Exploring the Protein Interactome Related to Hepatitis C Virus

Shifa Tariq Ashraf, Syeda Marriam Bakhtiar*

Department of Bioinformatics,Mohammad Ali Jinnah University,
Expressway, Kahuta Road, Zone-V,Islamabad, Pakistan

***Corresponding Author:** marriam@jinnah.edu.pk

ABSTRACT

Hepatitis C virus (HCV) stands as a health problem experienced across the globe leading to chronic or acute liver diseases such as cirrhosis, hepatocellular carcinoma and various others. It is a complex disease with extensive genetic heterogeneity with little known about the interactions of complex intra- and intercellular processes. The evolving tools in the application of network science to identify diseases have paved a way for the study of complex diseases at system level. This study focuses on identifying the significant proteins and the biological regulatory pathways involved in Hepatitis C virus and performing topological analysis of the PPIs derived by the proteins encoded by the susceptible genes in order to look for the molecular connectivity between these pathways.

Key words: Protein-protein interaction, network, regulatory pathways, analysis, construction of PPI and Hepatitis C Virus

1. INTRODUCTION

Hepatitis Inflammation of the liver is an infection which causes chronic liver diseases such as cirrhosis and is widely linked to hepatocellular carcinoma. The infection may be found in an acute or chronic form. Successful treatments are available for Hepatitis C through antiviral drugs but the chronicity of the disease may lead to a prolonged liver damage. Hepatitis can be classified in to five main types namely, Hep C (HCV), Hep A (HAV), Hep B (HBV), Hep E (HEV) and Hep D (HDV). All of these types are responsible for liver damage with few discrete differences (Sun et al., 2015).

Hepatitis C virus (HCV) belongs to the Flaviviridae viral family with a positive, single-strand RNA molecule as a genome with extensive genetic heterogeneity. It has an open reading frame responsible for encoding a large polyprotein of about 3000 amino acids . It transmits through infected blood, and becomes the most serious type of hepatitis amongst the others, consequently becoming the global health problem infecting approximately 200 million people around the globe (Isken et al., 2007).

Currently no vaccination is available for HCV and therapies have been failed to improve the health of patients suffering with disease (Pawlotsky, 2006).

According to Centre for disease control (CDC), in 2009 there were about 16,000 reported cases of acute HCV out of which about 60 to 70 percent are likely to develop chronic liver disease . Chronic infection leads to liver cirrhosis in approximately 30% of infected individuals (McDermott et al., 2012).

Protein–protein interactions (PPIs) establish physically between more than two proteins as a consequence of biochemical events and/or electrostatic forces. At cellular and systemic levels proteins are macromolecules of dynamic nature, but they seldom act alone. PPIs are an essential part of biological processes. Therefore, the construction of PPI networks provides a foundation for understanding protein function. The unidentified protein function can be predicted on the basis of their PPIs, whose function is already revealed (Rao et al., 2014).

In the recent years, high-throughput methods have been advanced in order to measure the transcript or protein levels, globally at system level. These methods are used to identify genes or proteins that are likely to be involved in a diseased process -in this case HCV, in order to direct further experimental investigation. Since the disease process consists of proteins of fundamental importance, the PPIs encoded by the susceptible genes were considered to be significant in the pathogenicity of HCV. Furthermore, recent advances in topological analysis have been applied to interactome networks comprising PPI networks, where physical interactions depicts nodes . This study is intended to recognize the significant proteins and the biological regulatory pathways involved in HCV pathogenicity and perform topological analysis of the PPIs derived by the proteins encoded by the susceptible genes in order to look for the interactions between these pathways at molecular level.

2. METHODS

The presented study comprised six steps which are as follows:

2.1. Manual Screening of Proteins Associated with HCV from the Literature

The first step included manual screening of the proteins associated with HCV by using the PolySearch online text mining system, producing genes/proteins related to "Hepatitis C virus" by analyzing multiple sources of information which includes PubMed, OMIM, Swiss-Prot and DrugBank. Other

than the gene/proteins, it covers concepts like diseases, drugs, pathways, Single Nucleotide Polymorphisms, metabolites and tissues . "Disease-Gene/protein association" query type and the query keyword "Hepatitis C Virus" were used. Initially the system returned 499 literatures which were further manually screened. Finally, 294 candidate genes were obtained by checking their relevance to HCV from the published resources and eliminating the genes that weren't related to HCV (Supplementary Information).

2.2. Scanning for Protein-protein Interactions

The candidate genes obtained in step one was used to scan PPIs in step two. In order to scan the PPIs, they were extracted from STRING database (http://string-db.org/), which is a source for discovering protein-protein interactions.

2.3. Construction of the PPIs Network

An extended network was constructed consisting obtained proteins and their protein-protein interaction (PPI) neighbors along with the interactions between these proteins. In order to construct this network a highly resourceful tool –Pajek, was used . One giant network and two separate small networks –consisting of four proteins were obtained from the extended network. Through this information, the study directed the discovery of HCV at system level, and due to the high number of nodes in the giant network large betweenness centrality (BC) values were expected to be in that parameter. Considering this fact, only the giant network was thoroughly studied.

2.4. Topology of PPI Network

The analysis of PPI network topology involves major properties of nodes which form the basis of the network analysis or process. Hence, to evaluate the nodes in a network, connectivity degree (k), BC and closeness centrality (CC) were implemented. Degree and BC are two basic parameters in the network theory , . Degree (k) is the interactions between a protein and its neighbors and is the fundamental property of a node. Closeness centrality (CC) points out the center of the network, as it is the inverse of the mean length of the shortest paths directing to and from all the other nodes in the graph. Betweeness centrality (BC) is the measure of nodes that occur on the shortest paths amongst other nodes i.e. number of shortest paths that pass through each node. BC is a highly useful property as it indicated the detection of bottleneck proteins in a network. Also, it greatly influences over what flows in the network. Furthermore, average degree, mean shortest path length and diameter are some of the measures of topology used globally to character network . Average degree (<k>) is the average of all degree values of nodes present in a network. Mean shortest path length (mspl) connects each pair of nodes via shortest path by calculating average of the steps. And, the longest amid all shortest paths is diameter (D). All these characteristics were used to characterize the network using Pajek.

2.5. Creation of Backbone Network

A backbone network is made up of high BC proteins and the links between them. In order to create a backbone, 10% of the total node set was set as high BC to get maximum number of nodes as the backbone in order to study the genes associated with HCV i.e. 146 in the network. Hence, the first 15 BC nodes and their connections were extracted from the giant network to extract a backbone network. In network, BC was initially used to calculate the centrality of the nodes. The shortest paths function as bottleneck to control the communication between other nodes in the network as they go through the nodes with high BC.

2.6. Construction of Betweenness Centrality (BC) Values

A subnetwork of the genes associated with HCV, either connected indirectly or directly with the shortest path between the genes, was constructed. This was done by measuring the shortest paths using Pajek, resulting in a subnetwork including nodes present in these paths.

3. RESULTS

3.1. Protein-protein Interaction Network

The extended network comprises one large network along with two separated small networks which resulted from four seed proteins, POLR2K, ITPA, IL28B and GPT2 (Figure 1). The large network comprised 146 nodes. The backbone network consisted of 25 nodes. The largest degree in the giant network is 44, whereas the average degree is 9.630.

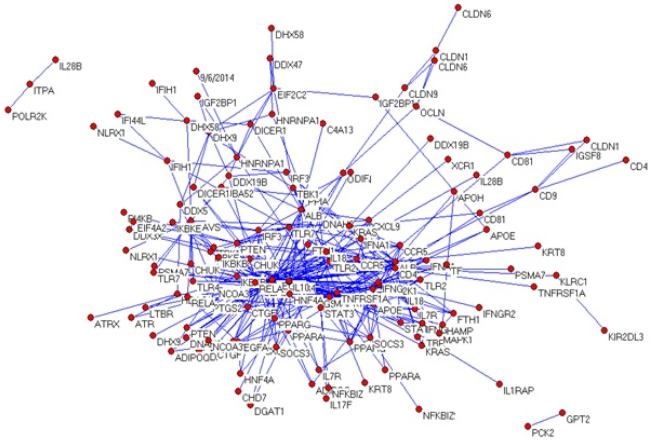

Figure 1. Extended network. The network comprises one large network along with two small networks resulting from four seed proteins, POLR2K, ITPA, IL28B and GPT2.

3.2. Key Nodes Identification

In this study, the key nodes were classified as the ones with high BC value or large degree, and 10% of the 146 nodes were used as the integral point of high BC and large degree nodes. Of 146 nodes, 15 have highest betweenness centrality (Table 1), 15 have high degree (Table 2) and 15 nodes were selected with large degree and high BC values (Table 3). Also, 4 proteins (CD4, VEGFA, MAVS, IFNG) were selected amongst the 15 nodes with BC value only (also mentioned in Table 3). So, to distinguish the different nodes and their roles in the network, different colors were assigned to each category (Figure 2). Pajek returned values calculated from the network for the degree, BC and CC value. MAPK1 (Mitogen-activated protein kinase 1) is a hub (center) protein with the largest degree, while CD4 is a bottleneck protein with the highest BC. Whereas, DIF is the central protein in the network

with the most interactions since it has the highest CC.

Table 1. The high betweeness centrality nodes along with their CC values

S. No.	Symbol	BC	CC value
1.	CD4	0.146060821	0.429224
2.	MAPK1	0.115570473	0.403598
3.	MAPK1	0.115570473	0.403598
4.	RELA	0.085469027	0.407245
5.	ALB	0.08177989	0.390767
6.	STAT3	0.077300083	0.396497
7.	ALB	0.070137334	0.385201
8.	PPARG	0.061925174	0.378726
9.	VEGFA	0.060126883	0.414741
10.	UBA52	0.056967454	0.414741
11.	TLR4	0.053361921	0.407245
12.	MAVS	0.050364462	0.325013
13.	IFNG	0.049459971	0.3919
14.	EIF2C2	0.046524672	0.281092
15.	TLR2	0.044640386	0.394185

Table 2. The large degree nodes along with their CC values

S. No.	Symbol	Degree	CC value
1.	MAPK1	44	0.403598
2.	RELA	43	0.407245
3.	DIF	38	0.441847
4.	CD4	38	0.429223
5.	STAT3	34	0.396497
6.	UBA52	33	0.414740
7.	TLR4	32	0.407245
8.	TNFRSF1A	32	0.385200
9.	MAVS	31	0.325013
10.	IFNG	30	0.391899
11.	PPARG	29	0.378726
12.	TLR2	28	0.394185
13.	TBK1	27	0.347571
14.	VEGFA	26	0.414740
15.	RXRA	26	0.318880

Table 3. List of high BC and large degree node along with their functions

Symbol	Function Description
MAPK1	Acts as a point that integrates various biochemical signals
RELA	REL-associated protein involved in the formation of NF-C heterodimer and the activation and translocation in the nucleus.
DIF	A dorsal-related gene found in Drosophilia, responsible for mediating an immune response.
CD4	Glycoprotein found in the immune system on the surface of T helper cells, monocytes, macrophages, and dendritic cells
STAT3	It directs the production of proteins which are involved in the pathways responsible for chemical signaling within the cells.
UBA52	A very conserved protein found in the nucleus and cytoplasm, that targets the cellular proteins for degradation
TLR4	Involved in recognizing pathogens and activating innate immunity
MAVS	Encodes an intermediary protein essential in the signaling pathways of beta interferon, triggered by virus
IFNG	Soluble cytokine with eminent properties like antiviral, immunoregulatory and anti-tumor and plays a role in activating macrophages
PPARG	Belongs to the nuclear receptor subfamily i.e. peroxisome proliferator-activated receptor (PPAR). PPARs play a vital role in formation of heterodimers with retinoid X receptors (RXRs) and regulate transcription of numerous genes
TLR2	Involved in recognizing pathogens and activating innate immunity
VEGFA	This protein is a glycosylated mitogen that specifically acts on endothelial cells and has effects like: mediation of increased vascular permeability, angiogenesis induction, cell growth of vasculogenesis and endothelial, cell migration promotion, and apoptosis inhibition

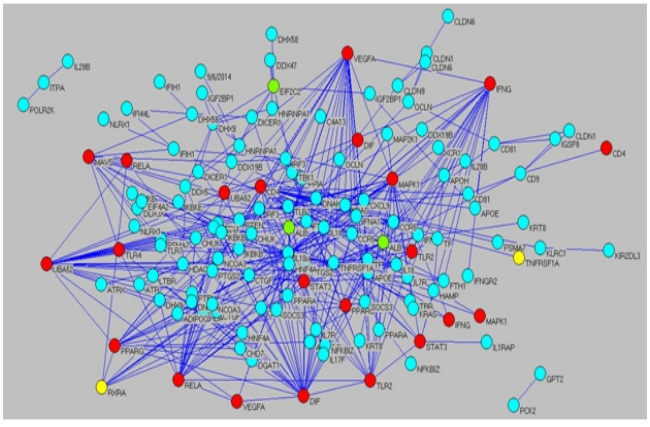

Figure 2. Giant network topology.

3.3. Cross-talk Between the Signaling Pathway in the High BC Network and Backbone Network Derived from them

A backbone network was constructed with high BC nodes. There 15 high BC nodes present in the backbone network

which corresponded to their sizes. DIF is the center protein and has 10 neighbors (Figure 3): CD4, ALG, PPARG, VEGFA, RELA, TLR4, STAT3, UBA52, IFNG and MAPK1. These proteins are involved in Leishmaniasis, Pancreatic cancer, Dorso-ventral axis formation, processing and presentation of antigen, T cell receptor signaling pathway, autophagy regulation and many more.

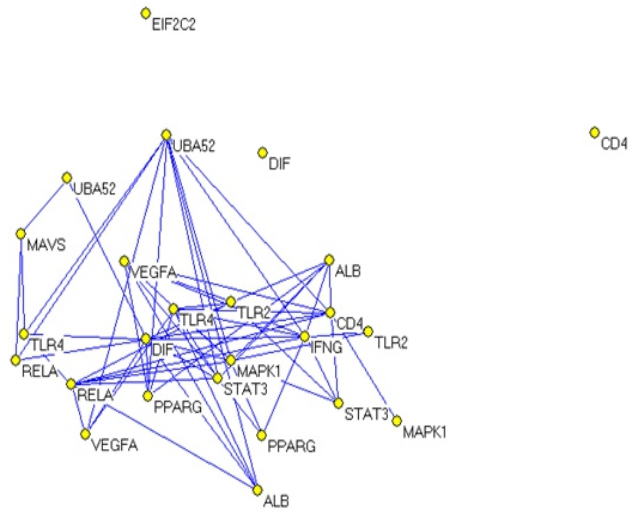

Figure 3. Backbone network topology.

3.4. Subnetwork Between the Candidate Genes Consisting of all Shortest Paths

The subnetwork consists of 146 nodes comprising 15 proteins that are not high BC nodes, 15 large BC nodes, and 116 seed proteins (Figure 4). It can be discovered that CD4 has the largest BC value and that the top 15 BC nodes in this network correspond with the backbone network nodes. There are only 15 proteins not in the list of 5 nodes with large BC value in the giant network. They are IL28B, C4A13, CHD7, PCK2 and POLR2K (Table 4).

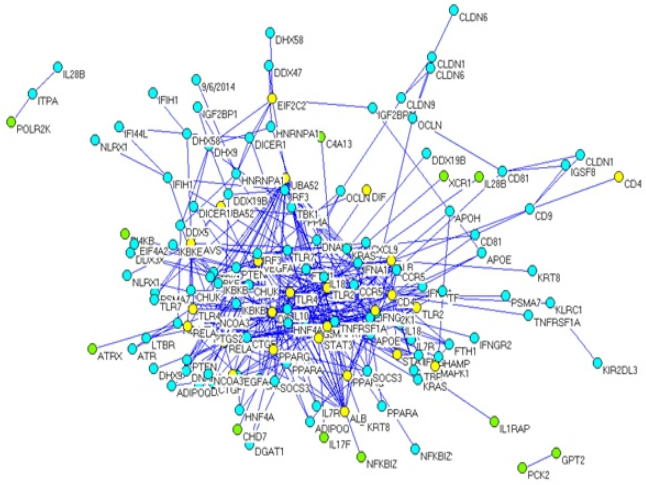

Figure 4. The subnetwork. It includes the paths shortest amongst the genes relevant to HCV. All shortest paths connect the candidate genes (yellow color nodes). There are 15 nodes that are without large BC (green color)

Table 4. The list of top 15 BC nodes in the subnetwork

S. No.	Symbol	BC
1.	CD4	0.146060821
2.	MAPK1	0.115570473
3.	DIF	0.103334907
4.	RELA	0.085469027
5.	ALB	0.08177989
6.	STAT3	0.077300083
7.	ALB	0.070137334
8.	PPARG	0.061925174
9.	VEGFA	0.060126883
10.	UBA52	0.056967454
11.	TLR4	0.053361921
12.	MAVS	0.050364462
13.	IFNG	0.049459971
14.	EIF2C2	0.046524672
15.	TLR2	0.044640386

4. DISCUSSION

Although a number of studies report the numerous genes relevant to HCV, it's pathogenesis still has room for further study. This study focuses on the analysis of how the proteins found contribute to the pathogenesis of HCV and determines other key proteins through topological analysis. The evaluation of proteins in PPIs related to the disease is done by the two fundamental properties of network theory i.e. degree and betweenness.

146 genes have been searched in this study as genes susceptible to HCV. The network constructed from the converted seed proteins, consists of a large giant network and two small networks (Figure 1). Five seed proteins (POLR2K, ITPA, IL28B, PCK2, and GPT2) included in the two separated networks indicate HCV variation between these proteins. As proposed by Ran et al. , at times some genes are missed from the literature search and new genes susceptible to the disease may be discovered. Hence, false interactions may lead to false recognition of nodes. MAPK1 with the highest degree is on 2nd rank in large BC proteins list. On the other hand, CD4 with largest BC is ranked 4th in the high degree proteins list.

The bottle neck proteins included UBA52, a very conserved protein found in the nucleus and cytoplasm, that targets the cellular proteins for degradation MAVS gene encodes a fundamental protein essential for the virus-triggered beta interferon signaling pathways . Protein encoded by TLR4 gene is a member of the Toll-like receptor (TLR) family including TLR2 gene is involved in recognizing pathogens and activating innate immunity . IFNG gene belongs to type II interferon family. It is a soluble cytokine with eminent properties like antiviral, immunoregulatory and anti-tumor and plays a role in activating macrophages. PPARG gene belongs to the nuclear receptor subfamily i.e. peroxisome

proliferator-activated receptor (PPAR). PPARs play a vital role in formation of heterodimers with retinoid X receptors (RXRs) and regulate transcription of numerous genes. VEGFA gene found as a disulfide linked homodimer, is a member of the growth factor family PDGF/VEGF. This protein is a glycosylated mitogen that specifically acts on endothelial cells and has effects like: mediation of increased vascular permeability, angiogenesis induction, cell growth of vasculogenesis and endothelial, cell migration promotion, and apoptosis inhibition . DIF (Dosral-related Immunity Factor) is a well-studied gene in Dorsophila Melanogaster. DIF gene is normally localized in cytoplasm of larva responsible for mediating an immune response. . DIF gene is involved in the differential activation of NF-kappaB. In nucleus DIF binds to sequence motifs like kappa B found in promoter regions of genes responsible for immunity . This protein is the center protein because of its highest CC value.

4.1. Differential Activation of NF-kappaB by DIF

HCV infection initiates innate antiviral responses comprising the production of IL-28A, IL-28B, and IL-29, known as type III interferon. But, the molecular mechanisms involved in expression regulation of IFN genes in hepatocytes infected with HCV remains unclear. The binding of specific transcription factors to promoter regions was further determined by the regulatory elements inducing the IFN genes resulting in hepatocytes infected with HCV. Interferon regulatory factor (IRF) -3 and -7 are the transcriptional factors required for the stimulation of interferons IL-28A and IL-28B genes, while NF-B required for the stimulation of the IL-29 gene. A decrease in viral replication was observed by the addition of IFN- to HCV-infected hepatocytes resulting in microRNA-122 (miR-122) reduction .

The family of NF-KappaB (Nuclear Factor-kappa B)/I-KappaB stimulates the expression of over a 100 proteins contributing in the host immune response. The target proteins comprise cytokine and chemokine receptors essential for immune recognition, antigen presentation proteins, and adhesion receptors. NF-KappaB has been labeled the central mediator of the immune response, because of this extensive role in immune action.

Multiple families of viruses, including HCV (Hepatitis C Virus) activate NF-KappaB. This activation may have several functions: to promote viral replication, prevent virus-induced apoptosis, and mediate the immune response.

5. CONCLUSION

Amongst the 146 candidate genes, most of them were associated with HCV and their protein-protein interaction neighbors connected to a large network. The backbone network gave an evident overview of all the important genes in HCV. The finding suggested the link between HCV and the PPI network which was centered at DIF. DIF gene is involved in the differential activation of NF-kappaB which in turn is relevant to HCV which activates this particular factor.

6. ACKNOWLEDGMENT

Administration of Mohammad Ali Jinnah University, Islamabad is acknowledged for providing the necessary facilities for the study. The authors have full access to all of the data in the study and take responsibility for the integrity of the data and accuracy of the analysis.

Conflict of Interest

The authors declare no conflict of interest related to this manuscript.

REFERENCES

Batagelj V, Mrvar A. 1998. "Batagelj V, Mrvar A: Pajek-Program for Large Network Analysis. Connections. 1998, 21:47–57." Connections. http://vlado.fmf.uni-lj.si/pub/networks/doc/pajek.pdf.

Brubaker, Sky W, Anna E Gauthier, Eric W Mills, Nicholas T Ingolia, and Jonathan C Kagan. 2014. "A Bicistronic MAVS Transcript Highlights a Class of Truncated Variants in Antiviral Immunity." Cell 156 (4): 800–811. doi:10.1016/j.cell.2014.01.021.

Cheng, Dean, Craig Knox, Nelson Young, Paul Stothard, Sambasivarao Damaraju, and David S Wishart. 2008. "PolySearch: A Web-Based Text Mining System for Extracting Relationships between Human Diseases, Genes, Mutations, Drugs and Metabolites." Nucleic Acids Research 36 (Web Server issue): W399–405. doi:10.1093/nar/gkn296.

Goñi, Joaquín, Francisco J Esteban, Nieves Vélez de Mendizábal, Jorge Sepulcre, Sergio Ardanza-Trevijano, Ion Agirrezabal, and Pablo Villoslada. 2008. "A Computational Analysis of Protein-Protein Interaction Networks in Neurodegenerative Diseases." BMC Systems Biology 2 (January): 52. doi:10.1186/1752-0509-2-52.

Hedengren-Olcott, M., M. C. Olcott, D. T. Mooney, S. Ekengren, B. L. Geller, and B. J. Taylor. 2004. "Differential Activation of the NF- B-like Factors Relish and Dif in Drosophila Melanogaster by Fungi and Gram-Positive Bacteria." Journal of Biological Chemistry 279 (20): 21121–27. doi:10.1074/jbc.M313856200.

Hwang, Sohyun, Seung-Woo Son, Sang Cheol Kim, Young Joo Kim, Hawoong Jeong, and Doheon Lee. 2008. "A Protein Interaction Network Associated with Asthma." Journal of Theoretical Biology 252 (4): 722–31. doi:10.1016/j.jtbi.2008.02.011.

Isken, Olaf, Martina Baroth, Claus W Grassmann, Susan Weinlich, Dirk H Ostareck, Antje Ostareck-Lederer, and Sven-Erik Behrens. 2007. "Nuclear Factors Are Involved in Hepatitis C Virus RNA Replication." RNA (New York, N.Y.) 13 (10): 1675–92. doi:10.1261/rna.594207.

Jovanović, Nataša, Snežana Žunić Božinovski, Slobodan Krstić, Radmila Obrenović, Jasna Trbojević-Stanković, and Biljana Stojimirović. 2014. "Vascular Endothelial Growth Factor in Peritoneal Dialysis Patients." Clinical Laboratory 60 (10): 1695–1701. http://www.ncbi.nlm.nih.gov/pubmed/25651716.

Lee, Hai-Chon, Sowmya Narayanan, Sung-Jae Park, Seung-Yong Seong, and Young S Hahn. 2014. "Transcriptional Regulation of IFN-Λ Genes in Hepatitis C Virus-Infected Hepatocytes via IRF-3·IRF-7·NF-κB Complex." The Journal of Biological Chemistry 289 (8): 5310–19. doi:10.1074/jbc.M113.536102.

Lyra, A.C., X. Fan, and A.M. Di Bisceglie. 2004. "Molecular Biology and Clinical Implication of Hepatitis C Virus." Brazilian Journal of Medical and Biological Research 37 (5). Brazilian Journal of Medical and Biological Research: 691–95. doi:10.1590/S0100-879X2004000500010.

Mansur, Ashham, Luisa von Gruben, Aron F Popov, Maximilian Steinau, Ingo Bergmann, Daniel Ross, Michael Ghadimi, Tim Beissbarth, Martin Bauer, and José Hinz. 2014. "The Regulatory Toll-like Receptor 4 Genetic Polymorphism rs11536889 Is Associated with Renal, Coagulation and Hepatic Organ Failure in Sepsis Patients." Journal of Translational Medicine 12 (January): 177. doi:10.1186/1479-5876-12-177.

McDermott, Jason E, Deborah L Diamond, Courtney Corley, Angela L Rasmussen, Michael G Katze, and Katrina M Waters. 2012. "Topological Analysis of Protein Co-Abundance Networks Identifies Novel Host Targets Important for HCV Infection and Pathogenesis." BMC Systems Biology 6 (1): 28. doi:10.1186/1752-0509-6-28.

Pawlotsky, Jean-Michel. 2006. "Therapy of Hepatitis C: From Empiricism to Eradication." Hepatology (Baltimore, Md.) 43 (2 Suppl 1): S207–20. doi:10.1002/hep.21064.

Raman, Karthik. 2010. "Construction and Analysis of Protein-Protein Interaction Networks." Automated Experimentation 2 (1): 2. doi:10.1186/1759-4499-2-2.

Ran, Jihua, Hui Li, Jianfeng Fu, Ling Liu, Yanchao Xing, Xiumei Li, Hongming Shen, et al. 2013. "Construction and Analysis of the Protein-Protein Interaction Network Related to Essential Hypertension." BMC Systems Biology 7 (1): 32. doi:10.1186/1752-0509-7-32.

Rao, V Srinivasa, K Srinivas, G N Sujini, and G N Sunand Kumar. 2014. "Protein-Protein Interaction Detection: Methods and Analysis." International Journal of Proteomics 2014: 147648. doi:10.1155/2014/147648.

Rual, Jean-François, Kavitha Venkatesan, Tong Hao, Tomoko Hirozane-Kishikawa, Amélie Dricot, Ning Li, Gabriel F Berriz, et al. 2005. "Towards a Proteome-Scale Map of the Human Protein-Protein Interaction Network." Nature 437 (7062): 1173–78. doi:10.1038/nature04209.

Sun, Xin, Carrie D Patnode, Clara Williams, Caitlyn A Senger, Tanya J Kapka, and Evelyn P Whitlock. 2015. "Introduction." Agency for Healthcare Research and Quality (US). Accessed March 27. http://www.ncbi.nlm.nih.gov/books/NBK116459/.