Developing A Virus-MicroRNA Interactome Using Cytoscape

*Meredith Hill[1], *Dayna Mason[1], *Tânia Monteiro Marques [2], Margarida Gama Carvalho[2, ☯], Nham Tran[1, 3, ☯]

1. School Biomedical Engineering. University of Technology Sydney, NSW, Australia

2. University of Lisboa, Faculty of Sciences, BioISI – Biosystems & Integrative Sciences Institute, Campo Grande, Lisboa, Portugal

3. The Sydney Head and Neck Cancer Institute, Sydney Cancer Centre, Royal Prince Alfred Hospital, NSW, Australia

*Equal First Author

☯Corresponding author

Email address:

nham.tran@uts.edu.au

mhcarvalho@fc.ul.pt

Abstract

Summary

It is currently difficult to determine the effect of oncogenic viruses on the global function and regulation of pathways within mammalian cells. 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 is freely available at https://cytoscape.org/

Supplementary information: Supplementary Data is available at *Bioinformatics* online.

1. Introduction

Viral infection is associated with multiple different types of cancer, and is a key risk factor in tumour development. In 2008, 15% of the 14 million identified cancer cases were attributed to carcinogenic infections (De Martel, et al., 2012), which highlights the impact of viruses on the global cancer burden.

These carcinogenic viruses belong to multiple viral families, which may have different routes of transmission and infection. 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. These interactions may occur between proteins, genes and non-coding RNAs such as miRNAs. MicroRNAs are short (21nt), single-stranded, non-coding RNAs, that are involved in the regulation of gene expression (Calin, et al., 2002). It is estimated that greater than 60% of all protein-coding genes are regulated by miRNAs, with a single miRNA having the ability to control thousands of gene targets (Bartel, 2004; Friedman, et al., 2009).

The high degree of miRNA involvement in the control of gene expression compounds its regulatory complexity. Previously, we mapped the interactions between HPV16 and miRNAs in Head and Neck Squamous Cell Carcinoma (HNSCC) (Mason, et al., 2018). This systematic mapping afforded a greater appreciation of the gene networks and was implemented in the identification potential regulatory mechanisms.

However, the visualisation and mapping of the many thousands of gene interactions on a comprehensive scale is currently a challenging task. At present, there is very little information or suitable technical papers dedicated to the interrogation and mapping of viral-gene interactions. This situation is further exacerbated, as molecular biologists are often not experts in the analysis and visualision of large biological datasets. Free software packages, such as Cytoscape (Shannon, et al., 2003; Smoot, et al., 2010), aid scientists in the visualisation and analysis of organism-specific molecular interactions. Specifically, Cytoscape can be used to map, filter, and interpret large amounts of data. Thus it may be used to uncover cellular processes involved in virally associated cancer development.

Our described methodology utilises Cytoscape and freely-available databases for the analysis and visualisation of large biological datasets, and aims to integrate and map the interactions between viral oncogenes, their cellular targets, and the impact of other regulators on the resultant network.

2. 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 particular emphasis on the effect of the viral oncoproteins E6 and E7 on miRNA and gene regulation. For this interactome, we specifically focused upon hsa-miR-33a, hsa-miR-496, and SREBF2. This method is shown in Figure 1, and can be adapted to a specific virus and miRNA(s) of interest.

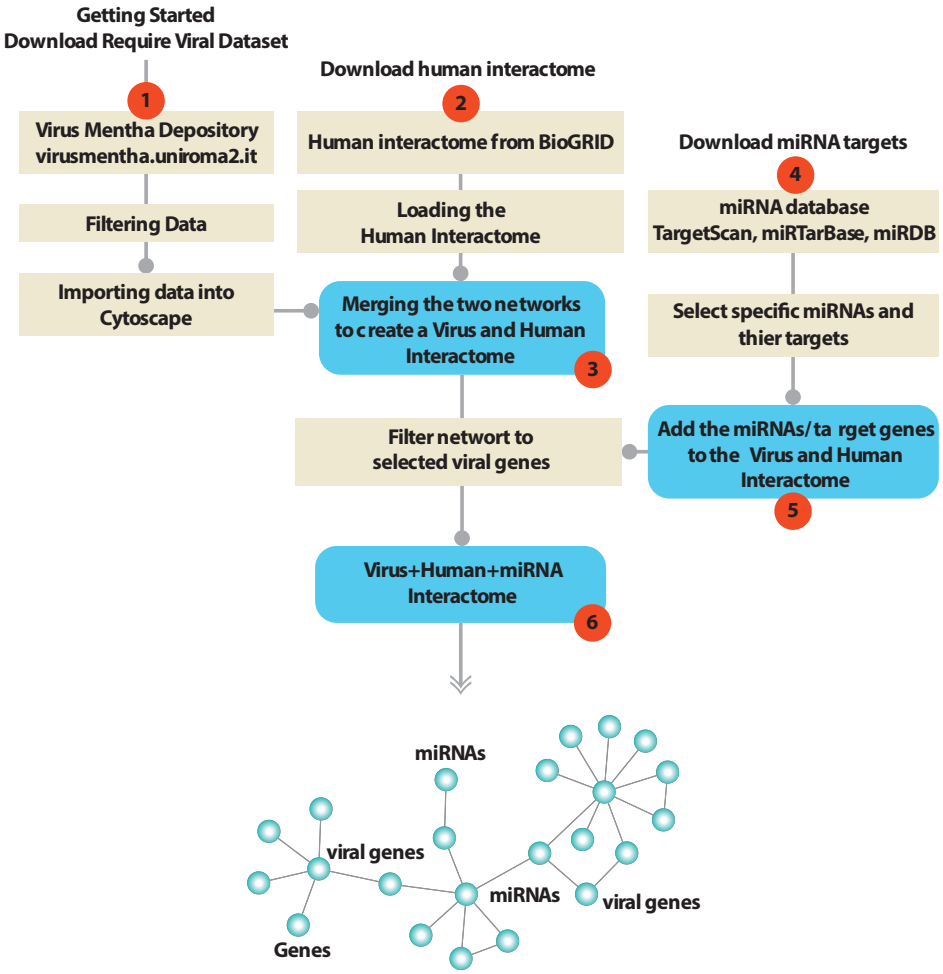

Figure 1: Flowchart describing the construction of a virus-miRNA interactome. Each of the numbers shown correspond to the step by step downloadable guide.

Downloading Raw Human and HPV Data

The HPV16 viral protein interactions were downloaded from the Virus Mentha website, and filtered in Excel by taxon 333760 (HPV16) and 9606 (H. sapiens) (Supplementary Figure 1A). This dataset was imported into Cytoscape as the HPV16 interactome (Supplementary Figure 1B). This network formed the basis for the final viral-human interactome. The BIOGRID human interactome was loaded from the Cytoscape menu (Supplementary Figure 1C). This was utilised to determine the direct and indirect effects of the HPV16 viral oncoproteins on gene expression.

Merged and Filtered Network

Both the HPV16 interactome and BIOGRID Human genome interactome were merged by ‘shared name’ and ‘PSMI-25.alias’ (Supplementary Figure 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. 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 for the subsequent creation of a smaller network.

Both E6 and E7 were selected, along with their direct interactors to produce a smaller secondary network, as shown in Supplementary Figure 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 (Supplementary Figure 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 (Supplementary Figure 3B). As the genes in the network were labelled by Entrez ID, the DAVID name conversion tool was used to convert this number identifier to 'Gene Name' (Supplementary Figure 3C). These gene names were then imported into Cytoscape to correspond with their respective Entrez ID. This allowed for the genes to be renamed within the table and resultant network.

Filtering the Network and Addition of miRNAs

The gene and transcription factor lists of hsa-miR-33a, hsa-miR-496 and SREBF2 were used to identify the appropriate nodes within the network. The secondary interactors of these nodes were also selected to identify potential indirect effects of viral oncoproteins. The selected nodes within the network are highlighted in Supplementary Figure 4. These interactions were used to create a new network, which was used for the remainder of the analysis.

Within this network, nodes were added and renamed 'miR-33a' and 'miR-496'. The colour of these nodes was modified using the style panel to differentiate them from the surrounding nodes. The targets and transcription factors of each miRNA were selected and connected to their respective miRNA through the addition of individual edges. This was repeated for each miRNA to visualise their regulation of the genes affected by HPV16 E6 and E7.

Annotation of Genes

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 (Supplementary Figure 5A).

To annotate the network, the created excel table was imported into Cytoscape, and the samples were matched according to ‘gene name’ (Supplementary Figure 5B). The inclusion of these annotations allows for the alteration of the visual properties of the network, such as varying the colour of nodes according to their targets (Supplementary Figure 5C), the shape of the nodes to indicate their biological role, and the colour of edges according to their regulatory interaction. The final interactome, which includes these features, in addition to edge annotations indicating the direction of the transcription factor interactions, is shown in Supplementary Figure 6. A step by step guide is also included in the supplementary download.

3. Discussion

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 the tumorigenic characteristics of a cell. 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.

In our working example, it was uncovered that the overexpression of the HPV16 oncoproteins, E6 and E7, targets specific transcription factors that possibly have a downstream effect on the expression levels of miR-33a, miR-496 and SREBF2. The identification of these pathways was made possible through the integration of the viral and human genome, and have since been found to have an impact in HNSCC (Mason, et al., 2018). Therefore, this Cytoscape methodology may be utilised in discovering novel viral-human interactions, but can also complement findings established from traditional laboratory methods. In this way, it provides a more rigorous and thorough understanding of the underlying pathways involved in virally driven carcinogenesis.

Our described methodology may be applied to a wide range of viruses. This is highly advantageous, as it can be applied to many different areas of study within the virology field, and is not just limited to those related to cancer. Previously, identifying pathways and genes to investigate has been a case of ‘luck-of-the-draw’. However, this style of investigation is limited with the use of our methodology, as researchers

can easily visualise targets of interest, thus reducing time performing initial investigative work. It may also be applied after certain techniques, such as microarrays, providing more information to the results obtained in relation to genome interactions and changes with experimental conditions.

The addition of miRNAs adds another layer to the specificity and uniqueness to this methodology. The modification of gene expression by the introduction of the virus would have a downstream influence on the levels of miRNAs. The result of this is the dysregulation of target regulation, which may aid in magnifying the oncogenic or disease specific effect of the virus. Again, this is observed in our example, whereby miR-496 and miR-33a regulate genes and transcription factors that are not directly affected by the HPV16 oncoproteins, but are secondary interactors of E6 and E7. Therefore, the inclusion of miRNAs to this network analysis provides a broader and more complete view of cellular changes through the addition of information on the downstream effects on genes that are not direct targets of the virus. Also, to our knowledge, no methodology has been previously published describing the integration of the viral and human genome with miRNA interactions, establishing our method as holistic and unique.

Although this method provides a structured way of both determining and visualising the impact of viruses on miRNAs and proteins within a cell, it does have several limitations. For one, the inclusion of large amounts of data and information to the network masks any potential functional pathways that may have a role in virally-induced disease. Thus, large interactomes are harder to visualise and interpret. On the other hand, filtering down a larger network may introduce unintentional bias

through the selection of genes that have a specific function, for example, transcription factors. This has the potential to exclude other important pathways that are altered by the virus, which may otherwise be identified. Therefore, several aspects of the network details need to be considered if the filtered network is to accurately reflect the pathways affected by viral expression.

The use of Cytoscape plug-ins may provide more information that is useful in determining influential viral targets, their regulation, and the identification of diagnostic or prognostic biomarkers. Although not applied in our example interactome, altering the size of the nodes according to their number of connecting edges, termed degree, is one method to determine key players within the created network. Additionally, the use of Gene Ontology (GO) analysis using plug-ins such as BiNGO (Maere, et al., 2005), allows for the identification of molecular and cellular pathways that are most affected by the introduction of the virus. By gathering more information about the network, targets can be included or excluded for further experimental analysis, which increases the chance of selecting a highly involved gene or miRNA for further study.

Overall, we established a method using Cytoscape to create and visualise interactions that occur between known viruses and the human genome, and the impact of miRNAs on these interactions. Our described method will enable researchers to more easily identify targets and pathways of interest in the context of human viral infection and the development of disease.

4. References

Bartel, D.P. MicroRNAs: genomics, biogenesis, mechanism, and function. *cell* 2004;116(2):281-297.

Calin, G.A.*, et al.* Frequent deletions and down-regulation of micro-RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proceedings of the National Academy of Sciences* 2002;99(24):15524-15529.

De Martel, C.*, et al.* Global burden of cancers attributable to infections in 2008: a review and synthetic analysis. *The lancet oncology* 2012;13(6):607-615.

Friedman, R.C.*, et al.* Most mammalian mRNAs are conserved targets of microRNAs. *Genome research* 2009;19(1):92-105.

Maere, S., Heymans, K. and Kuiper, M. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* 2005;21(16):3448-3449.

Mason, D.*, et al.* Human papillomavirus 16 E6 modulates the expression of miR-496 in oropharyngeal cancer. *Virology* 2018;521:149-157.

Shannon, P.*, et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome research* 2003;13(11):2498-2504.

Smoot, M.E.*, et al.* Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* 2010;27(3):431-432.