Dissertation
submitted to the
Combined Faculties for the Natural Sciences and for Mathematics
of the Ruperto-Carola University of Heidelberg, Germany
for the degree of Doctor of Natural Sciences

presented by
Sandeep Amberkar, M.Sc
born in **Mumbai, MH, India**
Oral Examination: **18.07.2014**

Integrative bioinformatics analyses of genome-wide RNAi screens

Referees: Prof. Dr. Roland Eils
Prof. Dr. Lars Kaderali

Abstract

In past few years, genome-wide RNAi screens have identified many novel genes involved in diseases for many viruses such as Human Immunodeficiency Virus-1 (HIV-1), Hepatitis C virus (HCV), West Nile Virus (WNV) and Influenza virus (IV)[1–8]. However, due to difference in experimental conditions, usage of different viral strains and inherent biological noise, these screens have shown low number of common or overlapping hits for a virus[9]. Moreover, this overlap gets poorer for similar studies on viruses of different families. Although these overlaps are significant, their lower size restricts a comprehensive insight from a comparative analysis. Thus, a direct comparison of gene hit-lists of RNAi screens may not always give meaningful results. To address this problem we propose an integrative bioinformatics pipeline that allows for network based meta-analysis of viral HT-RNAi screens. Initially, human protein interaction network (PIN) generated by collating data from various public repositories, is subjected to unsupervised clustering to determine functional modules. Those modules that are significantly enriched in host dependency factors (HDFs) and/or host restriction factors (HRFs) are then filtered based on network topology and semantic similarity measures. Modules passing all these criteria are then interpreted for their biological significance from enrichment analyses. With our approach we could predict Tankyrase-1 as a potential novel hit within the functional subnetworks, within the human PIN for Hepatitis C virus (HCV). and Human Immunodeficiency Virus-1 (HIV-1), based on HDFs and HRFs identified in the corresponding genome-wide RNAi screens of these viruses. Thus, our approach allows for a network based meta-analysis of genome-wide screens to develop plausible hypotheses for novel regulatory mechanisms in virus-host interactions based on RNAi screens.

Zusamenfassung

Durch genom-weite RNAi-Screenings konnten in den letzten Jahren viele Gene neu identifiziert werden, die an Infektionen beteiligt sind, die durch Viren wie das Humane Immundefizienz-Virus (HIV-I), das Hepatitis C Virus (HCV), das West-Nil-Virus (WNV) und das Influenza-Virus (IV) hervorgerufen werden [1–8]. Aufgrund unterschiedlicher experimenteller Bedingungen, der Verwendung verschiedener Erregerstämme und Rauschens in den Daten haben verschiedene Untersuchungen zu einem Virus jedoch nur wenige gemeinsame Gene hervorgebracht [9]. Die Anzahl gemeinsamer Treffer sinkt sogar noch weiter, wenn Studien zu Viren unterschiedlicher Gattungen betrachtet werden. Zwar sind die überlappungen signifikant, doch erlaubt die kleine Anzahl gemeinsamer Treffer keine tieferen Einblicke, und ein direkter Vergleich der detektieren Gene verschiedener RNAi-Screenings liefert womöglich keine belastbaren Ergebnisse. Zur Lösung dieses Problems stellen wir einen integrativen Bioinformatik-Ansatz zur netzwerkbasierten Meta-Analyse viraler HT-RNAi-Screenings vor. Zunächst werden öffentlich verfügbare Daten zusammengeführt und durch unüberwachte Clustering-Verfahren Protein-Interaktions netzwerke (PIN) generiert. Cluster, die signifikant erhöhte host dependeny factors(HDF) und/oder host restriction factors(HRF) aufweisen, werden anschließend auf Grundlage von Netzwerk-Topologie und semantischen ähnlichkeitsmaßen gefiltert. Nach dieser Filterung werden die verbleibenden Cluster auf ihre biologische Signifikanz untersucht. Mit diesem Verfahren konnten wir Tankyrase-1 als potentiellen neuen Treffer in funktionellen Unternetzwerk, innerhalb der humanen PIN fuer Hepatitis C und HIV, basierend auf HDFs and HRFs in genom-weiten RNAi-Screenings dieser Viren detektiert werden. Unser Verfahren erlaubt daher mittels netzwerkbasierter Meta-Analyse von genom-weiten Screenings die Entwicklung neuer Hypothesen zu regulativen Mechanismen von Virus-Wirt Interaktionen.

Acknowledgements

Although my name may appear on first but there have been substantial non-textual influences and contributions from a lot of people, that made this thesis possible. First and foremost, I would like to thank the HGS Mathcomp for providing me a doctoral fellowship to craft my research work. Dr. Michael Winckler and the office staff at the HGS Mathcomp office have been very kind and helpful to us first-batch students. Secondly, this work couldn't have been possible had it not been for the ambitious and multi-perspective outlook of Lars (Prof. Lars Kaderali). He helped me realise my ideas into statistical experiments and provided constructive critic that sharpened the rough edges of my results as well as the text in this thesis. I would also like to thank my many collaborators, especially Prof. Ralf Bartenschlager, Prof. Hans Georg Kräuesslich and Prof. Gualtiero Alvisi. Working with Marion, Alessia and Kathleen was very exciting, especially when we hypothesised on virus-host interactions from our RNAi screen analyses. It was a mutual give-n-take that helped us learn a lot from each other.

¯I am especially grateful to my colleagues who with their feedback, inputs and coffee-chats made working enjoyable. I cherish the discussions and the time spent with Gajendra, Nurgazy and Narsis who, in no time, erased the boundaries between colleagues and friends and shifted to the latter. I couldn't thank Prabhav and Bettina, whose R primers and debugs helped me reshape my working codes to a great extent. Johanna, Joanne, Diana and Narsis always gave me a physicist and mathematician's outlook on anything and in turn I used to give them one of a biologist. It was exhilarating to see inter-disciplinary science in practice. The umpteen number of Masters students who stayed for their short term theses added a sense of action with their novelty of topics and kept us on our toes to read.

¯Last and definitely never the least, I am indebted to my family, who, with their constant support, encouragement and faith in me helped me tread through some difficult times in my grad school. I owe more than a sincere thanks to my extended family, especially my hon. uncles, Mr. Vivek Chhatre and Mr. Mahesh Mohite, who looked after my ailing mother last year which tested my tenacity to its limit. I feel fortunate to have found an equally supportive and caring fiancée in Suchita, whose trust in me has made me continue my scientific pursuit even further. I won't forget my 2nd extended family here in Dresden who never made me realise the stress of the final days of my PhD; Avin, Debo, Sukhi, Mansi, Abhishek, Pranav, Pradnya and Atul. I also can't thank my Heidelberg friends enough, Maria, Pavan, Madhuri, Aprameyo, Gajendra, Anashua, Avik, Budhaditya and many more who joined in heartily to bitch about our grad-school life that made it spicier, livelier and helped each one of us tread through it, successfully.

¯In today's inter-disciplinary biology, it is often the case that the whole is greater than the sum of its parts. With scientific and inter-personal inputs like that of the above, this thesis is no exception to that.

Contents

Figure

List of Tables

Dedicated to my family, especially to my fiancee without whom I could have completed this thesis earlier...

Chapter 1

Introduction

In the past decade, the advent of high throughput technology and its application in biology has redefined our understanding of cellular and molecular biology. The completion of the human genome project in 2003[10, 11] opened the floodgates of data, after which the need to decipher biology computationally, became obligatory[12, 13]. To tackle the vast amount of sequence data, numerous algorithms were developed that made extraction of meaningful information possible[14].

The parallel discoveries in experimental molecular biology, particularly, that of RNA interference (RNAi) allowed researchers to probe molecular systems in an altogether different manner. Instead of determining functions of individual genes through cross-overs and biochemical techniques, this otherwise "*reverse*" approach to determine the cell's phenotype after "*shutting off,knocking down*" the expression of particular gene, gained momentum. Scientists could then envision this process to scale up in order to probe multiple genes in a single experiment in order to determine cellular fate, roughly mimicking a stress response or a disease state of a cell. This would prove to be a paradigm shift in the way we understood the cell itself; from complex cellular processes to molecular aspects of disease. This chapter introduces the concept of RNAi, its application to understand human diseases and experimental/bioinformatics approaches to tackle problems of this technology.

1.1 RNA Interference

RNA Interference can be defined as a biomolecular process triggered by double stranded RNA (dsRNA) molecules and controlled by the RNA-induced silencing complex (RISC) which reduces the RNA targets of dsRNA in a sequence specific manner. This phenomenon was first

observed in *C.elegans*[15, 16]. For this fundamental discovery, both, Craig Mello and Andrew Fire received the Nobel prize for Medicine in 2006.

The process is triggered by the introduction of dsRNA in the cell. The dsRNA is recognized by RNase III family member(e.g. Dicer in *Drosophila*) and then cleaved into siRNAs of 21-23 nucleotides[15, 16]. These siRNAs then form the RNA-induced silencing complex (RISC) which degrades any mRNA homologous with the incorporated siRNA from the previous step[17]. Specifically, the target mRNA is cleaved in the centre complimentary to the siRNA thus resulting into rapid degradation of the target mRNA and subsequently, reduced protein expression[18].

Over time RNAi has proven to be one of the powerful techniques to interrogate gene function in cellular systems, tissues and organisms such as *Caenorhabditis elegans,Drosophila melanogaster* and plant species like *Arabidopsis thaliana*. This process was up-scaled for high-throughput analyses, wherein thousands of genes could be analysed by loss-of-function. Till date there have been numerous examples highlighting the advantages of this technique in different cellular systems applied to probe different questions.

1.2 Genome-wide RNAi screens

Applying RNAi to determine gene function pushed researchers to scale up this technology in order to probe thousands of genes. The rationale behind this approach was to have an unbiased view of how genes interact with each other in a network to modulate stress response, transmit signals from receptors to effector molecules and how these networks alter during a pathogen attack.

1.2.1 Application to viral diseases

Viruses are obligate intracellular parasites which are also causative agents of many human diseases. These diseases now account for more than 3 million deaths per year worldwide (http://www.cdc.gov). With their small sized genomes, they rely heavily on the host system they attack in order to replicate. Thus, such host factors required in the viral life cycle become indispensable targets for antivirals[19, 20]. However, identifying host factors that are druggable yet non-lethal to the host is a difficult task. With the advances in RNAi screening technologies and design of siRNA libraries for majority of genes in the human genome, researchers could now probe genes in a high-throughput manner to determine their role in viral diseases. It was easy to set up genome-wide screens in the Drosophila cells as Drosophila melanogaster genome was completed earlier than the human genome[21] which allowed for synthesising

dsRNA libraries[22, 23]. Subsequently, the first genome wide screen performed on viruses was on Drosophila C virus[24]. This screen had a very simple setup; cells were incubated with a single RNAi specific to each gene in a 384 well setup and incubated for 3 days with DCV. A day later, the cells were processed for immunoflourescence against the capsid antigen and subjected to automated microscopic imaging. The experiment yielded 210 dsRNA species that reduced the ¯relative number of infected cells by $> 40\%$ and were identified by visual inspection. dsRNAs targeting these genes were re-synthesised and re-tested for their ability to decrease DCV infection. This "*Validation screening*" further identified 112 Host Dependency Factors (HDFs); 66 of them being ribosomal proteins specifically required for translation of DCV polyprotein. In contrast, these genes were not required by vesicular stomatitis virus whose genome, unlike DCV, doesn't contain a ribosomal entry site (IRES) mediating RNA translation in the absence of a 5' cap. Thus, the authors concluded that the 66 genes discovered in the 2nd step, are essential for DCV IRES mediated genome translation.
This approach has now been successfully applied to multiple viruses including Human Immunodeficiency Virus (HIV-1) , Hepatitis C virus (HCV), Dengue virus (DV), Influenza virus (IV) and West Nile Virus (WV) multiple screens to identify host factors for HIV, HCV, Influenza virus and West Nile virus [1–8, 25–30]. A common factor of all these screens has been that all these screens were able to detect novel host factors that reduced viral replication upon silencing, thus called *Host Dependency Factors* and those that increased viral replication upon silencing, thus called *Host Restriction Factors*. Discovery of these novel genes have allowed for designing novel drugs for these viral diseases.

1.2.2 Current Issues with the technology

As with any technology, RNAi is not without its own limitations. One of the major issues with RNAi, has been the "*Off-target*" effect. It refers to the non-specific gene silencing by siRNAs which often leads to false-positives. Recent studies have enabled scientists to understand the causes of these effects and eventually, devise strategies to overcome these effects[31]. Off-target effects can be of two types, namely:

[1] **Sequence dependent effects**: As the name suggests, these effects arise due to sequence homology of the siRNA with multiple targets, including unintended ones. Specifically, siRNAs regulate the expression of certain genes in a sequence-dependent manner by acting like a microRNA. This occurs when the exogenously introduced siRNA induces microRNA-like effects by interacting with its target sequences through its *seed region*; the 5' region of guide strand of an siRNA or miRNA sequence, extending from nucleotides 2-7 (hexamer) or 2-8 (heptamer)[32–35].

Sequence independent effects: Introduction of exogenous siRNA or short hairpin RNAs (shRNAs) may affect cellular physiology in various ways. For instance, exogenous expression of shRNA can interfere with endogenous processing of microRNAs, as the branches of siRNA and microRNA pathway converge at the Dicer cleavage stage in the RNAi pathway. This has been shown to have fatal consequences in mice, possibly due to saturation of the pathway that is used to export miRNA precursors from the nucleus[36]. A second manifestation of exogenous siRNA introduction can be that these siRNAs can displace endogenous microRNAs, which leads to alteration of normal patterns of gene expression. A bioinformatics analysis of published transcriptional profiling experiments revealed that the predicted targets of endogenous microRNAs show higher expression after transfection of a siRNA directed towards silencing of a specific gene[37]. A third type of cellular change involves activation of the non-specific immune response to the introduced double-stranded RNA (dsRNA) or from the transfect viruses used to express shRNAs[38, 39].
Either type of these effects influence the final outcome of a RNAi screen, which tends to misinterpretation of results. In general, it leads to higher false-positives. For instance, in a screen to identify novel regulators of the HIF-α transcription pathway, the top scoring hits were obtained through microRNA-like seed-match effects instead of specific knockdown of their targets[40]. This was also true in the case of a large scale screen to determine modulators of the TGF-β signaling pathway, where at least 1% of the sequences targeting the 6000 genes of the library showed measurable off-target effects occurring via microRNA-like regulation, on the upstream pathway components TGF-β receptors 1 and 2 (TGFBR1 and TGFBR2)[41]. The second type, false-negatives, pose another problem. Most of the time, they are missed due to low coverage of the RNAi library. Booker et al. performed a meta-analysis to determine false-negative rates in Drosophila screens, where they determined that this rate is at least 8% [42]. These rates are influences by RNAi reagents, namely, siRNA pools which are different than the endogenous Drosophila cellular dsRNAs.
For both scenarios, there are certain common remedial measures. For instance, using multiple siRNAs for the same gene cancels these error rates to some extent and effectively reducing the occurrences of false-positives and false-negatives. However, there is always a chance of the inherent error rates of these multiple siRNAs to add up instead of cancelling each other which may have the exact opposite effect. Thus, this measure comes with a certain trade-off of its own. Another approach is to complement the results of a RNAi screen with transcriptome data, where expression levels of strong hits can be checked in a particular cell-line or tissue [42]. This can, to some extent reduce the number of false-positives and false-negatives. However, for significant reduction and recovery of true-positives, additional measures and strategies have to be devised. We discuss these in the following subsection with emphasis on bioinformatics approaches.

1.3 Bioinformatics Approaches for analysis of genome-wide RNAi screens

Parallel to the developments in HT-RNAi technology, statistical and bioinformatics approaches have also been developed to overcome the shortcomings of this technology. A major focus of these algorithms has been to improve reliability of the hits and to allow for in-depth of the biology of host-virus interactions, with RNAi hits being the starting point. Based on the the methodology used, we broadly categorise these studies in 2 classes; "*Machine Learning Approaches*" and "*Network-based approaches*".

1.3.1 Machine-learning approaches

The increase in the number of genome-wide screens to determine novel host factors for viruses has helped strengthening confidence of theoretical approaches, as the novel host factors serve as high-confidence positive sets in machine-learning methods to predict novel host factors.
An example of such approach is the SinkSource algorithm by Murali et.al who developed a semi-supervised machine learning approach to predict novel HIV-1 HDFs using previously identified ones[1–3, 43]. HDFs identified in these 3 HIV screens were utilised in a classification algorithm that would *learn* from these hits and predict novel HDFs. Specifically, the host PPI network was modelled as a liquid flow network. Each node (protein) in the network acts as a reservoir and the edge (interaction) connecting them is a pipe. The weight of an edge is directly proportional to the amount of fluid that can flow through the pipe per unit time. When the liquid attains equilibrium (amount of liquid flowing in and out remains constant) the height of the node denotes its confidence of being a HDF. Known HDFs from the screens were assigned a node value of 1 while non-HDFs were assigned a node value of 0. The authors note that their algorithm is similar to Nabieva et al. which was formulated for functional prediction of genes, except that SinkSource accepts negative values in the form of non-lethal, non-HDFs[44]. Thus, the non-lethal, non-HDFs formed the negative set while the HDFs identified in the 3 screens and their intersections formed the 4 positive sets, which were then used for two-fold cross validation. Two-fold cross validation involved splitting of each of the positive and negative sets in halves and each half was used for prediction of genes from the other half. Ultimately, SinkSource predicted new 1394 HDFs in addition to 908 form the above 3 screens, with an accuracy > 80% based on the above mentioned two-fold cross validation. Combining these HDFs to form a protein network, dense subgraphs were detected using MCODE[45]. Enriched GO terms from these subgraphs included;spliceosome, kinetochore and mitochondrion, whereas GTPase mediated signal transduction, DNA replication initiation and MHC protein complex, all relevant to HIV-1 replication.

A major advantage of machine learning methods lies in the analysis of hi-dimensional image-based screens. For e.g. Walter et al. developed an automated classification to classify phenotypes in a genome wide screen by time-lapse imaging. They utilised 190 features from hi-content images to classify nuclei from these images into different stages of mitosis (interphase, prometaphase, metaphase and anaphase)[46]. They further introduced Event Order Maps, a visualisation tool for visualising phenotypic dynamics from time lapse RNAi screens, which could then specifically determine tendencies of causes and consequences of phenotypic classes.
Another important application of supervised classification is to recover false negatives from genome wide RNAi screens which is extremely difficult or impossible from experimental methods. Wang et al. developed a network-based phenotype scoring method that considers network topology and RNAi screening results to identify putative false positives and false negatives within a RNAi screen[47]. Using 24 genome-wide screens they observed that RNAi hits are tightly connected to each other in a protein interaction network when compared to random hits. Thus, they considered network centralities such as direct neighbour, shortest path, diffusion kernel and association analysis-based transformation along with gene similarities and developed a set of scoring functions called Network RNAi Phenotype (NePhe)[48]. Applying the guilt by-association principle, Wang et al., hypothesised that FNs should be scored higher by a scoring function over false positives (FPs), as they are linked by a greater number of true hits. Thus, an ideal gene classifier would always rank FNs with a higher rank compared to a non-hit. They applied this scoring function to Wnt and the Hedgehog signaling pathways and the NePhe scoring system could identify all regulators of these pathways which were even missed by follow-up Drosophila knockdown screens.

1.3.2 Network-based approaches

This section describes how using network data enhances and complements RNAi screens. This subsection is further subdivided into 3 subsections which describe usage of network data and network topology, applied for analysing RNAi screens.

1.3.2.1 Using heterogeneous data

The increased throughput in biology has not only developed the amount of data but also its types. These include genome-wide association studies, gene expression, protein protein interaction, gene regulatory interactions, transcription factor binding site data and so on. They individually answer specific questions but their combination further helps in improved understanding of biological systems. For instance, Zhu et al. combined multiple data types which included genotypic, expression, transcription factor binding site (TFBS), and protein-protein interaction (PPI) to reconstruct causal gene networks for yeast [49]. From these networks, Zhu

et al. could identify that this reconstructed network identified knockdown signatures better than networks constructed using the respective, singular dataset. Reiss et al. reiterate the advantages of such data combination wherein their algorithm, *cMonkey* derives biclusters from subsets of experimental conditions using gene expression data and multiple gene association networks[50]. cMonkey identified additional motifs in the bacteriorhodopsin regulon.

1.3.2.2 Using PPI data

In the past decade, there has been a constant rise in the identification of protein interactions across all species and the repositories housing them. Among these, repositories with virus-host protein interactions are an important subset. Virusmint, VirHostNet and the HIV-1 Human Protein Interaction Database (HHPID) at National Institute of Allergy and Infectious Diseases (NIAID) are few examples[51–53]. Of late, many approaches have come forth to analyse and interpret the genome-wide RNAi screens using protein interaction networks and host-virus interaction networks.

Another instance of an integrated network analysis is from Macpherson et al. who used host-virus protein interaction data from HIV-1 HHPID at NIAID[53, 54]. They identified individual enriched clusters from the host-virus interaction network using bi-clustering. For obtaining a hierarchical view of function, these clusters were organised in a functional cladogram where the distance between 2 clusters was determined by the number of overlaps between the clusters; clusters with more overlapping genes were closer to each other than clusters with fewer overlaps. GO enrichment analysis of these clusters indicated that 37 host subsystems were potentially important for HIV-1 infection. Moreover, hits from the 3 HIV screens were enriched in 10 of these subsystems and included proteasome core complex, regulation of apoptosis, mRNA transport, endosome, RNA polymerase activity, peptidase activity, regulation of transcription, ubiquitin camp-dependent protein kinase complex, and v-akt. The HIV-1 HHPID resource defines how a single viral protein interacts with a single host protein. Although important, it takes further computations to determine how viral proteins interact with host proteins. To that end, Macpherson et al. described how viral proteins interact with host subsystems over individual proteins. This aids the identification of novel hits and mechanisms in a theoretical framework.

The ultimate goal of viral RNAi screens has been to discover novel drug targets. Using drug-target information in conjunction with protein interaction networks allows to predict or hypothesise drug mechanism and adapt the therapeutic in case there are chances of side-effects. A recent study by de Chassey et al. used the Drugbank database (`http://www.drugbank.ca`, protein interaction networks and 6 Influenza(IFV) virus screens[7, 8, 27, 28, 55–58] to predict drug targets for Influenza virus. A total of 925 host factors were identified, which the authors termed as Essential Host Factors (EHFs). Network analysis using the PPI dataset from VirHostNet

revealed that 17 EHFs are directly targeted by a viral protein[52]. Interestingly, the neighbourhood of these directly targeted proteins was equally informative; which included 204 proteins targeted by at least one viral protein. Secondly, drug molecules interacting with EHFs were further retrieved from the DrugBank database. From these retrieved molecules, it was found that 100 EHFs could be targeted by 298 molecules comprising of 204 FDA-approved drugs and 94 experimental drugs. For further confidence, these 100 EHFs were further filtered for the following three criteria, such that a molecule satisfies at least one of these:

- EHF was directly targeted by a viral protein.
- EHF was connected to at least another EHF.
- EHF was connected to a non-EHF targeted by the virus.

This filtering yielded 33 EHFs, 32 of which could be targeted by 49 FDA approved molecules with an exception of HSP90AA1 which also satisfied the first 2 criteria mentioned above. HSP90AA1 can be targeted by 1 FDA-approved molecule and 5 experimental drugs. Amongst these experimental drugs, Geldanamycin has been proved to reduce IFV replication by 2 logs in cell culture[59]. Lastly, the authors suggest that a combination of drugs including Geldanamycin and Ribafutin(used as 1st line of treatment against tuberculosis) may represent a novel strategy to identify novel drugs to combat IFV infection.
These studies show that how usage of network data enhance the findings of a RNAi screen. Moreover, it also helps researchers to realise the ultimate goal of such RNAi screens which is to determine novel drug targets and suggest alternative therapeutic measures in addition to the known ones.

1.3.2.3 Using network topology

Integrating network data with RNAi screens and then interpreting screen hits seems a "*direct*" approach towards analysing a screen. There has been a complementary or "*reverse*" approach as well where researchers have studied network properties of hits targeted by pathogens (including viruses) and how usage of such properties has helped in interpreting virus-host interactions.
To this end, the most comprehensive study about network topological properties of pathogen targeted proteins has been published by Dyer et al[60]. The authors studied topological features of host factors involved in the life cycle of 190 different pathogens partitioned into 54 groups (35 viral, 17 bacterial, and two protozoan) pooled from 7 public databases[61–67]. The primary finding from this study has been that pathogens preferentially target **Bottleneck** or **Hub** proteins which indicates targeting central proteins in a network is a common property that is shared by many pathogens. Another important finding from this study has been that viral targeted host proteins also play a major role in different cancers, some of which are induced by viral infection

itself (For e.g., Herpesvirus and Papillomavirus). A similar study, specific to HIV, by van Dijk et al analysed the HIV-1-Human protein interaction data and reconfirmed that viral proteins attack **Bottlenecks** or **Hub** proteins. Furthermore, they also identified distinct network structures, otherwise known as *Network Motifs*, which allowed for dynamics interpretation of interactions. For instance, the 2-node feedback loop found in the HIV-host activation/inhibition network, shows the inhibitory nature of HIV proteins towards HRFs.

There have also been experimental approaches to study the virus-host interactome of HCV, DENV and HTLV-1/2[68–70]. A recent study also studied a comprehensive set of genes forming the interactome of 70 viral modulators of the innate immune response from 30 different viruses[71]. The primary conclusions from these experimental approaches reinstate a key feature of viral proteins; they target many host proteins that are *central* to the networks, target proteins that are key components of many cellular pathways as compared to an average human protein. Summing up, it is clear that a virus targets a protein that has the following properties:

1. Higher Betweenness and Degree or more *central* proteins
2. Smaller mean path lengths compared to whole networks
3. Proteins that are *closely bound*

The results of these studies can then be used to develop a classifier, that can utilise network topology to theoretically predict potential HDFs for a virus. Moreover, other network topologies can be probed to determine if they contribute towards a protein being a viral HDF/HRF. For e.g. Node PageRank centrality has been utilised to predict novel HIV hits using published data by Jaeger et al.[72]. They used node PageRank to identify 21 surface membrane proteins critical for HIV-1 infection of which 11 are novel predictions, 3 are confirmed hits (chemokine receptor CCR1, chemokine binding protein 2 and duffy antigen chemokine receptor). The remaining 7 proteins have been confirmed in other studies. These receptors are involved in different phases of HIV infection and thus influence progression of AIDS.

From both experimental and computational studies of network topology, it is clear that viral HDFs/HRFs possess distinct topological features from other proteins in the network. It would be worthwhile to study other network centralities to see if they too influence a protein being a viral HDF/HRF. We thus utilised as many as 7 network centralities and 2 semantic similarity measures to define subnetworks enriched with viral HDF/HRF. The details of their implementation and interpretation is described at length in the following chapters.

Chapter 2

Materials and Methods

This chapter gives a detailed description of the methodology developed in this thesis. Since the methodology involved usage of several scoring functions and enrichment tools, a short description of each of these is provided.

2.1 Defining the problem!

Genome-wide RNAi screens have provided ample opportunities to investigate host-virus interactions and to identify novel host factors, their mechanisms in a high-throughput manner. This is evident from the many genome-wide screens for viruses such as HIV, Influenza virus, HCV, WNV. These results have opened possibilities of comparing them and understand similarities-differences in viral infection. Other data sources such as gene expression, protein interaction networks, and small molecule or drug screens can also be utilised in tandem to add depth to such analyses. Thus, the availability of heterogeneous data on one hand and genome-wide RNAi screens on the other have opened up many avenues for a systems approach of analysing RNAi screens. However, incorporating this data and utilising it for a comprehensive systems analysis of RNAi screens, is a challenging task.
Based on this data, we posed 2 key questions;

1 Are there functional *regions* in the HPIN that are broadly targeted by many viruses and/or specific to only some viruses?

2 Using network topological properties of such *regions*, can we predict novel hit proteins or host-virus mechanisms?

We approached this problem in the following manner:

1. To identify functional regions within the HPIN, we initially constructed a HPIN using collated interactions from public repositories.
2. Functional modules were identified using a clustering algorithm, in this case, ClusterOne[73].
3. These modules were filtered in 2 steps:
    a. Modules significantly enriched with *Hits* from a RNAi screen(s). These *Hit-enriched* modules were further filtered based on network topology in the next step.
    b. Topological filtering of modules, wherein *Hit-enriched* modules from the previous step were filtered based on network topology.
4. These doubly filtered modules were then functionally characterised based on pathway and GO enrichment.
5. Novel host factors were predicted using gene-expression data, virus-host interaction data and known human protein complexes.

2.1.1 Datasets

The following subsections describe the different types of datasets used for this study.

2.1.1.1 Protein Interaction Networks

To build a comprehensive set of host protein interactions, we downloaded human protein interactions from multiple public resources. Protein interactions were pooled from 10 different public repositories that included computationally predicted interactions. These were downloaded from iRefIndex v9.0[74] that provided unique indices to protein interactions from public repositories; DIP[63], IntAct[61], MINT[62], BioGRID[75], BIND[67], CORUM[76], MPact[77], HPRD[64], MPPI[65], OPHID[78], summarised in table 2.1. We also included predicted interactions(that constitute 50% of the interactions of the final dataset) from STRING[79] in order to have enhanced coverage of the network. STRING assigns a score to each protein interaction pair based on the sources from which the interaction is referenced[79]. For interactions from STRING, we employed a filter by including interactions that had a combined score of 0.75 and above. This resulted in an interaction network from 15383 proteins and 337413 interactions. Hereafter, this network is referred as integrated human protein-protein interaction network Hu.PPI.

2.1.1.2 RNAi screen Hits

In general, ***Hits*** are defined as genes that are at least 2SDs $\geq$ plate mean, identified in the published studies[1–6, 30]. These were either called Host Dependency Factors (HDFs) or Host

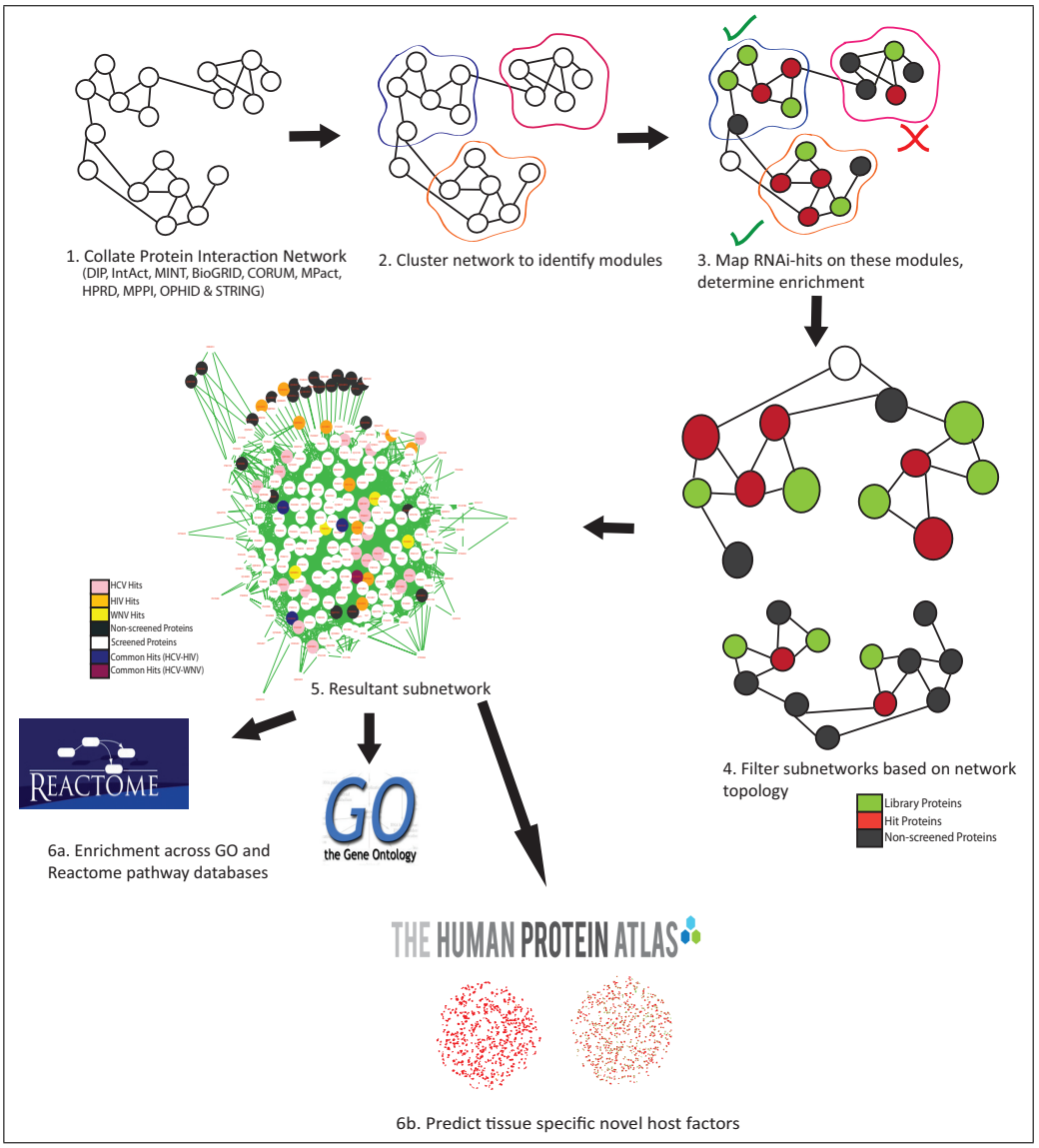


FIGURE 2.1: The figure summarises the work flow executed in this analysis. 1) In step 1, we build the Human Protein Interaction Network(HPIN) by collating interactions from public databases (See Datasets). 2) In this step, we apply ClusterONE on HPIN to identify functional modules.3) RNAi hits are then mapped to these modules and their enrichment in each of these modules is determined using Fisher's exact test. Modules statistically significant at 5%, are then subjected to topological filtering in the next step. 4) This step further filters the enriched modules from the previous step, using mean network centralities and semantic similarity scores (See Filtering subnetworks).5) Subnetworks obtained from these 2 filtering steps are then functionally analysed using, Reactome and GO enrichment.

Source	Version (date)
BIND	2005-05-25
BIND Translation	Version 1.0 (2010-12-15)
BioGRID	Version 3.1.81 (2011-10-01)
CORUM	02/12/09
DIP	10/10/10
HPRD	Release 9 (2010-04-13)
IntAct	2011-09-29
MINT	21/12/10
MPACT	10/01/08
MPPI	2004-06-01 (from archive)
OPHID	2006-07-07

TABLE 2.1: IrefIndex ver 9.0 and versions of the included databases. We also used STRING v9.0 to build the final host protein interaction network (HPIN).

Restriction Factors (HRFs). These hits were further grouped in 2 classes; **Virus-specific Hits** and **Combined Hits**. Virus-specific hits included hits from multiple screens for a virus (of the same species). For example, **HIV hits** comprised HDFs and HRFs from the 3 HIV-1 genome-wide RNAi screens[1–3]. **HCV hits** were pooled accordingly. Since there was only one screen for WNV, it was directly used as it was published. On the other hand, **Combined Hits** included hits from all screens, including inter-species(between hit-sets of virus species) or intra-species(within hit-sets virus species) overlapping hits. The pipeline was run on each of these hit-sets separately. All these screens used the Dharmacon siGENOME library consisting of 17745 proteins.

2.1.1.3 Tissue-specific Expression Data

Human tissue expression data was downloaded from http://www.proteinatlas.org/ version 10.0 - 2012.09.12, Ensembl version: 67.37. The flatfile consisted of Ensembl gene identifier ("Gene"), tissue name ("Tissue"), annotated cell type ("Cell type"), expression value ("Level"), the type of annotation (annotated protein expression (APE), based on more than one antibody, or staining, based on one antibody only) ("Expression type"), and the reliability or validation of the expression value ("Reliability"). Genes were filtered for moderate to high expression levels based on the expression value labels - "Medium/Moderate" or "High/Strong". The same filter was applied to the *Reliability* field of this dataset, wherein genes with high support were selected (Reliability values were "High" or "Supportive"). Using these filters we short-listed moderate-high expressed genes in macrophages (HIV) and liver (HCV) from this set.

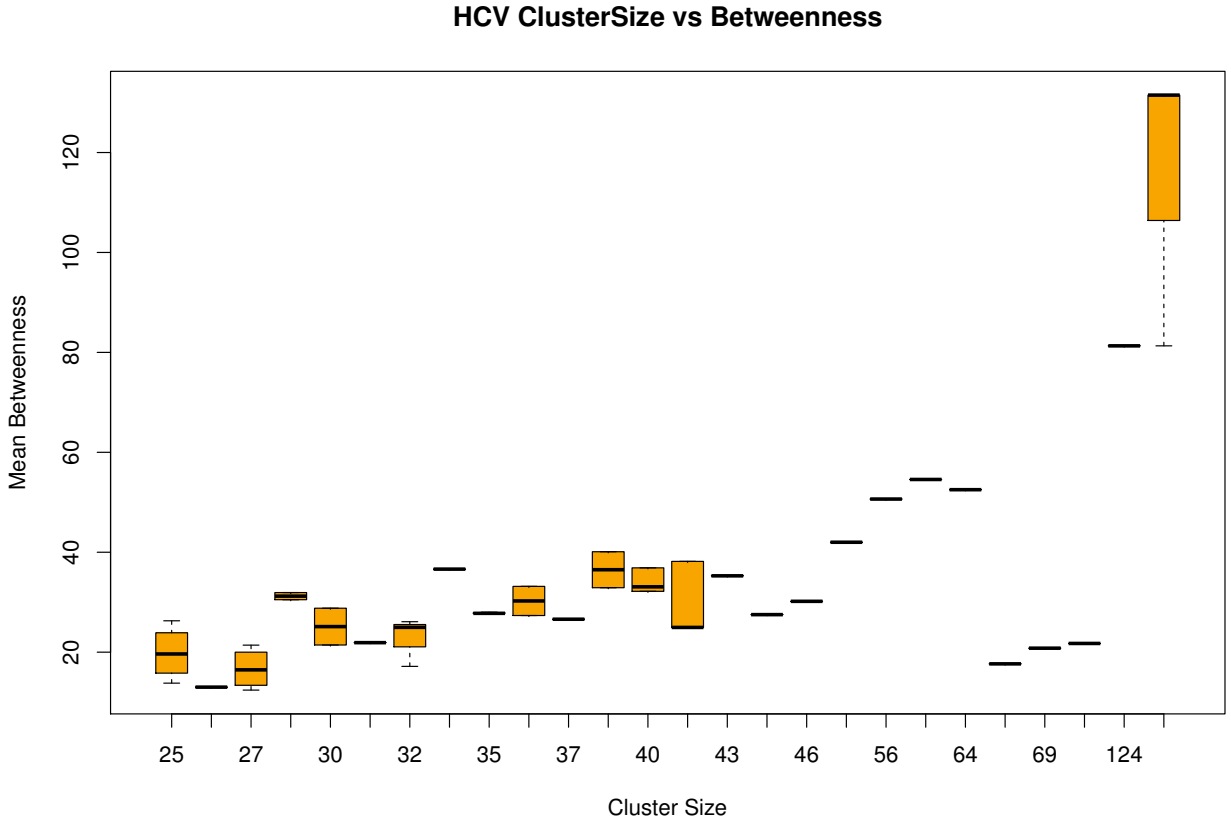


FIGURE 2.2: Boxplot displaying mean betweenness of subnetworks for all cluster sizes for all the HCV subnetworks. These values were further utilised to filter these subnetworks for topological enrichment, detailed in the Parameters section.

2.1.2 Clustering Hu.PPI Network

We used a clustering algorithm that allows for detection of overlapping protein complexes by neighbourhood expansion, called ClusterONE[73]. The usage of an algorithm that detects overlapping complexes was to identify multifunction proteins that play different roles in different complexes for the same virus as well as in different viruses. Such multifunction proteins can also help interpret the role of the protein complex, of which they are a part of, in multiple perspectives. One of the required input parameters to the algorithm was minimum cluster size. ClusterONE generates clusters of at least the size specified in the size parameter, so different clusters of sizes ranging from 25 to 100 were obtained. For values of cluster size 100 and above, the cluster size and the number of the generated clusters remained constant as shown in Fig 2.2.

2.1.2.1 ClusterONE algorithm

The algorithm is based on the concept of cohesiveness score and then uses a greedy growth process to find groups that could potentially be protein complexes. Cohesiveness measures the likelihood of a group of proteins to form a complex. It is defined as follows. Consider $w^{in}(V)$ as the total weight of the edges contained between the group or proteins V, and let $w^{bound}(V)$ denote the total weight of the edges that connect the group to the rest of the network. The cohesiveness of V is given by,

$$f(V) = \frac{w^{in}(V)}{w^{in}(V) + w^{bound}(V) + p|V|} \tag{2.1}$$

$p|V|$ is a penalty term that is used to model the uncertainty in the data based on the assumption that there exist yet undiscovered protein interactions. By allowing values of $p \geq 0$, offsets the boundary weight $w^{bound}(V)byp|V|$, that implies every protein in V has p additional boundary connections that could not be identified owing to limitations of experimental procedures. Thus, the authors suggest that, different values of p can be used for different biological assumptions. For instance, a well-studied protein can have lower p value as it is highly unlikely that it has undiscovered interactions. For sake of simplicity, we chose not to alter this penalty term and used the default value as set by the algorithm. The ClusterONE algorithm works in 5 major steps:

while $v \subset V \neq \phi$ **do**
  Step1 : let $V_0 = \{v_0\}, t = 0$;
  Step2 : Calculate cohesiveness for V_t using (2).
  Step3 : and let $V_{t+1} = V_t$;
  For every external vertex v on at least one boundary edge, cohesiveness of $V' = V_t \cup v_t$. If $f(V') > f(V_{t+1})$, let $V_{t+1} = V'$;
  Step4 : For every internal vertex v on at least one boundary edge, cohesiveness of $V" = V_t \cup v_t$. If $f(V") > f(V_{t+1})$, let $V_{t+1} = V"$;
  Step5 : if $V_t \neq V_{t+1}$, increase t and return to Step 2.;
  Else V_t a locally optimal cohesive group.
end

Algorithm 1: ClusterOne algorithm

An illustration for this greedy cohesive process of group detection is illustrated in figure 2.3. This graph consists of 12 nodes, marked from node0 to node11. Assuming $p = 0$, the cohesiveness of the shaded set is $7/14$. In steps 3 and 4, the algorithm can add either of the following nodes; node5, node6, node7 and node8. The best choice is to add node 7 as it increases the internal count by 4 as it converts these boundary edges to internal ones. Thus, the cohesiveness of the group would be $11/14$ while for node8 it would be $10/14$, for node6 will be $9/14$ and for node5 it will be $8/14$.

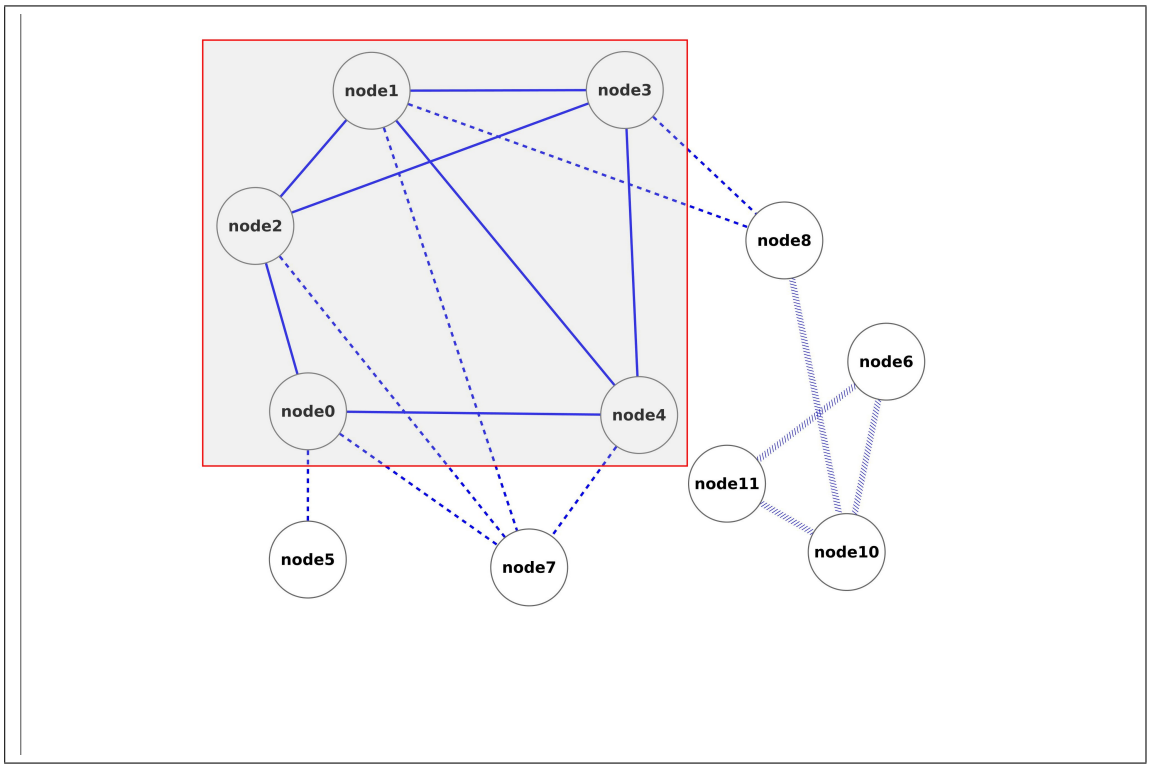


FIGURE 2.3: Illustration of greedy cohesive group detection in ClusterOne. Shaded area is indicated by a shaded area. Thick blue edges are completely internal, dashed edges are boundary edges and forward slash edges are completely external.

1. In the first step, the algorithm grows groups of proteins with high cohesiveness from selected seed proteins. First, a protein with highest degree is selected as the 1st seed and a greedy search procedure grows a cohesive group from this protein. When this growth process ends, the algorithm selects the next seed from the unassigned proteins and repeats the process. This process continuous till there are no proteins left to consider.

2. In this step, locally optimal cohesive groups are merged. The extent of overlaps between each group pair is computed and those groups that have the overlaps beyond a certain threshold($\omega > 0.8$), are merged.

$$\omega((A,B) = \frac{|A \cap B|^2}{|A||B|} \tag{2.2}$$

   ClusterONE performs these overlaps pair-wise; overlap scores for each group pair is calculated and an overlap graph is constructed where each vertex represents a cohesive group. Two groups are connected to each other are merged into protein complex candidates. If a group has no further connections to any other groups, it is then labelled as a protein complex candidate and no additional merging is performed.

3. In this final step, the predicted groups are pruned for size and density. Specifically, those groups smaller than 3 and a density δ below a certain threshold are discarded, where δ is

defined as,

$$\delta = \frac{\Sigma w^{in}(V)}{\frac{n(n-1)}{2}} \tag{2.3}$$

where n is the number of proteins.

2.1.3 Determine RNAi hits enrichment for all clusters

For every cluster generated, a p-value was calculated using the Fisher's exact test to determine significant enrichment of hits in a cluster[80]. This was repeated for all hit-sets mentioned above which resulted in 4 sets of clusters, namely: HIV, HCV, WNV and Combined. The Combined hit-set consisted of hits pooled for all viruses from all screens. For all subsequent references in the analysis, a *subnetwork* will refer to a visual representation of an enriched cluster. Each of these enriched cluster sets were analysed for network topology in the next step.
We demonstrate hit enrichment in a subnetwork with the following illustration. This is a contingency table used for the Fisher's exact test, to compute p-values. For the HCV hit-set, we obtained one of the subnetwork of size 40, of which 38 proteins are from the library of genes that were screened (in this case the library being Dharmacon siGENOME library which was common for all the 7 screens). Thus, the contingency table of this subnetwork will be:

Sr.No	**Hits**	**Non-Hits**	**TOTAL**
In-cluster	4	34	38
Not-in-cluster	0	2	2
TOTAL	4	36	40

TABLE 2.2: An example of RNAi hit enrichment in subnetwork

2.2 Parameters

2.2.1 Network Centralities

In order to understand how a network functions, it is worthwhile to study its structural properties and how these impacts its function. Intuitively, the manner in which individual components of the network interact provides a basis for quantifying such structural properties. Formally, these measures that allow to quantify and characterise network topology are referred to as, **Network Centralities**. This term was elaborated by Freeman in his seminal paper titled, *"Centrality*

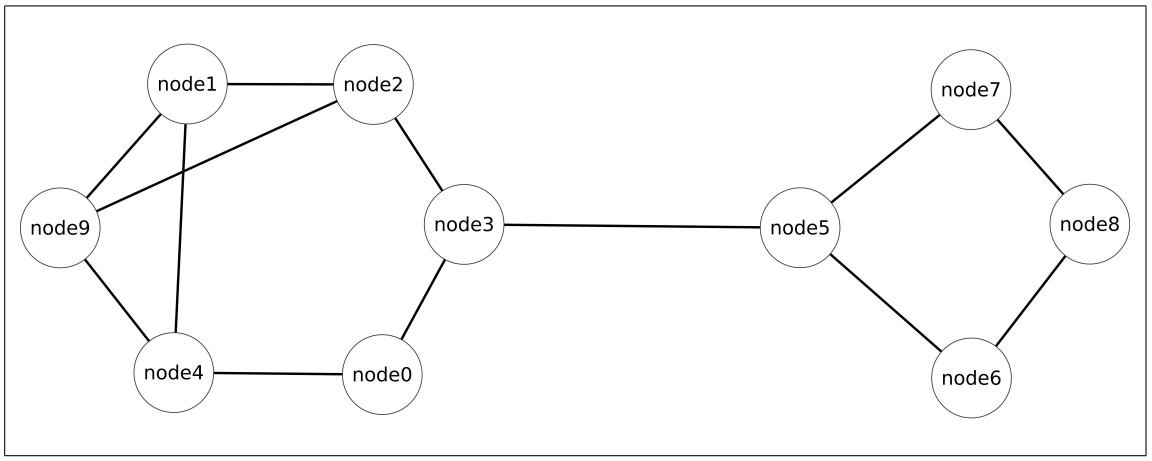


FIGURE 2.4: A sample protein interaction network (PIN), represented as an undirected, unweighted graph. The nodes represent proteins and the edges represent interactions between them.

in social networks conceptual clarification"[81]. Over the years, researchers developed a set of such centralities that characterise different aspects of a network. For e.g. **Betweenness** centrality gives the extent to which the shortest paths pass through a particular node. The higher number of paths passing through such a node, the higher is its *"Betweenness"*. In protein networks, this automatically translates to proteins that function as **Bottlenecks** which are more likely to be essential proteins and play an important role in gene expression[82, 83].
Before we describe these network centralities, we introduce some basic concepts of graph theory as defined in the literature, along with their mathematical representation[84].

1. **Undirected graph:** A graph G can be defined as a pair (V,E) where V is a set of vertices representing nodes and E is a set of edges representing the connections between the nodes. A single connection between nodes i and j is defined as $E = \{(i,j)|i,j \in V\}$, in which case i and j are ***neighbours***. If the edges connecting the vertices have no direction and numerical weights assigned to them, then such a graph is called an *undirected, unweighted graph.* Protein interaction networks are more often represented as these graphs.
2. **Directed graph:** A directed graph is defined as an ordered triple $G = (V,E,f)$, where f is a function that maps each element in E to an ordered pair of vertices in V. Such directed pairs of edges are called ***directed edges, arcs or arrows***[85]. In protein interaction networks, a directed connection usually represents sequential interaction of the elements and the flow of information in the network. For e.g. signaling networks that consists of signaling cascades where binding of a receptor triggers a certain pathway which is followed by their action on final effector molecules, the direction of an edge represents the flow of such a signal.

3. **Adjacency Matrix:** For a graph $G = (V, E)$, the adjacency matrix is represented by $|V|x|V| = nxn$ matrix $A = (a_{ij})$ such that $a_{ij} = 1$ if $(i, j) \in V$ or $a_{ij} = 0$ or otherwise

$$A = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix}, n = |V|$$

For the sample PIN listed above 2.4, the matrix will be,

$$A_{SampleNetwork} = \begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}, n = 10$$

In case of **undirected graphs**, the matrix is symmetric because $a_{ij} = a_{ji}$. This isn't true in the case of **directed graphs**, as the upper and lower triangle parts of the matrix reveal the direction of edges.

4. **Walk, Path and Trail** - A ***walk*** is a pass through a specific sequence of nodes $(v_1, v_2, \dots, v_N)$ and $(v_1, v_2), (v_2, v_3), \dots, (v_{N-1}, v_N) \subseteq E$. A ***simple path*** is a walk with no repeated nodes while a ***trail*** is a path where no edge is repeated.

5. **Shortest Path** - For a pair of nodes $i, j \in V$, the distance $\delta(i, j)$ from i to j is the length of the *shortest path* from i to j in G. If no such path exists, then $\delta(i, j) = \inf$. Shortest path problem refers to the process to determine a path between two nodes in a graph such that the sum of the weights of its constituent edges is minimised.

   We utilised 7 network centralities and 2 semantic similarity measures to filter subnetworks. With reference to the sample network in Fig. 2.4, we describe few basic graph theory terminologies and definitions of these measures as follows:

   1. **Degree centrality** - As defined by Freeman, degree of a node denotes the number of edges incident to a node or emanating from it [81]. For an undirected graph G, $G = (V, E)$ where V is the set of vertices/nodes and E is the set of edges, the degree

of a vertex v is denoted by, $deg(v)$. By this definition, the nodes *node1, node2, node3, node4, node5 and node9* of the sample network, have the highest degree, which is 3. Intuitively, higher degree also reflects a node's importance in a network, in this case, indicates the multifunction aspect of a protein. Such proteins are also referred to as **Hubs** and indeed, are essential for normal cellular function purely due to their abundant interaction partners[86].

2. **Closeness centrality** - indicates the extent to which a node is "*closer*" to other nodes in the network. It also determines the ability of a node to quickly communicate with other nodes with respect to flow of information. Mathematically, it is defined as the inverse of the shortest path between a pair of nodes. For an undirected graph $G = (V,E)$, closeness centrality of a node i $C_{closeness}(i)$ is defined as,

$$C_{closeness}(i) = \frac{1}{\sum_{j \in V}^{|V|} dist(i,j)} \tag{2.4}$$

where $dist(i,j)$ denotes the distance or the shortest path p between the nodes i and j. Mazurie et al. showed that, in a metabolic reaction network, a more *central* node has a faster rate of transfer of metabolites than the less *central* ones[87]. Specifically, nodes with higher closeness aid in faster transfer of metabolites in a metabolite reaction network. They also showed that as distance between metabolic pathways increased, it led to a consequent decrease in the closeness centrality of metabolites in these pathways in context of bacterial evolution. In the context of signaling networks, a special class of PINs, it indicates how a particular signal can transfer within a network.
Using the formula above, closeness centrality of the proteins in the sample network are as follows:

Centrality	**node1**	**node2**	**node3**	**node4**	**node5**	**node6**	**node7**	**node8**	**node9**	**node10**
CLS	0.0435	0.0526	0.0625	0.0435	0.0555	0.0417	0.05	0.033	0.0435	0.0417
BWN	0.5	10.0	21.0	2.0	18.5	3.5	5.0	0.5	0.5	3.5
PGR	0.106	0.107	0.112	0.108	0.121	0.087	0.077	0.089	0.106	0.087
CC	0.67	0.33	0.0	0.33	0.0	0.0	0.0	0.0	0.67	0.0

TABLE 2.3: Centralities for all nodes in the sample network. The 3-letter abbreviations in the first column denote centralities; CLS-Closeness, BWN-Betweenness, PGR-PageRank centrality and CC-Clustering coefficient.

From the table 2.3, it is clear that *"node3"* is *closer* to all other nodes in the network and thus, can quickly convey a signal from one protein to another.

3. **Betweenness centrality** - Betweenness of a node indicates the number of shortest paths passing through that node, for any two nodes in a network[88–90]. Such nodes thus act as **bridging nodes** between closely connected regions or communities within

a network. Removal of such nodes can disconnect these communities and break the network. These are also referred to as **Bottlenecks**. For unique nodes, $i,j,w \in V(G)$, σ_{ij} be the total number of shortest paths between i and j while $\sigma_{ij}(w)$ be the number of shortest paths between i and j passing through w. Furthermore, for $w \in V(G)$, let V(i) denote the set of ordered pairs, (i,j) in $V(G) \times V(G)$ such that i,j,w are distinct. Then, the Betweenness centrality of w is given by,

$$C_{Betweenness}(w) = \sum_{(i,j) \in V(w)} \frac{\sigma_{ij(\mathbf{w})}}{\sigma_{ij}} \tag{2.5}$$

In PINs, proteins with high betweenness(**Bottlenecks**), alike those with high degree(**Hubs**) are favoured by many pathogens.

Betweenness centrality values for the proteins in the sample network are mentioned in 2.3. It is clear from this table that *node3* and *node5* are the nodes with high betweenness.

4. **Mean path length** - As mentioned above, the *shortest path* between a node pair *(i,j)* in a graph G is given by, $\delta(i,j)$. Thus, the ***mean or average path length*** of a graph G is defined as the average of $\delta(i,j)$ taken over all distinct pairs of nodes, $i,j \in V(G)$ given that they are at least connected by one path. Specifically, the average path length of a network is the mean number of edges between nodes, which must span the shortest path between a node pair. It is given by,

$$\delta = \frac{2}{N(N-1)} \sum_{i=1}^{N} \sum_{j=1}^{N} \delta_{min}(i,j) \tag{2.6}$$

where $\delta_{min}(i,j)$ is the minimum distance between nodes i,j and N is the total number of nodes in a network. Some of the popular algorithms to determine shortest paths in a network, are the **Dijkstra's greedy algorithm**[91] and **Floyd's algorithm**[92].

For the sample PIN listed above 2.4, the shortest path matrix will be represented as, Thus, using the above formula, the average path length of the sample network is **2.44**.

5. **PageRank/Eigenvector centrality** - It uses the principle that, the *importance* of a node is determined by how *important* its neighbours are. Thus, a node connected to more *important* neighbours is ranked higher than a node with lower number of such neighbours. Mathematically, it is the largest eigenvalue of an adjacency matrix. Such an implementation was suggested by Bonacich, who deemed that it can be a good centrality measure[93]. For a graph *G(V,E)*, consisting of vertices V and edges E, let A be the adjacency matrix for this graph; $a_{ij} = 1$ if i and j are connected and $a_{ij} = 0$ otherwise. The matrix A is symmetrical and thus, its eigenvalues are real, its eigenvectors orthogonal. The largest eigenvalue of A is then represented by λ_{max}, given by, $\lambda C_{eiv} = AC_{eiv}$, where C_{eiv} is the eigenvector of λ_{max}. Thus, for a undirected graph

	node1	node2	node3	node4	node5	node6	node7	node8	node9	node10
node1	0	1	2	1	3	4	2	5	1	4
node2	1	0	1	2	2	3	2	4	1	3
node3	2	1	0	2	1	2	1	3	2	2
node4	1	2	2	0	3	4	1	5	1	4
node5	3	2	1	3	0	1	2	2	3	1
node6	4	3	2	4	1	0	3	1	4	2
node7	2	2	1	1	2	3	0	4	2	3
node8	5	4	3	5	2	1	4	0	5	1
node9	1	1	2	1	3	4	2	5	0	4
node10	4	3	2	4	1	2	3	1	4	0

TABLE 2.4: Shortest paths matrix for all against all nodes. The diagonal elements of this matrix are 0 while the upper and lower diagonal elements consists of pairwise distances.

G with its adjacency matrix A with $V(G) = \{v_1, \ldots, v_n\}$ and $\rho(A) = \max_{\lambda \in \sigma(A)} |\lambda|$, then the eigenvector centrality $C_{eiv}(V_i)$ of the node v_i is given by the i^{th} coordinate x_i of a normalised eigenvector that satisfies the condition $Ax = \rho(A)x$. This algorithm has been implemented to efficiently ranking of web pages, notably by Google. Using this measure, Jaeger et al. identified novel surface membrane receptors of HIV from a PIN[72].

From table 2.3, *node5* has the highest eigenvalue or PageRank value of **0.121** indicating that it is the most important node in the network.

6. **Clustering coefficient** - It's a measure to indicate the connectedness of a vertex to its neighbours, for the graph in consideration[94]. For an undirected graph G, if i is a vertex with degree *deg(i)=k* and the number of edges between the k neighbours of i in G are e, then the *Local clustering coefficient* is given by,

$$C_{(i)} = \frac{2e}{k(k-1)} \tag{2.7}$$

In other words, C_i measures the ratio of the number of edges between the neighbours of i to the total possible edges, which are $k(k-1)/2$ and ranges within $0 \leq C_i \leq 1$. The ***mean clustering coefficient*** of the whole network is then given by,

$$C_{mean} = \frac{1}{N} \sum_{i=1}^{N} \frac{E_i}{k_i(k_i-1)} \tag{2.8}$$

where, $N = |V|$ is equal to the number of vertices. The closer the value of C_{mean} is to 1, the greater is its tendency to form a cluster. Table 2.3, shows the local clustering coefficient values for nodes in the sample network. Using the equation for C_{mean}, the mean clustering coefficient of the sample network is **0.273**.

2.2.2 Semantic similarity indices

1. **Dice Similarity Coefficient** - For an undirected graph, G, such that $G = (V, E)$ where V is a set of vertices and E is a set of edges; then for a vertex pair, u,v, the Dice Similarity coefficient for the vertex pair is given by,

$$Dice_{u,v} = \frac{2C}{|Deg_u| + |Deg_v|} \tag{2.9}$$

   Here, C denotes common neighbours of u and v whereas Deg_u and Deg_v denotes degrees of nodes u and v. We iterate this formula over a whole subnetwork and then obtain the mean Dice similarity coefficient.

2. **Wang similarity coefficient** - A GO term A can formally be represented as a set of directed acyclic graph, $DAG_A = (A, T_A, E_A)$, where T_A is the set of GO terms in DAG_A including A and all its ancestor terms in the GO graph and E_A is the set of edges connecting the GO terms in DAG_A. For a quantitative comparison of GO terms, Wang defined that the semantic value of GO term A is the aggregate contribution of all terms in DAG_A. In other words, the contribution of term t to the semantics of GO term A is referred to as the semantic value or the S-value of GO term t related to A. So, for any of term t in $DAG_A = (A, T_A, E_A)$, its S-value related to term A, $S_A(t)$ is defined as:

$$\begin{cases} S_A(A) = 1 \\ S_A(t) = max\{w_e \times S_A(t') | t' \in childrenof(t)\} if t \neq A \end{cases} \tag{2.10}$$

   where w_e is the semantic contribution factor for edge $e \in E_A$ linking term t with its child term t'. In this function, contribution of term A towards itself is assigned a value of 1. After computing all S-values for all terms in DAG_A, the semantic value of GO term A, SV(A), is calculated as:

$$SV(A) = \sum_{t \in T_A} S_A(t) \tag{2.11}$$

   Thus for any given GO terms, A and B, the Wang similarity between A and B, $sim_{Wang}(A,B)$, is given by,

$$sim_{Wang(A,B)} = \frac{\sum_{t \in T_A \cap T_B} S_A(t) + S_B(t)}{SV(A) + SV(B)} \tag{2.12}$$

   wherein $S_A(t)$ is the semantic or S-value (SV) of GO term t related to term A and $S_B(t)$ is the S-value of GO term t related to term B. Wang's method determines the semantic similarity of two GO terms based on both the locations of these terms in the GO graph and their relations with the ancestor terms. For a group or protein like the ones in our subnetworks, we utilised the mean of the Wang similarity coefficient,

```
goSim("GO:0004022", "GO:0005515", ont = "MF", measure = "Wang")

## [1] 0.158

go1 = c("GO:0004022", "GO:0004024", "GO:0004174")
go2 = c("GO:0009055", "GO:0005515")
mgoSim(go1, go2, ont = "MF", measure = "Wang", combine = NULL)

##            GO:0009055 GO:0005515
## GO:0004022      0.205      0.158
## GO:0004024      0.185      0.141
## GO:0004174      0.205      0.158

mgoSim(go1, go2, ont = "MF", measure = "Wang", combine = "BMA")

## [1] 0.192
```

FIGURE 2.5: GOSemSim sample calculation

calculated pairwise for each GO term pair for a protein m with another protein n. Thus, the above equation can then be represented as,

$$sim_{Wang.avg(g1,g2)} = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n} sim(go_{1i}, go_{2j})}{m \times n} \tag{2.13}$$

where $g1$ and $g2$ are a gene pair and go_{1i} and go_{2j} are their corresponding GO terms, such that $GO_1 = \{go_{11}, go_{12}, \ldots, go_{1m}\}$ and $GO_2 = \{go_{21}, go_{22}, \ldots, go_{2n}\}$. For our subnetwork, we consider a protein pair and all its associated GO terms, for all ontologies. We then perform a ontology-wise calculation of the Wang similarity score for each combination of GO terms for this pair. This is iterated for all possible pairs within a subnetwork. Finally, we calculate the mean of Wang similarity for each GO ontology and assign it to the subnetwork. Thus, each subnetwork has 3 values that are the mean Wang similarity scores for the three GO ontologies. The following example shows a sample calculation of semantic similarity scores using the mgoSim and goSim functions of the GOSemSim package.

2.2.3 Filtering subnetworks

Cluster size is one of the important parameters in any graph based clustering algorithm (in our case, subnetwork size generated from these clusters). The 2 main/required inputs to ClusterOne [73] are: the network to be subjected for clustering and the minimum size of predicted complex. There are 11 other input parameters which are as follows:

1. **–input-format** - indicates the input format of the file. This can be either an edge list or simple interaction file (sif)

2. **–output-format** - indicates the output format of the file (plain, csv or genepro)
3. **–min-density** - used to set the minimum density of the complexes to be predicted.
4. **–debug** - initiates the debugging mode in the algorithm
5. **–fluff** - used to fluff the clusters as a post processing step. This parameter should be used in a case specific manner. The main purpose is to check if the external boundary nodes of each cluster connect to more than two third of the internal nodes; if such is the case, these external boundary nodes are added to the cluster. It must be applied before the size and density filters.
6. **–haircut** - This parameter is used to apply a haircut transformation as a post-processing step on the detected clusters. Basically this step removes any dangling nodes from a cluster provided if the total weight of connections from a node to the rest of the cluster is less than x times the average node weight in the cluster (where x is the argument of the switch), the node will be removed. The process is iterated until there are no further nodes to be removed. Alike fluff, this method is applied before the size and density filters.
7. **–max-overlap** - indicates the maximum allowed overlapping proteins between two clusters. This is measured by the match coefficient which takes the size of the overlap squared, divided by the product of the sizes of the two clusters being considered, as in the paper of Bader and Hogue[45].
8. **–similarity** - used to set the similarity function to be used in the merging step. Specifically, this argument controls which scoring function is to be used to decide whether two complexes overlap significantly or not. This option has the following possible values:

    a. **merge** calculates the intersection size squared, divided by the product of the sizes of the two complexes. This is also called the *matching score* and is the default value.

    b. **meet/min** or **simpson** calculates the Simpson coefficient i.e. the intersection size over the size of the smaller complex.

    c. **jaccard** calculates the Jaccard similarity coefficient i.e. the intersection size over the size of the union of two complexes.

    d. **dice** calculates the Dice similarity coefficient i.e. twice the intersection size over the sum of the sizes of the two complexes.
9. **–merge-method** - specifies the method to be used to merge highly overlapping clusters. This option has the following values:

a. **single** calculates similarity scores between all pairs of complexes and creates a graph where the nodes are the complexes and the two nodes are connected if the corresponding complexes are highly overlapping. Complexes in the same connected component of the graph will then be merged and by its designation, this is a single-pass method.

b. **multi** calculates similarity scores between all pairs of complexes and stores those pairs that have a score larger than a given threshold. The highest scoring pair is merged and the similarity of the merged complex towards its neighbours is recalculated. This process is iterated until there are no more highly overlapping complexes remaining. As its label suggests, this is a multi-pass method where similarities are recalculated after each merging step.

10. **–penalty** - sets a penalty value for the inclusion of each node. For any penalty value x, ClusterONE assumes that this node has an extra boundary weight of x when it considers the addition of the node to a cluster. This option can then be utilised to model the possibility of uncharted connections for each node, so nodes with only a single weak connection to a cluster will not be added to the cluster as the penalty value will nullify the benefits of adding the node. The default penalty value is 2.
11. **–seed-method** specifies the seed generation method to use. The following values are accepted:

    a. **nodes** - every node will be used as a seed.

    b. **unused_nodes** - nodes will be tested in the descending order of their weights (where the weight of a node is the sum of the weights on its incident edges). If a cluster is found in this search, these nodes will be excluded from the list of potential seeds. In other words, the node with the largest weight that does *not* participate in the clusters found thus far will be selected as the next seed.

    c. **edges** - every edge will be considered once, each yielding a seed consisting of the two endpoints of the edge.

    d. **cliques** - every *maximal* clique of the graph is considered as a seed.

    e. **file(*filename*)** - seeds will be generated from the nodes listed in the given file. Each line must contain a space separated list of node IDs that will be a part of the seed. If a line contains a single * character only, this means that besides the seeds given in the file, every node that is not part of any of the seeds will also be considered as a potential seed on its own.

    f. **single(*node1*,*node2*,...)** - a single seed will be used with the given nodes as members. Node names must be separated by commas or spaces.

g. **stdin** - seeds will be given on the standard input, one by line. Each line must contain a space-separated list of node IDs that will be a part of the seed. It may be useful to use this method together with **–no-merge** if the result of earlier seedings shouldn't interfere the result of the later ones.

Among these parameters, only the **–merge-method** was set to *multi*, to merge highly overlapping complexes. For all other parameters, their corresponding default values were used. We subjected the Hu.PPI network to different values of min.size, starting from 25 to 100, beyond which the number of clusters obtained didn't differ in their size. In order to define optimal size for the clustering, we utilised 6 network centralities and 2 semantic similarity measures.

The rationale for choosing these measures was 2 fold; firstly, incorporating network topology in the analysis allows for a topological perspective on the enriched subnetworks and secondly, the semantic measures to check the quality of clustering. Mean values of all these measures were calculated for every subnetwork. Initially, subnetworks significantly enriched with RNAi hits were filtered. From these, those subnetworks with size identical to their *random-hit-enriched* counterparts were considered for difference of mean test. These *random-hit-enriched* subnetworks were determined in the following manner. We randomly sampled identical number of hits for each virus from the HPIN and used them to compute *hit-enriched* modules in the $2^n d$ step (see 2.1). Alike their *hit-enriched* counterparts, mean values of all the 8 network centrality and semantic similarity measures were computed. These scores were then utilised to determine topologically significant subnetworks, from within *hit-enriched* and *random-hit-enriched* subnetworks, of identical sizes.

For this purpose, we used the Wilcox test to compute significance of mean between these 2 sets, for every network centrality. Subnetwork sizes that yielded significant difference of mean for every network centrality at a significance level of 5%, were considered for further analysis.

As mentioned before, the analysis was carried out for every hit-set, i.e. individually for each virus and also by combining hits for all the 3 viruses. Given the stringency of the criterion, the range of significant subnetwork sizes obtained, were few. Only subnetworks for HIV of size 52 and 66 passed this criterion(see 2.6). In other cases, subnetworks significant for most network centralities were chosen. For instance, in the case of HCV, subnetworks of size 62, 64 were significant for all network measures and Wang semantic similarity of GO.CC class but not for GO.BP and GO.MF. This trend was consistent with combined hits. With the exception of GO.MF, subnetworks of size 46 and 52 were significant for all measures, for the combined hits dataset. For WNV, none of the subnetworks were significant for any of these measures.

The motivation behind this topology-based filtering was that viruses alike other pathogens

Centrality	Hit enriched subnetwork	Random-hit-enriched subnetwork	p-value
Mean-Betweenness	35.279	34.9533	0.0247
Mean-Closeness	0.0092	0.0093	0.0247
Mean-Clust-Coefficient	0.365	0.238	0.0247
Mean-PGR	0.0233	0.0233	NA
Mean-Degree	20.698	21.163	0.0247
Mean-PathLength	2.68	2.66	0.0247
Mean-DiceSim	0.211	0.1882	0.0247
Mean-WangSim-GO.BP	0.221	0.206	0.592
Mean-WangSim-GO.CC	0.489	0.464	0.592
Mean-WangSim-GO.MF	0.385	0.402	0.0497

TABLE 2.5: Wilcox test results for significance of mean between hit-enriched and random-hit-enriched subnetworks. Both subnetworks are of size 43. As evident from the p-values, not all p-values are significant. These are highlighted by grey coloured cells.

target *centrally located* proteins in the HPIN[60, 71, 95]. In order to predict such proteins and more importantly, *hotspots* within the HPIN that harbour such proteins, we performed this test of significance for the 6 centrality measures mentioned above. We discuss the results of these *double-enriched* that is, subnetworks enriched with hits as well as topology, in the **Results** section.

2.2.4 Functional Analysis of filtered subnetworks

All Reactome pathway and GO based enrichments were computed using the Bioconductor packages,*"clusterProfiler"* and *"ReactomePA"*[66, 96–98]. Semantic similarities were computed using the *"GOSemSim"* package[99].

2.3 Novel Hit Prediction

One of the main objectives of this study was to develop a computational basis for prediction of novel host factors for these viruses. Since the subnetworks are already enriched with hits and are topologically significant, we employ a simple neighbourhood search to predict novel host factors. This step is supplemented with multiple data sources to put forth plausible hypothesis of the predicted host factors and their regulation.

2.3.1 Mapping Tissue-specific expression data on filtered subnetworks

As mentioned above, we utilised tissue-specific gene expression data from Human Protein in association with the enriched subnetworks, for novel hit prediction. In particular, we selected cell-lines/tissues, favoured by the viruses in consideration and

TABLE 2.6: p-values of Wilcox test to determine significance of mean values of network centralities and semantic measures for subnetworks.

Sr.No	Centrality	HIV_s66	HIV_s52	HCV_s62	HCV_s64	Combi_s46	Combi_s52
1	BWN	0.00004657	0.01312	0.0001375861	0.000549182	0.013123807	0.013123807
2	CLS	0.00004657	0.01312	0.0001375861	0.000549182	0.013123807	0.013123807
3	CC	0.00004657	0.01312	0.0001375861	0.000549182	0.013123807	0.013123807
4	PGR	0.00004657	0.01312	NaN	NaN	NaN	0.005741142
5	DEG	0.00004657	NA	0.0001375861	0.000549182	0.02107057	0.013123807
6	PL	0.00004657	0.01312	NaN	0.0002379966	0.013123807	0.013123807
7	DICE	0.00004657	0.01312	0.0001375861	0.000549182	0.013123807	0.013123807
8	GO.BP	0.0004038	0.02857	0.4291953	0.6008947	0.02842954	0.02857143
9	GO.CC	0.0004038	0.02857	0.01770607	0.03150926	0.02842954	0.02857143
10	GO.MF	0.0004038	0.02857	0.49324286	0.77134265	1	0.3428571

filtered for genes with moderate/high expression. These filtered genes were overlaid on the subnetworks and the overlaps were further assessed for their role in viral infection/replication. In some cases, the overlaps yielded well-studied proteins like JAK2(Janus Kinase 2) in case of HCV while others were yet uncharacterised. This validated our approach in a way and the latter, thus formed *"Putative Novel Hits"* or *"Predicted Novel Hits"*

2.3.2 Mapping virus-host interactions on filtered subnetworks

We used multiple sources to map virus-host interactions for their overlay on subnetworks. For the HCV subnetworks, we utilised the first experimentally determined viral protein- host protein interaction network by de Chassey et al.[70]. For the HIV subnetworks, we used the HIV-1 NIAID database[53]. For WNV, we utilised the VirHostNet knowledgebase[52]. However, the number of interactions between WNV proteins and human proteins were quite limited (17) and thus, we didn't find any WNV-human protein interactions with any subnetworks. This observation holds true only for the "Combi" subnetworks as WNV hit-set didn't lead to any enriched subnetworks, with the previously described steps.

Chapter 3

Results & Discussion

This chapter highlights the functional characteristics of all the subnetworks. Based on enriched pathways, GO ontologies and their interacting partners, we further postulate novel regulatory mechanisms and putative novel host factors for each virus.

3.1 Functional properties of subnetworks

3.1.1 Enrichment Specificity

From both analyses types (with virus-specific and Combined hit sets), we observed that all subnetworks showed specificity on the level of enriched pathway and GO terms. A general observation from these subnetworks is that their the enriched pathway and GO terms are **"functionally specific"**. In other words, these subnetworks show enriched pathway terms of a specific biological process when compared to enriched terms of their corresponding hit-list. For e.g, The Brass HIV screens' enriched Reactome pathways included; Immune System, HIV infection, Gene Expression and Disease (see Figure 3.3). On the other hand, the enriched Reactome pathways for the HIV-s52 network included; Transcriptional Regulation of White Adipocyte Differentiation,Developmental Biology, Generic Transcription pathway and SMAD2/SMAD3/SMAD4 heterotrimer regulates transcription and Gene expression. We also observed this trend in the enrichment analyses of HCV subnetworks, thus indicating that interpreting these subnetworks over individual hit-lists proves to be more informative.

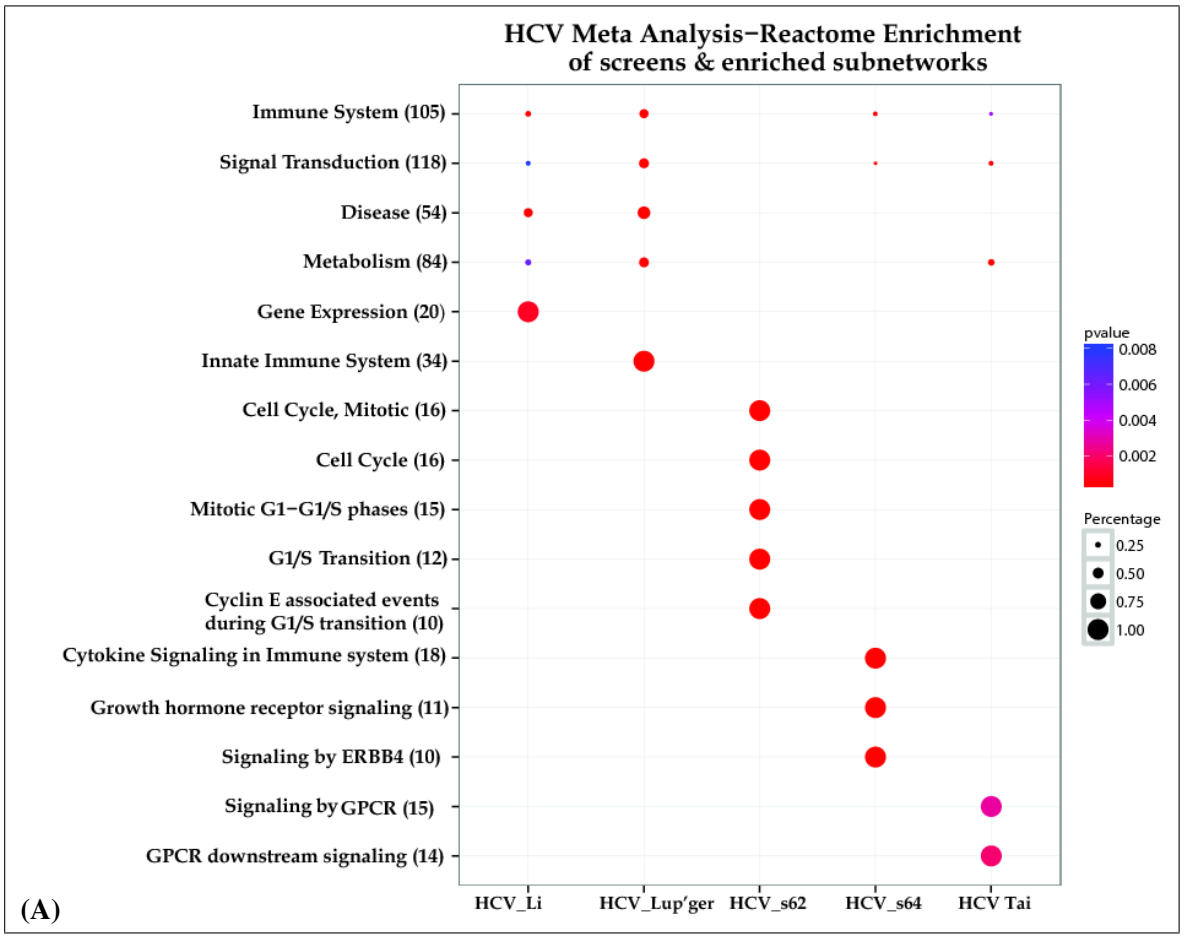


FIGURE 3.1: Reactome Pathway Enrichment: Comparative analysis of enriched Reactome pathway terms in subnetworks and corresponding screens. The size of data points indicate the percentage of genes present in a pathway term whereas color shades indicates p-values. **(A)**-Pathway enrichment comparisons in all RNAi screens vs. **"Combined" subnetworks**

3.2 HIV-1 Meta-analysis Results

For HIV, we identified 2 subnetworks, namely, the HIV_s52 and the HIV_s66 subnetwork. The suffix *s* is short for subnetwork size while the number denotes the numerical size of the subnetwork. Each of these subnetworks emphasise different aspects of HDFs involved in HIV infection. While HIV_s52 subnetwork predominates in transcription, the HIV_s66 subnetwork consists several small complexes that serve different processes during the HIV life-cycle. These are summarised in the enrichment plots of Reactome pathway and GO terms given below. As compared to individual screen hit-lists, the filtered subnetworks have specific biological processes/pathways that are enriched in them. We describe each of these subnetworks in detail, in the following subsections.

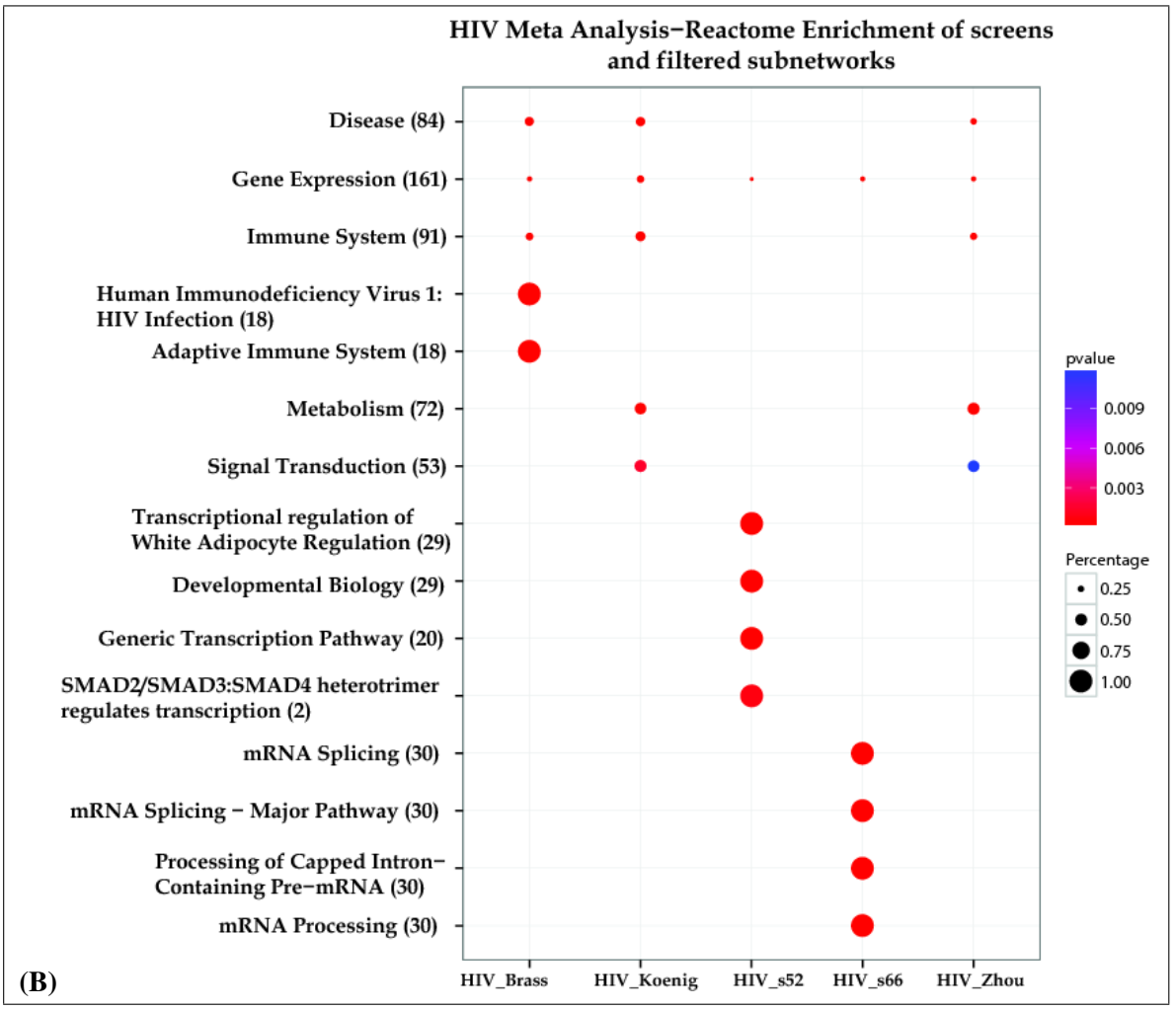


FIGURE 3.2: Reactome Pathway Enrichment: Comparative analysis of enriched Reactome pathway terms in subnetworks and corresponding screens. The size of data points indicate the percentage of genes present in a pathway term whereas color shades indicates p-values. **(B)**-Pathway enrichment comparisons in HIV-1 RNAi screens vs. **HIV-1 subnetworks**

3.2.1 HIV_s52 Subnetwork

3.2.1.1 Mediator Complex

The HIV_s52 subnetwork consisted primarily genes involved in the transcription, where Mediator complex subunits were the most dominant (Appendix-I). The Mediator complex is a large multi-subunit complex consisting of 26 subunits and weighs 1.2 Mda[100]. A unique feature of this complex is that it is present only in eukaryotes as evidence for Mediator-like activity in microbes is lacking. Additionally, the general transcription factors (GTFs) that include TFIIE, TFIIH, and TFIID are also absent in microbes. The parallel co-evolution of these GTFs along with mediator suggest that these complexes function coordinately to regulate expression of protein coding genes. Thus, this complex is a preferred target for viruses. This result is consistent with the fact that being retrovirus, HIV hijacks and employs the host transcription

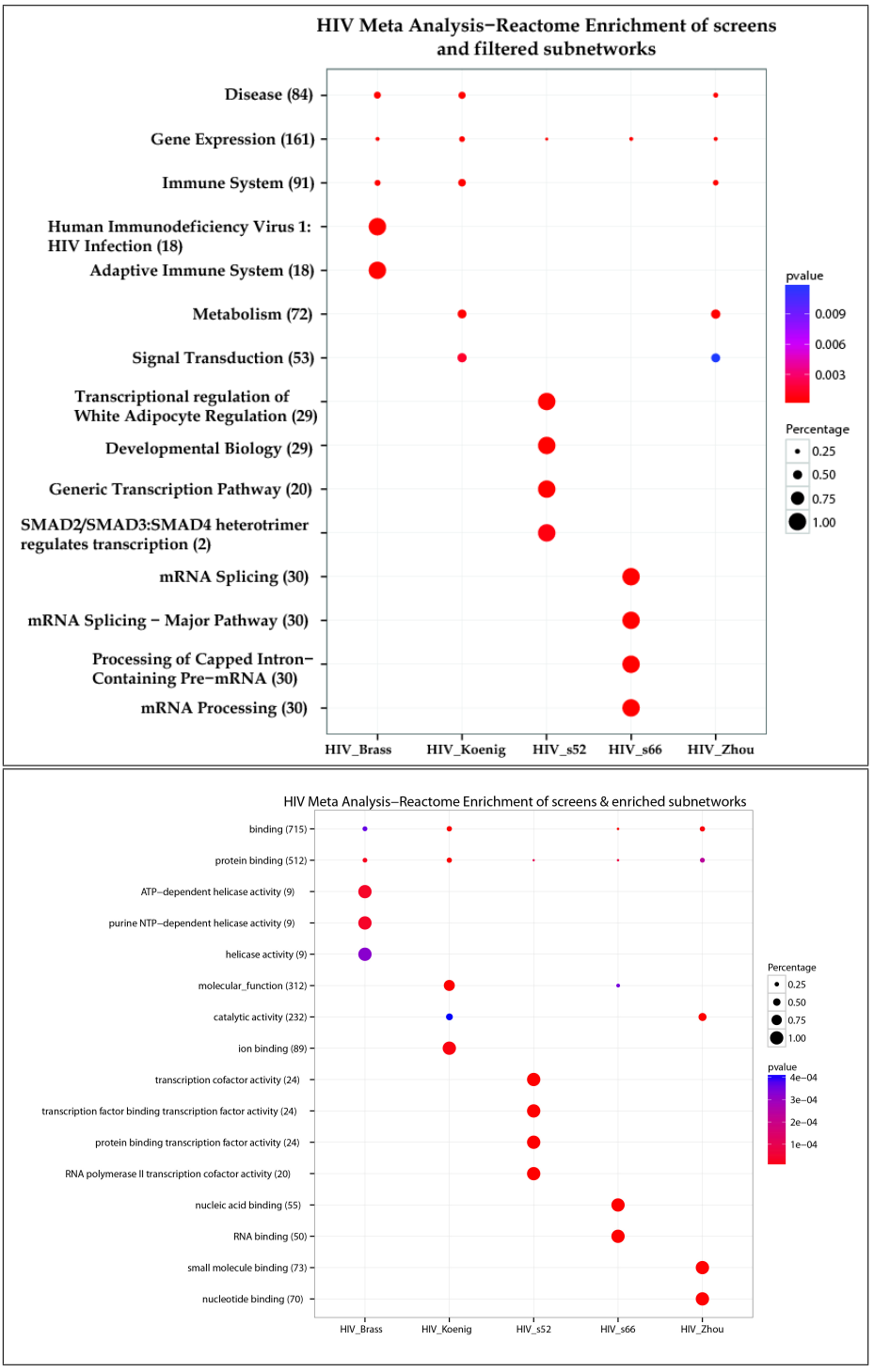
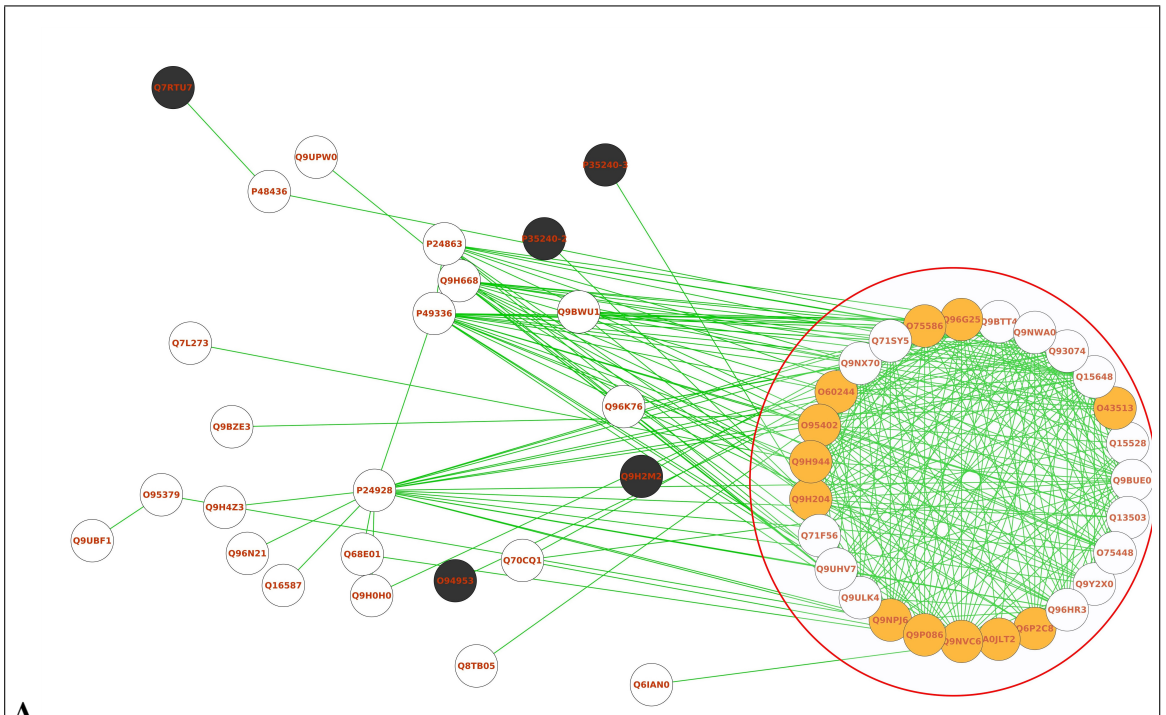


FIGURE 3.3: Reactome pathway and GO enrichment terms comparison between individual HIV-1 RNAi screens and filtered HIV-1 subnetworks, as obtained from our analysis. Both subnetworks show functional specificity at the pathway and GO terms, as compared to individual screen hit-lists.



(a) The HIV_s52 subnetwork.

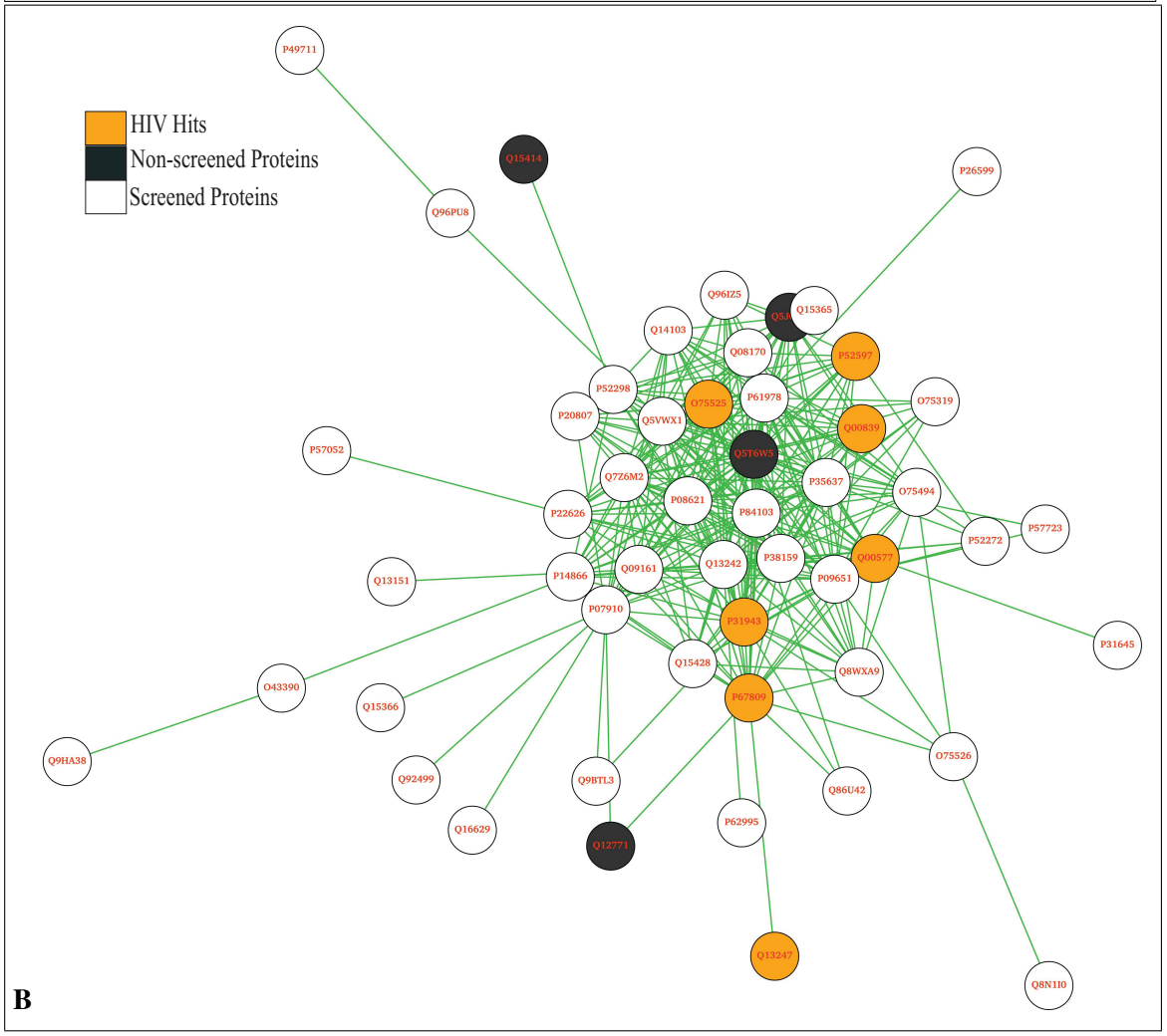


FIGURE 3.4: A) Mediator complex subunits in the HIV_s52 subnetwork, highlighted in the red circle. B) The HIV_s52 subnetwork

machinery for its replication. The Mediator complex subunits were major hits in the meta-analysis by Bushman et al.[9]. The discovery of Mediator complex as a major hit in these RNAi screens[2, 3] have provided different hypotheses of its involvement in HIV infection. Zhou et al. suggest that mediator complex subunits are required for Tat-activated transcription while Koenig et al. suggest that they may be required in reverse transcription.

An important submodule within the Mediator complex is the CDK8 submodule which also contains cyclin C, MED12, and MED13. CDK8 plays a repressive role in transcription wherein CDK-Mediator complex fail to activate transcription while the Mediator complex without CDK8 achieved transcriptional activation[100]. The presence of the CDK8 submodule in our subnetwork implies the repressive mechanism probably being employed by the host cell as a countermeasure against HIV infection. It might also be aided by another protein of this subnetwork, namely Cyclin C, which is known to activate CDK8 and inhibit transcription initiation.

However, the precise and detailed mechanism of how the full Mediator complex is hijacked by HIV is yet to be studied.

3.2.1.2 Host epigenetic mechanisms during HIV infection

In addition to the mediator complex, our subnetwork also included several other interacting proteins (with the Mediator complex subunits), that might shed more light in the details of transcription hijack mechanisms of HIV. For instance, the subnetwork contains a lysine-specific demethylase 4B(KDM4B). This hints to epigenetic regulation in HIV-1 infection. Indeed, recent studies show that methylation at the host and viral protein levels has been a strategic mechanism employed by HIV-1 for inducing latency. Several studies point out to this process [101–104]. Studies from the Karn lab demonstrate that histone lysine methyltransferases (HKMTs) play a specific role in latency of HIV-1 in the Jurkat cells [101]. The authors showed that concurrent knockdown of EZH2, a key component of the Polycomb repressive complex 2 (PRC2) silencing machinery, and the enzyme which is required for trimethyl histone lysine 27 (H3K27me3) synthesis lead to induction of upto 40% of the latent HIV proviruses. In a second study from the same lab, chromatin immunoprecipitation (ChIP) assays confirm that the levels of HIV-1 Tat are restricted in latently infected cells but they increase stepwise during reactivation of provirus [102]. Moreover, ChIP assays of latently infected cells showed that latent proviruses had high levels of deacetylated histones and trimethylated histones. These levels of trimethylated histone H3 and HP1-α associated with HIV proviruses reduced rapidly after tumour necrosis factor alpha(TNF-α) activation. Dimethylation at H3 Lys^9 (H3K9) methyltransferase

G9a[105] and monomethylation at lysine 51 (K51) at the RNA binding domain of Tat maintains viral latency [106]. Another interesting perspective of the involvement of long non-coding RNAs(lncRNAs) in transcriptional regulation including the mediator complexes comes from a recent study by Lai et al. [107]. They show that a special class of lncRNAs, called ncRNA-activating (ncRNA-a), that activate genes through a cis-mediated mechanism[108, 109], require subunits of Mediator complex for their mechanism. Given the highly intricate and complicated intervention by HIV proteins, interaction between mediator subunits could influence the genes regulated by this mechanism. Broadly speaking, this HIV-specific subnetwork is associated with transcription directly/indirectly targeted by HIV.

3.2.2 HIV_s66 Subnetwork

The HIV_s66 subnetwork consists of several functional complexes that are involved in varied biological processes of the HIV life-cycle. Predominant among these are the heterogeneous ribonuclear proteins (hnRNPs), RNA-binding proteins (RBMPs) and Serine/Arginine rich splicing factors (SRPs) which co-ordinately regulate RNA metabolism [110, 111]. Among these, hnRNPs play diverse roles in RNA metabolism; many hnRNPs participate in pre-mRNA processing such as splicing and influence mRNA export, localization, translation, and stability[111]. They are associated with mRNA during its different stages of transport that includes passing through nuclear pores, ribosomes, hence undergoing nuclear-cytoplasmic shuttling[111]. Being a retrovirus, it is thus evident why HIV-1 would target these host protein complexes during its infection course. Alike hnRNPs, the SRPs too have multiple roles in RNA regulation that profoundly include alternative splicing events and post splicing activities like mRNA nuclear export, nonsense-mediated mRNA decay and mRNA translation [112].
Overall, this subnetwork consists of proteins that seem to play a crucial role not only in RNA metabolism but RNA metabolism during HIV-1 infection.

3.2.2.1 Novel hnRNP subunits and their probable role in HIV infection

Our analysis agrees with Bushman et al. and Murali et al., who also revealed hnRNP subunits in their meta-analysis of the 3 HIV screens [9, 43]. Generally, if one or more subunits of a protein complex (here, hnRNPs) are exploited by the pathogen, the whole protein complex itself is considered to be vital. For e.g. If "Proteasome" complex is significant in a genome-wide studies, the interpretation would be that all subunits of the Proteasome contribute in identical fashion towards the disease state, as the complex as a whole is significant and the complex has one function i.e proteolytic degradation. The hnRNPs in our subnetwork may be an exception to this rule. A

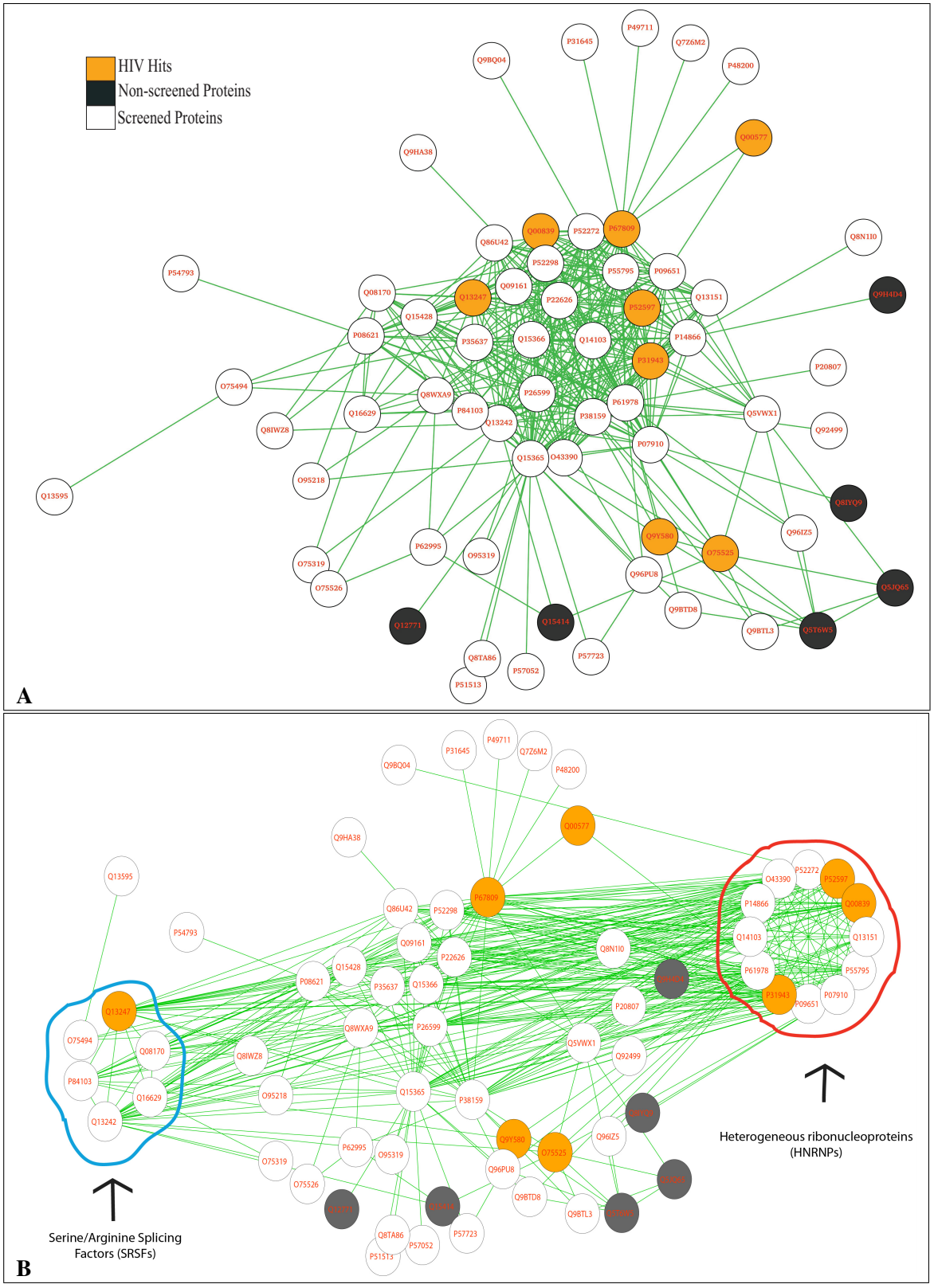


FIGURE 3.5: **A** The HIV_s66 subnetwork. **B** The HIV_s66 subnetwork with highlighted protein complexes; the Heterogeneous ribonucleoproteins (HNRNPs) and the Serine/Arginine splicing factors (SRSF)

recent RNAi screen by Lund et al.[113] showed mechanistic details of hnRNP complex subunits and their varied roles in HIV-1 infection. For e.g. they showed that hnRNP A1 subunit increases the expression of Gag and Env viral proteins but no subsequent increase of HIV-1 RNAs. On the other hand, silencing hnRNP A2 resulted in an increase of Gag protein expression followed by an increase with HIV-1 RNA. Furthermore, different isoforms of hnRNP D also play different roles in HIV-1 infection. Isoform p42, p45 are favourable HIV replication while isoform p37, p40 create a non-permissive state. However, other hnRNP subunits, namely, A2, E1 and E2 exhibit antiviral properties[114, 115]. Thus, hnRNP subunits were generally antiviral in nature but a detailed inspection shows that there are exceptions within the complex itself. For better hypothesis generation, we asked if these hnRNP subunits were expressed in HIV-1 susceptible tissues(see Materials and Methods), by overlaying tissue-specific expression data obtained from the Human Protein Atlas[116]. We found that 8 hnRNP subunits(C1/C2, D0, K, L, M, U, A1, A2/B1) of the subunits were highly expressed in macrophages. From these 8 subunits, 4 subunits (K, L, M, U) have not been characterised yet with respect to HIV-1 infection and thus, are putative novel hits, based on our results.

3.2.2.2 RNA-binding motif-proteins & Splicing Factors

This subnetwork also consisted of several RNA-binding proteins (RBM11, RBM41, RBM42, RBM4B, RBM7) that may have hitherto unexplored, specific roles in mRNA export during HIV replication. Cellular RNAs are bound to RNA-binding proteins (RBPs) to form ribonucleoprotein (RNP) complexes. These RBPs govern the structure and function of RNAs and as such, play a vital role in their biogenesis, stability, function, transport and cellular localization [117]. They are found in huge numbers in humans (> 500) and due to their involvement in so many aspects of RNA regulation, any disruption in their activity leads to complicated diseases [118]. Their dynamic interactions with RNAs in their coding, untranslated and non-protein-coding RNAs that comprise the RNPs allows the RBPs to bound stably to the RNA throughout its journey from synthesis to degradation[118]. Given such broad array of functions of the RBPs, it is likely that HIV utilises the functions of these RBM proteins for its own RNA metabolism.

The subnetwork further consists of several Serine/Arginine splicing factors (SRSF3, SRSF4, SRSF6, SRSF7, SRSF9, SRSF10) of which 3 (SRSF6, SRSF7, SRSF9) have direct interactions with viral proteins, as listed in the HIV-1 NIAID database[53] (see Table 3.1). It is clear from the table that despite belonging to the same functional class of proteins. Moreover, different viral proteins have different type of interactions with these proteins. The involvement of a splicing factor ASF/SF-2, which is a member of

Sr.No	Viral Protein	Host Protein	Interaction
1	Gag,Pr55	SRSF6	SR proteins, particularly SRp40 and SRp55, increase HIV-1 Gag translation from unspliced viral RNA. The second RNA recognition motif and the arginine-serine (RS) domain are determinants of SR protein activity
2	capsid	SRSF6	SRp55 induces production of extracellular p24gag from a rev-defective HIV-1 provirus
3	Tat,p14	SRSF7	HIV-1 Tat synergizes with type I activators, such as Sp1 and CTF, to enhance transcript elongation and exon skipping, suggesting Tat function leads to the inhibition of splicing factors SF2/ASF and 9G8
4	Vif,p23	SRSF4	A novel exonic splicing enhancer (ESE) element within the 5'-proximal region of HIV-1 mRNA exon 2 facilitates both exon inclusion and Vif expression. This ESE binds specifically to the cellular SR protein SRp75

TABLE 3.1: HIV-1 protein interactions with SRSF proteins of the HIV_s66 subnetwork, as extracted from the HIV-1 Human protein interactions database.

the SR (serine/arginine-rich) family of splicing factors, in HIV-1 infection was identified by a proteomic approach by Berro et al.[119]. Berro et al. also found p32 and other proteins preferentially bound to AcTat(acetylated Tat). p32 was recruited by the HIV genome, that suggest a mechanism by which Tat acetylation may inhibit HIV-1 splicing needed for the production of full length transcripts. Indeed, SR proteins has essential functions during spliceosome assembly and they interact with RNA regulatory sequences on the pre-mRNA as well as with multiple cofactors. Collectively, they recognise multiple splice sites with broad specificity and are at the core of regulation pathways that lead to the choice of alternative splice sites [120]. To that end, SRSF3, SRSF4 and SRSF10 might be involved in similar host-virus protein interactions with varied yet specific roles, just like the above mentioned members of the same family.

Besides these functional complexes, this subnetwork included several other proteins (See Appendix-I) that might serve an associative role in the above mentioned processes.

3.3 HCV Meta-analysis Results

Similar to the HIV subnetworks, we obtained 2 HCV subnetworks of different sizes; HCV_s43 and HCV_s66. These networks were also consisted of unique functional complexes that denote different aspects of cellular hijack during HCV infection.

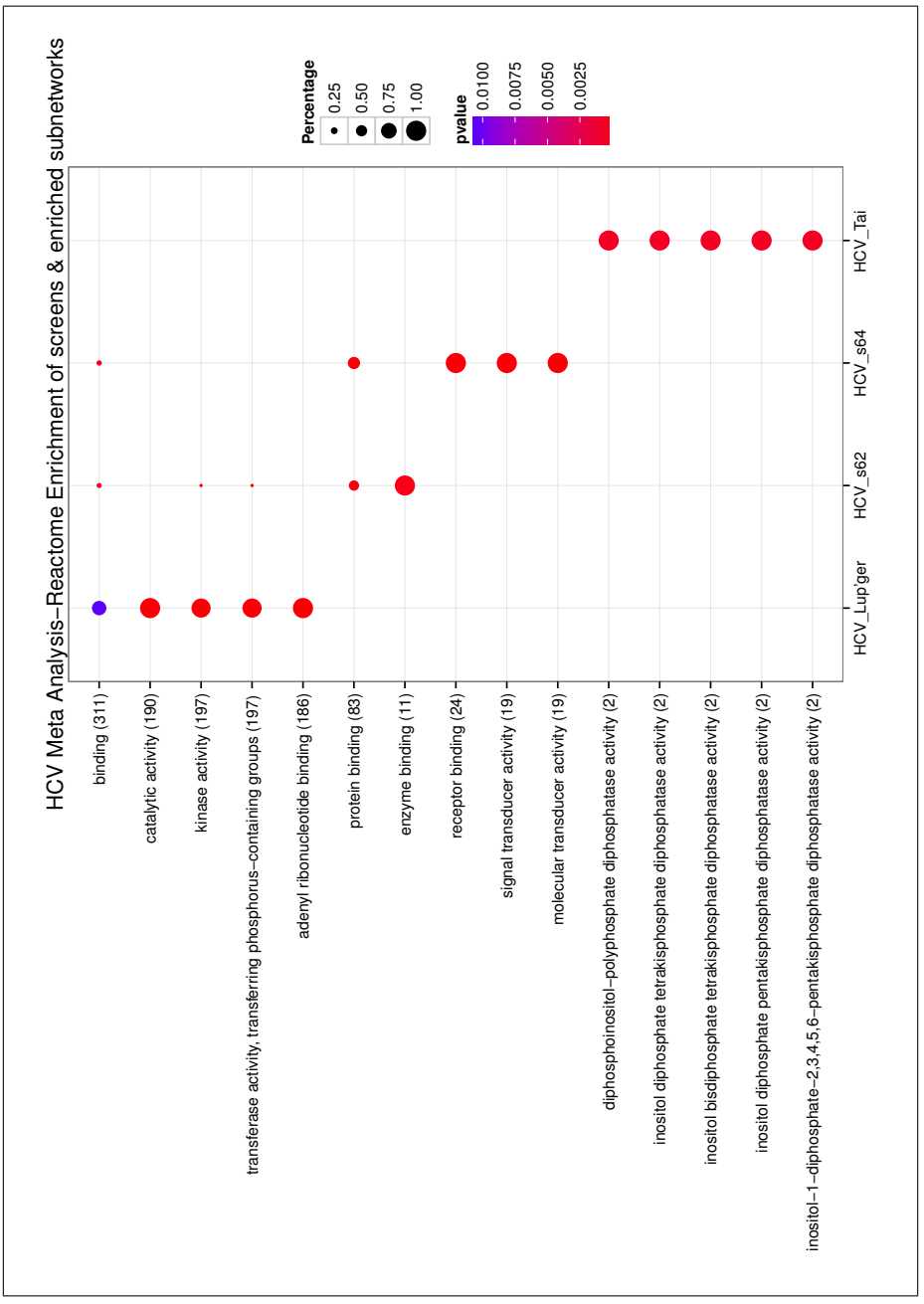


FIGURE 3.6: Reactome pathway and GO enrichment terms comparison between individual HCV-1 RNAi screens and filtered HCV-1 subnetworks, as obtained from our analysis. Both subnetworks show functional specificity at the pathway and GO terms, as compared to individual screen hit-lists.

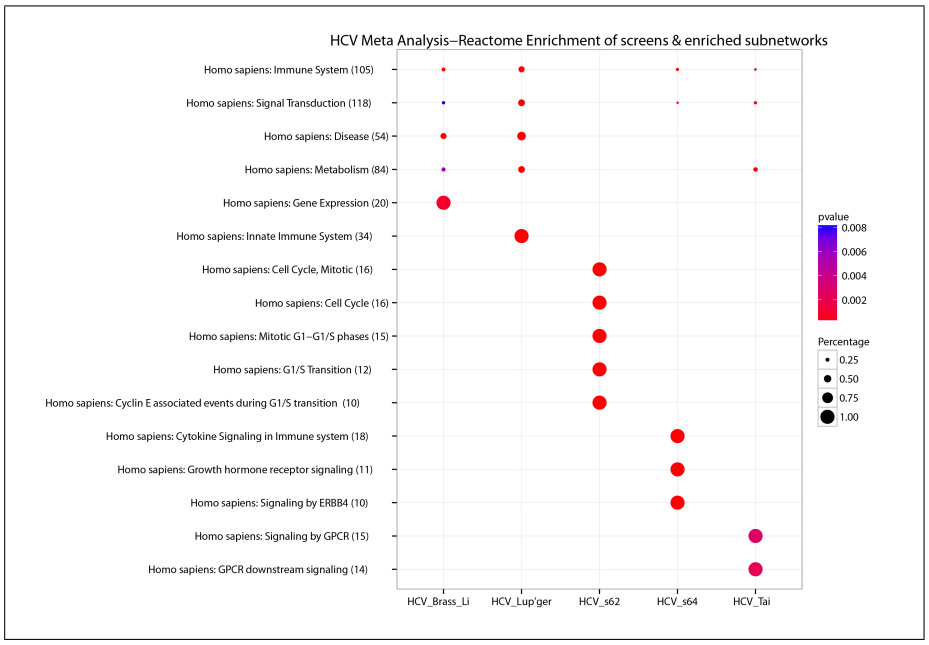


FIGURE 3.7: Reactome pathway and GO enrichment terms comparison between individual HCV-1 RNAi screens and filtered HCV-1 subnetworks, as obtained from our analysis. Both subnetworks show functional specificity at the pathway and GO terms, as compared to individual screen hit-lists.

3.3.1 HCV_s43 Subnetwork

The HCV_s43 subnetwork primarily consisted of proteins from these functional classes; Crystallin Proteins (Alpha, Beta & Gamma) Dual specificity protein phosphatase (Subunits 1, 2, 4, 6, 8-10, 16), Mitogen-activated protein kinases (MAPK-11, 12, 14) and MAP kinase-activated protein kinase (MAPKAPK-2, 3, 5). A common aspect of each of the above mentioned proteins, is that their are all associated with hepatocellular carcinoma, the end-stage of HCV infection. Each of these proteins had their unique role in the HCV life-cycle in the host cells, which we describe in the following subsections.

3.3.1.1 Intricacies of MAPK signaling during HCV infection

From the protein annotations, it is clear that this subnetwork consists of components of the Mitogen activated protein kinase (MAPK) signaling cascade (see figure 3.9,3.10). Mitogen activated protein kinases or MAPKs play an important role in multiple cellular functions, prominently cell growth and proliferation[121, 122]. MAPKs can be activated in a variety of ways; by hormones (e.g.,insulin), growth factors (e.g.,platelet-derived growth factor (PDGF), epidermal growth factor (EGF) and fibroblast growth factor (FGF) inflammatory cytokines of tumour necrosis factor (TNF) family and environmental stresses such as radiation,osmotic shock and ischemic injury[123]. With such important association of cellular processes with it, the MAPK signaling pathway is also a preferred target by HCV. Macdonald et al.[124] showed how NS5A inhibited the activity of the mitogenic and stress-activated transcription factor activating protein-1 (AP1). They showed that this inhibition is dependent upon a class II polyproline motif within NS5A. By combining dominant active and negative mutants of components of the MAPK signaling pathways, selective inhibitors, together with immunoblotting with phospho-specific and phosphorylation-independent antibodies, Macdonald et al. found that this inhibition is mediated via the ERK signaling pathway. This observation was consistent in both stable NS5A-expressing cells and Huh-7-derived cells harbouring subgenomic hepatitis C virus (HCV) replicons.

Ndjomou et al.[125] further showed that MEK/ERK signaling pathway, of which MAPKs are major players, act as a second line of defence against the virus, parallel to IFN-α signaling. They revealed how MEK/ERK inhibitors and negative mutants enhance HCV replication, RNA and protein syntheses. They also suggest that since inhibition of MEK signaling leads to up-regulation of HCV replication, usage of MEK inhibitors in treatment of hepatocellular carcinoma, induced in later stages of HCV infection, should be carefully monitored.

The downstream components of the ERK signaling cascade consists of ribosomal protein S6 kinases, two alpha subunits of these kinases are present in this subnetwork (RPS6KA4, RPS6KA5). Ribosomal protein S6 (rpS6) is a component of the 40s ribosome. Phosphorylation signals mitigated through mitogens from this cascade, at several Serine residues of the ribosomal protein S6 (rpS6), leads to translation initiation at the 7-methylguanosine cap complex[126]. However, rpS6 also requires phosphorylation at specific Serine residues, $Ser_{235/236}$, by a specific kinase, namely, ribosomal S6 kinase (RSK). Thus, a dual phosphorylation signal through the RAS/ERK signaling and the RSK is required for rpS6 function and thus, translation initiation. Roux et al.[126] also note that this mechanism is independent of the mammalian target of rapamycin (mTOR) pathway that modulate translation initiation. Intriguingly, the HCV non-structural protein NS5A, up-regulates host cap-dependent translation machinery

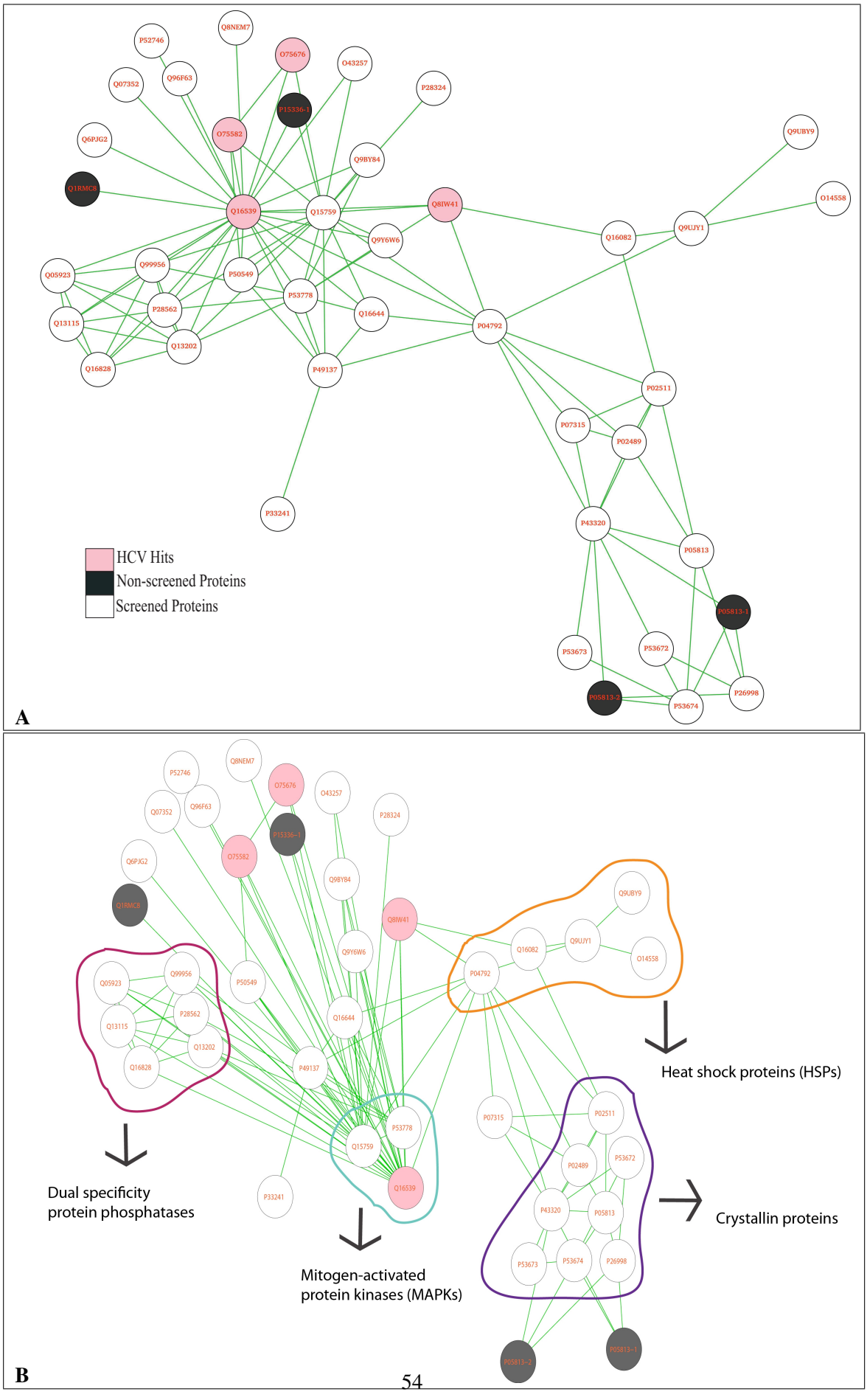


FIGURE 3.8: **A** The HCV_s43 subnetwork.**B** The HCV_s43 subnetwork and its protein complexes.

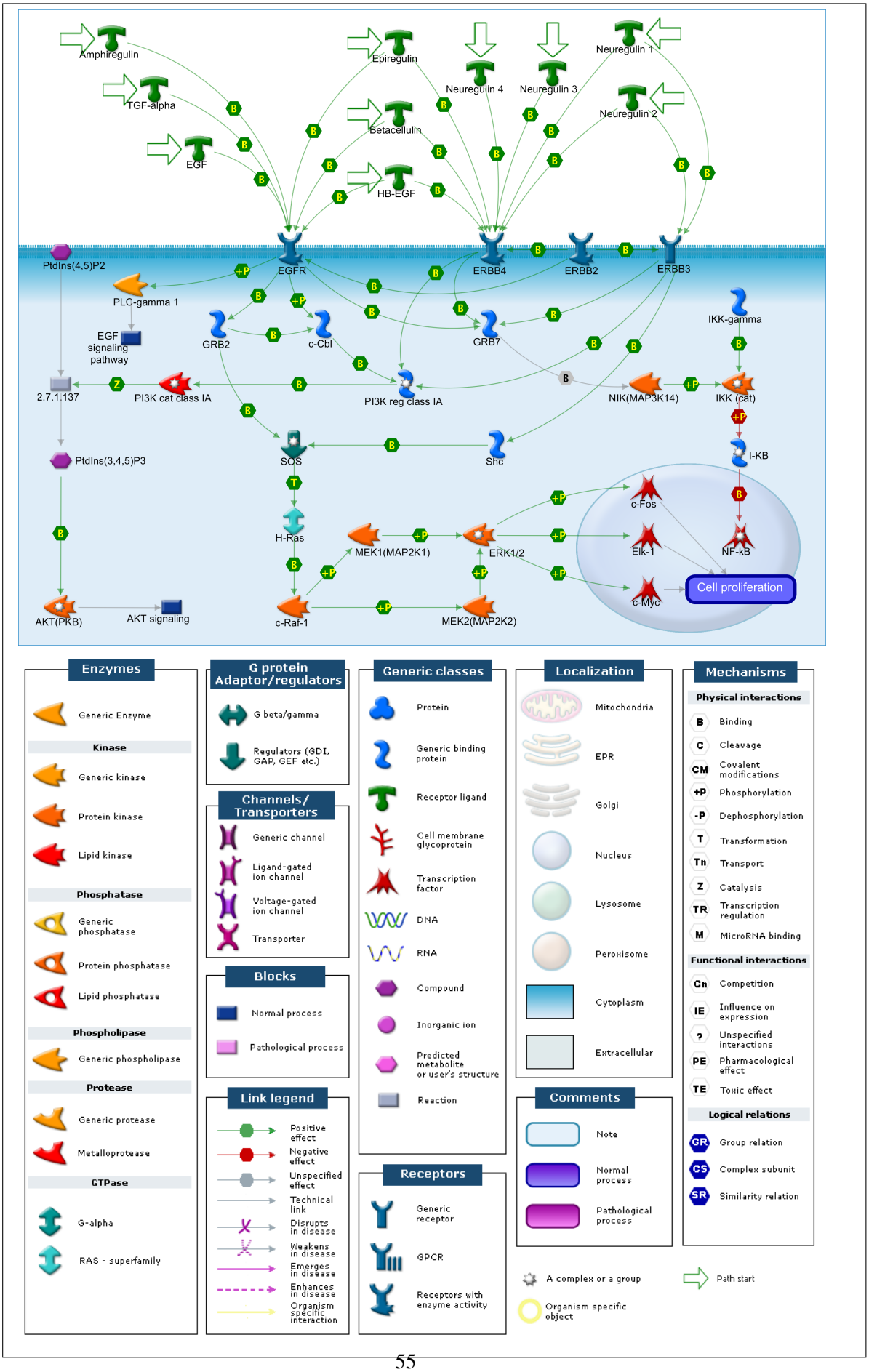


FIGURE 3.9: MAPK and ERK12 signaling from Metacore pathway maps© in uninfected cells.

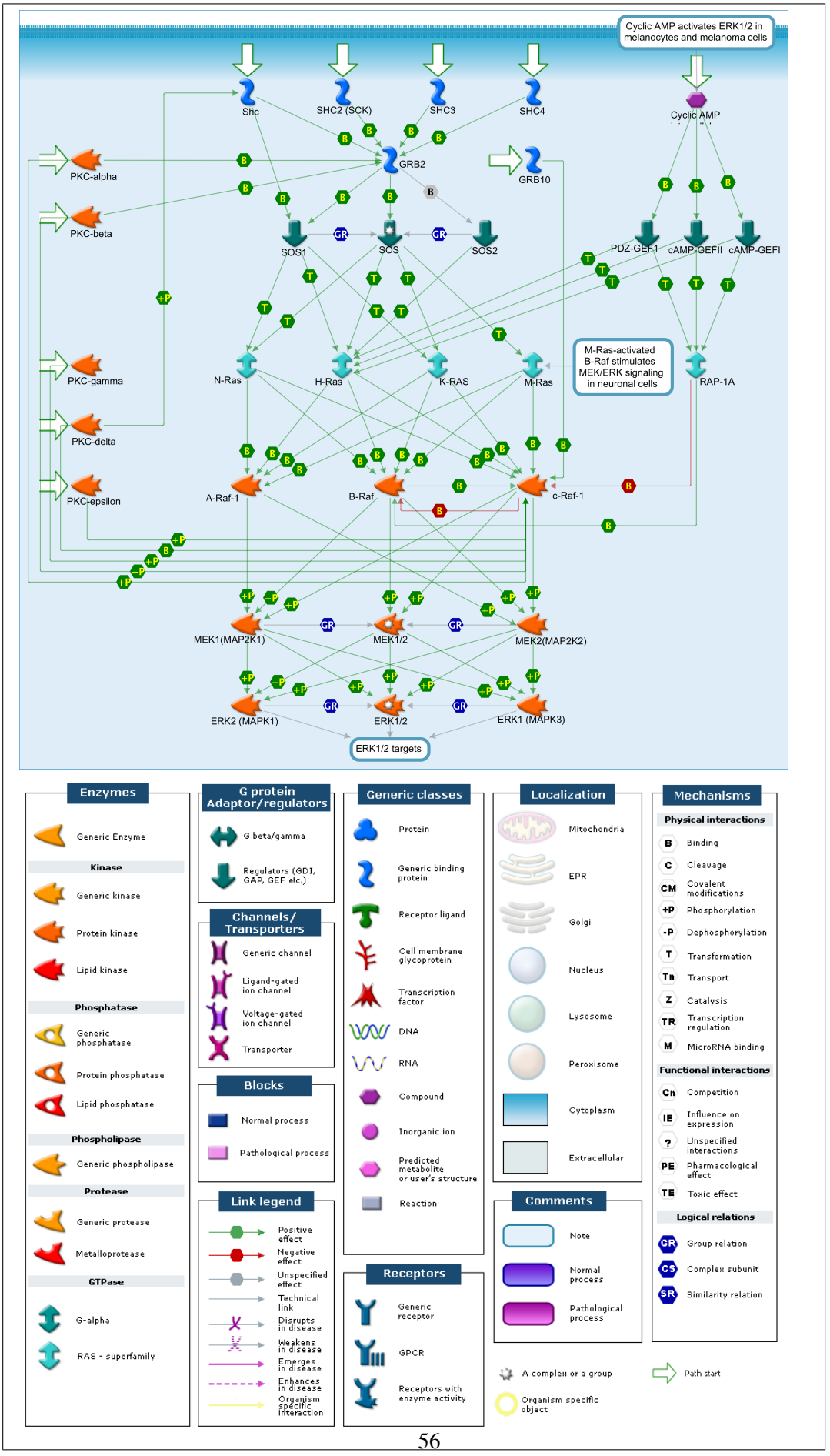


FIGURE 3.10: MAPK and ERK12 signaling from Metacore pathway maps© in uninfected cells.

in Huh7.5 cells, by simultaneous activation of mTORC1 and eukaryotic translation initiation factor 4E (eIF4E)[127]. Furthermore, NS5A phosphorylated eIF4E through the MAPK-MNK pathway. These 2 studies, in the context of our results, suggest 2 mechanisms of translation during HCV infection;

a) According to Roux et al., since rPS6 kinase mediated translation initiation is independent of mTOR-based modulation, it is likely that cellular translation is *redirected* during HCV infection, as mTOR-based translational initiation pathway is hijacked by HCV.

b) On the other hand, George et al. show a rather contrasting result, that involves rPS6 kinase itself in mTOR-based translation initiation and this process as up-regulated via NS5A interaction with 4E-binding protein (4EBP). Moreover, they also suggest a mTORC2 based activation of translational initiation, again via NS5A.

The HCV_s43 subnetwork consisted of 3 MAPKs, 3 MAPKAPKs, rPS6-kinase-α-4 and rPS6-kinase-α-5 subunit. It can be intriguing to explore which of the above routes of translation initiation occurs during HCV replication.

3.3.1.2 HCV induced stress induces molecular chaperones

This subnetwork further consisted of heat shock proteins (HSPs) and crystallin complexes, both act as *chaperones* or proteins that protect, fold or unfold protein substrates in a context-dependent manner[128–130]. Intuitively, this implies the induction of a defence mechanism of the cell, responding to the stress initiated from the viral replication. A couple of independent studies, indeed identified and described how the HCV viral protein, NS5A modulates Hsp72 levels. Chen et al. identified interacting proteins of NS5A by tandem affinity purification (TAP) from cells expressing NS5A and from mass spectrometry which included Hsp72[131]. Its association with the HCV replicase complex, that comprises of NS5A, NS5B and NS3, Hsp72 plays a positive regulatory role in HCV RNA replication as it increases levels of the replication complex. This was supported by increased stability of the viral proteins in the complex or to the enhanced translational activity of the internal ribosome entry site of HCV, both constitutive functions of HSPs. In the second study that identified Hsp72 and its interaction with NS5A, Lim et al. showed that NS5A increases levels of Hsp72 through 2 transcription factors, HSF1 and NFAT5[132]. Silencing the expression of the former, reduces HCV replication and viral release. Our subnetwork consists of several HSPs (see Appendix I) that may have similar function as Hsp72 in HCV replication.

A second class of HSPs are the crystallin proteins[129, 130].A recent study by Huang et al. shows that a second route of modulation occurs through the $\alpha\beta$-Crystallin complexes which subsequently promotes hepatocellular carcinoma (HCC) progression in vivo and in vitro[133]. $\alpha\beta$-Crystallin forms a complex with $14-3-3\zeta$

protein and also elevates its basal concentration. This leads to an up-regulation in ERK1/2 activity and accelerates HCC progression. Furthermore, overexpression of $\alpha\beta$-Crystallin impairs Sorafenib treatment, a multi-kinase inhibitor which is widely used in HCC treatment. HCC is triggered in patients at later stages of HCV infection and thus, $\alpha\beta$-Crystallin complexes are potential targets for reducing HCC progression and improve survival rates in patients. Our subnetwork also consisted several subunits of $\alpha\beta$-Crystallin (CRYBAA, CRYBAB, CRYBA1, CRYBA2, CRYBA4, CRYBA1, CRYBB1, CRYBB2, CRYBB3) which could play similar role as described in this study.

Either way, the presence of these kinase cascades in this subnetwork warrants a further detailed study of this cascade, in light of these studies. This specific subnetwork thus highlights the intricacies in cellular signaling during a viral infection.

3.3.2 HCV_s64 Subnetwork

The HCV_s64 subnetwork consisted functionally diverse set of proteins that included Interleukins, Insulin receptor substrates, Suppressor of cytokine signaling (SOCS) and multiple types of Tyrosine-protein phosphatase non-receptors. Being so diverse in function, the biological interpretations of this subnetwork were interesting in their own right.

3.3.2.1 Role of Interleukins in HCV therapies

Interleukins form an integral component of the cellular immune response. By definition, they are secreted proteins that bind to specific receptors and play a role in leukocyte communication. With such attributed function in cellular immunity, their role in host defence against infection, viral infection in particular, is of considerable importance.

This subnetwork predominantly included members of the IL-12 family of interleukins (See Appendix-I) [134]. IL-12 induces the production of IFN$-\gamma$ through T_H1 and NK cells and thus mediates development and maintenance of T_H1 cells. IFN$-\gamma$ itself is a potent inhibitor of HCV replication[135, 136]. In HCV infection and replication, interleukins play a key role in suppression of infection. For e.g. In patients suffering from chronic hepatitis C, IL-12 enhances cytokine production by PBMCs in some patients[137]. Furthermore, T cell immunoglobulin mucin domain protein 3 (Tim-3) expression has a direct effect on IL-12/IL-23 in human CD14 $^+$monocytes in patients with chronic HCV infection[138]. In these patients, Tim-3 are highly expressed and IL-17 is upregulated in $CD4^+$ T cells. Blocking Tim-3 signaling restores normal regulation of IL-12/IL-23 through STAT3 signaling and reduces the IL-17 levels both *ex vivo* and *in vitro*.

Additional resistance to IL-12 and its antiviral properties comes through the virus itself. Particularly, HCV core-gC1qR interaction with the cell surface of monocyte/-macrophages inhibits the production of IL-12p70 upon lipopolysaccharide stimulation. This response is specific to IL-12 as other interleukins, namely IL-6, IL-8 and IL-1β production were unaffected[139].Other interleukins, such as IL-10, whose blocking leads to a favourable balance of $CD4^+$ cells, that enhances their proliferation in response to HCV antigens[140]. Certain interleukins, like IL-27, have antiviral properties for multiple viruses. IL-27, which inhibits HIV-1 replication in $CD4^+$ cells, also inhibits HCV replication in HuH7.5 cell line. This happens via IFN-α like antiviral response which includes induction of antiviral genes without IFN-α induction[141]. On the other hand, a recent study suggests incorporation of interleukins in IFN-α therapy. For instance, a recombinant form of human interleukin 28B (rhIL28B) in combination with IFN-α inhibited HCV production in culture[142]. Inclusive or not,

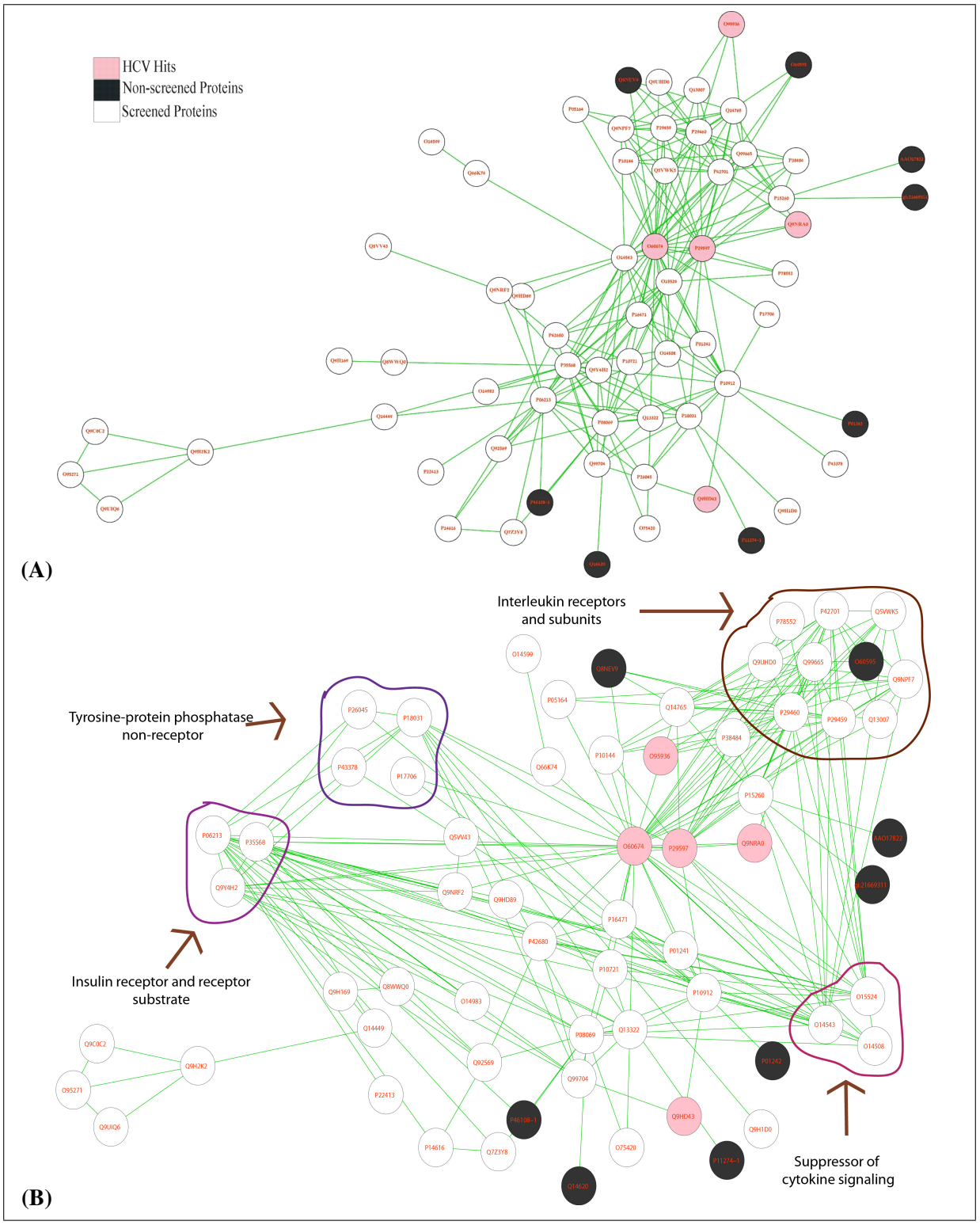


FIGURE 3.11: **(A)** HCV_s64 subnetwork **(B)** HCV_s64 with its protein complexes that include; Interleukin receptors and subunits, Suppressor of cytokine signaling, Insulin receptor with receptor substrates and Tyrosine protein phosphatase non-receptor

different family members of interleukins play an important role in eliciting an antiviral response that can be utilised in different combinations, depending upon the immunological repertoire of the patient and thus can lead to better therapeutics against HCV.

3.3.2.2 Insulin resistance and HCV

Insulin resistance is an outcome of chronic HCV infection in many patients[143–146]. Insulin resistance primarily affects IFN-α treatment and its associated with major pathway modulations. For instance, overexpression of the suppressor of cytokine signaling 3 (SOCS3) in liver tissue results in poor treatment outcome in patients with chronic hepatitis C viral genotype 1[143]. Since, antiviral therapy response is lower in patients with genotype-1 than those with genotype-2 of the virus, SOCS3 signaling should be monitored and lowered to basal levels in the former type of patients. This holds true for SOCS-1, where its silencing enhances IFN-signaling[147]. Contrary to this, SOCS-2 overexpression inhibits HCV replication in presence of a plant glycoside, Saponin which directly increases SOCS-2[148]. Thus, SOCS signaling may interfere with therapeutic IFN signaling for HCV but some SOCS genes function as antivirals when modulated by small molecules. This subnetwork includes all the three SOCS genes described above and thus the subnetwork also points to SOCS as an important component in affected signaling pathways during HCV.

Another level of insulin resistance occurs through the degradation of insulin receptor substrate 1(IRS-1) in human hepatoma cells (HuH-7) expressing core protein of HCV genotypes 3a and 1b[144]. This is also associated with upregulation of SOCS7 and downregulation of peroxisome proliferator-activated receptor γ (PPARγ). The second insulin receptor substrate (IRS-2), the knockdown of which induces insulin resistance[149]. Similar to IRS-1, the silencing of IRS-2 suppresses IFN-α response. This increases protein-tyrosine phosphatase (PTP) protein-tyrosine phosphatase 1B (PTP-1B) activity, which when silenced with Metformin improves the IFN-α response. In summary, SOCS signaling (through various SOCS family members), Insulin signaling and their regulation by PTPs form a crucial factor in determining the efficacy of IFN-α treatment.

This subnetwork consists of 4 PTPs (including PTP-1B described above, for others see Appendix-I), SOCS-1, SOCS-2, SOCS-3 described above) which could be attributed to similar roles in insulin resistance and IFN-α responses. Overall, this subnetwork shows tendency towards insulin resistance and how HCV modulates through various signaling pathways. It remains to be seen how these highly similar proteins, to the ones investigated in these aspects of HCV infection, play a role in the HCV life cycle and disease progression.

3.4 HIV, HCV & WNV Combined, Meta-analysis Results

3.4.1 Combi_s52 Subnetwork

This network is identical to the HIV_s52 subnetwork (see HIV _s52 Subnetwork 3.4).

3.4.2 Combi_s239 Subnetwork

The Combi _s239 Subnetwork is the largest subnetwork found in our analysis and as such contains many complexes that play specific roles in viral life cycle (see table A.7). We further split this section on virus specific interpretations, as the number of complexes found in this subnetwork point to some differential as well as common aspects of virus infection.

3.4.2.1 Src family of kinases and their role in virus infection

The Src family of kinases (SFKs) have different roles in the life cycles of all the 3 viruses in this study. A common observation is that SFKs are activated or upregulated in virus infected cells, although the effects of their activation differs between virus species. For instance, the SFK, c-Yes during WNV infection shows an increase in its expression[150]. This is followed by a decrease in 2-4 log decrease in viral titers which indicates that c-Yes activity is required for WNV replication. Inhibition by chemical inhibition (PP2) and by RNAi mediated silencing showed similar results. However, PP2 didn't reduce intracellular levels of either viral RNA or protein implying that the drug doesn't influence early stages of replication. Enzymatic digestion of viral envelope glycoprotein E by endoglycosidase H (endoH) revealed that E doesn't localise beyond the endoplasmic reticulum (ER). Electron microscopy of PP2-treated WNV-infected cells revealed that WNV virions accumulated within ER compartments as compared to their control counterparts. Thus, c-Yes inhibition didn't interfere with virus assembly but restricted transition of virions through the secretory pathway and it is an important problem in WNV assembly.
In case of HIV, SFKs are differentially activated which is achieved by Nef protein and in the presence of SFK-negative regulatory kinase Csk[151]. Expression of Fgr, Fyn, Hck, Lck, Lyn, and Yes as well as c-Src in yeast revealed that each kinase was active and induced growth expression in yeast. Co-expression of these SFKs with Nef showed that Nef strongly activated Hck, Lyn, and c-Src but didn't significantly alter expression levels of Fgr, Fyn, Lck, or Yes. Nef contains the P*XX*P motif essential for SH3 domain binding, mutagenesis of which reduced the effect of Nef on Hck, Lyn, and c-Src. This occurs due to allosteric displacement of intra-molecular SH3-linker interactions. Thus these selectively activated SFKs in HIV infected cells can serve as potential targets for HIV antivirals.

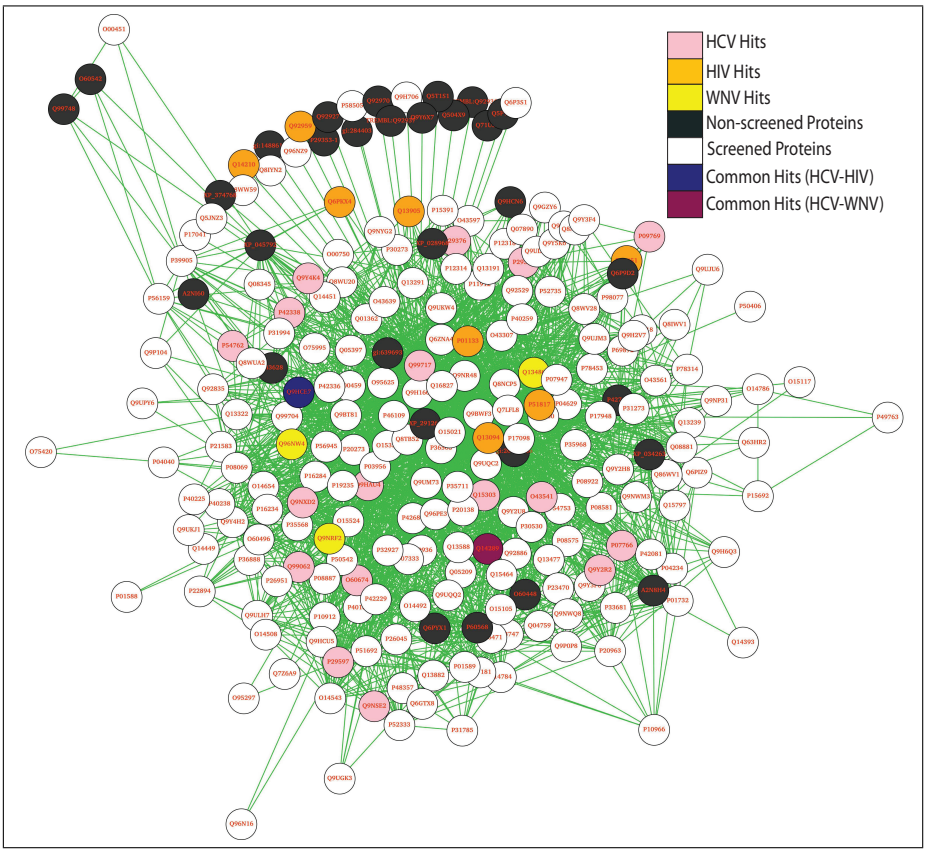


FIGURE 3.12: Combi_s239 subnetwork

Alike HIV, HCV also recruits more than one SFK at various stages of its life-cycle. Specifically, the nonstructural proteins, NS5A interacts with many SFKs through their respective SH3 domains and induces aberrant phosphorylation events[152–154]. Along with these proteins, the Combi_s239 contained several interleukin subunits, receptor tyrosine-protein phosphatases, receptor tyrosine protein kinases; their biology and predicted mechanisms have been discussed in subsections above.

3.5 Putative Novel Hits: Mapping tissue-specific expression data

One of the complications in secondary screening of RNAi screen hits, is the selection of *"interesting"* candidate genes. This is partly because in genome-wide screens, the number of false-positives is large[155, 156]. Secondly, the novel host factor has to have therapeutic potential and thus, tissue specificity plays an important role.We utilised the Human Protein Atlas dataset[116] in conjunction with subnetworks to propose novel host factors and regulators mechanisms. Specifically, we used moderate-highly expressing genes from cell-lines/tissues which were attacked by the viruses in consideration (For e.g. Hepatocytes in case of HCV, macrophages for HIV). These genes were then overlaid on virus-specific subnetwork to predict novel host factors and/or mechanisms.

For instance, the intersection between moderate-highly expressed proteins in hepatocytes and the HCV_s64 subnetwork yielded three proteins, namely, Tankyrase-1 (TNKS1), Sarcoplasmic/endoplasmic reticulum calcium ATPase1 (SERCA1) and Tyrosine-protein kinase (JAK-2), of which TNKS1 and SERCA1 are non-hits (weren't identified as hits in any HCV screen). HCV core protein induces endoplasmic reticulum(ER) stress and deregulates cellular apoptosis by modifying the calcium signaling pathway. Sarcoplasmic/endoplasmic reticulum calcium ATPase 2 (SERCA2), an ATPase of the same class as SERCA1, shows an impairment in its Ca2+ pumping leading to deregulated calcium uptake from the cytosol to the ER during the HCV-Core expression. As SERCA1 belongs to the same enzyme family as SERCA2, it would be interesting to verify SERCA1's role in similar conditions. HCV Core protein also modifies Janus kinase-(JAK)-signal transducer and activator transcription factor (STAT) pathway after stimulation from interleukin-6 (IL6) and interferon (IFN)-γ [157]. However, HCV core modifies this signaling cascade in different ways; it increased phosphorylation of JAK1-2, STAT1 and STAT1 mediated transcription under IFN-γ stimulation while JAK1-2, STAT3 and STAT3-mediated transcription were prevented under IL-6 stimulation. Moreover, HCV nonstructural proteins also induce oxidative stress via STAT3 activation, which is of a constitutive type in HCV replicon-expressing cells[158]. Amongst these proteins, is Tankyrase-1 which is a Poly-ADP-ribosyltransferase and its probable mechanisms in relation to HCV infection, we describe in the subsequent sections.

In the case of HIV, we observed that heterogeneous ribonuclear proteins (hnRNPs) were expressed in moderate-high levels in macrophages and were also part of the HIV_s66 subnetwork. As outlined in the subsection above (see figure 3.5), hnRNP subunits can play diverse roles in HIV-1 replication and infection processes and thus, need to be tested individually for understanding their involvement. Broadly, hnRNPs

are involved in mRNA metabolism that includes RNA export, transport among others, a stage-wise inhibition of these subunits can reduce viral replication in a cell.

3.6 Putative Novel Hits: Mapping CORUM protein complexes

Majortiy of graph based clustering algorithms, focused on protein complex prediction, utilise a gold-standard dataset of protein complexes to test biological relevance of their clustering[73, 159–161]. We use the human CORUM complexes in a similar fashion to overlay on these subnetworks to determine putative novel hits and mechanisms relating to the viral infection[76]. We computed the Jaccard index of each complex with each subnetwork and overlaid the complexes with Jaccard index > 0 . As ClusterOne allowed overlaps between complexes, all subnetworks contained overlapping complexes. For instance, The HCV-s64 subnetwork contained 3 complexes, viz. TRF1 telomere length regulation complex, TRF-Rap1 complex I, 2MD and Tankyrin 1-tankyrin 2-TRF1 complex (see figure 3.13). Interestingly, TNKS1 belonged to all three complexes. All these complexes were related to telomere length regulation and telomere dynamics. This implies that TNKS1 has a probable link in telomere length regulation and HCV infection.

Another example is the HIV_s66 subnetwork which included 40 CORUM complexes, majority of those implicating RNA splicing (see table A.8). Many protein from the subnetwork are heterogeneous nuclear ribonucleoprotein (hnRNP) which also belonged to multiple overlaid CORUM complexes. How do hnRNPs and associated splicing events play a role in HIV infection? Using these cues from these overlaid complexes, we propose probable mechanisms in the following subsections.

3.6.1 Alternative therapies: Tankyrase & HCV induced Hepatocellular Carcinoma(HCC)

Tankyrase 1(TANK1) is a human telomeric poly(ADP-ribose) polymerase(PARP) is a positive regulator of telomere length. By adding ribosyl residues to TRF-1, a telomeric DNA-binding protein, TANK1 inhibits the latter's binding to telomeric DNA, thus facilitating its elongation by telomerase[162–164].¯ Telomere shortening has been observed in leukocytes of HCV infected patients [165]. More importantly, HCV infected patients showed shorter telomere lengths than patients in remission. Telomere length is also associated with poor overall survival rates of patients [166]. Indeed, relative telomere length has been established as an independent prognostic marker for HCC patients. Secondly, chronic HCV infection leads to subsequent induction of

HCC, majority of which is associated with expression of HCV non-structural proteins [167, 168]. Taking these studies into account, it is evident that HCV infection leads to HCC in advanced stages of the disease. It is thus, important to determine molecular determinants that accelerate this progression or those that can be modulated to check it. The Wnt signaling pathway has been widely studied in this context, which controls the concentration of β-catenin, a transcriptional activator. β-catenin is regulated by proteasome-dependent degradation which is under the control of APC and Axin1[167]. Axin1 gene mutations are reported in ¯substantial portion of HCCs[169, 170]. Hence Satoh et.al suggest that Axin administration can be an effective therapeutic molecule to suppress growth of HCC.

A recent chemical genetic screen identified a small molecule, XAV939 which stimulates β-catenin degradation by stabilising axin[171]. XAV939 inhibits the PARP enzymes, TANK1 and TANK2 which are responsible for ubiquitin mediated degradation of axin. Inhibition of TANK1/2 results in accumulation of cytosolic axin which then forms the β-catenin destruction complex along with glycogen synthase kinase $3\alpha/\beta$(GSK3α/β) and adenomatous polyposis coli (APC). Putting this finding in the perspective of the overlaid complexes on the HCV_s64 subnetwork (see figure 3.13) we suggest that TANK1/2 inhibition can be an additional therapeutic alternative to the IFN-α-ribavarin therapy for HCV patients. Such combined therapies would ensure early measures to keep the secondary effects of a viral infection, in this case HCC, in check.

3.6.2 hnRNPs in the HIV life-cycle

One of the major complexes found in the HIV subnetworks were the heterogeneous ribonuclear proteins (hnRNPs). As mentioned in previous subsection, hnRNPs were prominent in the HIV_s66 subnetwork. Their importance is also validated by our functional analyses involving overlay of tissue-specific gene expression data and CORUM complexes. Despite belonging to the same complex, different subunits of hnRNPs have different roles in the HIV life cycle; majority of these are antiviral in nature while some, like hnRNP-A1 are proviral in nature[113]. Moreover, hnRNP H subunits also play in important role in mRNA splicing, particularly of viral gene transcripts. Schaub et al. have shown that hnRNPs H, H', F, 2H9, and GRSF-1 to form a multimeric complex by binding at the consensus motif DGGGD (where D can be either of U, G or A)[172]. This complex is involved in splicing of gene substrate derived from the HIV-1 *tat* gene via ATP-dependent spliceosomal complexes. Its important to note there that hnRNP subunits came up in both the secondary analyses and in the HIV_s66 subnetwork itself, thus pointing to multiple lines of evidence and

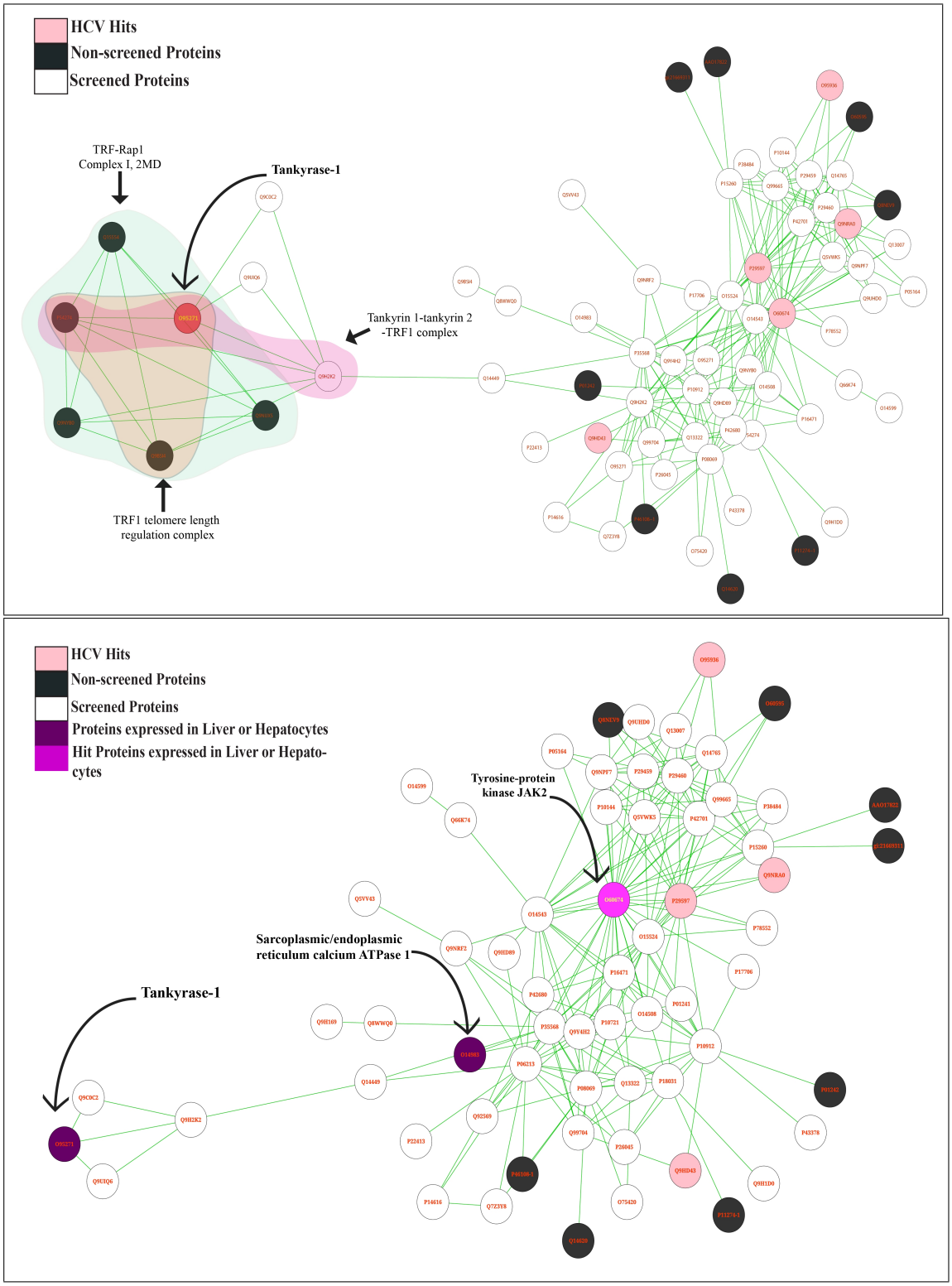


FIGURE 3.13: Tankyrase-1 is also one of the proteins expressed in hepatocytes and also part of the 3 CORUM complexes(TRF1 telomere length regulation complex, TRF-Rap1 complex I, 2MD and Tankyrin 1-tankyrin 2-TRF1 complex) overlaid on HCV_s64 subnetwork.

emphasising their presence in this subnetwork. Hence, it helps in prioritising candidate proteins/complexes from the already few, specific subnetworks obtained from our analysis. In this case, although the hnRNPs have been studied elsewhere, a double or multiple knockdown of these subunits in macrophage infected HIV cells may lead to interesting observations.

3.7 Putative Novel Hits: Mapping virus-host interaction data

A third approach to predict novel hits from within these subnetworks is to overlay virus-host interaction data. If a viral protein interacts with a host-protein that is a *non-hit*, such proteins along with *hits* can be interpreted collectively. The confidence of such *non-hits* increases if, from a group of such *non-hits*, some are *hits*.

3.7.1 Rev,p19 and its interactions with heterogeneous ribonuclear proteins (hnRNPs)

The HIV viral protein, Rev or otherwise known as p19, is an adaptor protein known for its function of nuclear export of HIV RNAs. To understand the details of its mechanisms, Hadian et al.[173] showed how Rev interacts with a a large family of multifunctional host factors call hnRNPs. Rev utilises amino acids 9-14, specifically to bind heterogeneous ribonucleoproteins (hnRNP) A1, Q, K, R and U. The HIV-1 NIAID Database [53] even lists several hnRNP subunits that interact with Rev. These include A/B isoform b, A1 isoform a, A3, D-like isoform a, D0 isoform d, F, H, H2, H3 isoform a, K isoform a, M isoform a, Q isoform 1, R isoform 2, U isoform b, A2/B1 isoform A2 and C1/C2 isoform b. The HIV_s66 subnetwork contains almost all of these subunits, if not the exact isoforms of these genes. We have already discussed the multifunctional properties of hnRNPs above (see figure 3.5).

3.7.2 Interactions of HCV NS3-4A protein

We utilised the virus-host interactions dataset of de Chassey et al.[70]. When overlaying the host genes detected to be interacting with viral proteins, on the HCV subnetworks, we found that NS3-4A protein interacted with with 2 proteins; RNA-binding protein 4 (RBM4) and E3 ubiquitin-protein ligase SMURF2 (SMURF2). Of these 2 proteins, SMURF2 is a hit while RBM4 is a non-hit. HCV nonstructural proteins interfere with TGF-β signaling via SMURF2, which is a negative regulator of this pathway. As described above, TGF-β stimulation led to an increase of SMAD-dependent genes[174]. However, this stimulated signaling was suppressed by SMURF2 while

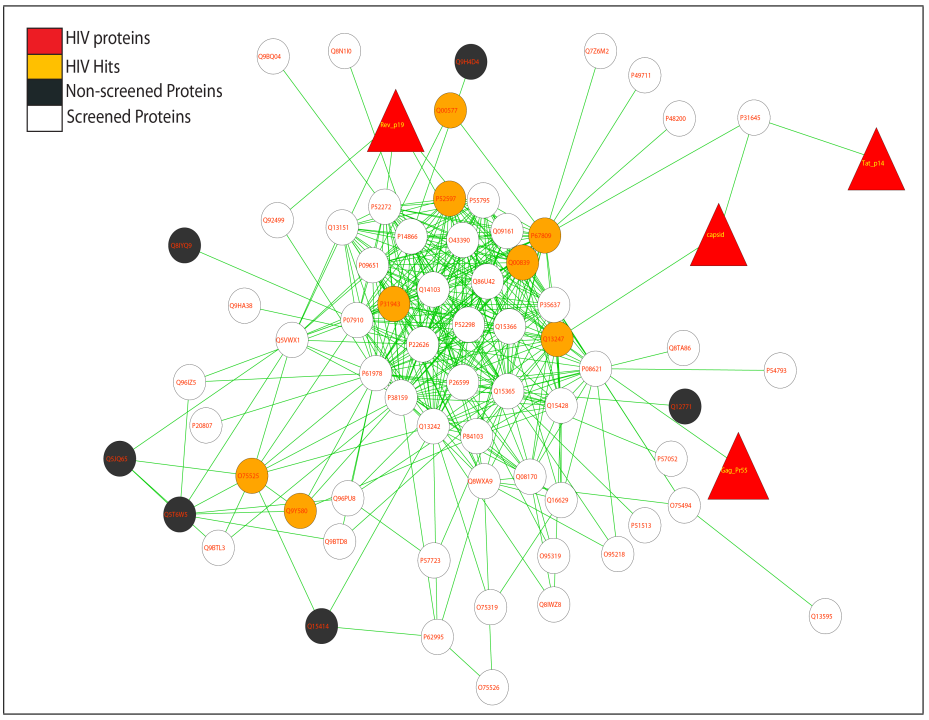


FIGURE 3.14: HIV_s66 subnetwork with interacting HIV-1 proteins

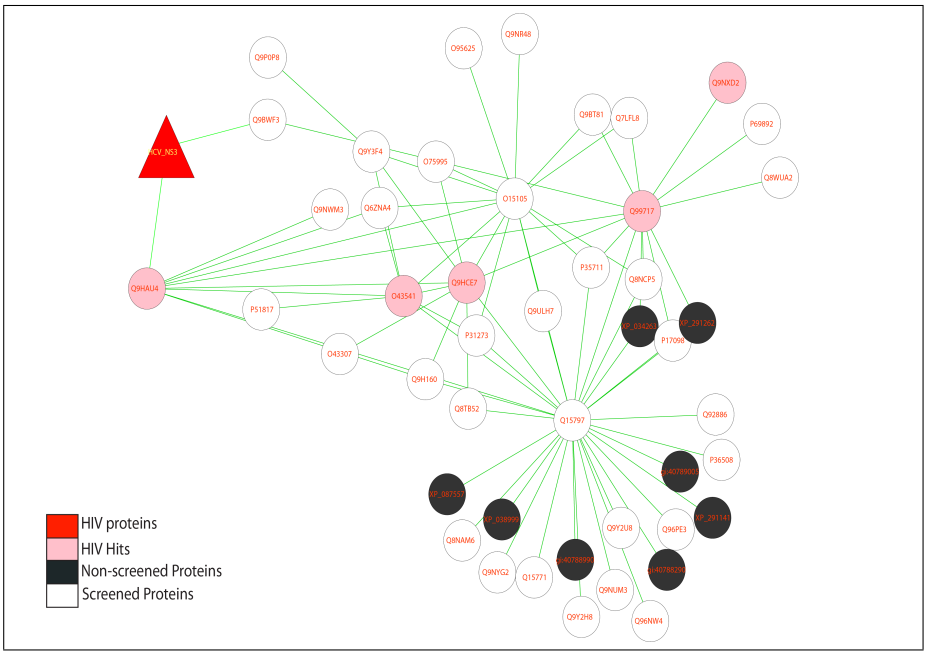


FIGURE 3.15: HCV_s46 subnetwork with interacting HCV-NS3 proteins

and mimicked upon SMURF2 silencing. Importantly, Verga-Gérard et al. showed that the ubiquitin ligase activity or NS3-4A protease activity wasn't required to modulate TGF-β signaling.

Chapter 4

Discussion

4.1 Integrative approaches reveal significant biology

Our results illustrate that using various kinds of datasets to analyse a genome-wide RNAi screens reveal multiple perspectives of underlying biological mechanisms. This is otherwise not possible from traditional enrichment analyses methods. Our analyses are one of the many that have used RNAi screen data to understand virus-host interactions and their biology. Noteworthy among these studies are from Bushman et al. who performed a comprehensive meta-analysis of all the published HIV-1 RNAi screens; Macpherson et al. who utilised the HIV-1 human protein interaction database (HHPID) in conjunction with the published HIV-1 RNAi screens to reveal perturbed host subsystems[1–3, 9, 53, 54]; Dickerson et al. who utilised the same dataset (HHPID) to reveal topological features of the most targeted host genes by HIV[175]; Murali et al. who developed a machine learning approach to predict novel HDFs using protein interaction network and the published RNAi screens [43] and finally, Schneider et al. who used a large number of RNAi screens and applied single cell analysis with some novel statistical functions to illustrate variation in identified hits [176]. Each of these analyses utilised RNAi screens with more than one data type (except Schneider et al.) to provide insights within the virus-host interactions. Since none of these studies have a common algorithmic basis except for the data used, it is not straightforward to compare their results. Particularly, except for Schneider et al., all these studies focus on HIV-1 screens whereas our analyses encompass HCV and WNV too. However, despite the difference in methodology and its application, if certain biological processes among these studies converge towards a specific set of biological processes/pathways for a particular virus, it provides a second level of validation about these processes/-pathways and helps in confirming their application as novel drug targets/therapeutics.

For instance, our analysis along with Bushman et al. and Murali et al., showed that

HNRNPs are important HDFs of HIV-1, using the published RNAi screen data. It thus validates the fact that among other hits identified from these studies, HNRNPs should be focused for a secondary experimental validation amongst other potentially novel hits, as they are predicted hits from all these studies. This also applies to the Mediator complex, again a prominent result from all the above mentioned studies as well as our analysis. Such multiple validations of certain biological processes/protein complexes is essential if these hits are to be considered for therapeutic use. Additionally, even ¯from our analysis; tissue-specific expression data and protein complex overlays identified HNRNPs as a prominent result. This is in contrast when compared exclusively to the computational studies mentioned above wherein we further highlight the importance of certain hits considering its expression. Theoretical predictions often leads to multiple hits with a fair chance of false-positives, similar to the experimental counterparts. However, by adjusting the stringency of scoring functions, it is relatively easy to control the rate of false-positives in such computational predictions. Despite such measures, more often there are still considerable number of *potential novel hits* to tackle with. Herein along with **importance** of hits their **relevance** should be considered, which, is the expression of such hits in a specific tissue. If a hit is highly expressed in host tissue most susceptible to virus as compared to other, such hit is more relevant than others in the list. This is where our network-based meta analysis differs from the other studies. We predicted tissue specific hits that may potentially have an important role in HCV infection and its progression towards HCC. We also hypothesize about possible small-molecule drug treatments for one of the identified enzymes, Tankyrase that might have a role in controlling hepatocytic progression of HCC.

In summary, multiple lines of evidence show that certain processes/pathways/protein complexes from within a hit-list of RNAi screen are *more important* over others. These may have poor statistical confidence ($p \leq 0.05$, yet towards the higher side) but when analysed with supplementary datasets, their biological relevance and significance becomes clearer. Thus, our approach allows for **Hit-prioritisation** from within hit-list(s) of a RNAi screen(s).

4.2 Host factors shared between virus species

A unique differentiation of our analysis as compared to others mentioned above, is comparative analysis of inter-species viral RNAi screens and identification of subnetworks enriched for HDFs for one or more virus species. Schneider et al. did utilise multiple screens of different virus species, however, they didn't point out functional subnetworks. From our combined analysis of hit-lists from RNAi screens of different virus species, we determined common HDFs required by HIV and HCV. In some

cases, such as the Src family of kinases, encompassing a large number of kinases with varied functions, our analysis shows that they are required by all three viruses (considered in our study); HIV, HCV and WNV. Such results open further avenues in therapeutics and at times are essential when it comes to cases of co-infection. For e.g. HIV patients co-infected with HCV. Although from different families, both HIV and HCV are RNA viruses and as such, determining these common HDFs from within these subnetwork helps in hypothesizing interactions within such patients and subsequently, devise unique therapeutic measures for them. Such strategies are otherwise extremely difficult to determine purely by experimentation and thus, such computational prediction leverage this task to a great extent .

Determining these common subnetworks becomes even more intriguing in the case where the viruses in consideration are from the same family. For instance, a particular subnetwork enriched for HDFs/HRFs for RNAi hits from viral screens of HCV, WNV an Dengue virus (family *Flaviviridae*), allows for studying finer level of interaction mechanisms between the commonly targeted host proteins by all three viruses. Conversely, other proteins that are specifically targeted by only one species also helps in understanding why different virus species of the same family target different proteins within the same functional subnetwork. Thus, such subnetwork becomes *more important* over other statistically significant hits, as identified from classical enrichment methods.

4.3 Multiple approaches of integrative analyses and results

In data-integrative meta-analyses of RNAi screens, some biological processes/pathways/protein complexes are *more significant* than others (HNRNPs in case of HIV), irrespective of the underlying methodology. An argument can be made about the reliability of such processes/pathways, that since such analyses involved protein interaction networks, which are susceptible to a large number of false positives. This is further marred by the fact that for many predicted interactions, interactions derived from literature and high throughput studies have ascertainment bias towards certain proteins over others[177, 178]. Macpherson et al. did account for this bias and applied a corrective measure in their analysis[54]. A stricter filtering of interactions can be at the level of interaction detection type wherein only experimentally verified reactions are considered. However, this is a compromise for coverage as such interactions aren't comprehensive and are neither mutually exclusive[177]. Furthermore, this also risks losing out false-negatives from the dataset. We observed this in our analyses too where we obtained few or no subnetworks using our 2-step filtering process, when

using experimentally-verified interactions alone. Thus, we chose for broader coverage to detect such functional subnetworks and subsequently, *"hot spots"* within the human PINs with a risk of having potential false-positives than missing out false negatives. Since we propose these subnetworks for therapeutics the risk of losing out true-negatives is more than having false-positives in the dataset. In order to enhance reliability and confidence of the subnetworks, we therefore utilised tissue-specific expression and known human protein complexes, to interpret these subnetworks. These steps keep false-positives in check as unrelated interactions and spurious interactions observed after these steps can then be tested with greater caution

In our methodology, a second possible factor that might influence the resulting subnetworks is the choice of the clustering algorithm, initially used to cluster the human PIN. Indeed, different clustering algorithms can lead to different clusters and as a consequence, might change the final resulting clusters. This can indeed be the case when the initial clustering is performed by an algorithm that doesn't account for overlaps. However, as we mentioned earlier, certain hits from a RNAi screen are robust and thus, despite change in clustering, scoring etc. didn't alter the results for these hits. It then becomes interesting to study interaction partners of such robust hits and hypothesize on their interaction mechanisms within the subnetwork that is enriched. On the other hand, the difference in the subnetworks obtained due to choice of clustering algorithms would then become a comparative analysis of the algorithms in question, which is beyond the scope of the thesis.

It also follows that in the topological filtering step in our methodology, larger subnetworks will have larger mean centrality values (in particular, Degree and Betweenness, see Appendix I) and thus, will lead to larger subnetworks that are diverse in function. But their filtering based on these mean centrality measures when compared to non-hit subnetworks of the same size (see Materials and methods) ensured that we got *reasonably* sized subnetworks. This is reflected in our results as well. With the exception of Combi_s239, all our subnetworks ranged from 40-64. Given these possible shortcomings, the biological hypotheses deduced from such analysis is evidently multi-faceted. Each facet of these subnetworks further enhances our understanding of virus-host interactions and provides a computational basis for drug design and effective therapy.

4.4 Applications of the methodology

Our methodology relies on large protein interaction networks and multiple gene hit lists from a study or gene hit list from multiple studies, aimed towards a biological process or identifying key determinants of a disease. Thus, practically our methodology can be applied to datasets that fulfil these criteria and allow for an integrative

meta-analysis of such datasets. One prominent application area being identification of up-regulated/down-regulated subnetworks for different cancer types. The predicted common subnetworks can then be targeted with specific inhibitor/enhancer molecules to regulate those cancers for which the subnetworks are common for. On the other hand, the subnetworks specific to a particular cancer can then be included with the common subnetworks' therapy for an even more effective and personalised treatment. Moreover, as described above, inclusion of multiple, relevant data sources (such as drug-target interactions, ligand-receptor interactions, expression data) can then be utilised to delve into further intricacies. Despite some of its shortcomings mentioned above, such approaches hold the key to utilise big, multiple data sources to understand biology and more so, develop new-age therapeutics for complex human diseases.

Chapter 5

Conclusion

We developed a data-driven, integrative bioinformatics framework for analysing genome-wide RNAi screens. Genome-wide knockdown screens have accelerated the rate at which novel host factors of viral diseases are identified. Although such screens are promising, they pose significant data-specific challenges of their own. In particular, screens with similar experimental setup for the same virus show relatively poor overlap between identified hit-sets. This leads one to infer that these screens aren't reproducible and are highly noisy. These inferences might be true but integrative statistical analyses can overcome this problem. Specifically, incorporating network information with RNAi hits can significantly improve our interpretation and understanding of the hits.

Initially we built a host protein interaction network collating interactions from public repositories. This network consisted of 15383 proteins and 337413 interactions. We then subject this network to an overlapping complex prediction algorithm, ClusterOne, to identify overlapping clusters. We chose an algorithm that predicts overlapping complexes to identify potential multifunction proteins. More so, as we chose to analyse multiple RNAi screens of different virus species and multifunction, druggable proteins can serve as drug targets for more than one virus. We included 7 genome-wide screens consisting of 3 HIV screens, 3 HCV screens and a WNV screen. From amongst the predicted clusters, we calculated hit enrichment in the clusters in 2 ways; with virus-specific hit-sets and combining hits from all screens. These clusters were then subjected to a 2-step filtering; in the 1st step, statistically significant clusters (at 5% significance) were filtered and in the 2nd step, was subjected to topology based filtering. We utilised 6 network topological measures, technically known as **Network Centralities** and 2 semantic similarity measures (Dice and Wang) for this 2nd step filtering. We calculated mean values for all these measures and used these mean values to filter subnetworks (i.e. clusters) based on size. Subnetworks of identical sizes, those enriched with hits and those not enriched with hits, were filtered based on the

mean values of network centralities and similarity measures using Wilcox test. Thus, subnetworks filtered in this step were statistically significant in 2 ways; enriched with *hits* and significant in topology compared to their non-enriched counterparts.

For further characterisation, GO and pathway enrichment of these subnetworks were performed to determine the role of these subnetworks in viral infection. We found that, for virus-specific subnetworks, GO and pathway terms were highly specific. This has to be attributed to the underlying clustering as well as the topological filtering step, as these subnetworks not only contained well characterised hits (e.g. hnRNPs in HIV infection) and identified as *hits* in RNAi screens but other potentially novel hit candidates in their neighbourhood. Using tissue-specific gene expression data and known protein complexes allowed for further specific characterisation and confidence for predicted novel hits. Moreover, it also allowed for hypothesising mechanisms of infection with known and unknown proteins. This helped to revisit known pathways and complexes for a detailed analysis to elucidate interaction mechanisms of such proteins in viral infection. In summary, our approach provides a computational, integrative frame-work for meta-analysis of genome-wide studies.

Appendix A

Appendix A

A.1 Subnetwork Protein Annotations

Entry	Entry name	Status	Protein names	Cross-reference GENEID
P54793	ARSF_HUMAN	reviewed	Arylsulfatase F (ASF) (EC 3.1.6.-)	416
P20807	CAN3_HUMAN	reviewed	Calpain-3 (EC 3.4.22.54) (Calcium-activated neutral proteinase 3) (CANP 3) (Calpain L3) (Calpain p94) (Muscle-specific calcium-activated neutral protease 3) (New calpain 1) (nCL-1)	825
O95319	CELF2_HUMAN	reviewed	CUGBP Elav-like family member 2 (CELF-2) (Bruno-like protein 3) (CUG triplet repeat RNA-binding protein 2) (CUG-BP2) (CUG-BP- and ETR-3-like factor 2) (ELAV-type RNA-binding protein 3) (ETR-3) (Neuroblastoma apoptosis-related RNA-binding protein) (hNAPOR) (RNA-binding protein BRUNOL-3)	10659
P49711	CTCF_HUMAN	reviewed	Transcriptional repressor CTCF (11-zinc finger protein) (CCCTC-binding factor) (CTCFL paralog)	10664
Q92499	DDX1_HUMAN	reviewed	ATP-dependent RNA helicase DDX1 (EC 3.6.4.13) (DEAD box protein 1) (DEAD box protein retinoblastoma) (DBP-RB)	1653
Q8N1I0	DOCK4_HUMAN	reviewed	Dedicator of cytokinesis protein 4	9732

O75319	DUS11_HUMAN	reviewed	RNA/RNP complex-1-interacting phosphatase (EC 3.1.3.-) (Dual specificity protein phosphatase 11) (Phosphatase that interacts with RNA/RNP complex 1)	8446
Q7Z6M2	FBX33_HUMAN	reviewed	F-box only protein 33	254170
P35637	FUS_HUMAN	reviewed	RNA-binding protein FUS (75 kDa DNA-pairing protein) (Oncogene FUS) (Oncogene TLS) (POMp75) (Translocated in liposarcoma protein)	2521
P31943	HNRH1_HUMAN	reviewed	Heterogeneous nuclear ribonucleoprotein H (hnRNP H) [Cleaved into: Heterogeneous nuclear ribonucleoprotein H, N-terminally processed]	3187
P55795	HNRH2_HUMAN	reviewed	Heterogeneous nuclear ribonucleoprotein H2 (hnRNP H2) (FTP-3) (Heterogeneous nuclear ribonucleoprotein H') (hnRNP H')	3188
P07910	HNRPC_HUMAN	reviewed	Heterogeneous nuclear ribonucleoproteins C1/C2 (hnRNP C1/C2)	3183
Q14103	HNRPD_HUMAN	reviewed	Heterogeneous nuclear ribonucleoprotein D0 (hnRNP D0) (AU-rich element RNA-binding protein 1)	3184
P52597	HNRPF_HUMAN	reviewed	Heterogeneous nuclear ribonucleoprotein F (hnRNP F) (Nucleolin-like protein mcs94-1) [Cleaved into: Heterogeneous nuclear ribonucleoprotein F, N-terminally processed]	3185
P61978	HNRPK_HUMAN	reviewed	Heterogeneous nuclear ribonucleoprotein K (hnRNP K) (Transformation up-regulated nuclear protein) (TUNP)	3190
P14866	HNRPL_HUMAN	reviewed	Heterogeneous nuclear ribonucleoprotein L (hnRNP L)	3191
P52272	HNRPM_HUMAN	reviewed	Heterogeneous nuclear ribonucleoprotein M (hnRNP M)	4670
O43390	HNRPR_HUMAN	reviewed	Heterogeneous nuclear ribonucleoprotein R (hnRNP R)	10236

Q00839	HNRPU_HUMAN	reviewed	Heterogeneous nuclear ribonucleoprotein U (hnRNP U) (Scaffold attachment factor A) (SAF-A) (p120) (pp120)	3192
P48200	IREB2_HUMAN	reviewed	Iron-responsive element-binding protein 2 (IRE-BP 2) (Iron regulatory protein 2) (IRP2)	3658
Q5VWX1	KHDR2_HUMAN	reviewed	KH domain-containing, RNA-binding, signal transduction-associated protein 2 (Sam68-like mammalian protein 1) (SLM-1) (hSLM-1)	202559
O75525	KHDR3_HUMAN	reviewed	KH domain-containing, RNA-binding, signal transduction-associated protein 3 (RNA-binding protein T-Star) (Sam68-like mammalian protein 2) (SLM-2) (Sam68-like phosphotyrosine protein)	10656
Q09161	NCBP1_HUMAN	reviewed	Nuclear cap-binding protein subunit 1 (80 kDa nuclear cap-binding protein) (CBP80) (NCBP 80 kDa subunit)	4686
P52298	NCBP2_HUMAN	reviewed	Nuclear cap-binding protein subunit 2 (20 kDa nuclear cap-binding protein) (Cell proliferation-inducing gene 55 protein) (NCBP 20 kDa subunit) (CBP20) (NCBP-interacting protein 1) (NIP1)	22916
P51513	NOVA1_HUMAN	reviewed	RNA-binding protein Nova-1 (Neuro-oncological ventral antigen 1) (Onconeural ventral antigen 1) (Paraneoplastic Ri antigen) (Ventral neuron-specific protein 1)	4857
Q86U42	PABP2_HUMAN	reviewed	Polyadenylate-binding protein 2 (PABP-2) (Poly(A)-binding protein 2) (Nuclear poly(A)-binding protein 1) (Poly(A)-binding protein II) (PABII) (Polyadenylate-binding nuclear protein 1)	8106
Q15365	PCBP1_HUMAN	reviewed	Poly(rC)-binding protein 1 (Alpha-CP1) (Heterogeneous nuclear ribonucleoprotein E1) (hnRNP E1) (Nucleic acid-binding protein SUB2.3)	5093

Q15366	PCBP2_HUMAN	reviewed	Poly(rC)-binding protein 2 (Alpha-CP2) (Heterogeneous nuclear ribonucleoprotein E2) (hnRNP E2)	5094
P57723	PCBP4_HUMAN	reviewed	Poly(rC)-binding protein 4 (Alpha-CP4)	57060
P26599	PTBP1_HUMAN	reviewed	Polypyrimidine tract-binding protein 1 (PTB) (57 kDa RNA-binding protein PPTB-1) (Heterogeneous nuclear ribonucleoprotein I) (hnRNP I)	5725
Q00577	PURA_HUMAN	reviewed	Transcriptional activator protein Pur-alpha (Purine-rich single-stranded DNA-binding protein alpha)	5813
Q96PU8	QKI_HUMAN	reviewed	Protein quaking (Hqk) (HqkI)	9444
Q9BTL3	RAM_HUMAN	reviewed	RNMT-activating mini protein (RAM) (Protein FAM103A1)	83640
P57052	RBM11_HUMAN	reviewed	Splicing regulator RBM11 (RNA-binding motif protein 11)	54033
Q96IZ5	RBM41_HUMAN	reviewed	RNA-binding protein 41 (RNA-binding motif protein 41)	55285
Q9BTD8	RBM42_HUMAN	reviewed	RNA-binding protein 42 (RNA-binding motif protein 42)	79171
Q9BQ04	RBM4B_HUMAN	reviewed	RNA-binding protein 4B (RNA-binding motif protein 30) (RNA-binding motif protein 4B) (RNA-binding protein 30)	83759
Q9Y580	RBM7_HUMAN	reviewed	RNA-binding protein 7 (RNA-binding motif protein 7)	10179
P38159	RBMX_HUMAN	reviewed	RNA-binding motif protein, X chromosome (Glycoprotein p43) (Heterogeneous nuclear ribonucleoprotein G) (hnRNP G) [Cleaved into: RNA-binding motif protein, X chromosome, N-terminally processed]	27316
P0DJD3	RBY1A_HUMAN	reviewed	RNA-binding motif protein, Y chromosome, family 1 member A1 (RNA-binding motif protein 1) (RNA-binding motif protein 2) (Y chromosome RNA recognition motif 1) (hRBMY)	5940

P0DJD4	RBY1C_HUMAN	reviewed	RNA-binding motif protein, Y chromosome, family 1 member C	
O75526	RMXL2_HUMAN	reviewed	RNA-binding motif protein, X-linked-like-2 (Testis-specific heterogeneous nuclear ribonucleoprotein G-T) (hnRNP G-T)	27288
Q13151	ROA0_HUMAN	reviewed	Heterogeneous nuclear ribonucleoprotein A0 (hnRNP A0)	10949
P09651	ROA1_HUMAN	reviewed	Heterogeneous nuclear ribonucleoprotein A1 (hnRNP A1) (Helix-destabilizing protein) (Single-strand RNA-binding protein) (hnRNP core protein A1)	3178
P22626	ROA2_HUMAN	reviewed	Heterogeneous nuclear ribonucleoproteins A2/B1 (hnRNP A2/B1)	3181
Q8TA86	RP9_HUMAN	reviewed	Retinitis pigmentosa 9 protein (Pim-1-associated protein) (PAP-1)	6100
P08621	RU17_HUMAN	reviewed	U1 small nuclear ribonucleoprotein 70 kDa (U1 snRNP 70 kDa) (U1-70K) (snRNP70)	6625
P31645	SC6A4_HUMAN	reviewed	Sodium-dependent serotonin transporter (5HT transporter) (5HTT) (Solute carrier family 6 member 4)	6532
Q15428	SF3A2_HUMAN	reviewed	Splicing factor 3A subunit 2 (SF3a66) (Spliceosome-associated protein 62) (SAP 62)	8175
Q8WXA9	SREK1_HUMAN	reviewed	Splicing regulatory glutamine/lysine-rich protein 1 (Serine/arginine-rich-splicing regulatory protein 86) (SRrp86) (Splicing factor, arginine/serine-rich 12) (Splicing regulatory protein 508) (SRrp508)	140890

O75494	SRS10_HUMAN	reviewed	Serine/arginine-rich splicing factor 10 (40 kDa SR-repressor protein) (SRrp40) (FUS-interacting serine-arginine-rich protein 1) (Splicing factor SRp38) (Splicing factor, arginine/serine-rich 13A) (TLS-associated protein with Ser-Arg repeats) (TASR) (TLS-associated protein with SR repeats) (TLS-associated serine-arginine protein) (TLS-associated SR protein)	100996657 10772
P84103	SRSF3_HUMAN	reviewed	Serine/arginine-rich splicing factor 3 (Pre-mRNA-splicing factor SRP20) (Splicing factor, arginine/serine-rich 3)	6428
Q08170	SRSF4_HUMAN	reviewed	Serine/arginine-rich splicing factor 4 (Pre-mRNA-splicing factor SRP75) (SRP001LB) (Splicing factor, arginine/serine-rich 4)	6429
Q13247	SRSF6_HUMAN	reviewed	Serine/arginine-rich splicing factor 6 (Pre-mRNA-splicing factor SRP55) (Splicing factor, arginine/serine-rich 6)	6431
Q16629	SRSF7_HUMAN	reviewed	Serine/arginine-rich splicing factor 7 (Splicing factor 9G8) (Splicing factor, arginine/serine-rich 7)	6432
Q13242	SRSF9_HUMAN	reviewed	Serine/arginine-rich splicing factor 9 (Pre-mRNA-splicing factor SRp30C) (Splicing factor, arginine/serine-rich 9)	8683
Q8IWZ8	SUGP1_HUMAN	reviewed	SURP and G-patch domain-containing protein 1 (RNA-binding protein RBP) (Splicing factor 4)	57794
Q13595	TRA2A_HUMAN	reviewed	Transformer-2 protein homolog alpha (TRA-2 alpha) (TRA2-alpha) (Transformer-2 protein homolog A)	29896
P62995	TRA2B_HUMAN	reviewed	Transformer-2 protein homolog beta (TRA-2 beta) (TRA2-beta) (hTRA2-beta) (Splicing factor, arginine/serine-rich 10) (Transformer-2 protein homolog B)	6434

P67809	YBOX1_HUMAN	reviewed	Nuclease-sensitive element-binding protein 1 (CCAAT-binding transcription factor I subunit A) (CBF-A) (DNA-binding protein B) (DBPB) (Enhancer factor I subunit A) (EFI-A) (Y-box transcription factor) (Y-box-binding protein 1) (YB-1)	4904
Q9HA38	ZMAT3_HUMAN	reviewed	Zinc finger matrin-type protein 3 (Zinc finger protein WIG-1) (p53-activated gene 608 protein)	64393
O95218	ZRAB2_HUMAN	reviewed	Zinc finger Ran-binding domain-containing protein 2 (Zinc finger protein 265) (Zinc finger, splicing)	9406
Q5JQ65	Q5JQ65_HUMAN	unreviewed	RNA binding motif protein, X-linked (Fragment)	
Q8IYQ9	Q8IYQ9_HUMAN	unreviewed	Importin subunit alpha	3839
Q5T6W5	Q5T6W5_HUMAN	unreviewed	Heterogeneous nuclear ribonucleoprotein K	
Q12771	Q12771_HUMAN	unreviewed	P37 AUF1	
Q9H4D4	Q9H4D4_HUMAN	unreviewed	ZNF143 protein (Fragment)	

TABLE A.1: HIV-s66 Protein Annotation and GeneIDs

Entry	**Entry name**	**Status**	**Protein names**	**Cross-reference GENEID**
Q96N21	AP4AT_HUMAN	reviewed	AP-4 complex accessory subunit tepsin (ENTH domain-containing protein 2) (Epsin for AP-4) (Tetra-epsin)	146705
Q9BZE3	BARH1_HUMAN	reviewed	BarH-like 1 homeobox protein	56751
P24863	CCNC_HUMAN	reviewed	Cyclin-C (SRB11 homolog) (hSRB11)	892
Q9BWU1	CDK19_HUMAN	reviewed	Cyclin-dependent kinase 19 (EC 2.7.11.22) (CDC2-related protein kinase 6) (Cell division cycle 2-like protein kinase 6) (Cell division protein kinase 19) (Cyclin-dependent kinase 11) (Death-preventing kinase)	23097

P49336	CDK8_HUMAN	reviewed	Cyclin-dependent kinase 8 (EC 2.7.11.22) (EC 2.7.11.23) (Cell division protein kinase 8) (Mediator complex subunit CDK8) (Mediator of RNA polymerase II transcription subunit CDK8) (Protein kinase K35)	1024
Q6IAN0	DRS7B_HUMAN	reviewed	Dehydrogenase/reductase SDR family member 7B (EC 1.1.-.-)	25979
Q9UPW0	FOXJ3_HUMAN	reviewed	Forkhead box protein J3	22887
Q9H0H0	INT2_HUMAN	reviewed	Integrator complex subunit 2 (Int2)	57508
Q68E01	INT3_HUMAN	reviewed	Integrator complex subunit 3 (Int3) (SOSS complex subunit A) (Sensor of single-strand DNA complex subunit A) (SOSS-A) (Sensor of ssDNA subunit A)	65123
Q7L273	KCTD9_HUMAN	reviewed	BTB/POZ domain-containing protein KCTD9	54793
O94953	KDM4B_HUMAN	reviewed	Lysine-specific demethylase 4B (EC 1.14.11.-) (JmjC domain-containing histone demethylation protein 3B) (Jumonji domain-containing protein 2B)	23030
Q9UBF1	MAGC2_HUMAN	reviewed	Melanoma-associated antigen C2 (Cancer/testis antigen 10) (CT10) (Hepatocellular carcinoma-associated antigen 587) (MAGE-C2 antigen) (MAGE-E1 antigen)	51438
Q71F56	MD13L_HUMAN	reviewed	Mediator of RNA polymerase II transcription subunit 13-like (Mediator complex subunit 13-like) (Thyroid hormone receptor-associated protein 2) (Thyroid hormone receptor-associated protein complex 240 kDa component-like)	23389
Q9BTT4	MED10_HUMAN	reviewed	Mediator of RNA polymerase II transcription subunit 10 (Mediator complex subunit 10) (Transformation-related gene 17 protein) (TRG-17) (Transformation-related gene 20 protein) (TRG-20)	84246

Q9P086	MED11_HUMAN	reviewed	Mediator of RNA polymerase II transcription subunit 11 (Mediator complex subunit 11)	400569
Q93074	MED12_HUMAN	reviewed	Mediator of RNA polymerase II transcription subunit 12 (Activator-recruited cofactor 240 kDa component) (ARC240) (CAG repeat protein 45) (Mediator complex subunit 12) (OPA-containing protein) (Thyroid hormone receptor-associated protein complex 230 kDa component) (Trap230) (Trinucleotide repeat-containing gene 11 protein)	9968
Q9UHV7	MED13_HUMAN	reviewed	Mediator of RNA polymerase II transcription subunit 13 (Activator-recruited cofactor 250 kDa component) (ARC250) (Mediator complex subunit 13) (Thyroid hormone receptor-associated protein 1) (Thyroid hormone receptor-associated protein complex 240 kDa component) (Trap240) (Vitamin D3 receptor-interacting protein complex component DRIP250) (DRIP250)	9969
O60244	MED14_HUMAN	reviewed	Mediator of RNA polymerase II transcription subunit 14 (Activator-recruited cofactor 150 kDa component) (ARC150) (Cofactor required for Sp1 transcriptional activation subunit 2) (CRSP complex subunit 2) (Mediator complex subunit 14) (RGR1 homolog) (hRGR1) (Thyroid hormone receptor-associated protein complex 170 kDa component) (Trap170) (Transcriptional coactivator CRSP150) (Vitamin D3 receptor-interacting protein complex 150 kDa component) (DRIP150)	9282

Q9Y2X0	MED16_HUMAN	reviewed	Mediator of RNA polymerase II transcription subunit 16 (Mediator complex subunit 16) (Thyroid hormone receptor-associated protein 5) (Thyroid hormone receptor-associated protein complex 95 kDa component) (Trap95) (Vitamin D3 receptor-interacting protein complex 92 kDa component) (DRIP92)	10025
Q9NVC6	MED17_HUMAN	reviewed	Mediator of RNA polymerase II transcription subunit 17 (Activator-recruited cofactor 77 kDa component) (ARC77) (Cofactor required for Sp1 transcriptional activation subunit 6) (CRSP complex subunit 6) (Mediator complex subunit 17) (Thyroid hormone receptor-associated protein complex 80 kDa component) (Trap80) (Transcriptional coactivator CRSP77) (Vitamin D3 receptor-interacting protein complex 80 kDa component) (DRIP80)	9440
Q9BUE0	MED18_HUMAN	reviewed	Mediator of RNA polymerase II transcription subunit 18 (Mediator complex subunit 18) (p28b)	54797
A0JLT2	MED19_HUMAN	reviewed	Mediator of RNA polymerase II transcription subunit 19 (Lung cancer metastasis-related protein 1) (Mediator complex subunit 19)	219541

Q15648	MED1_HUMAN	reviewed	Mediator of RNA polymerase II transcription subunit 1 (Activator-recruited cofactor 205 kDa component) (ARC205) (Mediator complex subunit 1) (Peroxisome proliferator-activated receptor-binding protein) (PBP) (PPAR-binding protein) (Thyroid hormone receptor-associated protein complex 220 kDa component) (Trap220) (Thyroid receptor-interacting protein 2) (TR-interacting protein 2) (TRIP-2) (Vitamin D receptor-interacting protein complex component DRIP205) (p53 regulatory protein RB18A)	5469
Q9H944	MED20_HUMAN	reviewed	Mediator of RNA polymerase II transcription subunit 20 (Mediator complex subunit 20) (TRF-proximal protein homolog) (hTRFP)	9477
Q13503	MED21_HUMAN	reviewed	Mediator of RNA polymerase II transcription subunit 21 (Mediator complex subunit 21) (RNA polymerase II holoenzyme component SRB7) (RNAPII complex component SRB7) (hSrb7)	9412
Q15528	MED22_HUMAN	reviewed	Mediator of RNA polymerase II transcription subunit 22 (Mediator complex subunit 22) (Surfeit locus protein 5) (Surf-5)	6837
Q9ULK4	MED23_HUMAN	reviewed	Mediator of RNA polymerase II transcription subunit 23 (Activator-recruited cofactor 130 kDa component) (ARC130) (Cofactor required for Sp1 transcriptional activation subunit 3) (CRSP complex subunit 3) (Mediator complex subunit 23) (Protein sur-2 homolog) (hSur-2) (Transcriptional coactivator CRSP130) (Vitamin D3 receptor-interacting protein complex 130 kDa component) (DRIP130)	9439

O75448	MED24_HUMAN	reviewed	Mediator of RNA polymerase II transcription subunit 24 (Activator-recruited cofactor 100 kDa component) (ARC100) (Cofactor required for Sp1 transcriptional activation subunit 4) (CRSP complex subunit 4) (Mediator complex subunit 24) (Thyroid hormone receptor-associated protein 4) (Thyroid hormone receptor-associated protein complex 100 kDa component) (Trap100) (hTRAP100) (Vitamin D3 receptor-interacting protein complex 100 kDa component) (DRIP100)	9862
Q71SY5	MED25_HUMAN	reviewed	Mediator of RNA polymerase II transcription subunit 25 (Activator interaction domain-containing protein 1) (Activator-recruited cofactor 92 kDa component) (ARC92) (Mediator complex subunit 25) (p78)	81857
O95402	MED26_HUMAN	reviewed	Mediator of RNA polymerase II transcription subunit 26 (Activator-recruited cofactor 70 kDa component) (ARC70) (Cofactor required for Sp1 transcriptional activation subunit 7) (CRSP complex subunit 7) (Mediator complex subunit 26) (Transcriptional coactivator CRSP70)	9441
Q6P2C8	MED27_HUMAN	reviewed	Mediator of RNA polymerase II transcription subunit 27 (Cofactor required for Sp1 transcriptional activation subunit 8) (CRSP complex subunit 8) (Mediator complex subunit 27) (P37 TRAP/SMC-C/PC2 subunit) (Transcriptional coactivator CRSP34)	9442
Q9H204	MED28_HUMAN	reviewed	Mediator of RNA polymerase II transcription subunit 28 (Endothelial-derived protein 1) (Mediator complex subunit 28) (Merlin and Grb2-interacting cytoskeletal protein) (Magicin) (Tumor angiogenesis marker EG-1)	80306

Q9NX70	MED29_HUMAN	reviewed	Mediator of RNA polymerase II transcription subunit 29 (Intersex-like protein) (Mediator complex subunit 29)	55588
Q96HR3	MED30_HUMAN	reviewed	Mediator of RNA polymerase II transcription subunit 30 (Mediator complex subunit 30) (TRAP/Mediator complex component TRAP25) (Thyroid hormone receptor-associated protein 6) (Thyroid hormone receptor-associated protein complex 25 kDa component) (Trap25)	90390
Q9NPJ6	MED4_HUMAN	reviewed	Mediator of RNA polymerase II transcription subunit 4 (Activator-recruited cofactor 36 kDa component) (ARC36) (Mediator complex subunit 4) (TRAP/SMCC/PC2 subunit p36 subunit) (Vitamin D3 receptor-interacting protein complex 36 kDa component) (DRIP36)	29079
O75586	MED6_HUMAN	reviewed	Mediator of RNA polymerase II transcription subunit 6 (Activator-recruited cofactor 33 kDa component) (ARC33) (Mediator complex subunit 6) (hMed6) (Renal carcinoma antigen NY-REN-28)	10001
O43513	MED7_HUMAN	reviewed	Mediator of RNA polymerase II transcription subunit 7 (hMED7) (Activator-recruited cofactor 34 kDa component) (ARC34) (Cofactor required for Sp1 transcriptional activation subunit 9) (CRSP complex subunit 9) (Mediator complex subunit 7) (RNA polymerase transcriptional regulation mediator subunit 7 homolog) (Transcriptional coactivator CRSP33)	9443
Q96G25	MED8_HUMAN	reviewed	Mediator of RNA polymerase II transcription subunit 8 (Activator-recruited cofactor 32 kDa component) (ARC32) (Mediator complex subunit 8)	112950

Q9NWA0	MED9_HUMAN	reviewed	Mediator of RNA polymerase II transcription subunit 9 (Mediator complex subunit 9)	55090
Q9H4Z3	PCIF1_HUMAN	reviewed	Phosphorylated CTD-interacting factor 1	63935
P24928	RPB1_HUMAN	reviewed	DNA-directed RNA polymerase II subunit RPB1 (RNA polymerase II subunit B1) (EC 2.7.7.6) (DNA-directed RNA polymerase II subunit A) (DNA-directed RNA polymerase III largest subunit) (RNA-directed RNA polymerase II subunit RPB1) (EC 2.7.7.48)	5430
Q7RTU7	SCX_HUMAN	reviewed	Basic helix-loop-helix transcription factor scleraxis (Class A basic helix-loop-helix protein 41) (bHLHa41) (Class A basic helix-loop-helix protein 48) (bHLHa48)	100129885,642658
P48436	SOX9_HUMAN	reviewed	Transcription factor SOX-9	6662
Q9H668	STN1_HUMAN	reviewed	CST complex subunit STN1 (Oligonucleotide/oligosaccharide-binding fold-containing protein 1) (Suppressor of cdc thirteen homolog)	79991
O95379	TFIP8_HUMAN	reviewed	Tumor necrosis factor alpha-induced protein 8 (TNF alpha-induced protein 8) (Head and neck tumor and metastasis-related protein) (MDC-3.13) (NF-kappa-B-inducible DED-containing protein) (NDED) (SCC-S2) (TNF-induced protein GG2-1)	25816
Q8TB05	UBAD1_HUMAN	reviewed	UBA-like domain-containing protein 1	124402
Q96K76	UBP47_HUMAN	reviewed	Ubiquitin carboxyl-terminal hydrolase 47 (EC 3.4.19.12) (Deubiquitinating enzyme 47) (Ubiquitin thioesterase 47) (Ubiquitin-specific-processing protease 47)	55031
Q70CQ1	UBP49_HUMAN	reviewed	Ubiquitin carboxyl-terminal hydrolase 49 (EC 3.4.19.12) (Deubiquitinating enzyme 49) (Ubiquitin thioesterase 49) (Ubiquitin-specific-processing protease 49)	25862

Q16587	ZNF74_HUMAN	reviewed	Zinc finger protein 74 (Zinc finger protein 520) (hZNF7)	7625
Q9H2M2	Q9H2M2_HUMAN	unreviewed	Estrogen receptor alpha	

TABLE A.2: HIV-s52 Protein Annotation and GeneIDs

Entry	**Entry name**	**Status**	**Protein names**	**Cross-reference GENEID**
O14983	AT2A1_HUMAN	reviewed	Sarcoplasmic/endoplasmic reticulum calcium ATPase 1 (SERCA1) (SR Ca(2+)-ATPase 1) (EC 3.6.3.8) (Calcium pump 1) (Calcium-transporting ATPase sarcoplasmic reticulum type, fast twitch skeletal muscle isoform) (Endoplasmic reticulum class 1/2 Ca(2+) ATPase)	487
Q99704	DOK1_HUMAN	reviewed	Docking protein 1 (Downstream of tyrosine kinase 1) (p62(dok)) (pp62)	1796
P22413	ENPP1_HUMAN	reviewed	Ectonucleotide pyrophosphatase/phosphodiesterase family member 1 (E-NPP 1) (Membrane component chromosome 6 surface marker 1) (Phosphodiesterase I/nucleotide pyrophosphatase 1) (Plasma-cell membrane glycoprotein PC-1) [Includes: Alkaline phosphodiesterase I (EC 3.1.4.1) Nucleotide pyrophosphatase (NPPase) (EC 3.6.1.9)]	5167
O95936	EOMES_HUMAN	reviewed	Eomesodermin homolog (T-box brain protein 2) (T-brain-2) (TBR-2)	8320
P10912	GHR_HUMAN	reviewed	Growth hormone receptor (GH receptor) (Somatotropin receptor) [Cleaved into: Growth hormone-binding protein (GH-binding protein) (GHBP) (Serum-binding protein)]	2690

P10144	GRAB_HUMAN	reviewed	Granzyme B (EC 3.4.21.79) (C11) (CTLA-1) (Cathepsin G-like 1) (CTSGL1) (Cytotoxic T-lymphocyte proteinase 2) (Lymphocyte protease) (Fragmentin-2) (Granzyme-2) (Human lymphocyte protein) (HLP) (SECT) (T-cell serine protease 1-3E)	3002
Q13322	GRB10_HUMAN	reviewed	Growth factor receptor-bound protein 10 (GRB10 adapter protein) (Insulin receptor-binding protein Grb-IR)	2887
Q14449	GRB14_HUMAN	reviewed	Growth factor receptor-bound protein 14 (GRB14 adapter protein)	2888
P42701	I12R1_HUMAN	reviewed	Interleukin-12 receptor subunit beta-1 (IL-12 receptor subunit beta-1) (IL-12R subunit beta-1) (IL-12R-beta-1) (IL-12RB1) (IL-12 receptor beta component) (CD antigen CD212)	3594
Q99665	I12R2_HUMAN	reviewed	Interleukin-12 receptor subunit beta-2 (IL-12 receptor subunit beta-2) (IL-12R subunit beta-2) (IL-12R-beta-2) (IL-12RB2)	3595
P78552	I13R1_HUMAN	reviewed	Interleukin-13 receptor subunit alpha-1 (IL-13 receptor subunit alpha-1) (IL-13R subunit alpha-1) (IL-13R-alpha-1) (IL-13RA1) (Cancer/testis antigen 19) (CT19) (CD antigen CD213a1)	3597
P08069	IGF1R_HUMAN	reviewed	Insulin-like growth factor 1 receptor (EC 2.7.10.1) (Insulin-like growth factor I receptor) (IGF-I receptor) (CD antigen CD221) [Cleaved into: Insulin-like growth factor 1 receptor alpha chain Insulin-like growth factor 1 receptor beta chain]	3480
P29459	IL12A_HUMAN	reviewed	Interleukin-12 subunit alpha (IL-12A) (Cytotoxic lymphocyte maturation factor 35 kDa subunit) (CLMF p35) (IL-12 subunit p35) (NK cell stimulatory factor chain 1) (NKSF1)	3592

P29460	IL12B_HUMAN	reviewed	Interleukin-12 subunit beta (IL-12B) (Cytotoxic lymphocyte maturation factor 40 kDa subunit) (CLMF p40) (IL-12 subunit p40) (NK cell stimulatory factor chain 2) (NKSF2)	3593
Q9UHD0	IL19_HUMAN	reviewed	Interleukin-19 (IL-19) (Melanoma differentiation-associated protein-like protein) (NG.1)	29949
Q9NPF7	IL23A_HUMAN	reviewed	Interleukin-23 subunit alpha (IL-23 subunit alpha) (IL-23-A) (Interleukin-23 subunit p19) (IL-23p19)	51561
Q5VWK5	IL23R_HUMAN	reviewed	Interleukin-23 receptor (IL-23 receptor) (IL-23R)	149233
Q13007	IL24_HUMAN	reviewed	Interleukin-24 (IL-24) (Melanoma differentiation-associated gene 7 protein) (MDA-7) (Suppression of tumorigenicity 16 protein)	11009
Q8NEV9	IL27A_HUMAN	reviewed	Interleukin-27 subunit alpha (IL-27 subunit alpha) (IL-27-A) (IL27-A) (p28)	246778
P15260	INGR1_HUMAN	reviewed	Interferon gamma receptor 1 (IFN-gamma receptor 1) (IFN-gamma-R1) (CDw119) (CD antigen CD119)	3459
P38484	INGR2_HUMAN	reviewed	Interferon gamma receptor 2 (IFN-gamma receptor 2) (IFN-gamma-R2) (Interferon gamma receptor accessory factor 1) (AF-1) (Interferon gamma transducer 1)	3460
P14616	INSRR_HUMAN	reviewed	Insulin receptor-related protein (IRR) (EC 2.7.10.1) (IR-related receptor) [Cleaved into: Insulin receptor-related protein alpha chain Insulin receptor-related protein beta chain]	3645
P06213	INSR_HUMAN	reviewed	Insulin receptor (IR) (EC 2.7.10.1) (CD antigen CD220) [Cleaved into: Insulin receptor subunit alpha Insulin receptor subunit beta]	3643
P35568	IRS1_HUMAN	reviewed	Insulin receptor substrate 1 (IRS-1)	3667
Q9Y4H2	IRS2_HUMAN	reviewed	Insulin receptor substrate 2 (IRS-2)	8660

O60674	JAK2_HUMAN	reviewed	Tyrosine-protein kinase JAK2 (EC 2.7.10.2) (Janus kinase 2) (JAK-2)	3717
Q5VV43	K0319_HUMAN	reviewed	Dyslexia-associated protein KIAA0319	9856
Q7Z3Y8	K1C27_HUMAN	reviewed	Keratin, type I cytoskeletal 27 (Cytokeratin-27) (CK-27) (Keratin-25C) (K25C) (Keratin-27) (K27) (Type I inner root sheath-specific keratin-K25irs3)	342574
P10721	KIT_HUMAN	reviewed	Mast/stem cell growth factor receptor Kit (SCFR) (EC 2.7.10.1) (Piebald trait protein) (PBT) (Proto-oncogene c-Kit) (Tyrosine-protein kinase Kit) (p145 c-kit) (v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog) (CD antigen CD117)	3815
Q9UIQ6	LCAP_HUMAN	reviewed	Leucyl-cystinyl aminopeptidase (Cystinyl aminopeptidase) (EC 3.4.11.3) (Insulin-regulated membrane aminopeptidase) (Insulin-responsive aminopeptidase) (IRAP) (Oxytocinase) (OTase) (Placental leucine aminopeptidase) (P-LAP) [Cleaved into: Leucyl-cystinyl aminopeptidase, pregnancy serum form]	4012
Q66K74	MAP1S_HUMAN	reviewed	Microtubule-associated protein 1S (MAP-1S) (BPY2-interacting protein 1) (Microtubule-associated protein 8) (Variable charge Y chromosome 2-interacting protein 1) (VCY2-interacting protein 1) (VCY2IP-1) [Cleaved into: MAP1S heavy chain MAP1S light chain]	55201

Q92569	P55G_HUMAN	reviewed	Phosphatidylinositol 3-kinase regulatory subunit gamma (PI3-kinase regulatory subunit gamma) (PI3K regulatory subunit gamma) (PtdIns-3-kinase regulatory subunit gamma) (Phosphatidylinositol 3-kinase 55 kDa regulatory subunit gamma) (PI3-kinase subunit p55-gamma) (PtdIns-3-kinase regulatory subunit p55-gamma) (p55PIK)	8503
P05164	PERM_HUMAN	reviewed	Myeloperoxidase (MPO) (EC 1.11.2.2) [Cleaved into: Myeloperoxidase 89 kDa myeloperoxidase 84 kDa myeloperoxidase Myeloperoxidase light chain Myeloperoxidase heavy chain]	4353
O75420	PERQ1_HUMAN	reviewed	PERQ amino acid-rich with GYF domain-containing protein 1 (GRB10-interacting GYF protein 1)	64599
Q8WWQ0	PHIP_HUMAN	reviewed	PH-interacting protein (PHIP) (IRS-1 PH domain-binding protein) (WD repeat-containing protein 11)	55023
P16471	PRLR_HUMAN	reviewed	Prolactin receptor (PRL-R)	5618
P18031	PTN1_HUMAN	reviewed	Tyrosine-protein phosphatase non-receptor type 1 (EC 3.1.3.48) (Protein-tyrosine phosphatase 1B) (PTP-1B)	5770
P17706	PTN2_HUMAN	reviewed	Tyrosine-protein phosphatase non-receptor type 2 (EC 3.1.3.48) (T-cell protein-tyrosine phosphatase) (TCPTP)	5771
P26045	PTN3_HUMAN	reviewed	Tyrosine-protein phosphatase non-receptor type 3 (EC 3.1.3.48) (Protein-tyrosine phosphatase H1) (PTP-H1)	5774
P43378	PTN9_HUMAN	reviewed	Tyrosine-protein phosphatase non-receptor type 9 (EC 3.1.3.48) (Protein-tyrosine phosphatase MEG2) (PTPase MEG2)	5780

Q9HD43	PTPRH_HUMAN	reviewed	Receptor-type tyrosine-protein phosphatase H (R-PTP-H) (EC 3.1.3.48) (Stomach cancer-associated protein tyrosine phosphatase 1) (SAP-1) (Transmembrane-type protein-tyrosine phosphatase type H)	5794
Q9HD89	RETN_HUMAN	reviewed	Resistin (Adipose tissue-specific secretory factor) (ADSF) (C/EBP-epsilon-regulated myeloid-specific secreted cysteine-rich protein) (Cysteine-rich secreted protein A12-alpha-like 2) (Cysteine-rich secreted protein FIZZ3)	56729
Q9NRF2	SH2B1_HUMAN	reviewed	SH2B adapter protein 1 (Pro-rich, PH and SH2 domain-containing signaling mediator) (PSM) (SH2 domain-containing protein 1B)	25970
O15524	SOCS1_HUMAN	reviewed	Suppressor of cytokine signaling 1 (SOCS-1) (JAK-binding protein) (JAB) (STAT-induced STAT inhibitor 1) (SSI-1) (Tec-interacting protein 3) (TIP-3)	8651
O14508	SOCS2_HUMAN	reviewed	Suppressor of cytokine signaling 2 (SOCS-2) (Cytokine-inducible SH2 protein 2) (CIS-2) (STAT-induced STAT inhibitor 2) (SSI-2)	8835
O14543	SOCS3_HUMAN	reviewed	Suppressor of cytokine signaling 3 (SOCS-3) (Cytokine-inducible SH2 protein 3) (CIS-3) (STAT-induced STAT inhibitor 3) (SSI-3)	9021
P01242	SOM2_HUMAN	reviewed	Growth hormone variant (GH-V) (Growth hormone 2) (Placenta-specific growth hormone)	2689
P01241	SOMA_HUMAN	reviewed	Somatotropin (Growth hormone) (GH) (GH-N) (Growth hormone 1) (Pituitary growth hormone)	2688
Q9NRA0	SPHK2_HUMAN	reviewed	Sphingosine kinase 2 (SK 2) (SPK 2) (EC 2.7.1.91)	56848

Q14765	STAT4_HUMAN	reviewed	Signal transducer and activator of transcription 4	6775
Q9H169	STMN4_HUMAN	reviewed	Stathmin-4 (Stathmin-like protein B3) (RB3)	81551
Q9C0C2	TB182_HUMAN	reviewed	182 kDa tankyrase-1-binding protein	85456
P42680	TEC_HUMAN	reviewed	Tyrosine-protein kinase Tec (EC 2.7.10.2)	7006
O95271	TNKS1_HUMAN	reviewed	Tankyrase-1 (TANK1) (EC 2.4.2.30) (ADP-ribosyltransferase diphtheria toxin-like 5) (ARTD5) (Poly [ADP-ribose] polymerase 5A) (TNKS-1) (TRF1-interacting ankyrin-related ADP-ribose polymerase) (Tankyrase I)	8658
Q9H2K2	TNKS2_HUMAN	reviewed	Tankyrase-2 (TANK2) (EC 2.4.2.30) (ADP-ribosyltransferase diphtheria toxin-like 6) (ARTD6) (Poly [ADP-ribose] polymerase 5B) (TNKS-2) (TRF1-interacting ankyrin-related ADP-ribose polymerase 2) (Tankyrase II) (Tankyrase-like protein) (Tankyrase-related protein)	80351
Q9H1D0	TRPV6_HUMAN	reviewed	Transient receptor potential cation channel subfamily V member 6 (TrpV6) (CaT-like) (CaT-L) (Calcium transport protein 1) (CaT1) (Epithelial calcium channel 2) (ECaC2)	55503
P29597	TYK2_HUMAN	reviewed	Non-receptor tyrosine-protein kinase TYK2 (EC 2.7.10.2)	7297
O14599	VCY2_HUMAN	reviewed	Testis-specific basic protein Y 2 (Basic charge, Y-linked 2) (Variably charged protein Y 2)	442867,442868908
O60595	O60595_HUMAN	unreviewed	Interleukin 12 (Interleukin 12, P35) (Interleukin 12A (Natural killer cell stimulatory factor 1, cytotoxic lymphocyte maturation factor 1, p35)) (Interleukin 12A (Natural killer cell stimulatory factor 1, cytotoxic lymphocyte maturation factor 1, p35), isoform CRA_a)	3592
Q14620	Q14620_HUMAN		Merged into P05019.	

TABLE A.3: HCV-s64 Protein Annotation and GeneIDs

Entry	Entry name	Status	Protein names	Cross-reference GENEID
Q96F63	CCD97_HUMAN	reviewed	Coiled-coil domain-containing protein 97	90324
P05813	CRBA1_HUMAN	reviewed	Beta-crystallin A3 [Cleaved into: Beta-crystallin A3, isoform A1, Delta4 form Beta-crystallin A3, isoform A1, Delta7 form Beta-crystallin A3, isoform A1, Delta8 form]	1411
P53672	CRBA2_HUMAN	reviewed	Beta-crystallin A2 (Beta-A2 crystallin)	1412
P53673	CRBA4_HUMAN	reviewed	Beta-crystallin A4 (Beta-A4 crystallin)	1413
P53674	CRBB1_HUMAN	reviewed	Beta-crystallin B1 (Beta-B1 crystallin)	1414
P43320	CRBB2_HUMAN	reviewed	Beta-crystallin B2 (Beta-B2 crystallin) (Beta-crystallin Bp)	1415
P26998	CRBB3_HUMAN	reviewed	Beta-crystallin B3 (Beta-B3 crystallin) [Cleaved into: Beta-crystallin B3, N-terminally processed]	1417
P07315	CRGC_HUMAN	reviewed	Gamma-crystallin C (Gamma-C-crystallin) (Gamma-crystallin 2-1) (Gamma-crystallin 3)	1420
P02489	CRYAA_HUMAN	reviewed	Alpha-crystallin A chain (Heat shock protein beta-4) (HspB4) [Cleaved into: Alpha-crystallin A chain, short form]	1409
P02511	CRYAB_HUMAN	reviewed	Alpha-crystallin B chain (Alpha(B)-crystallin) (Heat shock protein beta-5) (HspB5) (Renal carcinoma antigen NY-REN-27) (Rosenthal fiber component)	1410
Q9Y6W6	DUS10_HUMAN	reviewed	Dual specificity protein phosphatase 10 (EC 3.1.3.16) (EC 3.1.3.48) (Mitogen-activated protein kinase phosphatase 5) (MAP kinase phosphatase 5) (MKP-5)	11221
Q9BY84	DUS16_HUMAN	reviewed	Dual specificity protein phosphatase 16 (EC 3.1.3.16) (EC 3.1.3.48) (Mitogen-activated protein kinase phosphatase 7) (MAP kinase phosphatase 7) (MKP-7)	80824

P28562	DUS1_HUMAN	reviewed	Dual specificity protein phosphatase 1 (EC 3.1.3.16) (EC 3.1.3.48) (Dual specificity protein phosphatase hVH1) (Mitogen-activated protein kinase phosphatase 1) (MAP kinase phosphatase 1) (MKP-1) (Protein-tyrosine phosphatase CL100)	1843
Q05923	DUS2_HUMAN	reviewed	Dual specificity protein phosphatase 2 (EC 3.1.3.16) (EC 3.1.3.48) (Dual specificity protein phosphatase PAC-1)	1844
Q13115	DUS4_HUMAN	reviewed	Dual specificity protein phosphatase 4 (EC 3.1.3.16) (EC 3.1.3.48) (Dual specificity protein phosphatase hVH2) (Mitogen-activated protein kinase phosphatase 2) (MAP kinase phosphatase 2) (MKP-2)	1846
Q16828	DUS6_HUMAN	reviewed	Dual specificity protein phosphatase 6 (EC 3.1.3.16) (EC 3.1.3.48) (Dual specificity protein phosphatase PYST1) (Mitogen-activated protein kinase phosphatase 3) (MAP kinase phosphatase 3) (MKP-3)	1848
Q13202	DUS8_HUMAN	reviewed	Dual specificity protein phosphatase 8 (EC 3.1.3.16) (EC 3.1.3.48) (Dual specificity protein phosphatase hVH-5)	1850
Q99956	DUS9_HUMAN	reviewed	Dual specificity protein phosphatase 9 (EC 3.1.3.16) (EC 3.1.3.48) (Mitogen-activated protein kinase phosphatase 4) (MAP kinase phosphatase 4) (MKP-4)	1852
P28324	ELK4_HUMAN	reviewed	ETS domain-containing protein Elk-4 (Serum response factor accessory protein 1) (SAP-1) (SRF accessory protein 1)	2005
Q6PJG2	EMSA1_HUMAN	reviewed	ELM2 and SANT domain-containing protein 1	91748
P50549	ETV1_HUMAN	reviewed	ETS translocation variant 1 (Ets-related protein 81)	2115

P04792	HSPB1_HUMAN	reviewed	Heat shock protein beta-1 (HspB1) (28 kDa heat shock protein) (Estrogen-regulated 24 kDa protein) (Heat shock 27 kDa protein) (HSP 27) (Stress-responsive protein 27) (SRP27)	3315
Q16082	HSPB2_HUMAN	reviewed	Heat shock protein beta-2 (HspB2) (DMPK-binding protein) (MKBP)	3316
O14558	HSPB6_HUMAN	reviewed	Heat shock protein beta-6 (HspB6) (Heat shock 20 kDa-like protein p20)	126393
Q9UBY9	HSPB7_HUMAN	reviewed	Heat shock protein beta-7 (HspB7) (Cardiovascular heat shock protein) (cvHsp)	27129
Q9UJY1	HSPB8_HUMAN	reviewed	Heat shock protein beta-8 (HspB8) (Alpha-crystallin C chain) (E2-induced gene 1 protein) (Protein kinase H11) (Small stress protein-like protein HSP22)	26353
O75676	KS6A4_HUMAN	reviewed	Ribosomal protein S6 kinase alpha-4 (S6K-alpha-4) (EC 2.7.11.1) (90 kDa ribosomal protein S6 kinase 4) (Nuclear mitogen- and stress-activated protein kinase 2) (Ribosomal protein kinase B) (RSKB)	8986
O75582	KS6A5_HUMAN	reviewed	Ribosomal protein S6 kinase alpha-5 (S6K-alpha-5) (EC 2.7.11.1) (90 kDa ribosomal protein S6 kinase 5) (Nuclear mitogen- and stress-activated protein kinase 1) (RSK-like protein kinase) (RSKL)	9252
P33241	LSP1_HUMAN	reviewed	Lymphocyte-specific protein 1 (47 kDa actin-binding protein) (52 kDa phosphoprotein) (pp52) (Lymphocyte-specific antigen WP34)	4046
P49137	MAPK2_HUMAN	reviewed	MAP kinase-activated protein kinase 2 (MAPK-activated protein kinase 2) (MAPKAP kinase 2) (MAPKAP-K2) (MAPKAPK-2) (MK-2) (MK2) (EC 2.7.11.1)	9261

Q16644	MAPK3_HUMAN	reviewed	MAP kinase-activated protein kinase 3 (MAPK-activated protein kinase 3) (MAPKAP kinase 3) (MAPKAP-K3) (MAPKAPK-3) (MK-3) (EC 2.7.11.1) (Chromosome 3p kinase) (3pK)	7867
Q8IW41	MAPK5_HUMAN	reviewed	MAP kinase-activated protein kinase 5 (MAPK-activated protein kinase 5) (MAPKAP kinase 5) (MAPKAP-K5) (MAPKAPK-5) (MK-5) (MK5) (EC 2.7.11.1) (p38-regulated/activated protein kinase) (PRAK)	8550
Q15759	MK11_HUMAN	reviewed	Mitogen-activated protein kinase 11 (MAP kinase 11) (MAPK 11) (EC 2.7.11.24) (Mitogen-activated protein kinase p38 beta) (MAP kinase p38 beta) (p38b) (Stress-activated protein kinase 2b) (SAPK2b) (p38-2)	5600
P53778	MK12_HUMAN	reviewed	Mitogen-activated protein kinase 12 (MAP kinase 12) (MAPK 12) (EC 2.7.11.24) (Extracellular signal-regulated kinase 6) (ERK-6) (Mitogen-activated protein kinase p38 gamma) (MAP kinase p38 gamma) (Stress-activated protein kinase 3)	6300
Q16539	MK14_HUMAN	reviewed	Mitogen-activated protein kinase 14 (MAP kinase 14) (MAPK 14) (EC 2.7.11.24) (Cytokine suppressive anti-inflammatory drug-binding protein) (CSAID-binding protein) (CSBP) (MAP kinase MXI2) (MAX-interacting protein 2) (Mitogen-activated protein kinase p38 alpha) (MAP kinase p38 alpha) (Stress-activated protein kinase 2a) (SAPK2a)	1432
Q8NEM7	SP20H_HUMAN	reviewed	Transcription factor SPT20 homolog (p38-interacting protein) (p38IP)	55578

Q07352	TISB_HUMAN	reviewed	Zinc finger protein 36, C3H1 type-like 1 (Butyrate response factor 1) (EGF-response factor 1) (ERF-1) (Protein TIS11B)	677
P52746	ZN142_HUMAN	reviewed	Zinc finger protein 142 (HA4654)	7701
O43257	ZNHI1_HUMAN	reviewed	Zinc finger HIT domain-containing protein 1 (Cyclin-G1-binding protein 1) (Zinc finger protein subfamily 4A member 1) (p18 Hamlet)	10467
Q1RMC8	Q1RMC8_HUMAN	unreviewed	ROBO1 protein	6091

TABLE A.4: HCV-s43 Protein Annotation and GeneIDs

Entry	**Entry name**	**Status**	**Protein names**	**Cross-reference GENEID**
Q96N21	AP4AT_HUMAN	reviewed	AP-4 complex accessory subunit tepsin (ENTH domain-containing protein 2) (Epsin for AP-4) (Tetra-epsin)	146705
Q9BZE3	BARH1_HUMAN	reviewed	BarH-like 1 homeobox protein	56751
P24863	CCNC_HUMAN	reviewed	Cyclin-C (SRB11 homolog) (hSRB11)	892
Q9BWU1	CDK19_HUMAN	reviewed	Cyclin-dependent kinase 19 (EC 2.7.11.22) (CDC2-related protein kinase 6) (Cell division cycle 2-like protein kinase 6) (Cell division protein kinase 19) (Cyclin-dependent kinase 11) (Death-preventing kinase)	23097
P49336	CDK8_HUMAN	reviewed	Cyclin-dependent kinase 8 (EC 2.7.11.22) (EC 2.7.11.23) (Cell division protein kinase 8) (Mediator complex subunit CDK8) (Mediator of RNA polymerase II transcription subunit CDK8) (Protein kinase K35)	1024
Q6IAN0	DRS7B_HUMAN	reviewed	Dehydrogenase/reductase SDR family member 7B (EC 1.1.-.-)	25979
Q9UPW0	FOXJ3_HUMAN	reviewed	Forkhead box protein J3	22887
Q9H0H0	INT2_HUMAN	reviewed	Integrator complex subunit 2 (Int2)	57508

Q68E01	INT3_HUMAN	reviewed	Integrator complex subunit 3 (Int3) (SOSS complex subunit A) (Sensor of single-strand DNA complex subunit A) (SOSS-A) (Sensor of ssDNA subunit A)	65123
Q7L273	KCTD9_HUMAN	reviewed	BTB/POZ domain-containing protein KCTD9	54793
O94953	KDM4B_HUMAN	reviewed	Lysine-specific demethylase 4B (EC 1.14.11.-) (JmjC domain-containing histone demethylation protein 3B) (Jumonji domain-containing protein 2B)	23030
Q9UBF1	MAGC2_HUMAN	reviewed	Melanoma-associated antigen C2 (Cancer/testis antigen 10) (CT10) (Hepatocellular carcinoma-associated antigen 587) (MAGE-C2 antigen) (MAGE-E1 antigen)	51438
Q71F56	MD13L_HUMAN	reviewed	Mediator of RNA polymerase II transcription subunit 13-like (Mediator complex subunit 13-like) (Thyroid hormone receptor-associated protein 2) (Thyroid hormone receptor-associated protein complex 240 kDa component-like)	23389
Q9BTT4	MED10_HUMAN	reviewed	Mediator of RNA polymerase II transcription subunit 10 (Mediator complex subunit 10) (Transformation-related gene 17 protein) (TRG-17) (Transformation-related gene 20 protein) (TRG-20)	84246
Q9P086	MED11_HUMAN	reviewed	Mediator of RNA polymerase II transcription subunit 11 (Mediator complex subunit 11)	400569
Q93074	MED12_HUMAN	reviewed	Mediator of RNA polymerase II transcription subunit 12 (Activator-recruited cofactor 240 kDa component) (ARC240) (CAG repeat protein 45) (Mediator complex subunit 12) (OPA-containing protein) (Thyroid hormone receptor-associated protein complex 230 kDa component) (Trap230) (Trinucleotide repeat-containing gene 11 protein)	9968

Q9UHV7	MED13_HUMAN	reviewed	Mediator of RNA polymerase II transcription subunit 13 (Activator-recruited cofactor 250 kDa component) (ARC250) (Mediator complex subunit 13) (Thyroid hormone receptor-associated protein 1) (Thyroid hormone receptor-associated protein complex 240 kDa component) (Trap240) (Vitamin D3 receptor-interacting protein complex component DRIP250) (DRIP250)	9969
O60244	MED14_HUMAN	reviewed	Mediator of RNA polymerase II transcription subunit 14 (Activator-recruited cofactor 150 kDa component) (ARC150) (Cofactor required for Sp1 transcriptional activation subunit 2) (CRSP complex subunit 2) (Mediator complex subunit 14) (RGR1 homolog) (hRGR1) (Thyroid hormone receptor-associated protein complex 170 kDa component) (Trap170) (Transcriptional coactivator CRSP150) (Vitamin D3 receptor-interacting protein complex 150 kDa component) (DRIP150)	9282
Q9Y2X0	MED16_HUMAN	reviewed	Mediator of RNA polymerase II transcription subunit 16 (Mediator complex subunit 16) (Thyroid hormone receptor-associated protein 5) (Thyroid hormone receptor-associated protein complex 95 kDa component) (Trap95) (Vitamin D3 receptor-interacting protein complex 92 kDa component) (DRIP92)	10025

Q9NVC6	MED17_HUMAN	reviewed	Mediator of RNA polymerase II transcription subunit 17 (Activator-recruited cofactor 77 kDa component) (ARC77) (Cofactor required for Sp1 transcriptional activation subunit 6) (CRSP complex subunit 6) (Mediator complex subunit 17) (Thyroid hormone receptor-associated protein complex 80 kDa component) (Trap80) (Transcriptional coactivator CRSP77) (Vitamin D3 receptor-interacting protein complex 80 kDa component) (DRIP80)	9440
Q9BUE0	MED18_HUMAN	reviewed	Mediator of RNA polymerase II transcription subunit 18 (Mediator complex subunit 18) (p28b)	54797
A0JLT2	MED19_HUMAN	reviewed	Mediator of RNA polymerase II transcription subunit 19 (Lung cancer metastasis-related protein 1) (Mediator complex subunit 19)	219541
Q15648	MED1_HUMAN	reviewed	Mediator of RNA polymerase II transcription subunit 1 (Activator-recruited cofactor 205 kDa component) (ARC205) (Mediator complex subunit 1) (Peroxisome proliferator-activated receptor-binding protein) (PBP) (PPAR-binding protein) (Thyroid hormone receptor-associated protein complex 220 kDa component) (Trap220) (Thyroid receptor-interacting protein 2) (TR-interacting protein 2) (TRIP-2) (Vitamin D receptor-interacting protein complex component DRIP205) (p53 regulatory protein RB18A)	5469
Q9H944	MED20_HUMAN	reviewed	Mediator of RNA polymerase II transcription subunit 20 (Mediator complex subunit 20) (TRF-proximal protein homolog) (hTRFP)	9477

Q13503	MED21_HUMAN	reviewed	Mediator of RNA polymerase II transcription subunit 21 (Mediator complex subunit 21) (RNA polymerase II holoenzyme component SRB7) (RNAPII complex component SRB7) (hSrb7)	9412
Q15528	MED22_HUMAN	reviewed	Mediator of RNA polymerase II transcription subunit 22 (Mediator complex subunit 22) (Surfeit locus protein 5) (Surf-5)	6837
Q9ULK4	MED23_HUMAN	reviewed	Mediator of RNA polymerase II transcription subunit 23 (Activator-recruited cofactor 130 kDa component) (ARC130) (Cofactor required for Sp1 transcriptional activation subunit 3) (CRSP complex subunit 3) (Mediator complex subunit 23) (Protein sur-2 homolog) (hSur-2) (Transcriptional coactivator CRSP130) (Vitamin D3 receptor-interacting protein complex 130 kDa component) (DRIP130)	9439
O75448	MED24_HUMAN	reviewed	Mediator of RNA polymerase II transcription subunit 24 (Activator-recruited cofactor 100 kDa component) (ARC100) (Cofactor required for Sp1 transcriptional activation subunit 4) (CRSP complex subunit 4) (Mediator complex subunit 24) (Thyroid hormone receptor-associated protein 4) (Thyroid hormone receptor-associated protein complex 100 kDa component) (Trap100) (hTRAP100) (Vitamin D3 receptor-interacting protein complex 100 kDa component) (DRIP100)	9862
Q71SY5	MED25_HUMAN	reviewed	Mediator of RNA polymerase II transcription subunit 25 (Activator interaction domain-containing protein 1) (Activator-recruited cofactor 92 kDa component) (ARC92) (Mediator complex subunit 25) (p78)	81857

O95402	MED26_HUMAN	reviewed	Mediator of RNA polymerase II transcription subunit 26 (Activator-recruited cofactor 70 kDa component) (ARC70) (Cofactor required for Sp1 transcriptional activation subunit 7) (CRSP complex subunit 7) (Mediator complex subunit 26) (Transcriptional coactivator CRSP70)	9441
Q6P2C8	MED27_HUMAN	reviewed	Mediator of RNA polymerase II transcription subunit 27 (Cofactor required for Sp1 transcriptional activation subunit 8) (CRSP complex subunit 8) (Mediator complex subunit 27) (P37 TRAP/SMC-C/PC2 subunit) (Transcriptional coactivator CRSP34)	9442
Q9H204	MED28_HUMAN	reviewed	Mediator of RNA polymerase II transcription subunit 28 (Endothelial-derived protein 1) (Mediator complex subunit 28) (Merlin and Grb2-interacting cytoskeletal protein) (Magicin) (Tumor angiogenesis marker EG-1)	80306
Q9NX70	MED29_HUMAN	reviewed	Mediator of RNA polymerase II transcription subunit 29 (Intersex-like protein) (Mediator complex subunit 29)	55588
Q96HR3	MED30_HUMAN	reviewed	Mediator of RNA polymerase II transcription subunit 30 (Mediator complex subunit 30) (TRAP/Mediator complex component TRAP25) (Thyroid hormone receptor-associated protein 6) (Thyroid hormone receptor-associated protein complex 25 kDa component) (Trap25)	90390
Q9NPJ6	MED4_HUMAN	reviewed	Mediator of RNA polymerase II transcription subunit 4 (Activator-recruited cofactor 36 kDa component) (ARC36) (Mediator complex subunit 4) (TRAP/SMC-C/PC2 subunit p36 subunit) (Vitamin D3 receptor-interacting protein complex 36 kDa component) (DRIP36)	29079

O75586	MED6_HUMAN	reviewed	Mediator of RNA polymerase II transcription subunit 6 (Activator-recruited cofactor 33 kDa component) (ARC33) (Mediator complex subunit 6) (hMed6) (Renal carcinoma antigen NY-REN-28)	10001
O43513	MED7_HUMAN	reviewed	Mediator of RNA polymerase II transcription subunit 7 (hMED7) (Activator-recruited cofactor 34 kDa component) (ARC34) (Cofactor required for Sp1 transcriptional activation subunit 9) (CRSP complex subunit 9) (Mediator complex subunit 7) (RNA polymerase transcriptional regulation mediator subunit 7 homolog) (Transcriptional coactivator CRSP33)	9443
Q96G25	MED8_HUMAN	reviewed	Mediator of RNA polymerase II transcription subunit 8 (Activator-recruited cofactor 32 kDa component) (ARC32) (Mediator complex subunit 8)	112950
Q9NWA0	MED9_HUMAN	reviewed	Mediator of RNA polymerase II transcription subunit 9 (Mediator complex subunit 9)	55090
Q9H4Z3	PCIF1_HUMAN	reviewed	Phosphorylated CTD-interacting factor 1	63935
P24928	RPB1_HUMAN	reviewed	DNA-directed RNA polymerase II subunit RPB1 (RNA polymerase II subunit B1) (EC 2.7.7.6) (DNA-directed RNA polymerase II subunit A) (DNA-directed RNA polymerase III largest subunit) (RNA-directed RNA polymerase II subunit RPB1) (EC 2.7.7.48)	5430
Q7RTU7	SCX_HUMAN	reviewed	Basic helix-loop-helix transcription factor scleraxis (Class A basic helix-loop-helix protein 41) (bHLHa41) (Class A basic helix-loop-helix protein 48) (bHLHa48)	100129885,642658
P48436	SOX9_HUMAN	reviewed	Transcription factor SOX-9	6662

Q9H668	STN1_HUMAN	reviewed	CST complex subunit STN1 (Oligonucleotide/oligosaccharide-binding fold-containing protein 1) (Suppressor of cdc thirteen homolog)	79991
O95379	TFIP8_HUMAN	reviewed	Tumor necrosis factor alpha-induced protein 8 (TNF alpha-induced protein 8) (Head and neck tumor and metastasis-related protein) (MDC-3.13) (NF-kappa-B-inducible DED-containing protein) (NDED) (SCC-S2) (TNF-induced protein GG2-1)	25816
Q8TB05	UBAD1_HUMAN	reviewed	UBA-like domain-containing protein 1	124402
Q96K76	UBP47_HUMAN	reviewed	Ubiquitin carboxyl-terminal hydrolase 47 (EC 3.4.19.12) (Deubiquitinating enzyme 47) (Ubiquitin thioesterase 47) (Ubiquitin-specific-processing protease 47)	55031
Q70CQ1	UBP49_HUMAN	reviewed	Ubiquitin carboxyl-terminal hydrolase 49 (EC 3.4.19.12) (Deubiquitinating enzyme 49) (Ubiquitin thioesterase 49) (Ubiquitin-specific-processing protease 49)	25862
Q16587	ZNF74_HUMAN	reviewed	Zinc finger protein 74 (Zinc finger protein 520) (hZNF7)	7625
Q9H2M2	Q9H2M2_HUMAN	unreviewed	Estrogen receptor alpha	

TABLE A.5: Combi-s52 Protein Annotation and GeneIDs

Entry	**Entry name**	**Status**	**Protein names**	**Cross-reference GENEID**
Entry	Entry name	Status	Protein names	Cross-reference (GENEID)
Q96NW4	ANR27_HUMAN	reviewed	Ankyrin repeat domain-containing protein 27 (VPS9 domain-containing protein)	84079

O43307	ARHG9_HUMAN	reviewed	Rho guanine nucleotide exchange factor 9 (Collybistin) (PEM-2 homolog) (Rac/Cdc42 guanine nucleotide exchange factor 9)	23229
Q9NR48	ASH1L_HUMAN	reviewed	Histone-lysine N-methyltransferase ASH1L (EC 2.1.1.43) (ASH1-like protein) (huASH1) (Absent small and homeotic disks protein 1 homolog) (Lysine N-methyltransferase 2H)	55870
Q9P0P8	CF203_HUMAN	reviewed	Uncharacterized protein C6orf203	51250
Q9NWM3	CUED1_HUMAN	reviewed	CUE domain-containing protein 1	404093
Q7LFL8	CXXC5_HUMAN	reviewed	CXXC-type zinc finger protein 5 (CF5) (Putative MAPK-activating protein PM08) (Putative NF-kappa-B-activating protein 102) (Retinoid-inducible nuclear factor) (RINF)	51523
Q8TB52	FBX30_HUMAN	reviewed	F-box only protein 30	84085
P69892	HBG2_HUMAN	reviewed	Hemoglobin subunit gamma-2 (Gamma-2-globin) (Hb F Ggamma) (Hemoglobin gamma-2 chain) (Hemoglobin gamma-G chain)	3048
P31273	HXC8_HUMAN	reviewed	Homeobox protein Hox-C8 (Homeobox protein Hox-3A)	3224
Q9H160	ING2_HUMAN	reviewed	Inhibitor of growth protein 2 (Inhibitor of growth 1-like protein) (ING1Lp) (p32) (p33ING2)	3622
Q96PE3	INP4A_HUMAN	reviewed	Type I inositol 3,4-bisphosphate 4-phosphatase (EC 3.1.3.66) (Inositol polyphosphate 4-phosphatase type I)	3631
Q9Y2U8	MAN1_HUMAN	reviewed	Inner nuclear membrane protein Man1 (LEM domain-containing protein 3)	23592
Q9ULH7	MKL2_HUMAN	reviewed	MKL/myocardin-like protein 2 (Megakaryoblastic leukemia 2) (Myocardin-related transcription factor B) (MRTF-B)	57496
Q9NXD2	MTMRA_HUMAN	reviewed	Myotubularin-related protein 10	54893

Q92886	NGN1_HUMAN	reviewed	Neurogenin-1 (NGN-1) (Class A basic helix-loop-helix protein 6) (bHLHa6) (Neurogenic basic-helix-loop-helix protein) (Neurogenic differentiation factor 3) (NeuroD3)	4762
Q8WUA2	PPIL4_HUMAN	reviewed	Peptidyl-prolyl cis-trans isomerase-like 4 (PPIase) (EC 5.2.1.8) (Cyclophilin-like protein PPIL4) (Rotamase PPIL4)	85313
P51817	PRKX_HUMAN	reviewed	cAMP-dependent protein kinase catalytic subunit PRKX (PrKX) (Protein kinase X) (Protein kinase X-linked) (Serine/threonine-protein kinase PRKX) (EC 2.7.11.1) (Protein kinase PKX1)	5613
Q15771	RAB30_HUMAN	reviewed	Ras-related protein Rab-30	27314
Q9BWF3	RBM4_HUMAN	reviewed	RNA-binding protein 4 (Lark homolog) (hLark) (RNA-binding motif protein 4) (RNA-binding motif protein 4a)	5936
Q6ZNA4	RN111_HUMAN	reviewed	E3 ubiquitin-protein ligase Arkadia (EC 6.3.2.-) (RING finger protein 111)	54778
Q9NUM3	S39A9_HUMAN	reviewed	Zinc transporter ZIP9 (Solute carrier family 39 member 9) (Zrt- and Irt-like protein 9) (ZIP-9)	55334
O75995	SASH3_HUMAN	reviewed	SAM and SH3 domain-containing protein 3 (SH3 protein expressed in lymphocytes homolog)	54440
Q15797	SMAD1_HUMAN	reviewed	Mothers against decapentaplegic homolog 1 (MAD homolog 1) (Mothers against DPP homolog 1) (JV4-1) (Mad-related protein 1) (SMAD family member 1) (SMAD 1) (Smad1) (hSMAD1) (Transforming growth factor-beta-signaling protein 1) (BSP-1)	4086
Q99717	SMAD5_HUMAN	reviewed	Mothers against decapentaplegic homolog 5 (MAD homolog 5) (Mothers against DPP homolog 5) (JV5-1) (SMAD family member 5) (SMAD 5) (Smad5) (hSmad5)	4090

O43541	SMAD6_HUMAN	reviewed	Mothers against decapentaplegic homolog 6 (MAD homolog 6) (Mothers against DPP homolog 6) (SMAD family member 6) (SMAD 6) (Smad6) (hSMAD6)	4091
O15105	SMAD7_HUMAN	reviewed	Mothers against decapentaplegic homolog 7 (MAD homolog 7) (Mothers against DPP homolog 7) (Mothers against decapentaplegic homolog 8) (MAD homolog 8) (Mothers against DPP homolog 8) (SMAD family member 7) (SMAD 7) (Smad7) (hSMAD7)	4092
Q9HCE7	SMUF1_HUMAN	reviewed	E3 ubiquitin-protein ligase SMURF1 (hSMURF1) (EC 6.3.2.-) (SMAD ubiquitination regulatory factor 1) (SMAD-specific E3 ubiquitin-protein ligase 1)	57154
Q9HAU4	SMUF2_HUMAN	reviewed	E3 ubiquitin-protein ligase SMURF2 (hSMURF2) (EC 6.3.2.-) (SMAD ubiquitination regulatory factor 2) (SMAD-specific E3 ubiquitin-protein ligase 2)	64750
P35711	SOX5_HUMAN	reviewed	Transcription factor SOX-5	6660
Q9BT81	SOX7_HUMAN	reviewed	Transcription factor SOX-7	83595
Q9Y3F4	STRAP_HUMAN	reviewed	Serine-threonine kinase receptor-associated protein (MAP activator with WD repeats) (UNR-interacting protein) (WD-40 repeat protein PT-WD)	11171
O95625	ZBT11_HUMAN	reviewed	Zinc finger and BTB domain-containing protein 11	27107
Q8NCP5	ZBT44_HUMAN	reviewed	Zinc finger and BTB domain-containing protein 44 (BTB/POZ domain-containing protein 15) (Zinc finger protein 851)	29068
Q9NYG2	ZDHC3_HUMAN	reviewed	Palmitoyltransferase ZDHHC3 (EC 2.3.1.-) (Protein DHHC1) (Zinc finger DHHC domain-containing protein 3) (DHHC-3) (Zinc finger protein 373)	51304
Q9Y2H8	ZN510_HUMAN	reviewed	Zinc finger protein 510	22869

P36508	ZNF76_HUMAN	reviewed	Zinc finger protein 76 (Zinc finger protein 523)	7629
P17098	ZNF8_HUMAN	reviewed	Zinc finger protein 8 (Zinc finger protein HF.18)	7554
Q8NAM6	ZSCA4_HUMAN	reviewed	Zinc finger and SCAN domain-containing protein 4 (Zinc finger protein 494)	201516

TABLE A.6: Combi-s46 Protein Annotation and GeneIDs

Entry	**Entry name**	**Status**	**Protein names**	**Cross-reference GENEID**
Entry	Entry name	Status	Protein names	Cross-reference (GENEID)
P78314	3BP2_HUMAN	reviewed	SH3 domain-binding protein 2 (3BP-2)	6452
P50406	5HT6R_HUMAN	reviewed	5-hydroxytryptamine receptor 6 (5-HT-6) (5-HT6) (Serotonin receptor 6)	3362
P42684	ABL2_HUMAN	reviewed	Abelson tyrosine-protein kinase 2 (EC 2.7.10.2) (Abelson murine leukemia viral oncogene homolog 2) (Abelson-related gene protein) (Tyrosine-protein kinase ARG)	27
Q9UM73	ALK_HUMAN	reviewed	ALK tyrosine kinase receptor (EC 2.7.10.1) (Anaplastic lymphoma kinase) (CD antigen CD246)	238
Q9ULH1	ASAP1_HUMAN	reviewed	Arf-GAP with SH3 domain, ANK repeat and PH domain-containing protein 1 (130 kDa phosphatidylinositol 4,5-bisphosphate-dependent ARF1 GTPase-activating protein) (ADP-ribosylation factor-directed GTPase-activating protein 1) (ARF GTPase-activating protein 1) (Development and differentiation-enhancing factor 1) (DEF-1) (Differentiation-enhancing factor 1) (PIP2-dependent ARF1 GAP)	50807

P56945	BCAR1_HUMAN	reviewed	Breast cancer anti-estrogen resistance protein 1 (CRK-associated substrate) (Cas scaffolding protein family member 1) (p130cas)	9564
P11274	BCR_HUMAN	reviewed	Breakpoint cluster region protein (EC 2.7.11.1) (Renal carcinoma antigen NY-REN-26)	613
P51451	BLK_HUMAN	reviewed	Tyrosine-protein kinase Blk (EC 2.7.10.2) (B lymphocyte kinase) (p55-Blk)	640
Q8WV28	BLNK_HUMAN	reviewed	B-cell linker protein (B-cell adapter containing a SH2 domain protein) (B-cell adapter containing a Src homology 2 domain protein) (Cytoplasmic adapter protein) (Src homology 2 domain-containing leukocyte protein of 65 kDa) (SLP-65)	29760
Q06187	BTK_HUMAN	reviewed	Tyrosine-protein kinase BTK (EC 2.7.10.2) (Agammaglobulinemia tyrosine kinase) (ATK) (B-cell progenitor kinase) (BPK) (Bruton tyrosine kinase)	695
Q7Z6A9	BTLA_HUMAN	reviewed	B- and T-lymphocyte attenuator (B- and T-lymphocyte-associated protein) (CD antigen CD272)	151888
P04040	CATA_HUMAN	reviewed	Catalase (EC 1.11.1.6)	847
Q13191	CBLB_HUMAN	reviewed	E3 ubiquitin-protein ligase CBL-B (EC 6.3.2.-) (Casitas B-lineage lymphoma proto-oncogene b) (RING finger protein 56) (SH3-binding protein CBL-B) (Signal transduction protein CBL-B)	868
P22681	CBL_HUMAN	reviewed	E3 ubiquitin-protein ligase CBL (EC 6.3.2.-) (Casitas B-lineage lymphoma proto-oncogene) (Proto-oncogene c-Cbl) (RING finger protein 55) (Signal transduction protein CBL)	867
P15391	CD19_HUMAN	reviewed	B-lymphocyte antigen CD19 (B-lymphocyte surface antigen B4) (Differentiation antigen CD19) (T-cell surface antigen Leu-12) (CD antigen CD19)	930

P20273	CD22_HUMAN	reviewed	B-cell receptor CD22 (B-lymphocyte cell adhesion molecule) (BL-CAM) (Sialic acid-binding Ig-like lectin 2) (Siglec-2) (T-cell surface antigen Leu-14) (CD antigen CD22)	933
P10747	CD28_HUMAN	reviewed	T-cell-specific surface glycoprotein CD28 (TP44) (CD antigen CD28)	940
Q9Y5K6	CD2AP_HUMAN	reviewed	CD2-associated protein (Adapter protein CMS) (Cas ligand with multiple SH3 domains)	23607
P20138	CD33_HUMAN	reviewed	Myeloid cell surface antigen CD33 (Sialic acid-binding Ig-like lectin 3) (Siglec-3) (gp67) (CD antigen CD33)	945
P04234	CD3D_HUMAN	reviewed	T-cell surface glycoprotein CD3 delta chain (T-cell receptor T3 delta chain) (CD antigen CD3d)	915
P07766	CD3E_HUMAN	reviewed	T-cell surface glycoprotein CD3 epsilon chain (T-cell surface antigen T3/Leu-4 epsilon chain) (CD antigen CD3e)	916
P20963	CD3Z_HUMAN	reviewed	T-cell surface glycoprotein CD3 zeta chain (T-cell receptor T3 zeta chain) (CD antigen CD247)	919
P06127	CD5_HUMAN	reviewed	T-cell surface glycoprotein CD5 (Lymphocyte antigen T1/Leu-1) (CD antigen CD5)	921
P21854	CD72_HUMAN	reviewed	B-cell differentiation antigen CD72 (Lyb-2) (CD antigen CD72)	971
P11912	CD79A_HUMAN	reviewed	B-cell antigen receptor complex-associated protein alpha chain (Ig-alpha) (MB-1 membrane glycoprotein) (Membrane-bound immunoglobulin-associated protein) (Surface IgM-associated protein) (CD antigen CD79a)	973
P40259	CD79B_HUMAN	reviewed	B-cell antigen receptor complex-associated protein beta chain (B-cell-specific glycoprotein B29) (Ig-beta) (Immunoglobulin-associated B29 protein) (CD antigen CD79b)	974

P33681	CD80_HUMAN	reviewed	T-lymphocyte activation antigen CD80 (Activation B7-1 antigen) (BB1) (CTLA-4 counter-receptor B7.1) (B7) (CD antigen CD80)	941
P42081	CD86_HUMAN	reviewed	T-lymphocyte activation antigen CD86 (Activation B7-2 antigen) (B70) (BU63) (CTLA-4 counter-receptor B7.2) (FUN-1) (CD antigen CD86)	942
P01732	CD8A_HUMAN	reviewed	T-cell surface glycoprotein CD8 alpha chain (T-lymphocyte differentiation antigen T8/Leu-2) (CD antigen CD8a)	925
P10966	CD8B_HUMAN	reviewed	T-cell surface glycoprotein CD8 beta chain (CD antigen CD8b)	926
Q9NSE2	CISH_HUMAN	reviewed	Cytokine-inducible SH2-containing protein (CIS) (CIS-1) (Protein G18) (Suppressor of cytokine signaling) (SOCS)	1154
P46109	CRKL_HUMAN	reviewed	Crk-like protein	1399
P46108	CRK_HUMAN	reviewed	Adapter molecule crk (Proto-oncogene c-Crk) (p38)	1398
P07333	CSF1R_HUMAN	reviewed	Macrophage colony-stimulating factor 1 receptor (CSF-1 receptor) (CSF-1-R) (CSF-1R) (M-CSF-R) (EC 2.7.10.1) (Proto-oncogene c-Fms) (CD antigen CD115)	1436
Q99062	CSF3R_HUMAN	reviewed	Granulocyte colony-stimulating factor receptor (G-CSF receptor) (G-CSF-R) (CD antigen CD114)	1441
P41240	CSK_HUMAN	reviewed	Tyrosine-protein kinase CSK (EC 2.7.10.2) (C-Src kinase) (Protein-tyrosine kinase CYL)	1445
P16410	CTLA4_HUMAN	reviewed	Cytotoxic T-lymphocyte protein 4 (Cytotoxic T-lymphocyte-associated antigen 4) (CTLA-4) (CD antigen CD152)	1493
P58505	CU058_HUMAN	reviewed	Uncharacterized protein C21orf58	54058

Q9UN19	DAPP1_HUMAN	reviewed	Dual adapter for phosphotyrosine and 3-phosphotyrosine and 3-phosphoinositide (hDAPP1) (B lymphocyte adapter protein Bam32) (B-cell adapter molecule of 32 kDa)	27071
Q9UJU6	DBNL_HUMAN	reviewed	Drebrin-like protein (Cervical SH3P7) (Cervical mucin-associated protein) (Drebrin-F) (HPK1-interacting protein of 55 kDa) (HIP-55) (SH3 domain-containing protein 7)	28988
Q08345	DDR1_HUMAN	reviewed	Epithelial discoidin domain-containing receptor 1 (Epithelial discoidin domain receptor 1) (EC 2.7.10.1) (CD167 antigen-like family member A) (Cell adhesion kinase) (Discoidin receptor tyrosine kinase) (HGK2) (Mammary carcinoma kinase 10) (MCK-10) (Protein-tyrosine kinase 3A) (Protein-tyrosine kinase RTK-6) (TRK E) (Tyrosine kinase DDR) (Tyrosine-protein kinase CAK) (CD antigen CD167a)	780
Q6P3S1	DEN1B_HUMAN	reviewed	DENN domain-containing protein 1B (Connecdenn 2) (Protein FAM31B)	163486
Q99704	DOK1_HUMAN	reviewed	Docking protein 1 (Downstream of tyrosine kinase 1) (p62(dok)) (pp62)	1796
O60496	DOK2_HUMAN	reviewed	Docking protein 2 (Downstream of tyrosine kinase 2) (p56(dok-2))	9046
Q8TEW6	DOK4_HUMAN	reviewed	Docking protein 4 (Downstream of tyrosine kinase 4) (Insulin receptor substrate 5) (IRS-5) (IRS5)	55715
Q9P104	DOK5_HUMAN	reviewed	Docking protein 5 (Downstream of tyrosine kinase 5) (Insulin receptor substrate 6) (IRS-6) (IRS6)	55816
Q6PKX4	DOK6_HUMAN	reviewed	Docking protein 6 (Downstream of tyrosine kinase 6)	220164

P20827	EFNA1_HUMAN	reviewed	Ephrin-A1 (EPH-related receptor tyrosine kinase ligand 1) (LERK-1) (Immediate early response protein B61) (Tumor necrosis factor alpha-induced protein 4) (TNF alpha-induced protein 4) [Cleaved into: Ephrin-A1, secreted form]	1942
P00533	EGFR_HUMAN	reviewed	Epidermal growth factor receptor (EC 2.7.10.1) (Proto-oncogene c-ErbB-1) (Receptor tyrosine-protein kinase erbB-1)	1956
P01133	EGF_HUMAN	reviewed	Pro-epidermal growth factor (EGF) [Cleaved into: Epidermal growth factor (Urogastrone)]	1950
P29317	EPHA2_HUMAN	reviewed	Ephrin type-A receptor 2 (EC 2.7.10.1) (Epithelial cell kinase) (Tyrosine-protein kinase receptor ECK)	1969
P54762	EPHB1_HUMAN	reviewed	Ephrin type-B receptor 1 (EC 2.7.10.1) (ELK) (EPH tyrosine kinase 2) (EPH-like kinase 6) (EK6) (hEK6) (Neuronally-expressed EPH-related tyrosine kinase) (NET) (Tyrosine-protein kinase receptor EPH-2)	2047
P54753	EPHB3_HUMAN	reviewed	Ephrin type-B receptor 3 (EC 2.7.10.1) (EPH-like tyrosine kinase 2) (EPH-like kinase 2) (Embryonic kinase 2) (EK2) (hEK2) (Tyrosine-protein kinase TYRO6)	2049
P19235	EPOR_HUMAN	reviewed	Erythropoietin receptor (EPO-R)	2057
P01588	EPO_HUMAN	reviewed	Erythropoietin (Epoetin)	2056
P04626	ERBB2_HUMAN	reviewed	Receptor tyrosine-protein kinase erbB-2 (EC 2.7.10.1) (Metastatic lymph node gene 19 protein) (MLN 19) (Proto-oncogene Neu) (Proto-oncogene c-ErbB-2) (Tyrosine kinase-type cell surface receptor HER2) (p185erbB2) (CD antigen CD340)	2064

P21860	ERBB3_HUMAN	reviewed	Receptor tyrosine-protein kinase erbB-3 (EC 2.7.10.1) (Proto-oncogene-like protein c-ErbB-3) (Tyrosine kinase-type cell surface receptor HER3)	2065
Q15303	ERBB4_HUMAN	reviewed	Receptor tyrosine-protein kinase erbB-4 (EC 2.7.10.1) (Proto-oncogene-like protein c-ErbB-4) (Tyrosine kinase-type cell surface receptor HER4) (p180erbB4) [Cleaved into: ERBB4 intracellular domain (4ICD) (E4ICD) (s80HER4)]	2066
Q9UJM3	ERRFI_HUMAN	reviewed	ERBB receptor feedback inhibitor 1 (Mitogen-inducible gene 6 protein) (MIG-6)	54206
Q05397	FAK1_HUMAN	reviewed	Focal adhesion kinase 1 (FADK 1) (EC 2.7.10.2) (Focal adhesion kinase-related nonkinase) (FRNK) (Protein phosphatase 1 regulatory subunit 71) (PPP1R71) (Protein-tyrosine kinase 2) (p125FAK) (pp125FAK)	5747
Q14289	FAK2_HUMAN	reviewed	Protein-tyrosine kinase 2-beta (EC 2.7.10.2) (Calcium-dependent tyrosine kinase) (CADTK) (Calcium-regulated non-receptor proline-rich tyrosine kinase) (Cell adhesion kinase beta) (CAK-beta) (CAKB) (Focal adhesion kinase 2) (FADK 2) (Proline-rich tyrosine kinase 2) (Related adhesion focal tyrosine kinase) (RAFTK)	2185
P12319	FCERA_HUMAN	reviewed	High affinity immunoglobulin epsilon receptor subunit alpha (Fc-epsilon RI-alpha) (FcERI) (IgE Fc receptor subunit alpha)	2205
Q01362	FCERB_HUMAN	reviewed	High affinity immunoglobulin epsilon receptor subunit beta (FcERI) (Fc epsilon receptor I beta-chain) (IgE Fc receptor subunit beta) (Membrane-spanning 4-domains subfamily A member 2)	2206

P30273	FCERG_HUMAN	reviewed	High affinity immunoglobulin epsilon receptor subunit gamma (Fc receptor gamma-chain) (FcRgamma) (Fc-epsilon RI-gamma) (IgE Fc receptor subunit gamma) (FceRI gamma)	2207
P12318	FCG2A_HUMAN	reviewed	Low affinity immunoglobulin gamma Fc region receptor II-a (IgG Fc receptor II-a) (CDw32) (Fc-gamma RII-a) (Fc-gamma-RIIa) (FcRII-a) (CD antigen CD32)	2212
P31994	FCG2B_HUMAN	reviewed	Low affinity immunoglobulin gamma Fc region receptor II-b (IgG Fc receptor II-b) (CDw32) (Fc-gamma RII-b) (Fc-gamma-RIIb) (FcRII-b) (CD antigen CD32)	2213
P12314	FCGR1_HUMAN	reviewed	High affinity immunoglobulin gamma Fc receptor I (IgG Fc receptor I) (Fc-gamma RI) (FcRI) (Fc-gamma RIA) (Fcgam-maRIa) (CD antigen CD64)	2209
P09769	FGR_HUMAN	reviewed	Tyrosine-protein kinase Fgr (EC 2.7.10.2) (Gardner-Rasheed feline sarcoma viral (v-fgr) oncogene homolog) (Proto-oncogene c-Fgr) (p55-Fgr) (p58-Fgr) (p58c-Fgr)	2268
P36888	FLT3_HUMAN	reviewed	Receptor-type tyrosine-protein kinase FLT3 (EC 2.7.10.1) (FL cytokine receptor) (Fetal liver kinase-2) (FLK-2) (Fms-like tyrosine kinase 3) (FLT-3) (Stem cell tyrosine kinase 1) (STK-1) (CD antigen CD135)	2322
Q8WU20	FRS2_HUMAN	reviewed	Fibroblast growth factor receptor substrate 2 (FGFR substrate 2) (FGFR-signaling adaptor SNT) (Suc1-associated neurotrophic factor target 1) (SNT-1)	10818
O15117	FYB_HUMAN	reviewed	FYN-binding protein (Adhesion and degranulation promoting adaptor protein) (ADAP) (FYB-120/130) (p120/p130) (FYN-T-binding protein) (SLAP-130) (SLP-76-associated phosphoprotein)	2533

P06241	FYN_HUMAN	reviewed	Tyrosine-protein kinase Fyn (EC 2.7.10.2) (Proto-oncogene Syn) (Proto-oncogene c-Fyn) (Src-like kinase) (SLK) (p59-Fyn)	2534
Q13480	GAB1_HUMAN	reviewed	GRB2-associated-binding protein 1 (GRB2-associated binder 1) (Growth factor receptor bound protein 2-associated protein 1)	2549
Q9UQC2	GAB2_HUMAN	reviewed	GRB2-associated-binding protein 2 (GRB2-associated binder 2) (Growth factor receptor bound protein 2-associated protein 2) (pp100)	9846
Q9H706	GAREM_HUMAN	reviewed	GRB2-associated and regulator of MAPK protein (GRB2-associated and regulator of MAPK1)	64762
Q14393	GAS6_HUMAN	reviewed	Growth arrest-specific protein 6 (GAS-6) (AXL receptor tyrosine kinase ligand)	2621
P39905	GDNF_HUMAN	reviewed	Glial cell line-derived neurotrophic factor (hGDNF) (Astrocyte-derived trophic factor) (ATF)	2668
P56159	GFRA1_HUMAN	reviewed	GDNF family receptor alpha-1 (GDNF receptor alpha-1) (GDNFR-alpha-1) (GFR-alpha-1) (RET ligand 1) (TGF-beta-related neurotrophic factor receptor 1)	2674
O00451	GFRA2_HUMAN	reviewed	GDNF family receptor alpha-2 (GDNF receptor alpha-2) (GDNFR-alpha-2) (GFR-alpha-2) (GDNF receptor beta) (GDNFR-beta) (Neurturin receptor alpha) (NRTNR-alpha) (NTNR-alpha) (RET ligand 2) (TGF-beta-related neurotrophic factor receptor 2)	2675
P10912	GHR_HUMAN	reviewed	Growth hormone receptor (GH receptor) (Somatotropin receptor) [Cleaved into: Growth hormone-binding protein (GH-binding protein) (GHBP) (Serum-binding protein)]	2690
Q9HCN6	GPVI_HUMAN	reviewed	Platelet glycoprotein VI (GPVI) (Glycoprotein 6)	51206

O75791	GRAP2_HUMAN	reviewed	GRB2-related adapter protein 2 (Adapter protein GRID) (GRB-2-like protein) (GRB2L) (GRBLG) (GRBX) (Grf40 adapter protein) (Grf-40) (Growth factor receptor-binding protein) (Hematopoietic cell-associated adapter protein GrpL) (P38) (Protein GADS) (SH3-SH2-SH3 adapter Mona)	9402
Q13588	GRAP_HUMAN	reviewed	GRB2-related adapter protein	10750
Q13322	GRB10_HUMAN	reviewed	Growth factor receptor-bound protein 10 (GRB10 adapter protein) (Insulin receptor-binding protein Grb-IR)	2887
Q14449	GRB14_HUMAN	reviewed	Growth factor receptor-bound protein 14 (GRB14 adapter protein)	2888
P62993	GRB2_HUMAN	reviewed	Growth factor receptor-bound protein 2 (Adapter protein GRB2) (Protein Ash) (SH2/SH3 adapter GRB2)	2885
Q14451	GRB7_HUMAN	reviewed	Growth factor receptor-bound protein 7 (B47) (Epidermal growth factor receptor GRB-7) (GRB7 adapter protein)	2886
P08631	HCK_HUMAN	reviewed	Tyrosine-protein kinase HCK (EC 2.7.10.2) (Hematopoietic cell kinase) (Hemopoietic cell kinase) (p59-HCK/p60-HCK) (p59Hck) (p61Hck)	3055
P13747	HLAE_HUMAN	reviewed	HLA class I histocompatibility antigen, alpha chain E (MHC class I antigen E)	3133
P08069	IGF1R_HUMAN	reviewed	Insulin-like growth factor 1 receptor (EC 2.7.10.1) (Insulin-like growth factor I receptor) (IGF-I receptor) (CD antigen CD221) [Cleaved into: Insulin-like growth factor 1 receptor alpha chain Insulin-like growth factor 1 receptor beta chain]	3480
P01589	IL2RA_HUMAN	reviewed	Interleukin-2 receptor subunit alpha (IL-2 receptor subunit alpha) (IL-2-RA) (IL-2R subunit alpha) (IL2-RA) (TAC antigen) (p55) (CD antigen CD25)	3559

P14784	IL2RB_HUMAN	reviewed	Interleukin-2 receptor subunit beta (IL-2 receptor subunit beta) (IL-2R subunit beta) (IL-2RB) (High affinity IL-2 receptor subunit beta) (p70-75) (p75) (CD antigen CD122)	3560
P31785	IL2RG_HUMAN	reviewed	Cytokine receptor common subunit gamma (Interleukin-2 receptor subunit gamma) (IL-2 receptor subunit gamma) (IL-2R subunit gamma) (IL-2RG) (gammaC) (p64) (CD antigen CD132)	3561
P60568	IL2_HUMAN	reviewed	Interleukin-2 (IL-2) (T-cell growth factor) (TCGF) (Aldesleukin)	3558
P26951	IL3RA_HUMAN	reviewed	Interleukin-3 receptor subunit alpha (IL-3 receptor subunit alpha) (IL-3R subunit alpha) (IL-3R-alpha) (IL-3RA) (CD antigen CD123)	3563
P32927	IL3RB_HUMAN	reviewed	Cytokine receptor common subunit beta (CDw131) (GM-CSF/IL-3/IL-5 receptor common beta subunit) (CD antigen CD131)	1439
P24394	IL4RA_HUMAN	reviewed	Interleukin-4 receptor subunit alpha (IL-4 receptor subunit alpha) (IL-4R subunit alpha) (IL-4R-alpha) (IL-4RA) (CD antigen CD124) [Cleaved into: Soluble interleukin-4 receptor subunit alpha (Soluble IL-4 receptor subunit alpha) (Soluble IL-4R-alpha) (sIL4Ralpha/prot) (IL-4-binding protein) (IL4-BP)]	3566
P08887	IL6RA_HUMAN	reviewed	Interleukin-6 receptor subunit alpha (IL-6 receptor subunit alpha) (IL-6R subunit alpha) (IL-6R-alpha) (IL-6RA) (IL-6R 1) (Membrane glycoprotein 80) (gp80) (CD antigen CD126)	3570

P40189	IL6RB_HUMAN	reviewed	Interleukin-6 receptor subunit beta (IL-6 receptor subunit beta) (IL-6R subunit beta) (IL-6R-beta) (IL-6RB) (CDw130) (Interleukin-6 signal transducer) (Membrane glycoprotein 130) (gp130) (Oncostatin-M receptor subunit alpha) (CD antigen CD130)	3572
P17181	INAR1_HUMAN	reviewed	Interferon alpha/beta receptor 1 (IFN-R-1) (IFN-alpha/beta receptor 1) (Cytokine receptor class-II member 1) (Cytokine receptor family 2 member 1) (CRF2-1) (Type I interferon receptor 1)	3454
P06213	INSR_HUMAN	reviewed	Insulin receptor (IR) (EC 2.7.10.1) (CD antigen CD220) [Cleaved into: Insulin receptor subunit alpha Insulin receptor subunit beta]	3643
P35568	IRS1_HUMAN	reviewed	Insulin receptor substrate 1 (IRS-1)	3667
Q9Y4H2	IRS2_HUMAN	reviewed	Insulin receptor substrate 2 (IRS-2)	8660
O14654	IRS4_HUMAN	reviewed	Insulin receptor substrate 4 (IRS-4) (160 kDa phosphotyrosine protein) (py160) (Phosphoprotein of 160 kDa) (pp160)	8471
Q08881	ITK_HUMAN	reviewed	Tyrosine-protein kinase ITK/TSK (EC 2.7.10.2) (Interleukin-2-inducible T-cell kinase) (IL-2-inducible T-cell kinase) (Kinase EMT) (T-cell-specific kinase) (Tyrosine-protein kinase Lyk)	3702
P23458	JAK1_HUMAN	reviewed	Tyrosine-protein kinase JAK1 (EC 2.7.10.2) (Janus kinase 1) (JAK-1)	3716
O60674	JAK2_HUMAN	reviewed	Tyrosine-protein kinase JAK2 (EC 2.7.10.2) (Janus kinase 2) (JAK-2)	3717
P52333	JAK3_HUMAN	reviewed	Tyrosine-protein kinase JAK3 (EC 2.7.10.2) (Janus kinase 3) (JAK-3) (Leukocyte janus kinase) (L-JAK)	3718
Q96N16	JKIP1_HUMAN	reviewed	Janus kinase and microtubule-interacting protein 1 (GABA-B receptor-binding protein) (Multiple alpha-helices and RNA-linker protein 1) (Marlin-1)	152789

Q07666	KHDR1_HUMAN	reviewed	KH domain-containing, RNA-binding, signal transduction-associated protein 1 (GAP-associated tyrosine phosphoprotein p62) (Src-associated in mitosis 68 kDa protein) (Sam68) (p21 Ras GTPase-activating protein-associated p62) (p68)	10657
P43628	KI2L3_HUMAN	reviewed	Killer cell immunoglobulin-like receptor 2DL3 (CD158 antigen-like family member B2) (KIR-023GB) (Killer inhibitory receptor cl 2-3) (MHC class I NK cell receptor) (NKAT2a) (NKAT2b) (Natural killer-associated transcript 2) (NKAT-2) (p58 natural killer cell receptor clone CL-6) (p58 NK receptor CL-6) (p58.2 MHC class-I-specific NK receptor) (CD antigen CD158b2)	3804
P10721	KIT_HUMAN	reviewed	Mast/stem cell growth factor receptor Kit (SCFR) (EC 2.7.10.1) (Piebald trait protein) (PBT) (Proto-oncogene c-Kit) (Tyrosine-protein kinase Kit) (p145 c-kit) (v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog) (CD antigen CD117)	3815
Q04759	KPCT_HUMAN	reviewed	Protein kinase C theta type (EC 2.7.11.13) (nPKC-theta)	5588
P43405	KSYK_HUMAN	reviewed	Tyrosine-protein kinase SYK (EC 2.7.10.2) (Spleen tyrosine kinase) (p72-Syk)	6850
Q6GTX8	LAIR1_HUMAN	reviewed	Leukocyte-associated immunoglobulin-like receptor 1 (LAIR-1) (hLAIR1) (CD antigen CD305)	3903
O43561	LAT_HUMAN	reviewed	Linker for activation of T-cells family member 1 (36 kDa phospho-tyrosine adapter protein) (pp36) (p36-38)	27040

Q8IWV1	LAX1_HUMAN	reviewed	Lymphocyte transmembrane adapter 1 (Linker for activation of X cells) (Membrane-associated adapter protein LAX)	54900
P06239	LCK_HUMAN	reviewed	Tyrosine-protein kinase Lck (EC 2.7.10.2) (Leukocyte C-terminal Src kinase) (LSK) (Lymphocyte cell-specific protein-tyrosine kinase) (Protein YT16) (Proto-oncogene Lck) (T cell-specific protein-tyrosine kinase) (p56-LCK)	3932
Q13094	LCP2_HUMAN	reviewed	Lymphocyte cytosolic protein 2 (SH2 domain-containing leukocyte protein of 76 kDa) (SLP-76 tyrosine phosphoprotein) (SLP76)	3937
P48357	LEPR_HUMAN	reviewed	Leptin receptor (LEP-R) (HuB219) (OB receptor) (OB-R) (CD antigen CD295)	3953
P29376	LTK_HUMAN	reviewed	Leukocyte tyrosine kinase receptor (EC 2.7.10.1) (Protein tyrosine kinase 1)	4058
Q14210	LY6D_HUMAN	reviewed	Lymphocyte antigen 6D (Ly-6D) (E48 antigen)	8581
P07948	LYN_HUMAN	reviewed	Tyrosine-protein kinase Lyn (EC 2.7.10.2) (Lck/Yes-related novel protein tyrosine kinase) (V-yes-1 Yamaguchi sarcoma viral related oncogene homolog) (p53Lyn) (p56Lyn)	4067
Q92918	M4K1_HUMAN	reviewed	Mitogen-activated protein kinase kinase kinase kinase 1 (EC 2.7.11.1) (Hematopoietic progenitor kinase) (MAPK/ERK kinase kinase kinase 1) (MEK kinase kinase 1) (MEKKK 1)	11184
Q9Y4K4	M4K5_HUMAN	reviewed	Mitogen-activated protein kinase kinase kinase kinase 5 (EC 2.7.11.1) (Kinase homologous to SPS1/STE20) (KHS) (MAPK/ERK kinase kinase kinase 5) (MEK kinase kinase 5) (MEKKK 5)	11183
Q13477	MADCA_HUMAN	reviewed	Mucosal addressin cell adhesion molecule 1 (MAdCAM-1) (hMAdCAM-1)	8174

Q12866	MERTK_HUMAN	reviewed	Tyrosine-protein kinase Mer (EC 2.7.10.1) (Proto-oncogene c-Mer) (Receptor tyrosine kinase MerTK)	10461
P08581	MET_HUMAN	reviewed	Hepatocyte growth factor receptor (HGF receptor) (EC 2.7.10.1) (HGF/SF receptor) (Proto-oncogene c-Met) (Scatter factor receptor) (SF receptor) (Tyrosine-protein kinase Met)	4233
P03956	MMP1_HUMAN	reviewed	Interstitial collagenase (EC 3.4.24.7) (Fibroblast collagenase) (Matrix metalloproteinase-1) (MMP-1) [Cleaved into: 22 kDa interstitial collagenase 27 kDa interstitial collagenase]	4312
P22894	MMP8_HUMAN	reviewed	Neutrophil collagenase (EC 3.4.24.34) (Matrix metalloproteinase-8) (MMP-8) (PMNL collagenase) (PMNL-CL)	4317
O95297	MPZL1_HUMAN	reviewed	Myelin protein zero-like protein 1 (Protein zero-related)	9019
O43639	NCK2_HUMAN	reviewed	Cytoplasmic protein NCK2 (Growth factor receptor-bound protein 4) (NCK adaptor protein 2) (Nck-2) (SH2/SH3 adaptor protein NCK-beta)	8440
O14786	NRP1_HUMAN	reviewed	Neuropilin-1 (Vascular endothelial cell growth factor 165 receptor) (CD antigen CD304)	8829
Q99748	NRTN_HUMAN	reviewed	Neurturin	4902
Q9GZY6	NTAL_HUMAN	reviewed	Linker for activation of T-cells family member 2 (Linker for activation of B-cells) (Membrane-associated adapter molecule) (Non-T-cell activation linker) (Williams-Beuren syndrome chromosomal region 15 protein) (Williams-Beuren syndrome chromosomal region 5 protein)	7462

P04629	NTRK1_HUMAN	reviewed	High affinity nerve growth factor receptor (EC 2.7.10.1) (Neurotrophic tyrosine kinase receptor type 1) (TRK1-transforming tyrosine kinase protein) (Tropomyosin-related kinase A) (Tyrosine kinase receptor) (Tyrosine kinase receptor A) (Trk-A) (gp140trk) (p140-TrkA)	4914
O00750	P3C2B_HUMAN	reviewed	Phosphatidylinositol 4-phosphate 3-kinase C2 domain-containing subunit beta (PI3K-C2-beta) (PtdIns-3-kinase C2 subunit beta) (EC 2.7.1.154) (C2-PI3K) (Phosphoinositide 3-kinase-C2-beta)	5287
Q92569	P55G_HUMAN	reviewed	Phosphatidylinositol 3-kinase regulatory subunit gamma (PI3-kinase regulatory subunit gamma) (PI3K regulatory subunit gamma) (PtdIns-3-kinase regulatory subunit gamma) (Phosphatidylinositol 3-kinase 55 kDa regulatory subunit gamma) (PI3-kinase subunit p55-gamma) (PtdIns-3-kinase regulatory subunit p55-gamma) (p55PIK)	8503
P27986	P85A_HUMAN	reviewed	Phosphatidylinositol 3-kinase regulatory subunit alpha (PI3-kinase regulatory subunit alpha) (PI3K regulatory subunit alpha) (PtdIns-3-kinase regulatory subunit alpha) (Phosphatidylinositol 3-kinase 85 kDa regulatory subunit alpha) (PI3-kinase subunit p85-alpha) (PtdIns-3-kinase regulatory subunit p85-alpha)	5295
O00459	P85B_HUMAN	reviewed	Phosphatidylinositol 3-kinase regulatory subunit beta (PI3-kinase regulatory subunit beta) (PI3K regulatory subunit beta) (PtdIns-3-kinase regulatory subunit beta) (Phosphatidylinositol 3-kinase 85 kDa regulatory subunit beta) (PI3-kinase subunit p85-beta) (PtdIns-3-kinase regulatory subunit p85-beta)	5296

P16284	PECA1_HUMAN	reviewed	Platelet endothelial cell adhesion molecule (PECAM-1) (EndoCAM) (GPIIA') (PECA1) (CD antigen CD31)	5175
O75420	PERQ1_HUMAN	reviewed	PERQ amino acid-rich with GYF domain-containing protein 1 (GRB10-interacting GYF protein 1)	64599
P50542	PEX5_HUMAN	reviewed	Peroxisomal targeting signal 1 receptor (PTS1 receptor) (PTS1R) (PTS1-BP) (Peroxin-5) (Peroxisomal C-terminal targeting signal import receptor) (Peroxisome receptor 1)	5830
P16234	PGFRA_HUMAN	reviewed	Platelet-derived growth factor receptor alpha (PDGF-R-alpha) (PDGFR-alpha) (EC 2.7.10.1) (Alpha platelet-derived growth factor receptor) (Alpha-type platelet-derived growth factor receptor) (CD140 antigen-like family member A) (CD140a antigen) (Platelet-derived growth factor alpha receptor) (Platelet-derived growth factor receptor 2) (PDGFR-2) (CD antigen CD140a)	5156
P09619	PGFRB_HUMAN	reviewed	Platelet-derived growth factor receptor beta (PDGF-R-beta) (PDGFR-beta) (EC 2.7.10.1) (Beta platelet-derived growth factor receptor) (Beta-type platelet-derived growth factor receptor) (CD140 antigen-like family member B) (Platelet-derived growth factor receptor 1) (PDGFR-1) (CD antigen CD140b)	5159
Q9NWQ8	PHAG1_HUMAN	reviewed	Phosphoprotein associated with glycosphingolipid-enriched microdomains 1 (Csk-binding protein) (Transmembrane adapter protein PAG) (Transmembrane phosphoprotein Cbp)	55824
Q9UKJ1	PILRA_HUMAN	reviewed	Paired immunoglobulin-like type 2 receptor alpha (Cell surface receptor FDF03) (Inhibitory receptor PILR-alpha)	29992

P42336	PK3CA_HUMAN	reviewed	Phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit alpha isoform (PI3-kinase subunit alpha) (PI3K-alpha) (PI3Kalpha) (PtdIns-3-kinase subunit alpha) (EC 2.7.1.153) (Phosphatidylinositol 4,5-bisphosphate 3-kinase 110 kDa catalytic subunit alpha) (PtdIns-3-kinase subunit p110-alpha) (p110alpha) (Phosphoinositide-3-kinase catalytic alpha polypeptide) (Serine/threonine protein kinase PIK3CA) (EC 2.7.11.1)	5290
P42338	PK3CB_HUMAN	reviewed	Phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit beta isoform (PI3-kinase subunit beta) (PI3K-beta) (PI3Kbeta) (PtdIns-3-kinase subunit beta) (EC 2.7.1.153) (Phosphatidylinositol 4,5-bisphosphate 3-kinase 110 kDa catalytic subunit beta) (PtdIns-3-kinase subunit p110-beta) (p110beta)	5291
P19174	PLCG1_HUMAN	reviewed	1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase gamma-1 (EC 3.1.4.11) (PLC-148) (Phosphoinositide phospholipase C-gamma-1) (Phospholipase C-II) (PLC-II) (Phospholipase C-gamma-1) (PLC-gamma-1)	5335
P16885	PLCG2_HUMAN	reviewed	1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase gamma-2 (EC 3.1.4.11) (Phosphoinositide phospholipase C-gamma-2) (Phospholipase C-IV) (PLC-IV) (Phospholipase C-gamma-2) (PLC-gamma-2)	5336
P49763	PLGF_HUMAN	reviewed	Placenta growth factor (PlGF)	5228
Q96NZ9	PRAP1_HUMAN	reviewed	Proline-rich acidic protein 1 (Epididymis tissue protein Li 178) (Uterine-specific proline-rich acidic protein)	118471
P16471	PRLR_HUMAN	reviewed	Prolactin receptor (PRL-R)	5618
O60542	PSPN_HUMAN	reviewed	Persephin (PSP)	5623

Q13882	PTK6_HUMAN	reviewed	Protein-tyrosine kinase 6 (EC 2.7.10.2) (Breast tumor kinase) (Tyrosine-protein kinase BRK)	5753
Q06124	PTN11_HUMAN	reviewed	Tyrosine-protein phosphatase non-receptor type 11 (EC 3.1.3.48) (Protein-tyrosine phosphatase 1D) (PTP-1D) (Protein-tyrosine phosphatase 2C) (PTP-2C) (SH-PTP2) (SHP-2) (Shp2) (SH-PTP3)	5781
Q05209	PTN12_HUMAN	reviewed	Tyrosine-protein phosphatase non-receptor type 12 (EC 3.1.3.48) (PTP-PEST) (Protein-tyrosine phosphatase G1) (PTPG1)	5782
P18031	PTN1_HUMAN	reviewed	Tyrosine-protein phosphatase non-receptor type 1 (EC 3.1.3.48) (Protein-tyrosine phosphatase 1B) (PTP-1B)	5770
Q9Y2R2	PTN22_HUMAN	reviewed	Tyrosine-protein phosphatase non-receptor type 22 (EC 3.1.3.48) (Hematopoietic cell protein-tyrosine phosphatase 70Z-PEP) (Lymphoid phosphatase) (LyP) (PEST-domain phosphatase) (PEP)	26191
P26045	PTN3_HUMAN	reviewed	Tyrosine-protein phosphatase non-receptor type 3 (EC 3.1.3.48) (Protein-tyrosine phosphatase H1) (PTP-H1)	5774
P29350	PTN6_HUMAN	reviewed	Tyrosine-protein phosphatase non-receptor type 6 (EC 3.1.3.48) (Hematopoietic cell protein-tyrosine phosphatase) (Protein-tyrosine phosphatase 1C) (PTP-1C) (Protein-tyrosine phosphatase SHP-1) (SH-PTP1)	5777
P23467	PTPRB_HUMAN	reviewed	Receptor-type tyrosine-protein phosphatase beta (Protein-tyrosine phosphatase beta) (R-PTP-beta) (EC 3.1.3.48) (Vascular endothelial protein tyrosine phosphatase) (VE-PTP)	5787

P08575	PTPRC_HUMAN	reviewed	Receptor-type tyrosine-protein phosphatase C (EC 3.1.3.48) (Leukocyte common antigen) (L-CA) (T200) (CD antigen CD45)	5788
P23470	PTPRG_HUMAN	reviewed	Receptor-type tyrosine-protein phosphatase gamma (Protein-tyrosine phosphatase gamma) (R-PTP-gamma) (EC 3.1.3.48)	5793
Q12913	PTPRJ_HUMAN	reviewed	Receptor-type tyrosine-protein phosphatase eta (Protein-tyrosine phosphatase eta) (R-PTP-eta) (EC 3.1.3.48) (Density-enhanced phosphatase 1) (DEP-1) (HPTP eta) (Protein-tyrosine phosphatase receptor type J) (R-PTP-J) (CD antigen CD148)	5795
Q16827	PTPRO_HUMAN	reviewed	Receptor-type tyrosine-protein phosphatase O (R-PTP-O) (EC 3.1.3.48) (Glomerular epithelial protein 1) (Protein tyrosine phosphatase U2) (PTP-U2) (PTPase U2)	5800
P20936	RASA1_HUMAN	reviewed	Ras GTPase-activating protein 1 (GAP) (GTPase-activating protein) (RasGAP) (Ras p21 protein activator) (p120GAP)	5921
P07949	RET_HUMAN	reviewed	Proto-oncogene tyrosine-protein kinase receptor Ret (EC 2.7.10.1) (Cadherin family member 12) (Proto-oncogene c-Ret) [Cleaved into: Soluble RET kinase fragment Extracellular cell-membrane anchored RET cadherin 120 kDa fragment]	5979

A7KAX9	RHG32_HUMAN	reviewed	Rho GTPase-activating protein 32 (Brain-specific Rho GTPase-activating protein) (GAB-associated Cdc42/Rac GTPase-activating protein) (GC-GAP) (GTPase regulator interacting with TrkA) (Rho-type GTPase-activating protein 32) (Rho/Cdc42/Rac GTPase-activating protein RICS) (RhoGAP involved in the beta-catenin-N-cadherin and NMDA receptor signaling) (p200RhoGAP) (p250GAP)	9743
P08922	ROS1_HUMAN	reviewed	Proto-oncogene tyrosine-protein kinase ROS (EC 2.7.10.1) (Proto-oncogene c-Ros) (Proto-oncogene c-Ros-1) (Receptor tyrosine kinase c-ros oncogene 1) (c-Ros receptor tyrosine kinase)	6098
Q13905	RPGF1_HUMAN	reviewed	Rap guanine nucleotide exchange factor 1 (CRK SH3-binding GNRP) (Guanine nucleotide-releasing factor 2) (Protein C3G)	2889
P21583	SCF_HUMAN	reviewed	Kit ligand (Mast cell growth factor) (MGF) (Stem cell factor) (SCF) (c-Kit ligand) [Cleaved into: Soluble KIT ligand (sKITLG)]	4254
Q9NP31	SH22A_HUMAN	reviewed	SH2 domain-containing protein 2A (SH2 domain-containing adapter protein) (T cell-specific adapter protein) (TSAd) (VEGF receptor-associated protein)	9047
Q9NRF2	SH2B1_HUMAN	reviewed	SH2B adapter protein 1 (Pro-rich, PH and SH2 domain-containing signaling mediator) (PSM) (SH2 domain-containing protein 1B)	25970
O14492	SH2B2_HUMAN	reviewed	SH2B adapter protein 2 (Adapter protein with pleckstrin homology and Src homology 2 domains) (SH2 and PH domain-containing adapter protein APS)	10603

Q9UQQ2	SH2B3_HUMAN	reviewed	SH2B adapter protein 3 (Lymphocyte adapter protein) (Lymphocyte-specific adapter protein Lnk) (Signal transduction protein Lnk)	10019
Q96B97	SH3K1_HUMAN	reviewed	SH3 domain-containing kinase-binding protein 1 (CD2-binding protein 3) (CD2BP3) (Cbl-interacting protein of 85 kDa) (Human Src family kinase-binding protein 1) (HSB-1)	30011
Q15464	SHB_HUMAN	reviewed	SH2 domain-containing adapter protein B	6461
P29353	SHC1_HUMAN	reviewed	SHC-transforming protein 1 (SHC-transforming protein 3) (SHC-transforming protein A) (Src homology 2 domain-containing-transforming protein C1) (SH2 domain protein C1)	6464
P98077	SHC2_HUMAN	reviewed	SHC-transforming protein 2 (Protein Sck) (SHC-transforming protein B) (Src homology 2 domain-containing-transforming protein C2) (SH2 domain protein C2)	25759
Q92529	SHC3_HUMAN	reviewed	SHC-transforming protein 3 (Neuronal Shc) (N-Shc) (Protein Rai) (SHC-transforming protein C) (Src homology 2 domain-containing-transforming protein C3) (SH2 domain protein C3)	53358
Q92835	SHIP1_HUMAN	reviewed	Phosphatidylinositol 3,4,5-trisphosphate 5-phosphatase 1 (EC 3.1.3.86) (Inositol polyphosphate-5-phosphatase of 145 kDa) (SIP-145) (SH2 domain-containing inositol 5'-phosphatase 1) (SH2 domain-containing inositol phosphatase 1) (SHIP-1) (p150Ship) (hp51CN)	3635

O15357	SHIP2_HUMAN	reviewed	Phosphatidylinositol 3,4,5-trisphosphate 5-phosphatase 2 (EC 3.1.3.86) (Inositol polyphosphate phosphatase-like protein 1) (INPPL-1) (Protein 51C) (SH2 domain-containing inositol 5'-phosphatase 2) (SH2 domain-containing inositol phosphatase 2) (SHIP-2)	3636
Q9Y3P8	SIT1_HUMAN	reviewed	Signaling threshold-regulating transmembrane adapter 1 (SHP2-interacting transmembrane adapter protein) (Suppression-inducing transmembrane adapter 1) (gp30/40)	27240
Q86WV1	SKAP1_HUMAN	reviewed	Src kinase-associated phosphoprotein 1 (Src family-associated phosphoprotein 1) (Src kinase-associated phosphoprotein of 55 kDa) (SKAP-55) (pp55)	8631
Q13291	SLAF1_HUMAN	reviewed	Signaling lymphocytic activation molecule (CDw150) (IPO-3) (CD antigen CD150)	6504
Q13239	SLAP1_HUMAN	reviewed	Src-like-adapter (Src-like-adapter protein 1) (SLAP-1) (hSLAP)	6503
Q9H6Q3	SLAP2_HUMAN	reviewed	Src-like-adapter 2 (Modulator of antigen receptor signaling) (MARS) (Src-like adapter protein 2) (SLAP-2)	84174
Q92959	SO2A1_HUMAN	reviewed	Solute carrier organic anion transporter family member 2A1 (Prostaglandin transporter) (PGT) (Solute carrier family 21 member 2)	6578
O15524	SOCS1_HUMAN	reviewed	Suppressor of cytokine signaling 1 (SOCS-1) (JAK-binding protein) (JAB) (STAT-induced STAT inhibitor 1) (SSI-1) (Tec-interacting protein 3) (TIP-3)	8651
O14508	SOCS2_HUMAN	reviewed	Suppressor of cytokine signaling 2 (SOCS-2) (Cytokine-inducible SH2 protein 2) (CIS-2) (STAT-induced STAT inhibitor 2) (SSI-2)	8835

O14543	SOCS3_HUMAN	reviewed	Suppressor of cytokine signaling 3 (SOCS-3) (Cytokine-inducible SH2 protein 3) (CIS-3) (STAT-induced STAT inhibitor 3) (SSI-3)	9021
Q07889	SOS1_HUMAN	reviewed	Son of sevenless homolog 1 (SOS-1)	6654
Q07890	SOS2_HUMAN	reviewed	Son of sevenless homolog 2 (SOS-2)	6655
Q9H2V7	SPNS1_HUMAN	reviewed	Protein spinster homolog 1 (HSpin1) (Spinster-like protein 1)	83985
Q8WW59	SPRY4_HUMAN	reviewed	SPRY domain-containing protein 4	283377
O43597	SPY2_HUMAN	reviewed	Protein sprouty homolog 2 (Spry-2)	10253
P42229	STA5A_HUMAN	reviewed	Signal transducer and activator of transcription 5A	6776
P51692	STA5B_HUMAN	reviewed	Signal transducer and activator of transcription 5B	6777
Q9UGK3	STAP2_HUMAN	reviewed	Signal-transducing adaptor protein 2 (STAP-2) (Breast tumor kinase substrate) (BRK substrate)	55620
Q8IYN2	TCAL8_HUMAN	reviewed	Transcription elongation factor A protein-like 8 (TCEA-like protein 8) (Transcription elongation factor S-II protein-like 8)	90843
P42680	TEC_HUMAN	reviewed	Tyrosine-protein kinase Tec (EC 2.7.10.2)	7006
Q63HR2	TENC1_HUMAN	reviewed	Tensin-like C1 domain-containing phosphatase (EC 3.1.3.-) (C1 domain-containing phosphatase and tensin homolog) (C1-TEN) (Tensin-2)	23371
P40238	TPOR_HUMAN	reviewed	Thrombopoietin receptor (TPO-R) (Myeloproliferative leukemia protein) (Proto-oncogene c-Mpl) (CD antigen CD110)	4352
P40225	TPO_HUMAN	reviewed	Thrombopoietin (C-mpl ligand) (ML) (Megakaryocyte colony-stimulating factor) (Megakaryocyte growth and development factor) (MGDF) (Myeloproliferative leukemia virus oncogene ligand)	7066
Q6PIZ9	TRAT1_HUMAN	reviewed	T-cell receptor-associated transmembrane adapter 1 (T-cell receptor-interacting molecule) (TRIM) (pp29/30)	50852

P29597	TYK2_HUMAN	reviewed	Non-receptor tyrosine-protein kinase TYK2 (EC 2.7.10.2)	7297
P30530	UFO_HUMAN	reviewed	Tyrosine-protein kinase receptor UFO (EC 2.7.10.1) (AXL oncogene)	558
P52735	VAV2_HUMAN	reviewed	Guanine nucleotide exchange factor VAV2 (VAV-2)	7410
Q9UKW4	VAV3_HUMAN	reviewed	Guanine nucleotide exchange factor VAV3 (VAV-3)	10451
P15498	VAV_HUMAN	reviewed	Proto-oncogene vav	7409
P15692	VEGFA_HUMAN	reviewed	Vascular endothelial growth factor A (VEGF-A) (Vascular permeability factor) (VPF)	7422
P17948	VGFR1_HUMAN	reviewed	Vascular endothelial growth factor receptor 1 (VEGFR-1) (EC 2.7.10.1) (Fms-like tyrosine kinase 1) (FLT-1) (Tyrosine-protein kinase FRT) (Tyrosine-protein kinase receptor FLT) (FLT) (Vascular permeability factor receptor)	2321
P35968	VGFR2_HUMAN	reviewed	Vascular endothelial growth factor receptor 2 (VEGFR-2) (EC 2.7.10.1) (Fetal liver kinase 1) (FLK-1) (Kinase insert domain receptor) (KDR) (Protein-tyrosine kinase receptor flk-1) (CD antigen CD309)	3791
Q9UPY6	WASF3_HUMAN	reviewed	Wiskott-Aldrich syndrome protein family member 3 (WASP family protein member 3) (Protein WAVE-3) (Verprolin homology domain-containing protein 3)	10810
P42768	WASP_HUMAN	reviewed	Wiskott-Aldrich syndrome protein (WASp)	7454
O43516	WIPF1_HUMAN	reviewed	WAS/WASL-interacting protein family member 1 (Protein PRPL-2) (Wiskott-Aldrich syndrome protein-interacting protein) (WASP-interacting protein)	7456
P07947	YES_HUMAN	reviewed	Tyrosine-protein kinase Yes (EC 2.7.10.2) (Proto-oncogene c-Yes) (p61-Yes)	7525

P43403	ZAP70_HUMAN	reviewed	Tyrosine-protein kinase ZAP-70 (EC 2.7.10.2) (70 kDa zeta-chain associated protein) (Syk-related tyrosine kinase)	7535
Q5JNZ3	ZN311_HUMAN	reviewed	Zinc finger protein 311 (Zinc finger protein zfp-31)	282890
P17041	ZNF32_HUMAN	reviewed	Zinc finger protein 32 (C2H2-546) (Zinc finger protein KOX30)	7580
Q71UZ1	Q71UZ1_HUMAN	unreviewed	Interleukin 10 (Fragment)	
Q504X9	Q504X9_HUMAN	unreviewed	MAP1A protein (Fragment)	4130
A2NI60	A2NI60_HUMAN	unreviewed	BRE (Fragment)	
Q92970	Q92970_HUMAN	unreviewed	Zinc finger protein zfp30 (Fragment)	
Q6PYX1	Q6PYX1_HUMAN	unreviewed	Hepatitis B virus receptor binding protein (Fragment)	
Q9Y6X7	Q9Y6X7_HUMAN	unreviewed	KIAA0864 protein (Fragment)	
Q92927	Q92927_HUMAN	unreviewed	Small GTP-binding protein (Fragment)	
P78453	P78453_HUMAN	unreviewed	Tyrosine kinase (Fragment)	2268
Q5PY61	Q5PY61_HUMAN	unreviewed	Ubiquitin C splice variant	
A2N8H4	A2N8H4_HUMAN	unreviewed	TCR Ti gamma-chain V8-J2 region (Fragment)	
Q5T1S1	Q5T1S1_HUMAN		Deleted.	

TABLE A.7: Combi-s239 Protein Annotation and GeneIDs

TABLE A.8: Jaccard indices of HCV subnetworks and CORUM complexes

ComplexName	**HCV-s43**	**HCV-s64**
SNX complex (SNX1a, SNX2, SNX4, INSR)	0	0.0149
TRF1 telomere length regulation complex	0	0.0152
TRF-Rap1 complex I, 2MD	0	0.0145
Tankyrin 1-tankyrin 2-TRF1 complex	0	0.0308
IL12A homodimer complex	0	0.0156
IL12A-IL12B complex	0	0.0313
IL12A-IL12B-IL12RB1 complex	0	0.0469
IL12B-IL12RB1-IL12RB2 complex	0	0.0469
IL12A-IL12B-IL12RB2 complex	0	0.0469

IL12RB1-IL12RB2 complex	0	0.0313
JAK2-IL12RB2 complex	0	0.0313
ITGA1-ITGB1-PTPN2 complex	0	0.0152
Sam68-p85 P13K-IRS-1-IR signaling complex	0	0.0303
Prolactin (PRL) - PRL receptor (PRLR) complex	0	0.0154
PRL receptor (PRLR) dimer complex	0	0.0156
JAK2-PAFR-TYK2 complex	0	0.0308
IL-12 heterodimer complex	0	0.0313
IL-12 subunit p40 homodimer complex	0	0.0156
LMO4-gp130 complex	0	0.0147
SRCAP-associated chromatin remodeling complex	0.0192	0
BAG3-HSC70-HSPB8-CHIP complex	0.0217	0
ELK1-SRF-ELK4 complex	0.0222	0

TABLE A.9: Jaccard indices of HIV subnetworks and CORUM complexes

ComplexName	**HIV-s52**	**HIV-s66**
C complex spliceosome	0.000	0.081
Large Drosha complex	0.000	0.062
DGCR8 multiprotein complex	0.000	0.055
Spliceosome	0.000	0.050
TRA2B1-SRp30c-SRp55 complex	0.000	0.045
DCS complex (PTBP1, PTBP2, HNRPH1, HNRPF)	0.000	0.045
Emerin complex 52	0.000	0.035
CBC complex (cap binding complex)	0.000	0.030
CBC complex (cap binding complex)	0.000	0.030
HNRPF-HNRPH1 complex	0.000	0.030
SRp30c-SRp55 complex	0.000	0.030
PHAX-CBC complex (cap binding complex)	0.000	0.030
PGC-1-SRp40-SRp55-SRp75 complex	0.000	0.029
CRM1-RAN-PHAX-CBC complex (cap binding complex)	0.000	0.029
SRm160/300 complex	0.000	0.029

12S U11 snRNP	0.000	0.025
Emerin complex 25	0.000	0.025
Nop56p-associated pre-rRNA complex	0.000	0.024
NCBP-NIP1 complex	0.000	0.015
PIN1-AUF1 complex	0.000	0.015
YBX1-AKT1 complex	0.000	0.015
GR-hnRNP U complex	0.000	0.015
CTFC-TAF1 complex	0.000	0.015
CTCF-nucleophosmin complex	0.000	0.015
p23 protein complex	0.000	0.015
SF3A1-SF3A2-SF3A3 complex	0.000	0.015
SMAD3-SMAD4-CTCF protein-DNA complex	0.000	0.015
Actin-ribonucleoprotein complex (POLR2A, GTF2F1, HNRNPU)	0.019	0.015
Multiprotein complex (mRNA turnover)	0.000	0.014
PABPC1-HSPA8-HNRPD-EIF4G1 complex	0.000	0.014
Toposome	0.000	0.014
CTCF-nucleophosmin-PARP-HIS-KPNA-LMNA-TOP complex	0.000	0.014
H2AX complex, isolated from cells without IR exposure	0.000	0.013
Emerin complex 24	0.000	0.013
CF IIAm complex (Cleavage factor IIAm complex)	0.000	0.012
SNW1 complex	0.000	0.012
LARC complex (LCR-associated remodeling complex)	0.000	0.012
18S U11/U12 snRNP	0.000	0.011
CDC5L complex	0.000	0.011
17S U2 snRNP	0.000	0.010

TABLE A.10: Jaccard indices of Combi subnetworks and CORUM complexes

Complex Name	Combi_s239	Combi_ s52	Combi_s46
Emerin complex 52	0.000	0.000	0.015

Actin-ribonucleoprotein complex (POLR2A, GTF2F1, HNRNPU)	0.000	0.019	0.000
TRAP complex	0.000	0.111	0.000
CD28-transactivation complex	0.004	0.000	0.000
RNA polymerase II holoenzyme complex	0.000	0.013	0.000
RNA polymerase II core complex	0.000	0.016	0.000
CRSP complex	0.000	0.154	0.000
CRSP complex	0.000	0.154	0.000
SMCC complex	0.000	0.132	0.000
NAT complex	0.000	0.135	0.000
Mediator complex	0.000	0.556	0.000
ARC complex	0.000	0.241	0.000
CRSP complex	0.000	0.189	0.000
ARC-L complex	0.000	0.245	0.000
ARC complex	0.000	0.264	0.000
PC2 complex	0.000	0.185	0.000
SMCC complex	0.000	0.278	0.000
TFTC-type histone acetyl transferase complex	0.000	0.145	0.000
WIP-WASp-actin-myosin-IIa complex	0.008	0.000	0.000
TRAP complex	0.000	0.236	0.000
SMCC complex	0.000	0.245	0.000
DRIP complex	0.000	0.222	0.000
DRIP complex	0.000	0.222	0.000
IFNB1-IFNAR1-IFNAR2- complex	0.004	0.000	0.000
ASPP1-SAM68 complex	0.004	0.000	0.000
P2X7 receptor signalling complex	0.004	0.000	0.000
CSA-POLIIa complex	0.000	0.015	0.000
RICH1/AMOT polarity complex, Flag-Rich1 precipitated	0.008	0.000	0.000

TRAP-SMCC mediator complex	0.000	0.192	0.000
hMediator complex (MED23, CDK8, CCNC, MED7)	0.000	0.077	0.000
hMediator complex (MED23, CDK8, CCNC)	0.000	0.058	0.000
CDK8-CyclinC-Mediator complex	0.000	0.038	0.000
Mediator complex 1	0.000	0.096	0.000
Mediator complex 2	0.000	0.058	0.000
RET-Rai complex	0.008	0.000	0.000
SHC3-GAB1 complex	0.008	0.000	0.000
ARC92-Mediator complex	0.000	0.226	0.000
CRSP-Mediator 2 complex	0.000	0.170	0.000
CD8A-LCK complex	0.008	0.000	0.000
SNX complex (SNX1a, SNX2, SNX4, LEPR)	0.004	0.000	0.000
SNX complex (SNX1a, SNX2, SNX4, INSR)	0.004	0.000	0.000
SNX complex (SNX1a, SNX2, SNX4, EGFR)	0.004	0.000	0.000
SNX complex (SNX1,1a,2,4, PDGF receptor)	0.004	0.000	0.000
SMN complex	0.000	0.000	0.016
Integrator complex	0.000	0.032	0.000
DSS1 complex	0.000	0.016	0.000
Integrator-RNAPII complex	0.000	0.048	0.000
EGFR-containing signaling complex	0.013	0.000	0.000
RNA pol II containing coactivator complex Tat-SF	0.000	0.018	0.000
Nrp1-PlexinD1 complex	0.004	0.000	0.000
IL4-IL4R complex	0.004	0.000	0.000
IL4-IL4R-IL2RG complex	0.008	0.000	0.000
IL6ST-PRKCD-STAT3 complex	0.004	0.000	0.000
p300-SMAD1-STAT3 complex	0.000	0.000	0.021

G protein complex (BTK, GNG1, GNG2)	0.004	0.000	0.000
FAK-beta5 integrin complex, VEGF induced	0.004	0.000	0.000
ABL2-HRAS-RIN1 complex	0.004	0.000	0.000
IL2-IL2RA-IL2RB complex	0.013	0.000	0.000
SMN-PolII-RHA complex	0.000	0.016	0.000
SMN complex	0.000	0.000	0.018
CPSF6-ITCH-NUDT21-POLR2A-UBAP2L complex	0.000	0.018	0.000
CPSF6-EWSR1-ITCH-NUDT21-POLR2A-UBAP2L complex	0.000	0.018	0.000
CPSF6-ITCH-NUDT21-POLR2A complex	0.000	0.018	0.000
TGF-beta-receptor-SMAD7-SMURF2 complex	0.000	0.000	0.042
TGF-beta receptor I-SMAD7-SMURF1 complex	0.000	0.000	0.043
RNF11-SMURF2-STAMBP complex	0.000	0.000	0.021
CIN85-CBL-SH3GL2 complex	0.008	0.000	0.000
SH3P2/OSTF1-CBL-SRC complex	0.004	0.000	0.000
LEPR homodimer complex	0.004	0.000	0.000
AXL homodimer complex	0.004	0.000	0.000
STAT5B homodimer complex	0.004	0.000	0.000
STAT5A homodimer complex	0.004	0.000	0.000
JAK2-IL12RB2 complex	0.004	0.000	0.000
BRCA1-BARD1-POLR2A complex	0.000	0.019	0.000
ITGA6-ITGB4-FYN complex	0.004	0.000	0.000
ITGB6-FYN-FN1 complex	0.004	0.000	0.000
ITGAV-ITGB3-PXN-PTK2b complex	0.004	0.000	0.000
ITGAV-ITGB3-EGFR complex	0.004	0.000	0.000

Multiprotein complex (monoubiquitination)	0.013	0.000	0.000
CIN85-CBL-SH3GL2-EGFR complex, EGF stimulated	0.013	0.000	0.000
CIN85-SH3GL2 complex	0.004	0.000	0.000
MET-CIN85-SH3GL3-CBL complex, HGF stimulated	0.013	0.000	0.000
CIN85-SH3GL3 complex	0.004	0.000	0.000
p130Cas-ER-alpha-cSrc-kinase-PI3-kinase p85-subunit complex	0.008	0.000	0.000
CRKL-PDGFRA-CRK-RAPGEF1 complex	0.017	0.000	0.000
CIN85 complex (CIN85, CRK, BCAR1, CBL, PIK3R1, GRB2, SOS1)	0.029	0.000	0.000
GIPC1-NTRK1-RGS19 complex	0.004	0.000	0.000
NCR3-CD247 complex	0.004	0.000	0.000
ArgBP2a-CBL-PTK2B complex	0.008	0.000	0.000
ZAP70-CRKL-WIPF1-WAS complex	0.017	0.000	0.000
CRKL-WIPF1-WAS complex	0.013	0.000	0.000
CIN85-CBL complex	0.008	0.000	0.000
ERBB2-MEMO-SHC complex	0.008	0.000	0.000
LAT-PLC-gamma-1-p85-GRB2-CBL-VAV-SLP-76 signaling complex, C305 activated	0.029	0.000	0.000
Cbl-SLP-76-Grb2 complex, Fc receptor gamma-R1 stimulated	0.013	0.000	0.000
SLP-76-Cbl-Grb2-Shc complex, Fc receptor gamma-R1 stimulated	0.017	0.000	0.000
PLC-gamma-2-SLP-76-Lyn-Grb2 complex	0.017	0.000	0.000

PKC-alpha-PLD1-PLC-gamma-2 signaling complex, lacritin stimulated	0.004	0.000	0.000
BCR-ABL (p210 fusion protein)-GRB2 complex	0.004	0.000	0.000
HGF-Met complex	0.004	0.000	0.000
EGFR-CBL-GRB2 complex	0.013	0.000	0.000
PLC-gamma-1-SLP-76-SOS1-LAT complex	0.017	0.000	0.000
PDGFRA-PLC-gamma-1-PI3K-SHP-2 complex, PDGF stimulated	0.017	0.000	0.000
p56(LCK)-CAML complex	0.004	0.000	0.000
FGFR2-c-Cbl-Lyn-Fyn complex	0.013	0.000	0.000
p21(ras)GAP-Fyn-Lyn-Yes complex, thrombin stimulated	0.017	0.000	0.000
CD20-LCK-LYN-FYN-p75/80 complex, (Raji human B cell line)	0.013	0.000	0.000
CD19-Vav-PI 3-kinase (p85 subunit) complex	0.013	0.000	0.000
Sam68-p85 P13K-IRS-1-IR signaling complex	0.017	0.000	0.000
Sam68-p120GAP complex	0.008	0.000	0.000
RasGAP-AURKA/AURKB-survivin complex	0.004	0.000	0.000
POLR2A-CCNT1-CDK9-NCL-LEM6-CPSF2 complex	0.000	0.018	0.000
BRD4 complex	0.000	0.091	0.000
P-TEFb-BRD4-TRAP220 complex	0.000	0.018	0.000
CDK8-MED6-PARP1 complex	0.000	0.038	0.000
CCNC-CDK8-MED1-MED6-MED7 xcomplex	0.000	0.096	0.000
CCNC-CDK3 complex	0.000	0.019	0.000
HES1 promoter corepressor complex	0.000	0.018	0.000

HES1 promoter-Notch enhancer complex	0.000	0.048	0.000
SMAD1-P300 complex	0.000	0.000	0.021
RNA polymerase II (RNAPII)	0.000	0.015	0.000
BRCA1-core RNA polymerase II complex	0.000	0.016	0.000
PXN-ITGB5-PTK2 complex	0.004	0.000	0.000
FYB-CARMA1-BCL-10-MALT1 complex	0.004	0.000	0.000
BRCA1-RNA polymerase II complex	0.000	0.026	0.000
ING2 complex	0.000	0.000	0.018
CD20-LCK-FYN-p75/80 complex	0.008	0.000	0.000
BCR-ABL (p185 fusion protein)-GRB2 complex	0.004	0.000	0.000
BCR-ABL (p210 fusion protein)-GRB2-SOS1 complex	0.008	0.000	0.000
SHC-GRB2 complex	0.008	0.000	0.000
ITGA2b-ITGB3-CD47-FAK complex	0.004	0.000	0.000
PLC-gamma-2-Syk-LAT-FcR-gamma complex	0.017	0.000	0.000
PLC-gamma-2-Lyn-FcR-gamma complex	0.013	0.000	0.000
PLC-gamma-2-SLP-76 complex	0.008	0.000	0.000
PLC-gamma-2-LAT complex	0.008	0.000	0.000
Grb2-Sos complex, Fc receptor gamma-R1 stimulated	0.008	0.000	0.000
LAT-PLC-gamma-1-p85-GRB2-SOS signaling complex, C305 activated	0.021	0.000	0.000
Notch1-p56lck-PI3K complex	0.008	0.000	0.000
LCK-SLP76-PLC-gamma-1-LAT complex, pervanadate-activated	0.017	0.000	0.000

PLC-gamma-1-LAT-c-CBL complex, OKT3 stimulated	0.013	0.000	0.000
LAT-GRB2 complex, Fyn-mLck(KA) or Syk kinase activated	0.008	0.000	0.000
SMAD1-CBP complex	0.000	0.000	0.021
SMAD1-OAZ-HsN3 complex	0.000	0.000	0.021
SLP-76-PLC-gamma-1-ITK complex, alpha-TCR stimulated	0.013	0.000	0.000
SLP-76-PLC-gamma-1-VAV complex, alpha-TCR stimulated	0.013	0.000	0.000
CRK-BCAR1-DOCK1 complex	0.008	0.000	0.000
ITK-SLP-76 complex, anti-TCR stimulated	0.008	0.000	0.000
ITGA9-ITGB1-VEGFA complex	0.004	0.000	0.000
SMAD7-SMURF2 complex	0.000	0.000	0.043
SMAD7-SMURF1 complex	0.000	0.000	0.043
SMAD7-SMURF1-TGF-beta receptor complex	0.000	0.000	0.042
RNA polymerase II complex (RPB1, RAP74, CDK8, CYCC, SRB7, BAF190, BAF47), chromatin structure modifying	0.000	0.071	0.000
RNA polymerase II complex (CBP, PCAF, RPB1, BAF47, CYCC, CDK8), chromatin structure modifying	0.000	0.055	0.000
RNA polymerase II complex, incomplete (CBP, RPBI, PCAF, BAF47), chromatin structure modifying	0.000	0.018	0.000
RNA polymerase II complex, chromatin structure modifying	0.000	0.060	0.000
RNA polymerase II complex, chromatin structure modifying	0.000	0.050	0.000

RNA polymerase II complex, chromatin structure modifying	0.000	0.066	0.000
RNA polymerase II complex, incomplete (CDK8 complex), chromatin structure modifying	0.000	0.053	0.000
ITGA6-ITGB4-SHC1-GRB2 complex	0.008	0.000	0.000
ITGB1-NRP1 complex	0.004	0.000	0.000
Phosphatidylinositol 3-kinase (PIK3CA, PIK3R1)	0.008	0.000	0.000
MASH1 promoter-coactivator complex	0.000	0.016	0.000
CAMK2-delta-MASH1 promoter-coactivator complex	0.000	0.017	0.000
Sos1-Grb2 complex	0.008	0.000	0.000
Prolactin (PRL) - PRL receptor (PRLR) complex	0.004	0.000	0.000
PRL receptor (PRLR) dimer complex	0.004	0.000	0.000
CIN85 homotetramer complex	0.004	0.000	0.000
CIN85-BLNK complex	0.008	0.000	0.000
CIN85-c-CBL complex	0.008	0.000	0.000
PDGFRA-SHP-2 complex, PDGF stimulated	0.008	0.000	0.000
GRB2-SHP-2 complex, PDGF stimulated	0.008	0.000	0.000
Heterodimer complex (CDK9, IL6ST)	0.004	0.000	0.000
SMN complex (GEMIN6,7, UNRIP), SMN-independent intermediate	0.000	0.000	0.021
RIN1-STAM2-EGFR complex, EGF stimulated	0.004	0.000	0.000
SMURF2-SMAD2 complex, TGF(beta)-dependent	0.000	0.000	0.021

SMURF2-SMAD3 complex, TGF(beta)-dependent	0.000	0.000	0.021
SMURF2-SMAD3-SnoN complex, TGF(beta)-dependent	0.000	0.000	0.021
NRP1-VEGFR2-VEGF(165) complex	0.013	0.000	0.000
SMAD6-HOXC8 complex	0.000	0.000	0.043
SMAD6-HOXA9 complex	0.000	0.000	0.021
ZO1-(beta)cadherin-(VE)cadherin-VEGFR2 complex	0.004	0.000	0.000
SH3KBP1-CBLB-EGFR complex	0.013	0.000	0.000
Polycystin-1 multiprotein complex (ACTN1, CDH1, SRC, JUP, VCL, CTNNB1, PXN, BCAR1, PKD1, PTK2, TLN1)	0.008	0.000	0.000
JAK2-PAFR-TYK2 complex	0.008	0.000	0.000
TIAM1-EFNB1-EPHA2 complex	0.004	0.000	0.000
CAS-SRC-FAK complex	0.008	0.000	0.000
DDEF1-CTTN-PXN complex	0.004	0.000	0.000
ELMO1-DOCK1-CRKII complex	0.004	0.000	0.000
NGF-TrkA complex	0.004	0.000	0.000
EPOR receptor complex	0.004	0.000	0.000
CRKII-C3G complex	0.008	0.000	0.000
EPO-EPOR complex	0.008	0.000	0.000
Mediator complex	0.000	0.436	0.000
MED18-MED20-MED29 mediator subcomplex	0.000	0.058	0.000
LMO4-gp130 complex	0.017	0.000	0.000
CNTF-CNTFR-gp130-LIFR complex	0.004	0.000	0.000
LIFR-LIF-gp130 complex	0.004	0.000	0.000
Emerin complex 32	0.000	0.014	0.000
PSD95-FYN-NR2A complex	0.004	0.000	0.000

FARP2-NRP1-PlexinA1 complex	0.004	0.000	0.000
FARP2-NRP1-PlexinA2 complex	0.004	0.000	0.000
FARP2-NRP1-PlexinA3 complex	0.004	0.000	0.000
FARP2-NRP1-PlexinA4 complex	0.004	0.000	0.000
SEMA3C-PlexinD1-Nrp1 complex	0.004	0.000	0.000
PlexinA1-Nrp1 complex	0.004	0.000	0.000
PlexinA3-Nrp1 complex	0.004	0.000	0.000
PlexinB1-Nrp1 complex	0.004	0.000	0.000
SEMA6D-PlexinA1-NRP1 complex	0.004	0.000	0.000
VEGFA(165)-KDR-NRP1 complex	0.013	0.000	0.000
VEGFA(165)-KDR-NRP1 complex	0.004	0.000	0.000
VEGFA(165)-VEGFR2-NRP1 complex	0.013	0.000	0.000
NRP1-VEGF(165/121) complex	0.008	0.000	0.000
CIN85-SH3GL3-CBL complex	0.008	0.000	0.000
NRP1-VEGFC complex, heparin dependent	0.004	0.000	0.000
NRP1-VEGFD complex, heparin dependent	0.004	0.000	0.000
Pre-initiation complex (PIC)	0.000	0.019	0.000
PlexinA1-NRP1 complex	0.004	0.000	0.000
PlexinA1-NRP1-SEMA3A complex	0.004	0.000	0.000
PARK2-EPS15-EGFR	0.004	0.000	0.000

-

Bibliography

[1] Brass, A. L. *et al.* Identification of Host Proteins Required for HIV Infection Through a Functional Genomic Screen. *Science* **319**, 921–926 (2008). URL http://www.sciencemag.org/content/319/5865/921.

[2] König, R. *et al.* Global Analysis of Host-Pathogen Interactions that Regulate Early-Stage HIV-1 Replication. *Cell* **135**, 49–60 (2008). URL http://www.sciencedirect.com/science/article/pii/S0092867408009525.

[3] Zhou, H. *et al.* Genome-Scale RNAi Screen for Host Factors Required for HIV Replication. *Cell Host & Microbe* **4**, 495–504 (2008). URL http://www.sciencedirect.com/science/article/pii/S1931312808003302.

[4] Li, Q. *et al.* A Genome-Wide Genetic Screen for Host Factors Required for Hepatitis C Virus Propagation. *Proceedings of the National Academy of Sciences* **106**, 16410–16415 (2009). URL http://www.pnas.org/content/106/38/16410.

[5] Tai, A. W. *et al.* A Functional Genomic Screen Identifies Cellular Cofactors of Hepatitis C Virus Replication. *Cell Host & Microbe* **5**, 298–307 (2009). URL http://www.sciencedirect.com/science/article/pii/S1931312809000614.

[6] Lupberger, J. *et al.* EGFR and EphA2 are host factors for hepatitis C virus entry and possible targets for antiviral therapy. *Nature Medicine* **17**, 589–595 (2011). URL http://www.nature.com/nm/journal/v17/n5/full/nm.2341.html.

[7] Karlas, A. *et al.* Genome-wide RNAi screen identifies human host factors crucial for influenza virus replication. *Nature* **463**, 818–822 (2010). URL http://www.nature.com/nature/journal/v463/n7282/abs/nature08760.html.

[8] König, R. *et al.* Human host factors required for influenza virus replication. *Nature* **463**, 813–817 (2009). URL http://www.nature.com/nature/journal/v463/n7282/full/nature08699.html.

[9] Bushman, F. D. *et al.* Host Cell Factors in HIV Replication: Meta-Analysis of Genome-Wide Studies. *PLoS Pathog* **5**, e1000437 (2009). URL http://dx.doi.org/10.1371/journal.ppat.1000437.

[10] Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001). URL http://www.nature.com/nature/journal/v409/n6822/full/409860a0.html, cited by 0000.

[11] Venter, J. C. *et al.* The Sequence of the Human Genome. *Science* **291**, 1304–1351 (2001). URL http://www.sciencemag.org/content/291/5507/1304.

[12] Wu, T. D. Bioinformatics in the post-genomic era. *Trends in Biotechnology* **19**, 479–480 (2001). URL http://www.sciencedirect.com/science/article/pii/S0167779901018078.

[13] Yu, U., Lee, S. H., Kim, Y. J. & Kim, S. Bioinformatics in the post-genome era. *J. Biochem. Mol. Biol.* **37**, 75–82 (2004). PMID: 14761305.

[14] Kanehisa, M. & Bork, P. Bioinformatics in the post-sequence era. *Nat Genet* **33**, 305–310 (2003). URL http://www.nature.com/ng/journal/v33/n3s/full/ng1109.html, cited by 0170.

[15] Fire, A. *et al.* Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **391**, 806–811 (1998). URL http://www.nature.com/nature/journal/v391/n6669/full/391806a0.html.

[16] Montgomery, M. K., Xu, S. & Fire, A. RNA as a target of double-stranded RNA-mediated genetic interference in Caenorhabditis elegans. *Proceedings of the National Academy of Sciences* **95**, 15502 –15507 (1998). URL http://www.pnas.org/content/95/26/15502.abstract.

[17] Hammond, S. M., Bernstein, E., Beach, D. & Hannon, G. J. An RNA-directed nuclease mediates post-transcriptional gene silencing in Drosophila cells. *Nature* **404**, 293–296 (2000). URL http://www.nature.com/nature/journal/v404/n6775/full/404293a0.html.

[18] Hannon, G. J. RNA interference. *Nature* **418**, 244–251 (2002). URL http://www.nature.com/nature/journal/v418/n6894/full/418244a.html.

[19] Pinzone, M. R., Cacopardo, B., Condorelli, F., Rosa, M. D. & Nunnari, G. Sirtuin-1 and HIV-1: An Overview. *Current Drug Targets* **14**, 648–652 (2013).

[20] Pond, S. L. K., Murrell, B. & Poon, A. F. Y. Evolution of Viral Genomes: Interplay Between Selection, Recombination, and Other Forces. In *Evolutionary Genomics* (ed. Anisimova, M.), no. 856 in Methods in Molecular Biology, 239–272 (Humana Press, 2012). URL http://link.springer.com/protocol/10.1007/978-1-61779-585-5_10.

[21] Adams, M. D. *et al.* The Genome Sequence of Drosophila melanogaster. *Science* **287**, 2185–2195 (2000). URL http://www.sciencemag.org/content/287/5461/2185, PMID: 10731132.

[22] Kiger, A. A. *et al.* A functional genomic analysis of cell morphology using RNA interference. *J Biol* **2**, 1–15 (2003). URL http://link.springer.com/article/10.1186/1475-4924-2-27.

[23] Lum, L. *et al.* Identification of Hedgehog Pathway Components by RNAi in Drosophila Cultured Cells. *Science* **299**, 2039–2045 (2003). URL http://www.sciencemag.org/content/299/5615/2039, PMID: 12663920.

[24] Cherry, S. *et al.* Genome-wide RNAi screen reveals a specific sensitivity of IRES-containing RNA viruses to host translation inhibition. *Genes Dev.* **19**, 445–452 (2005). URL http://genesdev.cshlp.org/content/19/4/445, PMID: 15713840.

[25] Liu, L. *et al.* A whole genome screen for HIV restriction factors. *Retrovirology* **8**, 94 (2011). URL http://www.retrovirology.com/content/8/1/94/abstract, PMID: 22082156.

[26] Zhao, H. *et al.* A functional genomic screen reveals novel host genes that mediate interferon-alpha's effects against hepatitis C virus. *Journal of Hepatology* **56**, 326–333 (2012). URL http://www.sciencedirect.com/science/article/pii/S0168827811006611.

[27] Hao, L. *et al.* Drosophila RNAi screen identifies host genes important for influenza virus replication. *Nature* **454**, 890–893 (2008). URL http://www.nature.com/nature/journal/v454/n7206/abs/nature07151.html.

[28] Brass, A. L. *et al.* The IFITM Proteins Mediate Cellular Resistance to Influenza A H1N1 Virus, West Nile Virus, and Dengue Virus. *Cell* **139**, 1243–1254

(2009). URL http://www.cell.com/abstract/S0092-8674(09)01564-5.

[29] Ward, S. E. *et al.* Host Modulators of H1N1 Cytopathogenicity. *PLoS ONE* **7**, e39284 (2012). URL http://dx.doi.org/10.1371/journal.pone.0039284.

[30] Krishnan, M. N. *et al.* RNA interference screen for human genes associated with West Nile virus infection. *Nature* **455**, 242–245 (2008). URL http://www.nature.com/nature/journal/v455/n7210/abs/nature07207.html.

[31] Sigoillot, F. D. & King, R. W. Vigilance and Validation: Keys to Success in RNAi Screening. *ACS Chem. Biol.* **6**, 47–60 (2011). URL http://dx.doi.org/10.1021/cb100358f.

[32] Burchard, J. *et al.* MicroRNA-like off-target transcript regulation by siRNAs is species specific. *RNA* **15**, 308–315 (2009). URL http://rnajournal.cshlp.org/content/15/2/308, cited by 0052.

[33] Doench, J. G., Petersen, C. P. & Sharp, P. A. siRNAs can function as miRNAs. *Genes Dev.* **17**, 438–442 (2003). URL http://genesdev.cshlp.org/content/17/4/438, cited by 1075.

[34] Hammond, S. M. Dicing and slicing: The core machinery of the RNA interference pathway. *FEBS Letters* **579**, 5822–5829 (2005). URL http://www.sciencedirect.com/science/article/pii/S0014579305010884, cited by 0399.

[35] Carthew, R. W. & Sontheimer, E. J. Origins and Mechanisms of miRNAs and siRNAs. *Cell* **136**, 642–655 (2009). URL http://www.cell.com/abstract/S0092-8674(09)00083-X, cited by 1354.

[36] Grimm, D. *et al.* Fatality in mice due to oversaturation of cellular microRNA/short hairpin RNA pathways. *Nature* **441**, 537–541 (2006). URL http://www.nature.com/nature/journal/v441/n7092/full/nature04791.html, cited by 1064.

[37] Khan, A. A. *et al.* Transfection of small RNAs globally perturbs gene regulation by endogenous microRNAs. *Nat Biotech* **27**, 549–555 (2009). URL http://www.nature.com/nbt/journal/v27/n6/abs/nbt.1543.html, cited by 0184.

[38] Robbins, M. *et al.* Misinterpreting the Therapeutic Effects of Small Interfering RNA Caused by Immune Stimulation. *Human Gene Therapy* **19**, 991–999 (2008). URL http://online.liebertpub.com/doi/abs/10.1089/hum.2008.131.

[39] Sledz, C. A., Holko, M., de Veer, M. J., Silverman, R. H. & Williams, B. R. G. Activation of the interferon system by short-interfering RNAs. *Nat Cell Biol* **5**, 834–839 (2003). URL http://www.nature.com/ncb/journal/v5/n9/full/ncb1038.html, cited by 1346.

[40] Lin, X. *et al.* siRNA-mediated off-target gene silencing triggered by a 7 nt complementation. *Nucl. Acids Res.* **33**, 4527–4535 (2005). URL http://nar.oxfordjournals.org/content/33/14/4527, cited by 0272.

[41] Schultz, N. *et al.* Off-target effects dominate a large-scale RNAi screen for modulators of the TGF-β pathway and reveal microRNA regulation of TGFBR2. *Silence* **2**, 3 (2011). URL http://www.silencejournal.com/content/2/1/3/abstract, PMID: 21401928.

[42] Booker, M. *et al.* False negative rates in Drosophila cell-based RNAi screens: a case study. *BMC Genomics* **12**, 50 (2011). URL http://www.biomedcentral.com/1471-2164/12/50/abstract, cited by 0013.

[43] Murali, T. M., Dyer, M. D., Badger, D., Tyler, B. M. & Katze, M. G. Network-Based Prediction and Analysis of HIV Dependency Factors. *PLoS Comput Biol* **7**, e1002164 (2011). URL http://dx.doi.org/10.1371/journal.pcbi.1002164.

[44] Nabieva, E., Jim, K., Agarwal, A., Chazelle, B. & Singh, M. Whole-proteome prediction of protein function via graph-theoretic analysis of interaction maps. *Bioinformatics* **21**, i302–i310 (2005). URL http://bioinformatics.oxfordjournals.org/content/21/suppl_1/i302.abstract, cited by 0278.

[45] Bader, G. D. & Hogue, C. W. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* **4**, 2 (2003). URL http://www.biomedcentral.com/1471-2105/4/2/abstract.

[46] Walter, T. *et al.* Automatic identification and clustering of chromosome phenotypes in a genome wide RNAi screen by time-lapse imaging. *Journal of Structural Biology* **170**, 1–9 (2010). URL http://www.sciencedirect.com/science/article/pii/S1047847709002731.

[47] Wang, L., Tu, Z. & Sun, F. A network-based integrative approach to prioritize reliable hits from multiple genome-wide RNAi screens in Drosophila. *BMC Genomics* **10**, 220 (2009). URL http://www.biomedcentral.com/1471-2164/10/220/abstract.

[48] Pandey, G., Steinbach, M., Gupta, R., Garg, T. & Kumar, V. Association analysis-based transformations for protein interaction networks. 540 (ACM Press, 2007). URL http://portal.acm.org/citation.cfm?doid=1281192.1281251.

[49] Zhu, J. *et al.* Integrating large-scale functional genomic data to dissect the complexity of yeast regulatory networks. *Nat Genet* **40**, 854–861 (2008). URL http://www.nature.com/ng/journal/v40/n7/abs/ng.167.html.

[50] Reiss, D. J., Baliga, N. S. & Bonneau, R. Integrated biclustering of heterogeneous genome-wide datasets for the inference of global regulatory networks. *BMC Bioinformatics* **7**, 280 (2006). URL http://www.biomedcentral.com/1471-2105/7/280/abstract, PMID: 16749936.

[51] Chatr-aryamontri, A. *et al.* VirusMINT: a viral protein interaction database. *Nucleic Acids Research* **37**, D669–D673 (2009). URL http://nar.oxfordjournals.org/content/37/suppl_1/D669.long.

[52] Navratil, V. *et al.* VirHostNet: a knowledge base for the management and the analysis of proteome-wide virus-host interaction networks. *Nucleic Acids Research* **37**, D661–D668 (2009). URL http://nar.oxfordjournals.org/content/37/suppl_1/D661.long.

[53] Fu, W. *et al.* Human immunodeficiency virus type 1, human protein interaction database at NCBI. *Nucleic Acids Research* **37**, D417–D422 (2009). URL http://nar.oxfordjournals.org/content/37/suppl_1/D417.long.

[54] MacPherson, J. I., Dickerson, J. E., Pinney, J. W. & Robertson, D. L. Patterns of HIV-1 Protein Interaction Identify Perturbed Host-Cellular Subsystems. *PLoS Comput Biol* **6**, e1000863 (2010). URL http://dx.doi.org/10.1371/journal.pcbi.1000863.

[55] Chassey, B. d., Meyniel-Schicklin, L., Aublin-Gex, A., André, P. & Lotteau, V. Genetic screens for the control of influenza virus replication: from meta-analysis to drug discovery. *Mol. BioSyst.* **8**,

1297–1303 (2012). URL http://pubs.rsc.org/en/content/articlelanding/2012/mb/c2mb05416g.

[56] Watanabe, T., Watanabe, S. & Kawaoka, Y. Cellular Networks Involved in the Influenza Virus Life Cycle. *Cell Host & Microbe* **7**, 427–439 (2010). URL http://www.sciencedirect.com/science/article/pii/S193131281000171X.

[57] Stertz, S. & Shaw, M. L. Uncovering the global host cell requirements for influenza virus replication via RNAi screening. *Microbes and Infection* **13**, 516–525 (2011). URL http://www.sciencedirect.com/science/article/pii/S1286457911000426.

[58] Shaw, M. L. The Host Interactome of Influenza Virus Presents New Potential Targets for Antiviral Drugs. *Rev Med Virol* **21**, 358–369 (2011). URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3207218/, PMID: 21823192 PMCID: PMC3207218.

[59] Chase, G. *et al.* Hsp90 inhibitors reduce influenza virus replication in cell culture. *Virology* **377**, 431–439 (2008). URL http://www.sciencedirect.com/science/article/pii/S0042682208002973.

[60] Dyer, M. D., Murali, T. M. & Sobral, B. W. The Landscape of Human Proteins Interacting with Viruses and Other Pathogens. *PLoS Pathog* **4**, e32 (2008). URL http://dx.plos.org/10.1371/journal.ppat.0040032.

[61] Aranda, B. *et al.* The IntAct molecular interaction database in 2010. *Nucleic Acids Research* **38**, D525–D531 (2009). URL http://nar.oxfordjournals.org/content/38/suppl_1/D525.full?sid=810dc836-6ee0-4374-9079-41e5f9c4b933.

[62] Chatr-aryamontri, A. *et al.* MINT: the Molecular INTeraction database. *Nucleic Acids Research* **35**, D572–D574 (2007). URL http://nar.oxfordjournals.org/content/35/suppl_1/D572.long, cited by 0518.

[63] Xenarios, I. *et al.* DIP: The Database of Interacting Proteins. *Nucleic Acids Research* **28**, 289–291 (2000). URL http://nar.oxfordjournals.org/content/28/1/289.

[64] Keshava Prasad, T. S. *et al.* Human Protein Reference Database–2009 update. *Nucleic Acids Research* **37**, D767–D772 (2009). URL http://nar.oxfordjournals.org/content/37/suppl_1/D767.long.

[65] Pagel, P. *et al.* The MIPS Mammalian Protein–protein Interaction Database. *Bioinformatics* **21**, 832–834 (2005). URL http://bioinformatics.oxfordjournals.org/content/21/6/832.

[66] Matthews, L. *et al.* Reactome knowledgebase of human biological pathways and processes. *Nucl. Acids Res.* **37**, D619–D622 (2009). URL http://nar.oxfordjournals.org/content/37/suppl_1/D619.

[67] Bader, G. D., Betel, D. & Hogue, C. W. V. BIND: The Biomolecular Interaction Network Database. *Nucl. Acids Res.* **31**, 248–250 (2003). URL http://nar.oxfordjournals.org/content/31/1/248.

[68] Khadka, S. *et al.* A Physical Interaction Network of Dengue Virus and Human Proteins. *Mol Cell Proteomics* **10** (2011). URL http://www.mcponline.org/content/10/12/M111.012187.

[69] Simonis, N. *et al.* Host-pathogen interactome mapping for HTLV-1 and -2 retroviruses. *Retrovirology* **9**, 26 (2012). URL http://www.retrovirology.com/content/9/1/26/abstract.

[70] de Chassey, B. *et al.* Hepatitis C virus infection protein network. *Mol Syst Biol* **4** (2008). URL http://www.nature.com/msb/journal/v4/n1/full/msb200866.html.

[71] Pichlmair, A. *et al.* Viral immune modulators perturb the human molecular network by common and unique strategies. *Nature* **487**, 486–490 (2012). URL http://www.nature.com/nature/journal/v487/n7408/full/nature11289.html?WT.ec_id=NATURE-20120726.

[72] Jaeger, S., Ertaylan, G., van Dijk, D., Leser, U. & Sloot, P. Inference of Surface Membrane Factors of HIV-1 Infection through Functional Interaction Networks. *PLoS ONE* **5**, e13139 (2010). URL http://dx.doi.org/10.1371/journal.pone.0013139.

[73] Nepusz, T., Yu, H. & Paccanaro, A. Detecting overlapping protein complexes in protein-protein interaction networks. *Nature Methods* **9**, 471 (2012). URL http://www.nature.com/nmeth/journal/v9/n5/abs/nmeth.1938.html.

[74] Razick, S., Magklaras, G. & Donaldson, I. M. iRefIndex: A consolidated protein interaction database with provenance. *BMC Bioinformatics* **9**, 405 (2008). URL zotero://attachment/747/.

[75] Stark, C. *et al.* The BioGRID Interaction Database: 2011 Update. *Nucl. Acids Res.* (2010). URL http://nar.oxfordjournals.org/content/early/2010/11/11/nar.gkq1116.

[76] Ruepp, A. *et al.* CORUM: the comprehensive resource of mammalian protein complexes–2009. *Nucleic Acids Research* **38**, D497–D501 (2009). URL http://nar.oxfordjournals.org/content/38/suppl_1/D497.full.

[77] Guldener, U. MPact: the MIPS protein interaction resource on yeast. *Nucleic Acids Research* **34**, D436–D441 (2006). URL http://nar.oxfordjournals.org/content/34/suppl_1/D436.long.

[78] Brown, K. R. & Jurisica, I. Online Predicted Human Interaction Database. *Bioinformatics* **21**, 2076–2082 (2005). URL http://bioinformatics.oxfordjournals.org/content/21/9/2076.

[79] Mering, C. V. *et al.* STRING: A Database of Predicted Functional Associations Between Proteins. *Nucl. Acids Res.* **31**, 258–261 (2003). URL http://nar.oxfordjournals.org/content/31/1/258.

[80] Fisher, R. A. On the Interpretation of $\chi 2$ from Contingency Tables, and the Calculation of P. *Journal of the Royal Statistical Society* **85**, pp. 87–94 (1922). URL http://www.jstor.org/stable/2340521.

[81] Freeman, L. C. Centrality in social networks conceptual clarification. *Social Networks* **1**, 215–239 (1978). URL http://www.sciencedirect.com/science/article/pii/0378873378900217.

[82] Yu, H., Kim, P. M., Sprecher, E., Trifonov, V. & Gerstein, M. The Importance of Bottlenecks in Protein Networks: Correlation with Gene Essentiality and Expression Dynamics. *PLoS Comput Biol* **3**, e59 (2007). URL http://dx.plos.org/10.1371/journal.pcbi.0030059, cited by 0301.

[83] Joy, M. P., Brock, A., Ingber, D. E. & Huang, S. High-Betweenness Proteins in the Yeast Protein Interaction Network. *BioMed Research International* **2005**, 96–103 (2005). URL http://www.hindawi.com/journals/bmri/2005/594674/abs/, cited by 0159.

[84] Huber, W., Carey, V. J., Long, L., Falcon, S. & Gentleman, R. Graphs in molecular biology. *BMC Bioinformatics* **8**, S8 (2007). URL http://www.biomedcentral.com/1471-2105/8/S6/S8/abstract, cited by 0054.

[85] Pavlopoulos, G. A. *et al.* Using graph theory to analyze biological networks. *BioData Mining* **4**, 10 (2011). URL http://www.biodatamining.org/content/4/1/10/abstract, cited by 0036.

[86] He, X. & Zhang, J. Why Do Hubs Tend to Be Essential in Protein Networks? *PLoS Genet* **2**, e88 (2006). URL http://dx.plos.org/10.1371/journal.pgen.0020088, cited by 0209.

[87] Mazurie, A., Bonchev, D., Schwikowski, B. & Buck, G. A. Evolution of metabolic network organization. *BMC Systems Biology* **4**, 59 (2010). URL http://www.biomedcentral.com/1752-0509/4/59/abstract, cited by 0019.

[88] Marsden, P. V. Egocentric and sociocentric measures of network centrality. *Social Networks* **24**, 407–422 (2002). URL http://www.sciencedirect.com/science/article/pii/S0378873302000163, cited by 0184.

[89] Opsahl, T., Agneessens, F. & Skvoretz, J. Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks* **32**, 245–251 (2010). URL http://www.sciencedirect.com/science/article/pii/S0378873310000183, cited by 0224.

[90] White, D. R. & Borgatti, S. P. Betweenness centrality measures for directed graphs. *Social Networks* **16**, 335–346 (1994). URL http://www.sciencedirect.com/science/article/pii/0378873394900159, cited by 0073.

[91] Dijkstra, E. W. A note on two problems in connexion with graphs. *Numer. Math.* **1**, 269–271 (1959). URL http://link.springer.com/article/10.1007/BF01386390, cited by 11787.

[92] Floyd, R. W. Algorithm 97: Shortest path. *Commun. ACM* **5**, 345– (1962). URL http://doi.acm.org/10.1145/367766.368168, cited by 2180.

[93] Bonacich, P. Some unique properties of eigenvector centrality. *Social Networks* **29**, 555–564 (2007). URL http://www.sciencedirect.com/science/article/pii/S0378873307000342.

[94] Watts, D. J. & Strogatz, S. H. Collective dynamics of 'small-world' networks. *Nature* **393**, 440–442 (1998). URL http://www.nature.com/nature/journal/v393/n6684/full/393440a0.html, cited by 19957.

[95] Gulbahce, N. *et al.* Viral Perturbations of Host Networks Reflect Disease Etiology. *PLoS Comput Biol* **8**, e1002531 (2012). URL http://dx.doi.org/10.1371/journal.pcbi.1002531.

[96] Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS: A Journal of Integrative Biology* **16**, 284–287 (2012). URL http://online.liebertpub.com/doi/abs/10.1089/omi.2011.0118, 00014.

[97] Yu, G. ReactomePA: Reactome Pathway Analysis (????). 00000.

[98] Ashburner, M. *et al.* Gene Ontology: tool for the unification of biology. *Nat Genet* **25**, 25–29 (2000). URL http://www.nature.com/ng/journal/v25/n1/full/ng0500_25.html, 11409.

[99] Yu, G. *et al.* GOSemSim: an R package for measuring semantic similarity among GO terms and gene products. *Bioinformatics* **26**, 976–978 (2010). URL http://bioinformatics.oxfordjournals.org/content/26/7/976.

[100] Taatjes, D. J., Näär, A. M., Andel, F., Nogales, E. & Tjian, R. Structure, Function, and Activator-Induced Conformations of the CRSP Coactivator. *Science* **295**, 1058–1062 (2002). URL http://www.sciencemag.org/content/295/5557/1058, PMID: 11834832.

[101] Friedman, J. *et al.* Epigenetic Silencing of HIV-1 by the Histone H3 Lysine 27 Methyltransferase Enhancer of Zeste 2. *J. Virol.* **85**, 9078–9089 (2011). URL http://jvi.asm.org/content/85/17/9078, cited by 0032.

[102] Pearson, R. *et al.* Epigenetic Silencing of Human Immunodeficiency Virus (HIV) Transcription by Formation of Restrictive Chromatin Structures at the Viral Long Terminal Repeat Drives the Progressive Entry of HIV into Latency. *J. Virol.* **82**, 12291–12303 (2008). URL http://jvi.asm.org/content/82/24/12291, PMID: 18829756.

[103] Kauder, S. E., Bosque, A., Lindqvist, A., Planelles, V. & Verdin, E. Epigenetic Regulation of HIV-1 Latency by Cytosine Methylation. *PLoS Pathog* **5**, e1000495 (2009). URL http://dx.doi.org/10.1371/journal.ppat.1000495, cited by 0128.

[104] Marban, C. *et al.* Recruitment of chromatin-modifying enzymes by CTIP2 promotes HIV-1 transcriptional silencing. *EMBO J* **26**, 412–423 (2007). URL http://www.nature.com/emboj/journal/v26/n2/full/7601516a.html, cited by 0134.

[105] Imai, K., Togami, H. & Okamoto, T. Involvement of Histone H3 Lysine 9 (H3K9) Methyltransferase G9a in the Maintenance of HIV-1 Latency and Its Reactivation by BIX01294. *J. Biol. Chem.* **285**, 16538–16545 (2010). URL http://www.jbc.org/content/285/22/16538.

[106] Sakane, N. *et al.* Activation of HIV Transcription by the Viral Tat Protein Requires a Demethylation Step Mediated by Lysine-specific Demethylase 1 (LSD1/KDM1). *PLoS Pathogens* **7**, e1002184 (2011). URL http://dx.plos.org/10.1371/journal.ppat.1002184.

[107] Lai, F. *et al.* Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. *Nature* **494**, 497–501 (2013). URL http://www.nature.com/nature/journal/v494/n7438/full/nature11884.html#ref21.

[108] Ørom, U. A. *et al.* Long Noncoding RNAs with Enhancer-like Function in Human Cells. *Cell* **143**, 46–58 (2010). URL http://www.sciencedirect.com/science/article/pii/S0092867410010111.

[109] Wang, K. C. *et al.* A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. *Nature* **472**, 120–124 (2011). URL http://www.nature.com/nature/journal/v472/n7341/full/nature09819.html.

[110] Chaudhury, A., Chander, P. & Howe, P. H. Heterogeneous nuclear ribonucleoproteins (hnRNPs) in cellular processes: Focus on hnRNP E1's multifunctional regulatory roles. *RNA* **16**, 1449–1462 (2010). URL http://rnajournal.cshlp.org/content/16/8/1449, cited by 0050.

[111] Dreyfuss, G., Matunis, M. J., Pinol-Roma, S. & Burd, C. G. hnRNP Proteins and the Biogenesis of mRNA. *Annual Review of Biochemistry* **62**, 289–321 (1993). URL http://www.annualreviews.org/doi/abs/10.1146/annurev.bi.62.070193.001445, cited by 1283.

[112] Long, J. C. & Caceres, J. F. The SR protein family of splicing factors: master regulators of gene expression. *Biochemical Journal* **417**, 15 (2009). URL http://www.biochemj.org/bj/417/bj4170015.htm, 00328.

[113] Lund, N. *et al.* Differential effects of hnRNP D/AUF1 isoforms on HIV-1 gene expression. *Nucl. Acids Res.* **40**, 3663–3675 (2012). URL http://nar.oxfordjournals.org/content/40/8/3663.

[114] Bériault, V. *et al.* A Late Role for the Association of hnRNP A2 with the HIV-1 hnRNP A2 Response Elements in Genomic RNA, Gag, and Vpr Localization.

J. Biol. Chem. **279**, 44141–44153 (2004). URL http://www.jbc.org/content/279/42/44141.

[115] Woolaway, K., Asai, K., Emili, A. & Cochrane, A. hnRNP E1 and E2 have distinct roles in modulating HIV-1 gene expression. *Retrovirology* **4**, 28 (2007). URL http://www.retrovirology.com/content/4/1/28/abstract, 00026.

[116] Uhlen, M. *et al.* Towards a knowledge-based Human Protein Atlas. *Nature Biotechnology* **28**, 1248–1250 (2010). URL http://www.nature.com/nbt/journal/v28/n12/full/nbt1210-1248.html.

[117] Glisovic, T., Bachorik, J. L., Yong, J. & Dreyfuss, G. RNA-binding proteins and post-transcriptional gene regulation. *FEBS Letters* **582**, 1977–1986 (2008). URL http://www.febsletters.org/article/S0014-5793(08)00207-X/abstract, 00215.

[118] Lukong, K. E., Chang, K.-w., Khandjian, E. W. & Richard, S. RNA-binding proteins in human genetic disease. *Trends in Genetics* **24**, 416–425 (2008). URL http://www.sciencedirect.com/science/article/pii/S016895250800173X, 00122.

[119] Berro, R. *et al.* Acetylated Tat Regulates Human Immunodeficiency Virus Type 1 Splicing through Its Interaction with the Splicing Regulator p32. *J. Virol.* **80**, 3189–3204 (2006). URL http://jvi.asm.org/content/80/7/3189, 00062.

[120] Bourgeois, C. F., Lejeune, F. & Stévenin, J. Broad Specificity of SR (Serine?Arginine) Proteins in the Regulation of Alternative Splicing of Pre-Messenger RNA. In *Progress in Nucleic Acid Research and Molecular Biology*, vol. Volume 78, 37–88 (Academic Press, 2004). URL http://www.sciencedirect.com/science/article/pii/S0079660304780022, 00151.

[121] Katz, M., Amit, I. & Yarden, Y. Regulation of MAPKs by growth factors and receptor tyrosine kinases. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research* **1773**, 1161–1176 (2007). URL http://www.sciencedirect.com/science/article/pii/S0167488907000043.

[122] Farooq, A. & Zhou, M.-M. Structure and regulation of MAPK phosphatases. *Cellular Signalling* **16**, 769–779 (2004). URL

http://www.sciencedirect.com/science/article/pii/S089865680400021X.

[123] Krishna, M. & Narang, H. The complexity of mitogen-activated protein kinases (MAPKs) made simple. *Cell. Mol. Life Sci.* **65**, 3525–3544 (2008). URL http://link.springer.com/article/10.1007/s00018-008-8170-7, 00161.

[124] Macdonald, A. *et al.* The Hepatitis C Virus Non-structural NS5A Protein Inhibits Activating Protein–1 Function by Perturbing Ras-ERK Pathway Signaling. *J. Biol. Chem.* **278**, 17775–17784 (2003). URL http://www.jbc.org/content/278/20/17775, PMID: 12621033.

[125] Ndjomou, J., Park, I.-w., Liu, Y., Mayo, L. D. & He, J. J. Up-Regulation of Hepatitis C Virus Replication and Production by Inhibition of MEK/ERK Signaling. *PLoS ONE* **4**, e7498 (2009). URL http://www.plosone.org/article/info:doi/10.1371/journal.pone.0007498.

[126] Roux, P. P. *et al.* RAS/ERK Signaling Promotes Site-specific Ribosomal Protein S6 Phosphorylation via RSK and Stimulates Cap-dependent Translation. *J. Biol. Chem.* **282**, 14056–14064 (2007). URL http://www.jbc.org/content/282/19/14056.

[127] George, A. *et al.* Hepatitis C Virus NS5A Binds to the mRNA Cap-binding Eukaryotic Translation Initiation 4F (eIF4F) Complex and Up-regulates Host Translation Initiation Machinery through eIF4E-binding Protein 1 Inactivation. *J. Biol. Chem.* **287**, 5042–5058 (2012). URL http://www.jbc.org/content/287/7/5042.

[128] Verghese, J., Abrams, J., Wang, Y. & Morano, K. A. Biology of the Heat Shock Response and Protein Chaperones: Budding Yeast (Saccharomyces cerevisiae) as a Model System. *Microbiol. Mol. Biol. Rev.* **76**, 115–158 (2012). URL http://mmbr.asm.org/content/76/2/115, 00011.

[129] Boyle, D. & Takemoto, L. Characterization of the α-γ and α-β Complex: Evidence for an In Vivo Functional Role of α-Crystallin as a Molecular Chaperone. *Experimental Eye Research* **58**, 9–16 (1994). URL http://www.sciencedirect.com/science/article/pii/S0014483584711900, 00085.

[130] Klemenz, R., Fröhli, E., Steiger, R. H., Schäfer, R. & Aoyama, A. Alpha B-crystallin is a small heat shock protein. *PNAS* **88**, 3652–3656 (1991). URL http://www.pnas.org/content/88/9/3652, 00512.

[131] Chen, Y.-J. *et al.* Heat Shock Protein 72 Is Associated with the Hepatitis C Virus Replicase Complex and Enhances Viral RNA Replication. *J. Biol. Chem.* **285**, 28183–28190 (2010). URL http://www.jbc.org/content/285/36/28183, PMID: 20601427.

[132] Lim, Y. S. *et al.* Nonstructural 5A protein of hepatitis C virus regulates heat shock protein 72 for its own propagation. *Journal of Viral Hepatitis* **19**, 353–363 (2012). URL http://onlinelibrary.wiley.com/doi/10.1111/j.1365-2893.2011.01556.x/abstract, 00000.

[133] Huang, X.-Y. *et al.* αB-Crystallin complexes with 14-3-3ζ to induce epithelial-mesenchymal transition and resistance to sorafenib in hepatocellular carcinoma. *Hepatology* n/a–n/a (2013). URL http://onlinelibrary.wiley.com/doi/10.1002/hep.26255/abstract.

[134] Akdis, M. *et al.* Interleukins, from 1 to 37, and interferon-γ: Receptors, functions, and roles in diseases. *Journal of Allergy and Clinical Immunology* **127**, 701–721.e70 (2011). URL http://www.sciencedirect.com/science/article/pii/S0091674910019494.

[135] Wei, X. *et al.* Inhibition of Hepatitis C Virus Infection by Interferon-γ Through Downregulating Claudin-1. *Journal of Interferon & Cytokine Research* **29**, 171–178 (2009). URL http://online.liebertpub.com/doi/abs/10.1089/jir.2008.0040.

[136] Kondo, Y., Sung, V. M., Machida, K., Liu, M. & Lai, M. M. Hepatitis C virus infects T cells and affects interferon-γ signaling in T cell lines. *Virology* **361**, 161–173 (2007). URL http://www.sciencedirect.com/science/article/pii/S0042682206008385.

[137] Carreño, V. & Quiroga, J. A. Biological properties of interleukin-12 and its therapeutic use in persistent hepatitis B virus and hepatitis C virus infection. *Journal of Viral Hepatitis* **4**, 83–86 (1997). URL http://onlinelibrary.wiley.com/doi/10.1111/j.1365-2893.1997.tb00184.x/abstract.

[138] Wang, J. M. *et al.* Differential Regulation of Interleukin-12 (IL-12)/IL-23 by Tim-3 Drives TH17 Cell Development during Hepatitis C Virus Infection. *J. Virol.* **87**, 4372–4383 (2013). URL http://jvi.asm.org/content/87/8/4372, PMID: 23388728.

[139] Eisen-Vandervelde, A. L. *et al.* Hepatitis C Virus Core Selectively Suppresses Interleukin-12 Synthesis in Human Macrophages by Interfering with

AP-1 Activation. *J. Biol. Chem.* **279**, 43479–43486 (2004). URL http://www.jbc.org/content/279/42/43479, PMID: 15292184.

[140] Rigopoulou, E. I., Abbott, W. G., Haigh, P. & Naoumov, N. V. Blocking of interleukin-10 receptor—a novel approach to stimulate T-helper cell type 1 responses to hepatitis C virus. *Clinical Immunology* **117**, 57–64 (2005). URL http://www.sciencedirect.com/science/article/pii/S1521661605001816.

[141] Frank, A. C. *et al.* Interleukin-27, an Anti-HIV-1 Cytokine, Inhibits Replication of Hepatitis C Virus. *Journal of Interferon & Cytokine Research* **30**, 427–431 (2010). URL http://online.liebertpub.com/doi/abs/10.1089/jir.2009.0093.

[142] Cheng, M. *et al.* Recombinant human interleukin 28B: anti-HCV potency, receptor usage and restricted cell-type responsiveness. *J. Antimicrob. Chemother.* **67**, 1080–1087 (2012). URL http://jac.oxfordjournals.org/content/67/5/1080, PMID: 22323501.

[143] Persico, M. *et al.* Suppressor of cytokine signaling 3 (SOCS3) expression and hepatitis C virus–related chronic hepatitis: Insulin resistance and response to antiviral therapy. *Hepatology* **46**, 1009–1015 (2007). URL http://onlinelibrary.wiley.com/doi/10.1002/hep.21782/abstract.

[144] Pazienza, V. *et al.* The hepatitis C virus core protein of genotypes 3a and 1b downregulates insulin receptor substrate 1 through genotype-specific mechanisms. *Hepatology* **45**, 1164–1171 (2007). URL http://onlinelibrary.wiley.com/doi/10.1002/hep.21634/abstract.

[145] Brandman, D., Bacchetti, P., Ayala, C. E., Maher, J. J. & Khalili, M. Impact of Insulin Resistance on HCV Treatment Response and Impact of HCV Treatment on Insulin Sensitivity Using Direct Measurements of Insulin Action. *Dia Care* **35**, 1090–1094 (2012). URL http://care.diabetesjournals.org/content/35/5/1090.

[146] del Campo, J. A., García-Valdecasas, M., Rojas, L., Rojas, Á. & Romero-Gómez, M. The Hepatitis C Virus Modulates Insulin Signaling Pathway In Vitro Promoting Insulin Resistance. *PLoS ONE* **7**, e47904 (2012). URL http://dx.doi.org/10.1371/journal.pone.0047904.

[147] Shao, R.-X. *et al.* SOCS1 abrogates IFN's antiviral effect on hepatitis C virus replication. *Antiviral Research* **97**, 101–107 (2013). URL http://www.sciencedirect.com/science/article/pii/S016635421200280X.

[148] Lee, J. *et al.* Saponin Inhibits Hepatitis C Virus Propagation by Up-regulating Suppressor of Cytokine Signaling 2. *PLoS ONE* **7**, e39366 (2012). URL http://dx.doi.org/10.1371/journal.pone.0039366.

[149] García-Ruiz, I. *et al.* Protein-tyrosine Phosphatases Are Involved in Interferon Resistance Associated with Insulin Resistance in HepG2 Cells and Obese Mice. *J. Biol. Chem.* **287**, 19564–19573 (2012). URL http://www.jbc.org/content/287/23/19564.

[150] Hirsch, A. J. *et al.* The Src Family Kinase c-Yes Is Required for Maturation of West Nile Virus Particles. *J. Virol.* **79**, 11943–11951 (2005). URL http://jvi.asm.org/content/79/18/11943, PMID: 16140770.

[151] Trible, R. P. HIV-1 Nef Selectively Activates Src Family Kinases Hck, Lyn, and c-Src through Direct SH3 Domain Interaction. *Journal of Biological Chemistry* **281**, 27029–27038 (2006). URL http://europepmc.org/abstract/MED/16849330/reload=0;jsessionid=8xLUhg2JDgWr8lMEdB6m.10.

[152] Nakashima, K. *et al.* HCV NS5A Protein Containing Potential Ligands for Both Src Homology 2 and 3 Domains Enhances Autophosphorylation of Src Family Kinase Fyn in B Cells. *PLoS ONE* **7**, e46634 (2012). URL http://dx.doi.org/10.1371/journal.pone.0046634.

[153] Pfannkuche, A. *et al.* C-Src is required for complex formation between the hepatitis C virus–encoded proteins NS5A and NS5B: A prerequisite for replication. *Hepatology* **53**, 1127–1136 (2011). URL http://onlinelibrary.wiley.com/doi/10.1002/hep.24214/abstract.

[154] Martin-Garcia, J. M., Luque, I., Ruiz-Sanz, J. & Camara-Artigas, A. The promiscuous binding of the Fyn SH3 domain to a peptide from the NS5A protein. *Acta Cryst D, Acta Cryst Sect D, Acta Crystallogr D, Acta Crystallogr Sect D, Acta Crystallogr D Biol Crystallogr, Acta Crystallogr Sect D Biol Crystallogr* **68**, 1030–1040 (2012). URL http://scripts.iucr.org/cgi-bin/paper?mh5062, the hepatitis C virus nonstructural 5A (NS5A) protein is a large zinc-binding phosphoprotein that plays an important role in viral RNA replication and is involved in altering signal transduction pathways

in the host cell. This protein interacts with Fyn tyrosine kinase in vivo and regulates its kinase activity. The 1.5 Å resolution crystal structure of a complex between the SH3 domain of the Fyn tyrosine kinase and the C–terminal proline-rich motif of the NS5A-derived peptide APPIPPPRRKR has been solved. Crystals were obtained in the presence of ZnCl2 and belonged to the tetragonal space group P41212. The asymmetric unit is composed of four SH3 domains and two NS5A peptide molecules; only three of the domain molecules contain a bound peptide, while the fourth molecule seems to correspond to a free form of the domain. Additionally, two of the SH3 domains are bound to the same peptide chain and form a ternary complex. The proline-rich motif present in the NS5A protein seems to be important for RNA replication and virus assembly, and the promiscuous interaction of the Fyn SH3 domain with the NS5A C-terminal proline-rich peptide found in this crystallographic structure may be important in the virus infection cycle.

[155] Perrimon, N. & Mathey-Prevot, B. Applications of High-Throughput RNA Interference Screens to Problems in Cell and Developmental Biology. *Genetics* **175**, 7–16 (2007). URL http://www.genetics.org/content/175/1/7.

[156] Mohr, S., Bakal, C. & Perrimon, N. Genomic Screening with RNAi: Results and Challenges. *Annual Review of Biochemistry* **79**, 37–64 (2010). URL http://www.annualreviews.org/doi/abs/10.1146/annurev-biochem-060408-092949, PMID: 20367032.

[157] Hosui, A. *et al.* Hepatitis C Virus Core Protein Differently Regulates the JAK-STAT Signaling Pathway under Interleukin-6 and Interferon-γ Stimuli. *J. Biol. Chem.* **278**, 28562–28571 (2003). URL http://www.jbc.org/content/278/31/28562, PMID: 12764155.

[158] Waris, G., Turkson, J., Hassanein, T. & Siddiqui, A. Hepatitis C Virus (HCV) Constitutively Activates STAT-3 via Oxidative Stress: Role of STAT-3 in HCV Replication. *J. Virol.* **79**, 1569–1580 (2005). URL http://jvi.asm.org/content/79/3/1569, PMID: 15650183.

[159] Chen, B., Shi, J., Zhang, S. & Wu, F.-X. Identifying protein complexes in protein–protein interaction networks by using clique seeds and graph entropy. *PROTEOMICS* **13**, 269–277 (2013). URL http://onlinelibrary.wiley.com/doi/10.1002/pmic.201200336/abstract, cited by 0001.

[160] Yong, C. H., Liu, G., Chua, H. N. & Wong, L. Supervised maximum-likelihood weighting of composite protein networks for complex prediction. *BMC Systems Biology* **6**, S13 (2012). URL http://www.biomedcentral.com/1752-0509/6/S2/S13/abstract, PMID: 23281936.

[161] Saha, S., Dazard, J.-E., Xu, H. & Ewing, R. M. Computational Framework for Analysis of Prey–Prey Associations in Interaction Proteomics Identifies Novel Human Protein–Protein Interactions and Networks. *J. Proteome Res.* **11**, 4476–4487 (2012). URL http://dx.doi.org/10.1021/pr300227y, cited by 0000.

[162] Smith, S. & de Lange, T. Tankyrase promotes telomere elongation in human cells. *Current Biology* **10**, 1299–1302 (2000). URL http://www.sciencedirect.com/science/article/pii/S0960982200007521.

[163] Cook, B. D., Dynek, J. N., Chang, W., Shostak, G. & Smith, S. Role for the Related Poly(ADP-Ribose) Polymerases Tankyrase 1 and 2 at Human Telomeres. *Mol Cell Biol* **22**, 332–342 (2002). URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC134233/, PMID: 11739745 PMCID: PMC134233.

[164] Smith, S., Giriat, I., Schmitt, A. & Lange, T. d. Tankyrase, a Poly(ADP-Ribose) Polymerase at Human Telomeres. *Science* **282**, 1484–1487 (1998). URL http://www.sciencemag.org/content/282/5393/1484, PMID: 9822378.

[165] Kitay-Cohen, Y., Goldberg-Bittman, L., Hadary, R., Fejgin, M. & Amiel, A. Telomere length in Hepatitis C. *Cancer Genetics and Cytogenetics* **187**, 34–38 (2008). URL http://www.sciencedirect.com/science/article/pii/S0165460808004792.

[166] Zhang, Y., Shen, J., Ming-Whei, Lee, Y. P.-H. & Santella, R. M. Telomere length in hepatocellular carcinoma and paired adjacent non-tumor tissues by quantitative PCR. *Cancer Invest.* **25**, 668–677 (2007). Cited by 0006.

[167] Levrero, M. Viral hepatitis and liver cancer: the case of hepatitis C. *Oncogene* **25**, 3834–3847 (2006). URL http://www.nature.com/onc/journal/v25/n27/abs/1209562a.html.

[168] Saito, I. *et al.* Hepatitis C virus infection is associated with the development of hepatocellular carcinoma. *Proc Natl Acad Sci U S A* **87**, 6547–6549

(1990). URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC54573/, PMID: 2168552 PMCID: PMC54573.

[169] Clevers, H. Axin and hepatocellular carcinomas. *Nat Genet* **24**, 206–208 (2000). URL http://www.nature.com/ng/journal/v24/n3/full/ng0300_206.html.

[170] Satoh, S. *et al.* AXIN1 mutations in hepatocellular carcinomas, and growth suppression in cancer cells by virus-mediated transfer of AXIN1. *Nat Genet* **24**, 245–250 (2000). URL http://www.nature.com/ng/journal/v24/n3/full/ng0300_245.html.

[171] Huang, S.-M. A. *et al.* Tankyrase inhibition stabilizes axin and antagonizes Wnt signalling. *Nature* **461**, 614–620 (2009). URL http://www.nature.com/nature/journal/v461/n7264/full/nature08356.html.

[172] Schaub, M. C., Lopez, S. R. & Caputi, M. Members of the Heterogeneous Nuclear Ribonucleoprotein H Family Activate Splicing of an HIV-1 Splicing Substrate by Promoting Formation of ATP-dependent Spliceosomal Complexes. *J. Biol. Chem.* **282**, 13617–13626 (2007). URL http://www.jbc.org/content/282/18/13617, PMID: 17337441.

[173] Hadian, K. *et al.* Identification of a Heterogeneous Nuclear Ribonucleoprotein-recognition Region in the HIV Rev Protein. *J. Biol. Chem.* **284**, 33384–33391 (2009). URL http://www.jbc.org/content/284/48/33384, PMID: 19808671.

[174] Verga-Gérard, A. *et al.* Hepatitis C virus/human interactome identifies SMURF2 and the viral protease as critical elements for the control of TGF-β signaling. *FASEB J* (2013). URL http://www.fasebj.org/content/early/2013/06/17/fj.13-229187, PMID: 23781096.

[175] Dickerson, J. E., Pinney, J. W. & Robertson, D. L. The biological context of HIV-1 host interactions reveals subtle insights into a system hijack. *BMC Systems Biology* **4**, 80 (2010). URL http://www.biomedcentral.com/1752-0509/4/80/abstract.

[176] Snijder, B. *et al.* Single-cell analysis of population context advances RNAi screening at multiple levels. *Molecular Systems Biology* **8** (2012). URL http://www.nature.com/msb/journal/v8/n1/full/msb20129.html.

[177] Hakes, L., Pinney, J. W., Robertson, D. L. & Lovell, S. C. Protein-protein interaction networks and biology—what's the connection? *Nat Biotech* **26**,

69–72 (2008). URL http://www.nature.com/nbt/journal/v26/n1/full/nbt0108-69.html.

[178] Yu, X., Ivanic, J., Memišević, V., Wallqvist, A. & Reifman, J. Categorizing Biases in High-Confidence High-Throughput Protein-Protein Interaction Data Sets. *Mol Cell Proteomics* **10**, M111.012500 (2011). URL http://www.mcponline.org/content/10/12/M111.012500, PMID: 21876202.