



ELSEVIER

Contents lists available at ScienceDirect

MethodsX

journal homepage: www.elsevier.com/locate/mex

Method Article

Developing a virus-microRNA interactome using cytoscape

Meredith Hill^{a,1}, Dayna Mason^{a,1}, Tânia Monteiro Marques^{b,1}, Margarida Gama Carvalho^{b,**}, Nham Tran^{a,c,*}^a School Biomedical Engineering, University of Technology Sydney, NSW, Australia^b University of Lisboa, Faculty of Sciences, BiolSI – Biosystems & Integrative Sciences Institute, Campo Grande, Lisboa, Portugal^c The Sydney Head and Neck Cancer Institute, Sydney Cancer Centre, Royal Prince Alfred Hospital, NSW, Australia

A B S T R A C T

It is currently difficult to determine the effect of oncogenic viruses on the global function and regulation of pathways within mammalian cells. A thorough understanding of the molecular pathways and individual genes altered by oncogenic viruses is needed for the identification of targets that can be utilised for early diagnosis, prevention, and treatment methods. We detail a logical step-by-step guide to uncover viral-protein-miRNA interactions using publically available datasets and the network building program, Cytoscape. This method may be applied to identify specific pathways that are altered in viral infection, and contribute to the oncogenic transformation of cells. To demonstrate this, we constructed a gene regulatory interactome encompassing Human Papillomavirus Type 16 (HPV16) and its control of specific miRNAs. This approach can be broadly applied to understand and map the regulatory functions of other oncogenic viruses, and determine their role in altering the cellular environment in cancer.

Availability and Implementation Cytoscape (Shannon et al. (2003), Smoot et al. (2010)) is freely available at <https://cytoscape.org/>.

- This method allows for the analysis and visualization of large datasets to generate an interactome that integrates key players of molecular biology
- This approach may be applied to any oncogenic virus to map its regulatory functions, and its secondary impact on gene regulation via microRNAs.

© 2019 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

* Corresponding author at: School Biomedical Engineering, University of Technology Sydney, NSW, Australia.

** Corresponding author.

E-mail addresses: mhcarvalho@fc.ul.pt (M. Gama Carvalho), nham.tran@uts.edu.au (N. Tran).

¹ Equal first author.

<https://doi.org/10.1016/j.mex.2019.10.011>

2215-0161/© 2019 Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

ARTICLE INFO

Method name: Developing a virus-microRNA interactome

Keywords: Gene expression, Network analysis, Regulatory networks, MicroRNA, Visualisation

Article history: Received 8 March 2019; Accepted 6 October 2019; Available online 11 October 2019

Specifications Table

| | |
|-----------------------------|--|
| Subject Area: | Biochemistry, Genetics and Molecular Biology |
| More specific subject area: | This method describes the mapping of large datasets depicting oncogenic viruses, their impact on human genes and the involvement of microRNAs that regulate those genes. |
| Method name: | Developing a Virus-MicroRNA interactome |
| Resource availability: | Cytoscape [1,2] is freely available at https://cytoscape.org/ Virus mentha [3,4] virusmentha.uniroma2.it BioGRID [5] https://downloads.thebiogrid.org/BioGRID/Release-Archive/BIOGRID-3.4.163/ DAVID ID [6,7] https://david.ncifcrf.gov/conversion.jsp UCSC Genome Browser [11] https://genome.ucsc.edu |

Method details*Building a viral miRNA human interactome*

An example of a viral-human miRNA interactome was created using the guidelines described in this paper. The virus of choice was HPV16, with a emphasis on the effect of the viral oncoproteins E6 and E7 on microRNA (miRNA) and gene regulation. For this interactome, we specifically focused upon hsa-miR-33a, hsa-miR-496, and the transcription factor SREBF2. This method can be adapted to investigate a specific virus and miRNA(s) of interest.

Downloading required viral dataset

To build this network, publicly available datasets were downloaded from several websites. The Virus Mentha depository (virusmentha.uniroma2.it) contains data listing the interactions between a virus or specific set of viruses with human proteins [3,4].

The HPV16 viral protein interactions were downloaded from the Virus Mentha website, and filtered in Excel by taxon 333760 (HPV16) and 9606 (H. sapiens) (Fig. 1A).

Importing data into cytoscape. Files can be directly uploaded into Cytoscape using the import network tool (**File > Import > Network > Choose File**). When importing our data, the gene names listed as human were selected as the source nodes, and the HPV16 oncogene names as the target nodes. As there is no information about the directionality of the interactions, this choice is arbitrary and the source and target nodes can be attributed to either human-virus or virus-human. The created network only contains the direct interactions between the viral and host proteins (Fig. 1B). Next the viral interactome is merged with the human interactome to create a more complete network.

Downloading the human interactome

Cytoscape version 3.3.0 already contains the Human interactome from BioGRID. If using a newer version, the human interactome will need to be downloaded from BioGRID [5] (<https://downloads.thebiogrid.org/BioGRID/Release-Archive/BIOGRID-3.4.163/>) and imported into Cytoscape. To load the Human interactome in Cytoscape v3.3.0, go to the tool bar and select **Help > Show Welcome Screen > Select H. sapiens**. Import the Human interactome as a new collection network. This will load

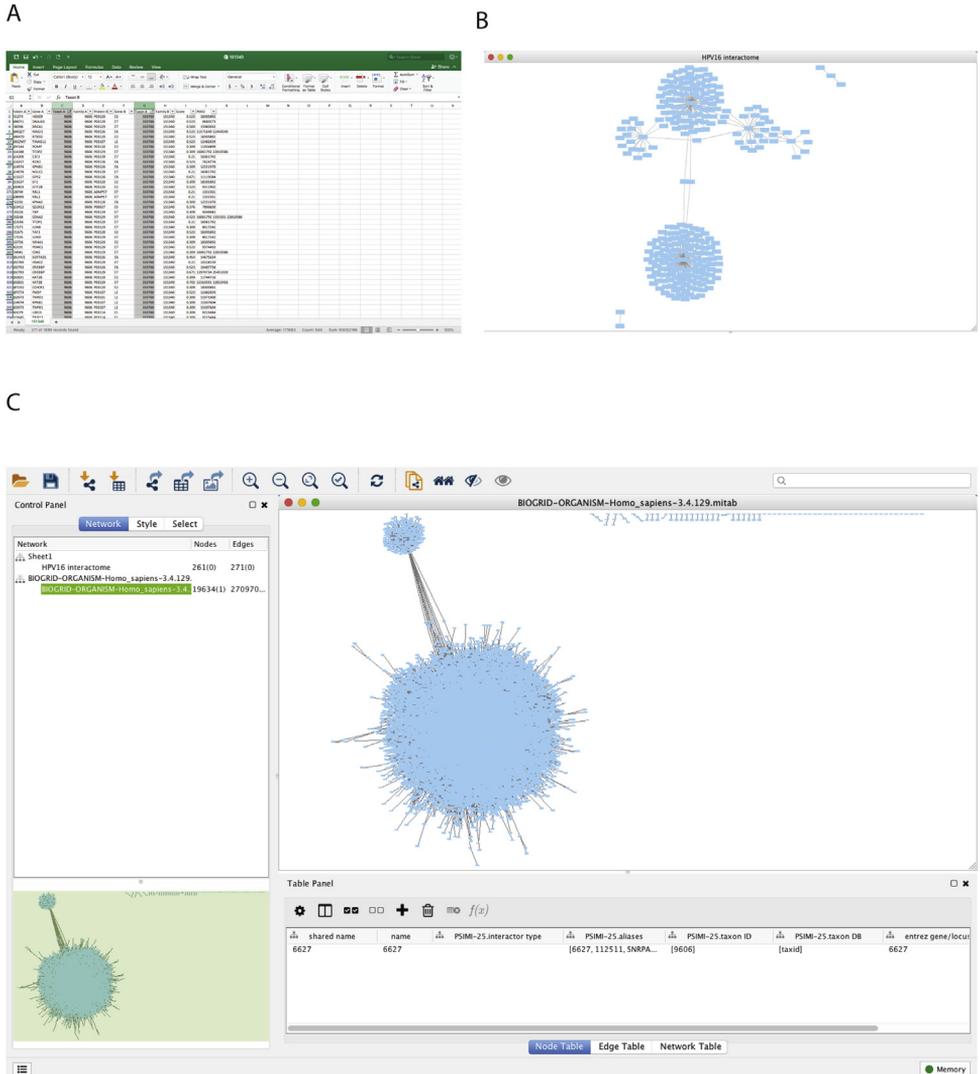


Fig. 1. Importing Raw Human and HPV Data. A) Screen capture of Excel workbook highlighting the filtered columns for Taxon A and Taxon B to select for HPV16 protein interactions specific for *H. sapiens*. B) Raw HPV16 interactome as visualised in Cytoscape. C) BIOGRID complete Human interactome as viewed in Cytoscape.

the entire human interactome (Fig. 1C). This was used to determine the direct and indirect effects of the HPV16 viral oncoproteins on gene expression.

Merging the two networks to create a virus and human interactome

The human-viral and BioGRID *H. sapiens* interactomes need to be merged together to visualise their connections. To merge the two networks, go to the option **Tools > Merge > Networks**. 'Shared name' and 'PSMI-25.alias' were selected as the attributes used to merge the HPV16 interactome and BioGRID Human genome interactome (Fig. 2A). The merge type was assigned 'Union', as the aim of this project was to integrate the HPV16 genes with those of the human genome.

A

B

| shared name | name | Matching Attribute | PSM-25 aliases | PSM-25 interactor type | PSM-25 Screen DB | PSM-25 Screen ID |
|-------------|---------------|---------------------------|---------------------------|------------------------|------------------|------------------|
| CM2 | 10172CM2 | CM2, 1017, 107452, CM... | 10107, 107451, CM2, ... | [None] | [None] | [None] |
| CM4 | 10103CM4 | CM4, 10103, 107453, CM... | 10103, 107454, CM4, ... | [None] | [None] | [None] |
| 1010 | 10103CM5 | 10102, 107453, CM5, PS... | 10102, 107453, CM5, ... | [None] | [None] | [None] |
| 10211 | 10211, IC... | 10211, 111570, ICAN... | 10211, 111570, ICAN... | [None] | [None] | [None] |
| 2081 | 104442041 | Z081, 10444, 111709, C... | 110444, 111709, Z081, ... | [None] | [None] | [None] |
| 10151 | 10151, Z08... | 10151, 111709, N040... | 10151, 111709, N040... | [None] | [None] | [None] |
| 11108 | 11016CM4 | 11108, 107513, CM4, CM... | 11108, 107513, CM4, ... | [None] | [None] | [None] |

C

| shared name | Matching Attribute | PSM-25 aliases | PSM-25 interactor type | PSM-25 Screen DB | PSM-25 Screen ID |
|-------------|--------------------|----------------|------------------------|------------------|------------------|
| E6(E7) | [None] | [None] | [None] | [None] | [None] |
| E7(E6) | [None] | [None] | [None] | [None] | [None] |
| E6(E7) | [None] | [None] | [None] | [None] | [None] |
| E7(E6) | [None] | [None] | [None] | [None] | [None] |
| E6(E7) | [None] | [None] | [None] | [None] | [None] |
| E7(E6) | [None] | [None] | [None] | [None] | [None] |
| E6(E7) | [None] | [None] | [None] | [None] | [None] |
| E7(E6) | [None] | [None] | [None] | [None] | [None] |
| E6(E7) | [None] | [None] | [None] | [None] | [None] |
| E7(E6) | [None] | [None] | [None] | [None] | [None] |

Fig. 2. Merged and Filtered Network. A) The pop-up screen containing the settings for the merging of the HPV16 interactome and the BIOGRID human interactome. B) The merged HPV16 and BIOGRID human interactome network. To the right of the network visualisation panel are the two oncogenes, E6 and E7. The yellow coloured nodes represent those genes that are direct interactors of E6 and E7. C) The smaller network containing E6 and E7, and their association with human genes.

Filtering the viral/human interactome. The HPV16 viral oncoproteins, E6 and E7, were searched for using the search network tool, and their respective nodes were moved away from the main interactome. This ensures for easier identification of the two genes in subsequent network creation.

For our example on HPV16, the two major oncoproteins E6 and E7 were selected, along with their direct interactors (first neighbors). This was done by searching for these two HPV16 oncogenes within the network, and selecting their first neighbors. Once all nodes of interest are selected, a new network is created by selecting **File > New > Network > from Selected nodes, Selected edges**. To tidy up this new network, select **Edit > Remove Duplicated edges**, and **Edit > Remove Self-Loops**.

This produces a smaller secondary network, as shown in Fig. 2B and C. This interactome contains the genes that directly interact with the viral oncoproteins, and their relationship with other human genes. It was this smaller interactome that was used for further network filtering and annotation, after the removal of duplicated edges and self-loops.

Assigning gene names and transcription factors. The gene targets of our miRNAs of interest, hsa-miR-33a and hsa-miR-496, were extracted from miRanda and saved as a .txt file (Fig. 3A). The transcription factors for hsa-miR-33a, hsa-miR-496 and SREBF2 were inferred from the UCSC Genome Browser and saved as a .txt file (Fig. 3B).

To select nodes from the miRNA target lists, which are downloaded as gene names, all nodes need to be changed to their respective gene name. This was conducted using the online tool DAVID ID (<https://david.ncifcrf.gov/conversion.jsp>) [6,7]. Export the merged interactome node list from Cytoscape and remove all columns except the 'Entrez ID'. Import the 'Entrez ID' list into DAVID to identify the gene names and abbreviations for each node. Download the gene names from DAVID and import into Cytoscape using **file > import**. Choose "name" as the key identifier. Once completed, all nodes in the interactome should now show their gene name (Fig. 3C). Alternatively, if using the human interactome from the BioGRID website the gene names are used automatically, rather than Entrez ID.

Downloading the miRNAs and their targets

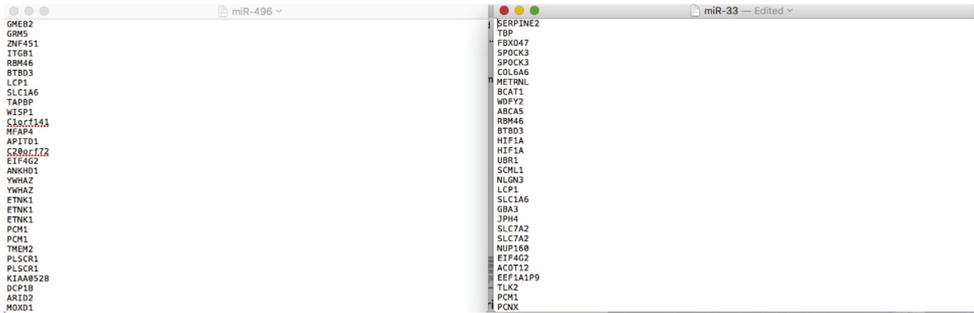
MiRNAs are important gene regulators, therefore it is essential include them in the interactome. The targets for the miRNAs of interest were retrieved from www.microRNA.org using the miRanda algorithm. Other data repositories can be used to obtain miRNA target information, such as TargetScan [8], miRTarBase [9] and miRDB [10]. Download the most recent file for human miRNA targets that are considered "highly conserved and good". Convert the miRNA-target files to a .txt file. For our example, we investigated the miRNAs miR-496 and miR-33a.

The transcription factors for each miRNA were inferred using the UCSC Genome Browser (<https://genome.ucsc.edu>) [11]. The UCSC predicted transcription factors are kept in separate lists to the miRNA targets. Each miRNA is searched individually using the FEB2009 (GRCH37/hg19) dataset. The track containing the transcription factors is titled "transcription factor chip-seq (162 factors) from ENCODE with Factor book motifs". To view the full track right click and select "full". The potential transcription factors were extracted 6 kbp upstream of the coding regions of the miRNAs of interest (Fig. 3B). The produced list of transcription factors for each miRNA were saved as a .txt file. In our example, we noted that miR-33a was encoded within the gene SREBF2, therefore, transcription factors regulating SREBF2 were also identified using the UCSC genome browser.

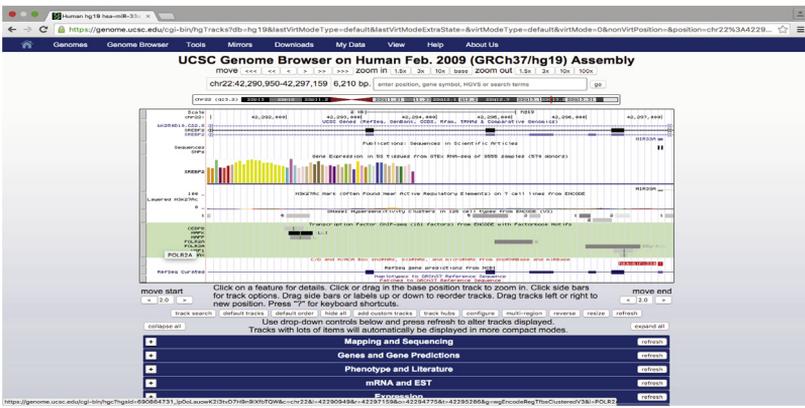
Addition of miRNAs and target genes to the virus and human interactome

The viral-human interactome needs to be filtered further to contain only targets and transcription factors of our miRNAs of interest. In our HPV16 (E6/E7) example, the gene lists containing the targets and transcription factors of miR-496, miR-33a and SREBF2 were used to select nodes. This was performed by going to the Cytoscape menu **Select > Nodes > From ID lists file** and choosing the gene list of interest. After this step, the adjacent edges of these nodes are selected by clicking **Select > Edges > Adjacent edges**. The nodes connected to those edges are then selected by going to **Select > Nodes > Nodes connected by selected edges** (Fig. 4A). Repeat this process for each miRNA or target of interest, without unselecting the previous nodes. After this step, a new network containing only the nodes and edges of interest can be created by selecting **File > New > Network > from selected nodes, selected edges**.

A



B

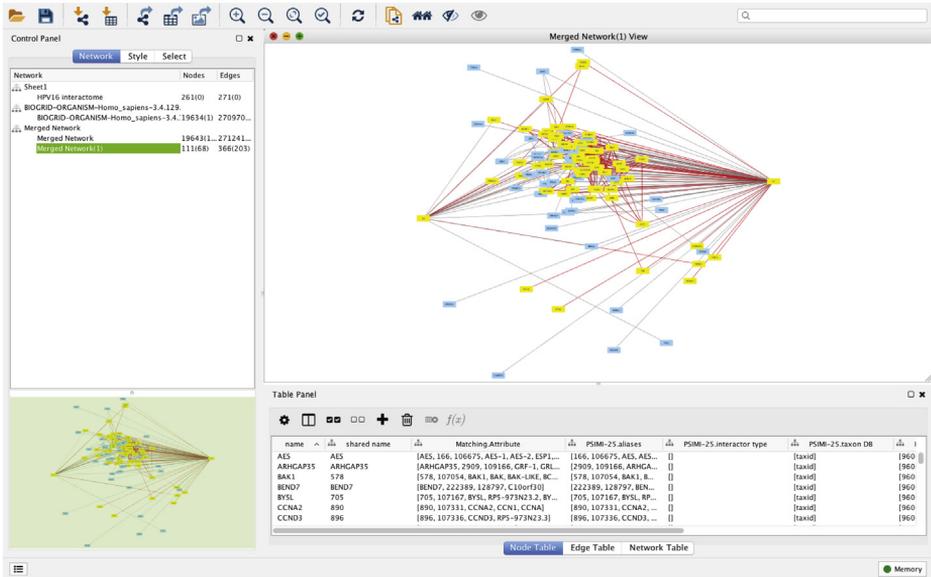


C



Fig. 3. Screenshots of the .txt files created for the miRNAs of interest. A) Gene target list for hsa-miR-33a (Left Panel). Transcription factor list of hsa-miR-496 as extracted from the UCSC Genome Browser (Right Panel). B) Screenshot of the UCSC Genome Browser highlighting the location of the transcription factors of the selected gene or miRNA. C) Screenshot of the DAVID output of the conversion from Entrez ID to Gene Name.

A



B

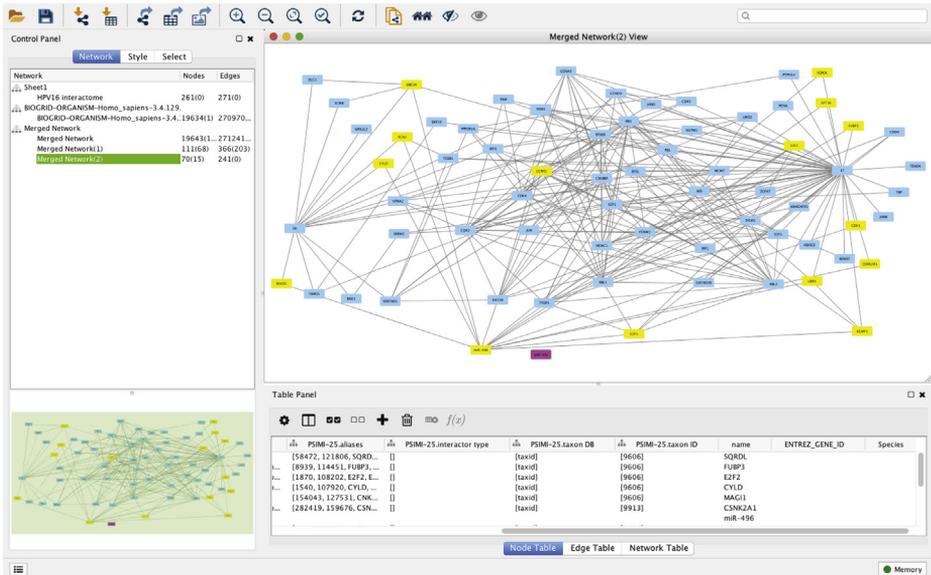


Fig. 4. A) Screenshot of the created viral interactome highlighting the nodes selected from the gene target and transcription factor lists and their direct interactors. B) Screenshot of the network, highlighting miR-496 and its connections to the highlighted nodes (yellow). The other miRNA, miR-33a, is shown in purple.

Secondary neighbors (or more) of these genes can also be included, if desired, by repeating the previous steps. Once the network is filtered to contain the relevant genes, the nodes for the miRNAs of interest need to be added manually by **right clicking** > **add** > **node**. The edges connecting these miRNAs to their targets and transcription factors also need to be added manually by **right**

clicking > add > edge connecting nodes. This process needs to be repeated for each miRNA (Fig. 4B). The target and transcription factor ID lists can be used to identify the nodes that require the addition of edges connecting to the miRNA of interest.

Annotation of genes and final virus/human/miRNA interactome

Once the miRNAs were added and connected to their respective targets, node annotation was performed to visually delineate between transcription factors, gene targets, and miRNAs. The 'gene name' column from the exported node table was used in a separate document to classify the nodes according to their connection to the HPV16 oncoproteins, and their characteristics (Fig. 5A).

To annotate the network, the created excel table was imported into Cytoscape, and the samples were matched according to 'gene name' (Fig. 5B). The inclusion of these annotations allows for the alteration of the visual properties of the network, such as varying the colour of nodes in relation to their targets (Fig. 5C), the shape of the nodes to indicate their biological role, and the colour of edges according to their regulatory interaction. The final interactome includes these features, in addition to edge annotations that indicate the direction of the transcription factor interactions (Fig. 6).

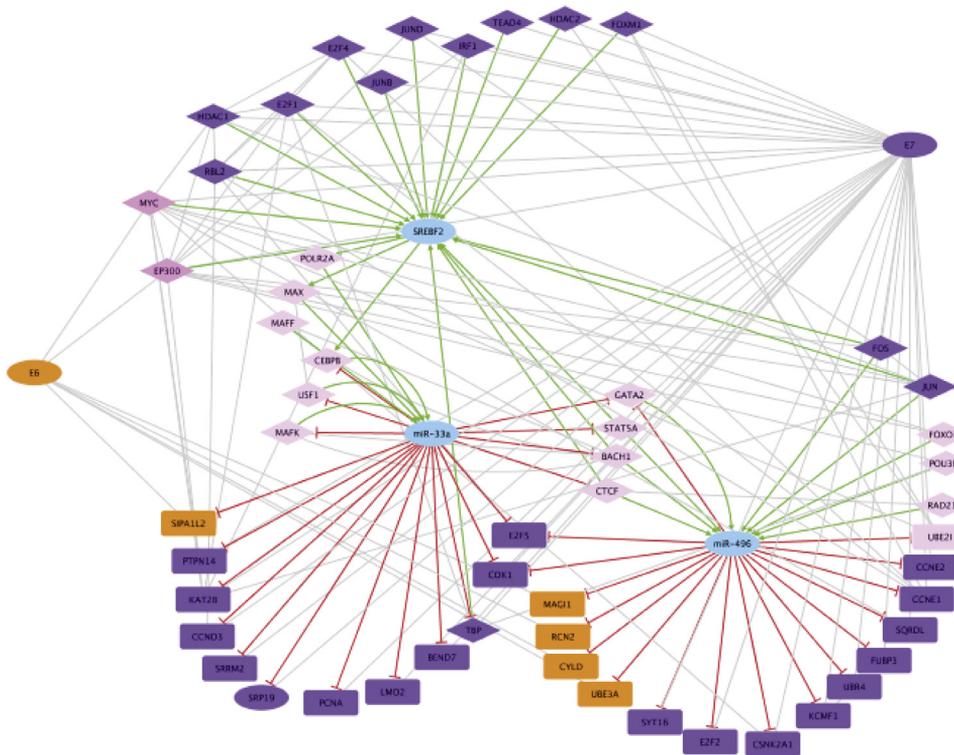


Fig. 6. The final HPV16-miRNA interactome created using the described Cytoscape methodology, indicating both the impact of transcription factors and genes on mRNA and miRNA expression. Orange represents targets of E6, blue indicates those targeted by E7, and the pink nodes are targeted by both E6 and E7. The lighter pink nodes represent secondary interactions to both E6 and E7, those proteins that are downstream targets of proteins directly targeted by these oncogenes. Additionally, the diamonds, rectangles and ovals represent the transcription factors, genes, and targets of interest respectively. Edge annotations were also added to indicate the direction of transcription factor control as either activating (green) or inhibitory (red).

Additional information

Changes to the human genome and the expression of its regulators, such as miRNAs, in response to viruses is highly complex. This is of particular importance in the case of virally driven oncogenesis, where further modifications to the regulatory network may compound pre-existing tumorigenic characteristics. Using the mapping software, Cytoscape, we developed a method to integrate the viral and human genome, along with miRNA regulators, which can be used to identify novel pathways and interactions. Our described method will enable researchers to readily identify targets and pathways of interest in the context of human-viral infection and the development of disease.

References

- [1] P. Shannon, et al., Cytoscape: a software environment for integrated models of biomolecular interaction networks, *Genome Res.* 13 (11) (2003) 2498–2504.
- [2] M.E. Smoot, et al., Cytoscape 2.8: new features for data integration and network visualization, *Bioinformatics* 27 (3) (2010) 431–432.
- [3] A. Calderone, et al., Mentha: a resource for browsing integrated protein-interaction networks, *Nat. Methods* 10 (8) (2013) 690.
- [4] A. Calderone, et al., VirusMentha: a new resource for virus-host protein interactions, *Nucleic Acids Res.* 43 (D1) (2014) D588–D592.
- [5] C. Stark, et al., BioGRID: a general repository for interaction datasets, *Nucleic Acids Res.* 34 (Suppl. 1) (2006) D535–D539.
- [6] D.W. Huang, et al., Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists, *Nucleic Acids Res.* 37 (1) (2008) 1–13.
- [7] D.W. Huang, et al., Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources, *Nat. Protoc.* 4 (1) (2008) 44.
- [8] V. Agarwal, et al., Predicting effective microRNA target sites in mammalian mRNAs, *elife* 4 (2015) e05005.
- [9] C.-H. Chou, et al., miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions, *Nucleic Acids Res.* 46 (D1) (2017) D296–D302.
- [10] N. Wong, X. Wang, miRDB: an online resource for microRNA target prediction and functional annotations, *Nucleic Acids Res.* 43 (D1) (2014) D146–D152.
- [11] W.J. Kent, et al., The human genome browser at UCSC, *Genome Res.* 12 (6) (2002) 996–1006.