariharan, E. A. Mukamel, D. Leung, N. Rajagopal, J. R. Nery, M. A. Urich, H. Chen, S. Lin, Y. Lin, I. Jung, A. D. Schmitt, S. Selvaraj, B. Ren, T. J. Sejnowski, W. Wang and J. R. Ecker, *Nature*, 2015, **advance on**.

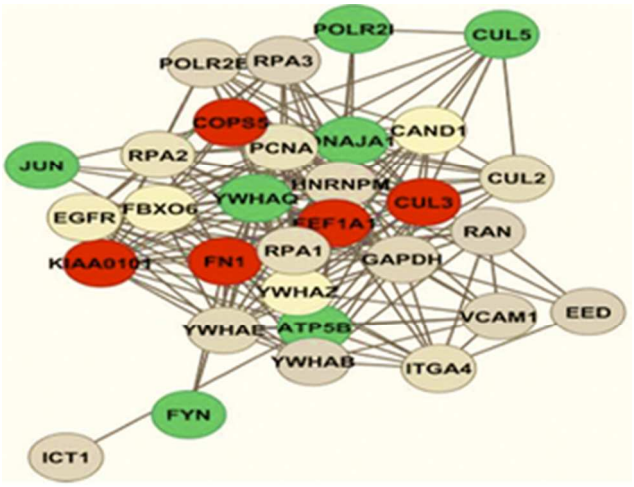
102 V. Hovestadt, D. T. W. Jones, S. Picelli, W. Wang, M. Kool, P. A. Northcott, M. Sultan, K. Stachurski, M. Ryzhova, H.-J. Warnatz, M. Ralser, S. Brun, J. Bunt, N. Jäger, K. Kleinheinz, S. Erkek, U. D. Weber, C. C. Bartholomae, C. von Kalle, C. Lawerenz, J. Eils, J. Koster, R. Versteeg, T. Milde, O. Witt, S. Schmidt, S. Wolf, T. Pietsch, S. Rutkowski, W. Scheurlen, M. D. Taylor, B. Brors, J. Felsberg, G. Reifenberger, A. Borkhardt, H. Lehrach, R. J. Wechsler-Reya, R. Eils, M.-L. Yaspo, P. Landgraf, A. Korshunov, M. Zapatka, B. Radlwimmer, S. M. Pfister and P. Lichter, *Nature*, 2014, **510**, 537–41.

103 W. G. Kaelin and S. L. McKnight, *Cell*, 2013, **153**, 56–69.

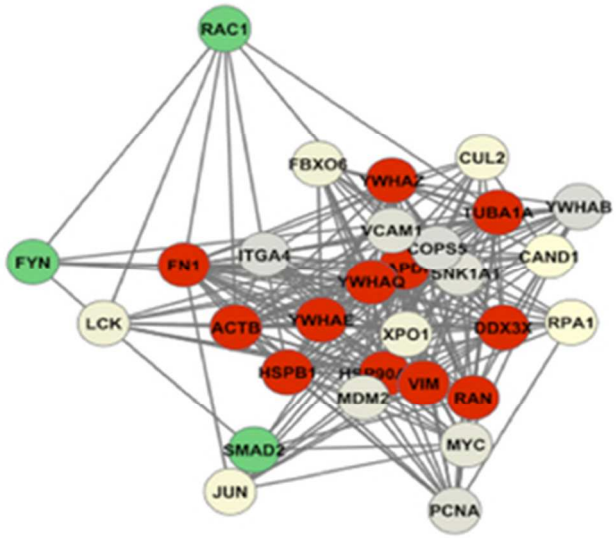
104 R. J. DeBerardinis and C. B. Thompson, *Cell*, 2012, **148**, 1132–44.



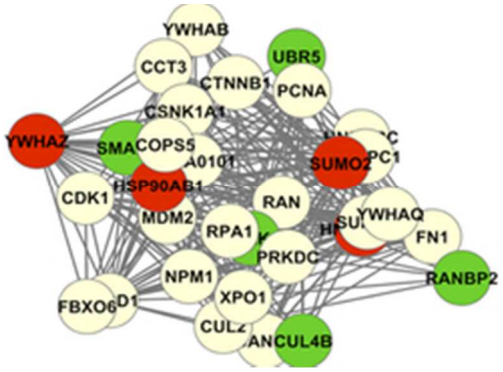
Venn diagram of all cases showing exclusive and recurrent genes in the different pathological stages: HCV (2748 genes) HCV Cirrhosis (2187 genes) HCC with HCV etiology (2471 genes) and HepG2 (651 genes). 14x6mm (300 x 300 DPI)



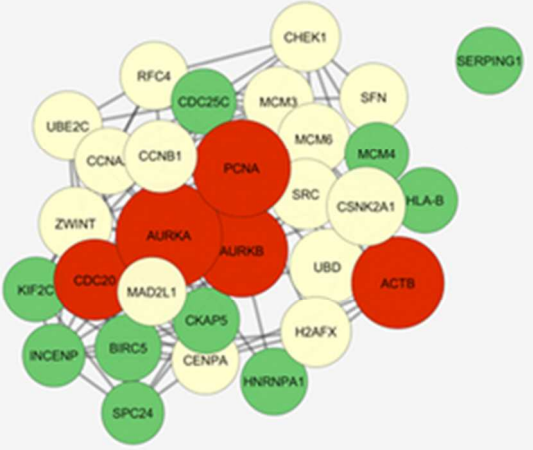
Sub-network specific for the HUB-HUB interactions in the HCV network. Nodes with similar colors are members of the same sub-network.
26x20mm (300 x 300 DPI)



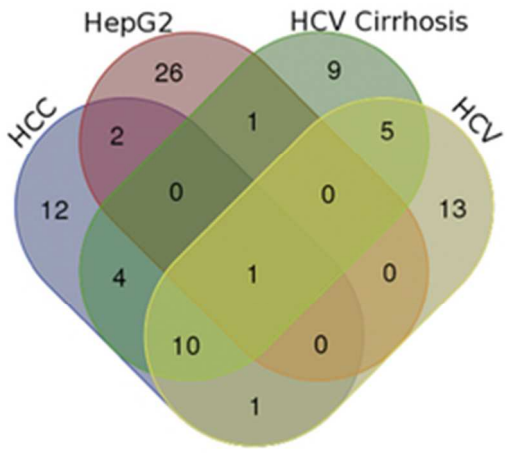
Sub-network specific for the HUB-HUB interactions in the HCV-related cirrhosis network. Nodes with similar colors are members of the same sub-network.
26x22mm (300 x 300 DPI)



Sub-network specific for the HUB-HUB interactions in the HCC with HCV-related cirrhosis network. Nodes with similar colors are members of the same sub-network.
20x15mm (300 x 300 DPI)



Sub-network specific for the HUB-HUB interactions in the HepG2 network. Nodes with similar colors are members of the same sub-network.
22x19mm (300 x 300 DPI)



Venn diagram of the interacting 84 HUB nodes in HCV, HCV-related cirrhosis, HCC with HCV-related cirrhosis, and HepG2.
22x19mm (300 x 300 DPI)