Table of content

- Supplementary methods
- Figure S1
- Figure S2
- Figure S3
- Table legend
Supplementary methods

Yeast two-hybrid bait basal transactivation test

Because bait constructs sometimes self-transactivate reporter genes, SD-L-H culture medium were supplemented with 3-aminotriazole (3-AT). Appropriate concentrations of this drug were determined by growing bait strains on SD-L-H medium supplemented with increasing concentrations of 3-AT. Self-transactivation by NS5A without its membrane anchor was too high to be titrated with 3-AT and was not further tested.

IST (Interactor Sequence Tag) Analysis

We have developed a bioinformatic pipeline that assigns each IST to its native human genome transcript. First, ISTs were filtered by using PHRED (Ewing and Green, 1998; Ewing et al., 1998) at a quality score superior to the conventional 20 threshold value (less than 1% sequence errors). Gal4 motif was searched (last 87 bases of GAL4-AD), sequences downstream of this motif were translated into peptides and aligned using BLASTP against the REFSEQ (http://www.ncbi.nlm.nih.gov/RefSeq/) human protein sequence database (release 04/2007). Low-confidence alignments (E value $>10^{-10}$, identity $< 80\%$) and premature STOP codon containing sequences were eliminated. Only in-frame proteins and high quality sequences were further considered.

Network visualization

Large Graph Layout (http://bioinformatics.icmb.utexas.edu/lgl/) was applied to visualize the H-H network in figure 2A. Guess tool (http://graphexploration.cond.org/) was used to graphically represent H-H$_{HCV}$ infection network in figure 2B and the IJT network in figure 4A. Figure 2b and 4a are available in a GUESS interactive format (GUESS Data Format) in

**Definition of graph theoretic concepts**

**Connected component:**

In an undirected network, a connected component is a maximal connected sub-network. Two nodes are in the same connected component if and only if there exist a path between them. We also included in the set of connected components those proteins that are not connected to any other protein (disconnected proteins), according to igraph R package.

Assume $G = (V,E)$, the graph representing the human interactome where $V$ is the set of nodes (proteins) and $E$ is the set of edges (interactions) between this nodes. We use $n$ and $m$ to denote the number of nodes and edges of $G$, respectively.

**Degree:**

The degree of a node ($k$) is the number of edges incident to the node. The degree distribution of human proteins was compared to the degree distribution of $H_{HCV}$.

**Shortest Path:**

The shortest path problem is the finding of a path between two nodes such that the sum of the weights of its constituent edges is minimized. The shortest paths (also called geodesics) are calculated here by using breath-first search in the graph. Edge weights are not used here, i.e every edge weight is one. The shortest path length ($l$) distribution of $H_{HCV}$ was compared to the shortest path length distribution of the human interactome.

**Betweenness:**

The node betweenness $b(v)$ can be defined by the number of shortest paths going through a node $v$ and was normalised by twice the total number of protein pairs in the network ($n*(n-1)$). The equation used to compute betweenness centrality, $b(v)$, for a node $v$ is:
Where $g_{ij}$ is the number of shortest paths going from node $i$ to $j$, $I$ and $j \in V$ and $g_{ij}(v)$ the number of shortest paths from $i$ to $j$ that pass through the node $v$. The betweenness distribution of $H_{HCV}$ was compared to the betweenness distribution of the human interactome.

**Generalised linear model**

The generalized linear model was constructed as follow:

$$Y = a0 + a1*k + a2*b$$

to test the effect of both predictor measures ($k$ and $b$, the independent variable) on the property of cellular protein to interact with viral proteins described by $Y$, the variable to predict.

In this logistic regression model:

- $Y$ is assumed to follows a binomial law. Indeed, as the response data $Y$ are binary, *i.e.* cellular proteins interact (equal to 1) or not (equal to 0) with viral proteins, it is generally assumed that $Y$ follows a binomial distribution of parameters $n$ and $p$, where the numbers of Bernoulli trials $n$ and the probabilities of success, *i.e.* that the cellular protein interact with viral proteins, $p$;

- The logistic regression is based on the fundamental hypothesis that $Y = a0 + a1*k + a2*b = \log(p/1-p)$, also called the logistic function;

- $k$ is the degree of the cellular proteins;

- $b$ is the betweenness of the cellular proteins.
**IJT network construction**

First, all proteins annotated in insulin (blue), Jak/STAT (red) and TGFβ (green) pathways were extracted from KEGG database. Second, a network was constructed by connecting all these proteins according to the H-H interactome dataset to create the backbone of the IJT network (colored network). Then, direct neighbors of each member of the three pathways were extracted from the H-H interactome. From this pool of proteins, only those interacting with viral proteins were retained and connected to the IJT network (black and grey nodes). The rationale for this selection is that targeted neighbors at the interface could be co-regulators or perturbators of these pathways.
Supplementary references


Figure S1

HCV 1b genome (strain Con1, AJ238799)

HCV 1b polyprotein

<table>
<thead>
<tr>
<th>ORF</th>
<th>Protein/domain</th>
<th>Genome coordinates</th>
<th>polyprotein coordinates</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Core full-length</td>
<td>342 914</td>
<td>1 191</td>
</tr>
<tr>
<td>2</td>
<td>Core mature</td>
<td>342 854</td>
<td>1 171</td>
</tr>
<tr>
<td>3</td>
<td>Core domain 1</td>
<td>342 695</td>
<td>1 118</td>
</tr>
<tr>
<td>4</td>
<td>Core domain 2</td>
<td>696 854</td>
<td>1 119 171</td>
</tr>
<tr>
<td>5</td>
<td>E1 full-length</td>
<td>915 1490</td>
<td>192 383</td>
</tr>
<tr>
<td>6</td>
<td>E1 ectodomain</td>
<td>915 1397</td>
<td>192 352</td>
</tr>
<tr>
<td>7</td>
<td>E2 full-length</td>
<td>1491 2579</td>
<td>384 746</td>
</tr>
<tr>
<td>8</td>
<td>E2 ectodomain</td>
<td>1491 2465</td>
<td>384 708</td>
</tr>
<tr>
<td>9</td>
<td>p7 full-length</td>
<td>2580 2768</td>
<td>747 809</td>
</tr>
<tr>
<td>10</td>
<td>p7 N-term helix</td>
<td>2580 2633</td>
<td>747 764</td>
</tr>
<tr>
<td>11</td>
<td>NS2 full-length</td>
<td>2769 3419</td>
<td>810 1026</td>
</tr>
<tr>
<td>12</td>
<td>NS2 ectodomain</td>
<td>3048 3419</td>
<td>903 1026</td>
</tr>
<tr>
<td>13</td>
<td>NS3 full-length</td>
<td>3420 5312</td>
<td>1027 1657</td>
</tr>
<tr>
<td>14</td>
<td>NS3 protease</td>
<td>3420 3959</td>
<td>1027 1206</td>
</tr>
<tr>
<td>15</td>
<td>NS3 helicase</td>
<td>3960 5312</td>
<td>1207 1657</td>
</tr>
<tr>
<td>16</td>
<td>NS4A full-length</td>
<td>5313 5474</td>
<td>1658 1711</td>
</tr>
<tr>
<td>17</td>
<td>NS4B full-length</td>
<td>5475 6257</td>
<td>1712 1972</td>
</tr>
<tr>
<td>18</td>
<td>NS5A full-length</td>
<td>6258 7598</td>
<td>1973 2419</td>
</tr>
<tr>
<td>19</td>
<td>NS5A cytosolic domain</td>
<td>6344 7598</td>
<td>2003 2419</td>
</tr>
<tr>
<td>20</td>
<td>NS5A membrane anchor</td>
<td>6258 6347</td>
<td>1973 2002</td>
</tr>
<tr>
<td>21</td>
<td>NS5A domain 1</td>
<td>6344 6896</td>
<td>2003 2185</td>
</tr>
<tr>
<td>22</td>
<td>NS5A domain 2</td>
<td>6897 7283</td>
<td>2186 2314</td>
</tr>
<tr>
<td>23</td>
<td>NS5A domain 3</td>
<td>7284 7598</td>
<td>2315 2419</td>
</tr>
<tr>
<td>24</td>
<td>NS5B full-length</td>
<td>7599 9374</td>
<td>2420 3010</td>
</tr>
<tr>
<td>25</td>
<td>NS5B cytosolic domain</td>
<td>7599 9308</td>
<td>2420 2989</td>
</tr>
<tr>
<td>26</td>
<td>NS4A-NS3 full-length</td>
<td>3420 5312</td>
<td>1027 1657</td>
</tr>
<tr>
<td>27</td>
<td>NS4A-NS3 protease</td>
<td>3420 3959</td>
<td>1027 1657</td>
</tr>
</tbody>
</table>

- Full length protein
- NS5A membrane anchor
- Domain
- Functional fusion
Figure S1: The HCV ORFeome.
Schematic representation of the HCV genome and definition of ORFs designed for Y2H screens. The HCV positive strand genome (purple) encodes a polyprotein (orange) which is co-translationally processed in 10 proteins. Red: full length protein; blue: domain; yellow + green: NS4A + NS3 chimeric fusion; pink: NS5A membrane anchor. Genome and polyprotein coordinates of each construction are given in the table.
Figure S2

Probability distribution (Log–Log Scale)
Figure S2: Degree and betweenness distributions of H, \( H_{HCV} \) and \( H_{EBV} \) proteins in H-H network.

log degree (left) and log betweenness (right) distributions of H proteins (blue), \( H_{HCV} \) (red) and \( H_{EBV} \) (green). Solid lines represent linear regression fits. Vertical dashed lines give mean degree and betweenness values.
Figure S3
Figure S3: Functional validation of focal adhesion perturbation by NS3 and NS5A

96-well plates were coated with fibronectin (left) or poly-L-lysine (right) at various concentrations. 293T cells expressing NS2, NS3, NS3/4A or NS5A were plated on the matrix for 30min (concentration range: 0,04 to 20μg/ml matrix). Adherent cells were stained with crystal violet. Values represent means of triplicates with standard deviation.
Supplementary Table legends

**Table SI: H_{HCV} listing.**

HCV proteins are referenced according to their NCBI mature peptide product name (column 1). Human proteins are referenced with their cognate NCBI gene name and gene ID (columns 2 and 3). The number of IST for IMAP Y2H (IMAP1 and IMAP2, according to the method of screening) is given in columns 4 and 5. IMAP LCI (Literature Curated Interactions) from text-mining and BIND database associated PubMed IDs are given in columns 6 and 7. Co-affinity purification (CoAP) or Y2H pairwise matrices validations are indicated in columns 8 and 9 (+: IMAP validation, -: not validated, NA: non assayable due to default of protein expression or to cellular protein directly interacting with GST). Links for PubMed IDs and Gene IDs are given at [http://www.ncbi.nlm.nih.gov/sites/entrez](http://www.ncbi.nlm.nih.gov/sites/entrez) (choose PubMed or Gene respectively). Gene expression in the liver, according to EST data (http://www.ncbi.nlm.nih.gov/sites/entrez?db=unigene) is given in column 10.

**Table SII: characteristics of IMAP1 and IMAP2 screens**

Numbers of transformants and yeast colonies are given for IMAP1 screens for each bait.

Number of diploids and yeast colonies are given for IMAP2 screens for each bait.

**Table SIII: Listing of human proteins interacting with more than one viral protein.**

Human proteins are referenced with their cognate NCBI gene name (column 1). HCV proteins are referenced according to their NCBI mature peptide product name (column 2). Origin of the dataset (IMAP Y2H, IMAP LCI, column 3).

**Table SIV: List of human protein-protein interactions**
Interacting human proteins are referenced with their cognate NCBI gene name (columns 1 and 2). These physical and direct binary protein-protein interactions were retrieved from BIND, BioGRID, DIP, GeneRIF, HPRD, IntAct, MINT, and Reactome databases.

Table SV: Topological analysis of the HCV-human network

Connected components of the HCV-Human network.

The size of the largest component and the number of connected components of V-H_{HCV} sub-network were computed (IMAP dataset, column 2). In order to test the significance of observed values, we computed the mean of the largest component size and the mean number of connected components obtained after 1000 simulations of random sub-networks (IMAP Sim column 3). The differences between the observed and the simulated mean values were highly significant (two sided z-test ***: p-value <10^{-10}).

Topological properties of the H_{HCV} and H_{EBV} proteins in the human interactome.

Full interactome (A), high-confidence interactome (B, containing only PPIs with at least two PMIDs or validated by two different methods). The number of proteins and PPIs that can be integrated into the human interactome are given for H_{HCV} and H_{EBV}. Percentage of H_{HCV} and H_{EBV} that are present in the human interactome are given according to the origin of the dataset. Average degree \(k\), betweenness \(b\) and shortest path \(l\) were computed for H_{HCV} and H_{EBV} in both full and high-confidence interactomes (Calderwood et al., 2007).

Table SVI: HCV Protein distribution and enrichment in IJT network.

A. H_{HCV} enrichment in IJT network for each viral protein. Number of H_{HCV} is given in V-H_{HCV} and IJT networks. Enrichment of H_{HCV} in IJT network was tested with exact Fisher test for each viral protein. Associated odd ratios and p-values are given.
B. \(H_{HCV}\) enrichment in Jak/STAT, TGF\(\beta\) and Insulin pathways for each viral protein. Number of \(H_{HCV}\) is given in V-H\(H_{HCV}\) network and Jak/STAT, TGF\(\beta\) and Insulin pathways (as defined in KEGG database). Enrichment of \(H_{HCV}\) in Jak/STAT, TGF\(\beta\) and Insulin pathways was tested with exact Fisher test for each viral protein. Associated odd ratios and p-values are given.

C. \(H_{HCV}\) enrichment in Jak/STAT, TGF\(\beta\) and Insulin inter-pathways for each viral protein. Inter-pathways are defined as the \(H_{HCV}\) connecting two or three KEGG pathways. Number of \(H_{HCV}\) is given in V-H\(H_{HCV}\) network and Jak/STAT, TGF\(\beta\) and Insulin inter-pathways. Enrichment of \(H_{HCV}\) in Jak/STAT, TGF\(\beta\) and Insulin inter-pathways was tested with exact Fisher test for each viral protein. Associated odd ratios and p-values are given.