IDENTIFICATION OF MIRNA-TARGETED CELLULAR PATHWAYS IN FLAVIVIRUS-INFECTED CELLS

By

Chantel C. McCoy

THESIS

Presented to the Department of Medical Informatics and Clinical Epidemiology
and the Oregon Health & Science University
in partial fulfillment of
the requirements for the degree of

Master of Science

June 2012

School of Medicine

Oregon Health & Science University

Certificate of Approval

This is to certify that the Master's Thesis of

Chantel C. McCoy

"Identification of miRNA-targeted cellular pathways in Flavivirus-infected cells"

Has been approved

_____

Thesis Advisor

_____

Committee Member

_____

Committee Member

_____

Committee Member

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGMENTS

I would like to thank Dr. Armand Bankhead for advising me on this thesis and for being a strong and reliable reference for questions and brainstorming ideas with. Second, I want to thank Dr. Shannon McWeeney for her willingness to guide me through this process and for reminding me to keep up the good work. I also want to thank Dr. Jay Nelson for his instrumental biological perspective and constructive input. Members of the Nelson lab, especially Dr. Alec Hirsch and Dr. Jessica Smith have been great resources throughout this process. Thank you.

Abstract.

MicroRNAs (miRNAs) are small non-coding RNAs that are responsible for post-transcriptional gene silencing. These miRNAs are associated with the RISC (RNA-Induced Silencing Complex) that uses a seed sequence to target specific genes. Both expression of miRNAs and genes within virus-infected cells provide independent data in identifying biological hypotheses concerning regulation changes during infection. To do this, three data types are available for integration: gene expression microarray, miRNA expression microarray and RISC-Immunoprecipitation (RISC-IP) microarray. RISC-IP data provides a look at what genes are associated with the complex during infection versus uninfected cells (mock). Incorporating three independent data types allows a more complete representation of host response to Flavivirus infection that is key to identifying miRNA regulators. This thesis integrates all three types of data for Flavivirus-infected cells using the statistical programming environment R to identify statistically enriched miRNAs regulating host response. First, the integration of gene expression changes and enriched RISC-associated genes identify significant miRNA regulators overlapping with miRNA expression data. Dengue gene expression was found to have a significant positive association with one or more differentially expressed miRNAs. More specifically, Dengue-infected cells' up-regulated genes are significantly associated with down-regulated miRNAs. No significant associations were found in West Nile-infected cells. Results suggest evidence Dengue gene expression and miRNA expression are complementary and suggests that miRNAs and genes are being co-regulated during infection. Secondly, regulatory networks were built using differentially expressed genes as identifiers of statistically common miRNA regulators between West Nile and Dengue

network modules. Comparing all Dengue modules with all West Nile gene network modules, 20 of 114 comparisons contained miRNA regulators statistically significantly overlapping. Eight miRNAs regulate these modules: hsa-let-7b, hsa-miR-1, hsa-miR-124, hsa-miR-155, hsa-miR-16, hsa-miR-29c, hsa-miR-30, and hsa-miR-373. These miRNAs, common to both Dengue and West Nile regulatory networks, suggest co-regulation of genes that are changing during host response to infection. Focusing on more informative miRNAs, those regulating genes during infection, will direct experimental research efforts to key miRNAs for experimental validation with the end-goal of guiding gene therapy and non-viral drug development in hopes to reduce the number of Flavivirus induced deaths.

CHAPTER 1: Introduction

The *Flavivirus* genus includes both West Nile and Dengue viruses, both of which cause severe human disease such as encephalitis and Dengue fever (1). However, the Center for Disease Control report about 90% of West Nile virus and 50% or more of Dengue cases are asymptomatic (2). West Nile and Dengue viruses are considered an endemic of some sub-tropical countries. Hundreds of thousands of deaths occur from Flavivirus infections each year because of the lack of antiviral drugs or vaccines (3). During infection, West Nile and Dengue viruses infect individuals primarily through mosquito bites. Treatment for infection of Flaviviruses such as West Nile and Dengue are needed in subtropical areas of the world where viral-infected mosquitoes are most prevalent. In order to understand the biology of Flavivirus infection to guide the development of antiviral treatment, the molecular mechanisms of the viruses must be explored.

The goal of this thesis is to associate one or more miRNAs with changes in host response to West Nile and Dengue viruses within networks by integrating three different data types. The research question that we are asking is: Which miRNAs regulate infection response in West Nile- and Dengue-infected cells? As a result of the research and fulfillment of the goals in this thesis, the integration of three independent data types can be achieved in order to create a more informative picture of Flavivirus infection. The miRNAs that regulate the response to infection, both from direct expression measurements and computational predictions based on published target genes, will allow for focused validation *in-vitro*. Understanding which miRNAs target transcripts changing in the cell after viral infection guides research of vaccine and gene therapy development

to specific miRNAs and their target genes. Rather than predicting miRNAs and their targets using sequence matching techniques, our approach incorporates several independent types of data during viral infection to identify key regulation within the cell which is imperative for developing vaccines or other gene therapy solutions in the interest of advancing research and reducing cost.

Background.

Members of the *Flavivirus* genus contain a positive sense single stranded RNA genome (4) and use host cellular machinery for genomic replication and produce progeny virions. One way that the host cell responds to invasion of a virus is through altering expression of gene transcripts in the cytoplasm.

Post-transcriptional changes in gene expression can be controlled by microRNAs (miRNAs) which are small non-coding RNAs produced by the cell to silence gene expression (5). The precursors of miRNAs are longer RNA structures that fold back onto themselves and form a hairpin structure. This double-stranded RNA is then cleaved by Dicer in the cytoplasm into ~22 nucleotides (6). One strand is degraded while the other becomes associated with the RISC (RNA-induced silencing complex). To silence expression, miRNAs bind to target transcripts as part of the complex by complementarity. For silencing to occur, a six to eight nucleotide seed region most often found in the 3-prime untranslated region (3'UTR) initiates binding between a particular miRNA with a target transcript (7). This process results in a considerable decrease in protein translation, thus turning off the gene.

This process is better known as RNAi or RNA-interference, a conserved defense

mechanism that suppresses gene translation directed by either endogenous miRNAs or exogenous double stranded RNA(8). RNAi serves as a system to protect the cell from viral infection including strategies for regulating gene expression, occurring post-transcriptionally to degrade mRNA transcripts (9).

In response to infection, the host cell may initiate the use of one miRNA or a combination of miRNAs to target genes. One example of using miRNAs as an avenue for gene therapy in Flaviviruses is explained in a study from 2011 which investigated insertion of cell-specific miRNA targets in the Flavivirus genome as an approach to control pathogenesis and virus attenuation(10). When response of the host to infection is understood, development of useful vaccines may be made possible for clinical cases.

Furthermore, the change in expression of transcripts provides information on which genes are changing in response to infection. Transcripts common to a pathway or more specifically, protein-protein interactions can enhance the comparison of West Nile and Dengue virus infections at a systems biology level. By using gene expression data of infected cells versus mock (uninfected cells), gene regulatory networks can be derived to model gene regulation within the cell. Viewing gene regulation within infected cells using a combination of different data types is key to understanding virus infection at a systems level because it creates a more complete picture of gene regulation as opposed to using one or multiple data types independently.

Specific Aims/Thesis statement.

The goal of this thesis is to associate a miRNA or group of miRNAs with changes in viral host response networks. The aims for this research are: 1) Combine RISC-

associated and differentially expressed genes to identify statistically enriched miRNAs regulating host response in West Nile and Dengue infections, and 2) Build regulatory networks for West Nile and Dengue virus infections using differentially expressed genes to identify common miRNA regulators.

The research question behind the first specific aim is: Can we more accurately predict miRNA-mRNA interactions using RISC-associated genes and differentially expressed genes together? In other words: Do the genes that associate with the RISC complex and the genes up- and down-regulated in the cell tell us more than the two independently. We hypothesize that using both data types will be more informative than the two data types individually. The second aim asks the question: Do the miRNA regulators shared between West Nile and Dengue virus regulatory networks overlap significantly? The hypothesis is that there will be significant commonality between miRNA regulators. Both specific aims were tested using the methodology described in the next section.

CHAPTER 2: Methods

Data.

Fortunately, comparable data describing processes of both West Nile and Dengue virus during infection of host cells is available. Jay Nelson's laboratory at Oregon Health and Sciences University performed analogous experiments using both viruses and collected three types of data: 1) gene expression microarray data, 2) RISC-IP microarray data, and 3) miRNA microarray expression data on both West Nile- and Dengue-infected cells versus mock (uninfected) cells. To address the aims described above, West Nile and Dengue virus data was analyzed according to data type (see Table 1 below for list of data types and platforms).

	Data type	Platform	Time (Hrs post-infection)
Dengue	Gene expression	Illumina human ref-8	48
	RISC-IP miRNA expression	Illumina HT-12v4 miRCURY LNA miRNA array 6th generation	48 8,24,56
West Nile	RISC-IP used for Gene expression RISC-IP	Illumina human ref-6 Illumina human ref-8	48 48
	miRNA expression	Illumina Human HT-12	48

Table 1. West Nile- and Dengue-infection experimental design, data types and platforms for data analyzed in this study

In order to address either of our aims, each of the data-sets had to be analyzed separately to identify differentially expressed genes (for the gene expression arrays), differentially expressed miRNAs (for the miRNA arrays) or enriched genes associated with the RISC complex (for the RISC-IP arrays). Microarray data was collected from Dengue- or West Nile-infected HEK-293 cells in comparison to mock-infected HEK-293 cells and analyses were performed using the statistical programming environment R

(version 2.14.1). All of the datasets came from an Illumina platform, with the exception of the Dengue miRNA expression microarray. Illumina data was imported into R to be processed using the Lumi Bioconductor package (11). The R Bioconductor package Limma(12) was used to create linear models based on contrasts between infected cells versus mock-infected cells for gene expression data and for RISC-IP data to compare virus immunoprecipitation (IP) with virus total expression versus mock IP with mock total expression. In the Dengue miRNA expression dataset, there are multiple time points: 8, 24 and 56 hours post-infection (see Table 1). The largest difference in infection time was included in our analyses: 56 versus 8 hours post-infection in virus-infected cells versus mock-infected cells.

From Limma, we produced a table of top-ranked genes bases on significance specified by a false-discovery rate (FDR) adjusted p-value. The West Nile RISC-IP array was not done in duplicate samples and thus a statistical analysis could not be conducted for this data set. Instead, enriched genes were determined by a $\log_2$ fold change cutoff of 0.58 which corresponds to a 1.5 fold change for non-$\log_2$ expression values.

Flagging probes.

Negative control probes were used within a data-set to control for background noise. A value representing the ninetieth percentile for $\log_2$ expression values of negative controls within each sample was used as a threshold for background noise. Any probes falling below this value were flagged with a ‘0’ and noted as potentially insignificant. All probes went through the linear model and those probes that may have been differentially expressed or enriched in the final probe list were labeled by either a flag ‘0’ for below

background noise or a flag of ‘1’ for passing the background filter. For data sets without negative control probes to control for background noise, probes expressed below the twenty-fifth percentile across all samples were labeled as potentially insignificant.

Aim 1 Analyses: Associating two independent data types with differentially expressed miRNAs

The goal of the first aim is to identify association of mRNA gene expression, RISC-IP, and mRNA gene expression/RISC-IP, with differentially expressed miRNAs regulating host response in West Nile and Dengue infections. Analytical tools can be used to associate miRNAs with gene lists. Several tools are described in the literature (13) including Sigterms which consists of a set of Excel macros, MMIA (“MicroRNA and mRNA integrated analysis”) which is web-based (14)(15) and CORNA, an open-source R package (13). Since the three data types are imported and processed within the statistical programming environment R, the CORNA package was originally chosen for the analysis. CORNA takes gene lists as input and analyzes significant miRNA targets using the seed sequence from miRBase and microRNA.org (16). Seed sequences are contained in the miRBase and miRNA databases within the listed websites and define how CORNA maps transcripts as targets of a particular miRNA through base complementarity. Although CORNA is a good tool for identifying potential gene targets for miRNAs, it does not restrict based on validated miRNAs from the literature. Instead, it includes predictions which may be based only on the seed sequence. As an alternative, TarBase was selected to map transcripts to miRNAs because it is based on confirmed target transcripts with their associated miRNAs from evidence through various biological

methods (17). TarBase provides a downloadable file that contains validated human miRNAs and the genes they target. Although this significantly lowers the number of genes and miRNAs that can be used in the analysis, the results are more biologically meaningful and reliable for modeling infection in the cell.

To test the research question related to specific aim #1: Can we more accurately predict miRNA-mRNA interactions using RISC-associated genes and differentially expressed genes together? TarBase was utilized to test three separately generated gene lists for West Nile and Dengue independently. The first list of genes is comprised of the transcripts differentially expressed from the gene expression data. The second list contains genes significantly associated (enriched) with the RISC complex during infection from the RISC-IP data set. Those genes that are common to both of these lists were combined to create a third gene set. Additionally, lists were constructed using directionality in differential expression of genes. Biologically, genes that are differentially down-regulated are more likely associated with up-regulated miRNAs (18)and we hypothesize genes that are up-regulated are more likely associated with down-regulated miRNAs. Addressing direction of expression in both genes and miRNAs tests the hypothesis that miRNAs are regulating genes during infection.

We examined the proportion of overlap of miRNAs via Fisher's exact test in order to assess whether there is a positive association between the differentially expressed gene list and differentially expressed miRNAs. Level of significance for all tests was 0.05.

Our hypothesis is that there is positive association between differentially expressed target genes who map to one or more miRNAs. Testing each gene list, we hypothesize that the genes that are both differentially expressed and enriched from the

RISC-IP (gene list #3) have a higher association with the differentially expressed miRNAs than the other two gene lists. Assuming the association is significant, the two data types are anticipated, when combined, to be more beneficial in analyzing viral infection processes than independently. To address directionality of the gene and miRNA expression, additional contingency tables were constructed. Target genes that are down-regulated were tested to overlap significantly with miRNAs being up-regulated and similarly, target genes that are up-regulated were tested to significantly overlap with miRNAs that are down-regulated. Since the transcripts from the RISC-IP data are enriched genes only, they were tested to be positively associated with up-regulated miRNAs. To determine how to classify up- or down-regulated differentially expressed genes, a $\log_2$ expression fold change cutoff of +/- 0.58 was set. Genes that had a $\log_2$ expression fold change of greater than or equal to 0.58 were categorized as up-regulated, and those with $\log_2$ expression values less than or equal to -0.58 categorized as down-regulated genes corresponding to a 1.5 non-$\log_2$ fold change in expression.

Aim 2 Analyses: Build regulatory using differentially expressed genes to identify common miRNA regulators

The research question of specific aim #2 asks: Are the miRNA regulators shared between West Nile and Dengue virus regulatory networks? This aim uses the miRNAs mapped from differentially expressed genes from mRNA expression microarrays of West Nile and Dengue virus infections. Two interaction networks were created using protein-protein interactions from the Pathway Commons database (19) which was incorporated into R using an interaction file of protein-protein interactions for genes in human. These

large protein-interaction networks were created for each virus independently. Input for the networks came from the differentially expressed genes for each virus and restricted the interactions to only those genes. The network was then broken into modules following the methods of (20). Modules are organized based on protein relationships in the form of an adjacency matrix of binary values made to represent the differentially expressed genes' interactions with themselves where a '1' indicates interaction between the two proteins and a '0' indicates no interaction. This was then organized into hierarchical clusters and visualized by a dendrogram graph structure (see Appendix: Figure A-1) (21). Colors at the bottom of the dendrogram represent putative gene modules, whereas the grey represents genes not belonging to a specific module or genes that did not interact with any other genes in the list. Branches on the dendrogram are clustered groups of genes which can be detected through a variety of methods. The 'Dynamic Tree Cut' (21), has been used in other species such as yeast (22) (20) and mouse (23) and uses a top-down algorithm to create clusters using the dendrogram and breaking them down and rebuilding until the cluster is stable and distinct from other clusters. Resulting modules within a network were used to define pathways in each virus for subsequent comparison.

Just as used for testing specific aim #1, contingency tables were constructed to quantify the overlap of miRNAs within modules between West Nile and Dengue. Pairwise comparisons of genes within West Nile modules with genes in Dengue modules as well as similarly comparing the miRNAs from those target genes provide counts for the contingency tables. The significance of two modules overlapping was determined using the Fisher's exact test such that the contingency table contained: 1) the number of

genes shared between a West Nile and Dengue module comparison, 2) the number of genes not in West Nile's current module but in Dengue's current module, 3) the number of genes not in Dengue's current module but in West Nile's current module, and 4) the number of genes that were in neither. The genes that comprised the fourth category (neither) are all remaining genes in the interactome used for the protein-protein interactions from Pathway Commons. All genes within the interactome were chosen to comprise the 'gene universe' or reference set, but also done similarly using genes for all differentially expressed probes as the reference set, and again using all probes that are shared on the Dengue and West Nile gene expression arrays. All three approaches are relevant and valuable for examining significance of gene overlap between Flavivirus module overlap.

A similar approach was taken to compare the overlap of miRNA regulators by using the confirmed target genes within a module and mapping to miRNAs. The 'miRNA universe' was limited to confirmed target gene/miRNA pairs from TarBase. For both comparison of genes and miRNAs across modules, p-values were FDR adjusted. For those significant module comparisons between the two Flaviviruses, a list of miRNAs was found to regulate overlapping modules. Evidence for co-regulation of miRNAs was also explored. A table of miRNAs regulating the significant module comparisons was constructed to determine whether the same miRNAs regulate multiple modules. Modules were also overlaid with Gene Ontology (GO) ontologies for pathway annotation using CORNA, an R package (16). To visually represent the network overlap between a particular Dengue and West Nile module overlap, the software program Cytoscape v.2.8.3 was used to produce gene networks showing target genes and overlapping genes,

which were tested for GO enrichment using the BINGO plug-in (24). BINGO compared the overlapping genes with annotated *Homo sapiens* genes testing for enrichment using a Hypergeometric test, and adjusting for false-discovery rate.

CHAPTER 3: Results and Evaluation

Aim 1.

The Dengue mRNA expression data-set contained 4,396 differentially expressed genes and the West Nile RISC-IP data-set used as a substitute for expression data produced 2,521 differentially expressed genes. Differentially expression was determined by the methodology explained in the Data section of Chapter 2: Methods. More specifically, Dengue differentially expression was defined by a Bonferroni adjusted p-value of 0.05 while West Nile differential expression by an unadjusted p-value of 0.1. The only statistically significant association for each of the three gene lists in the Flaviviruses was for differentially expressed Dengue genes targeted by one or more differentially expressed miRNAs (p-value = 0.0268; Table 2A). Tables 3 and 4 shown below did not test significant for any association between differentially expressed and/or enriched genes being targeted by one or more differentially expressed miRNAs.

A) Dengue p-value = 0.0268		Genes targeted by one or more DE miRNAs		
		Yes	**No**	**Total**
Dengue DE genes	**Yes**	96	129	225
	No	159	299	458
	Total	255	428	683

B) West Nile p-value = 0.9044		Genes targeted by one or more DE miRNAs		
		Yes	**No**	**Total**
West Nile DE genes	**Yes**	7	113	120
	No	53	544	597
	Total	60	657	717

Table 2. Contingency table for Dengue (A) and West Nile (B) showing the number of DE (differentially expressed) genes and non-DE genes that are targets of one or more DE miRNAs, or no DE miRNAs; Fisher's exact test for positive association is significant in Dengue only.

A) Dengue p-value = 0.9991		Genes targeted by one or more DE miRNAs		
		Yes	**No**	**Total**
Dengue enriched genes	**Yes**	46	121	167
	No	212	313	525
	Total	258	434	692

B) West Nile p-value = 0.9980		Genes targeted by one or more DE miRNAs		
		Yes	**No**	**Total**
West Nile enriched genes	**Yes**	6	152	158
	No	54	476	530
	Total	60	628	688

Table 3. Contingency table for Dengue (A) and West Nile (B) showing the number of enriched genes and non-enriched genes that are targets of one or more DE miRNAs, or no DE miRNAs; Fisher's exact test for a positive association is not significant.

A) Dengue p-value = 0.9688		Genes targeted by one or more DE miRNAs		
		Yes	**No**	**Total**
Dengue DE/enriched genes	**Yes**	19	49	68
	No	234	374	608
	Total	253	423	676

B) West Nile p-value = 0.9763		Genes targeted by one or more DE miRNAs		
		Yes	**No**	**Total**
West Nile DE/enriched genes	**Yes**	1	40	41
	No	57	587	644
	Total	58	627	685

Table 4. Contingency table for Dengue (A) and West Nile (B) showing the number of DE/enriched genes and all other genes that are targets of one or more DE miRNAs, or no DE miRNAs; Fisher's exact test for association is not significant.

Using directionality of expression in both the gene list and miRNA expression, 27

Dengue up-regulated target genes are significantly positively associated with one or more Dengue down-regulated differentially expressed miRNA (p-value = 6.906e-15; Table 5A). On the other hand, only one of the 125 Dengue down-regulated target genes were found to be targeted by one or more up-regulated miRNAs and found to be insignificant(p-value: 0.5556; Table 5B).

A)

p-value = 6.906e-15

		Genes targeted by one or more down-regulated DE miRNAs		
		Yes	**No**	**Total**
Dengue up-regulated DE genes	**Yes**	27	43	70
	No	1	154	155
	Total	28	197	225

B)

p-value = 0.5556

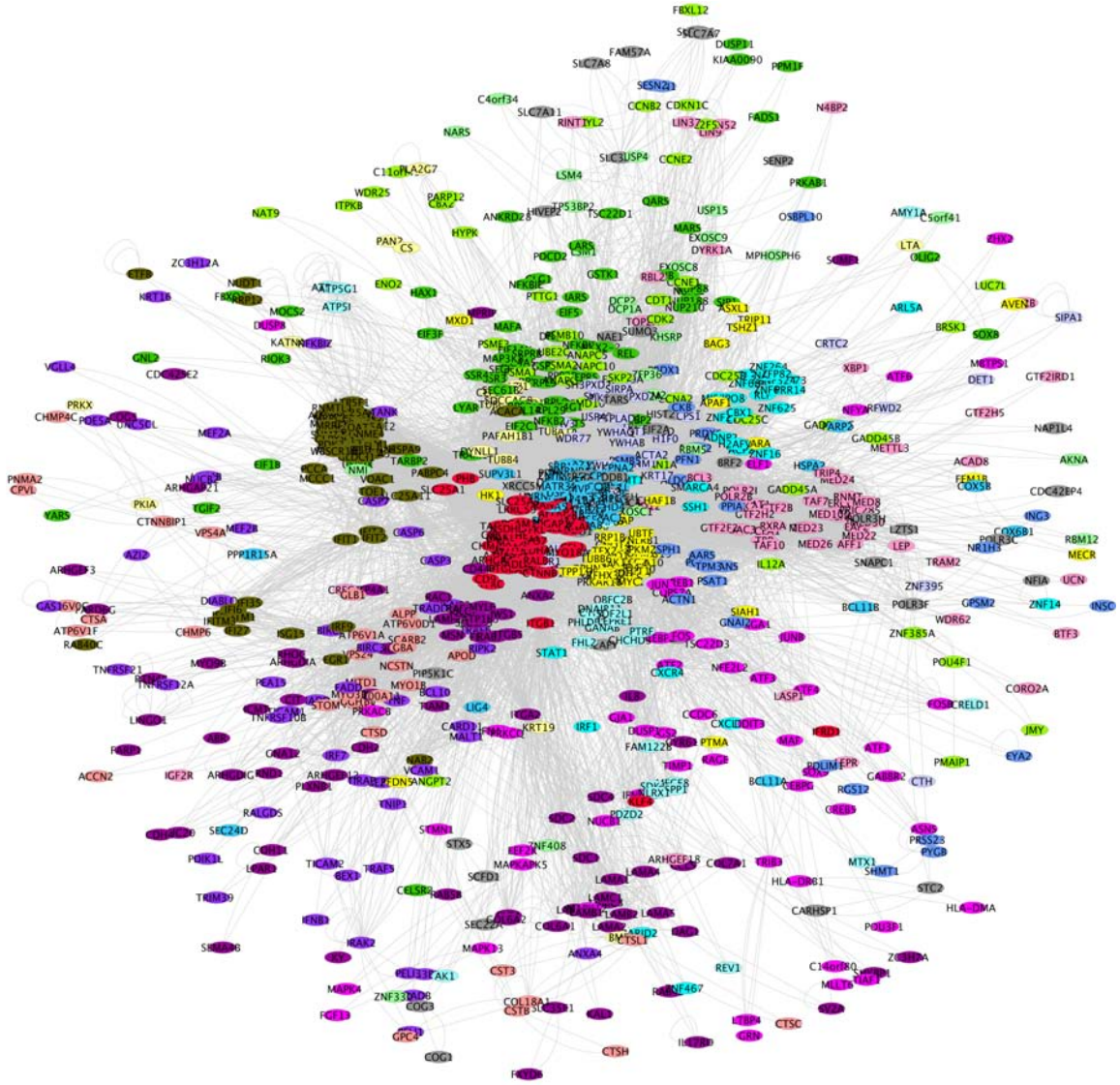
		Genes targeted by one or more up-regulated DE miRNAs		
		Yes	**No**	**Total**
Dengue Down-regulated DE genes	**Yes**	1	124	125
	No	0	100	100
	Total	1	224	225

Table 5. Contingency table for Dengue up-regulated genes with down-regulated miRNAs (A) and Dengue down-regulated genes with up-regulated miRNAs (B); Fisher's exact test for association is significant for up-regulated genes and down-regulated miRNAs only.

Aim 2.

Of the 2,521 genes in the West Nile network, 265 genes grouped into six modules and the remainder into the grey module. Sizes of the modules were between 24 and 79 genes. Of the 4,396 genes in the Dengue network, 955 genes grouped into nineteen modules and the remainder into the grey module. Colored Dengue modules were between twenty and 110 genes.

(A)



(B)

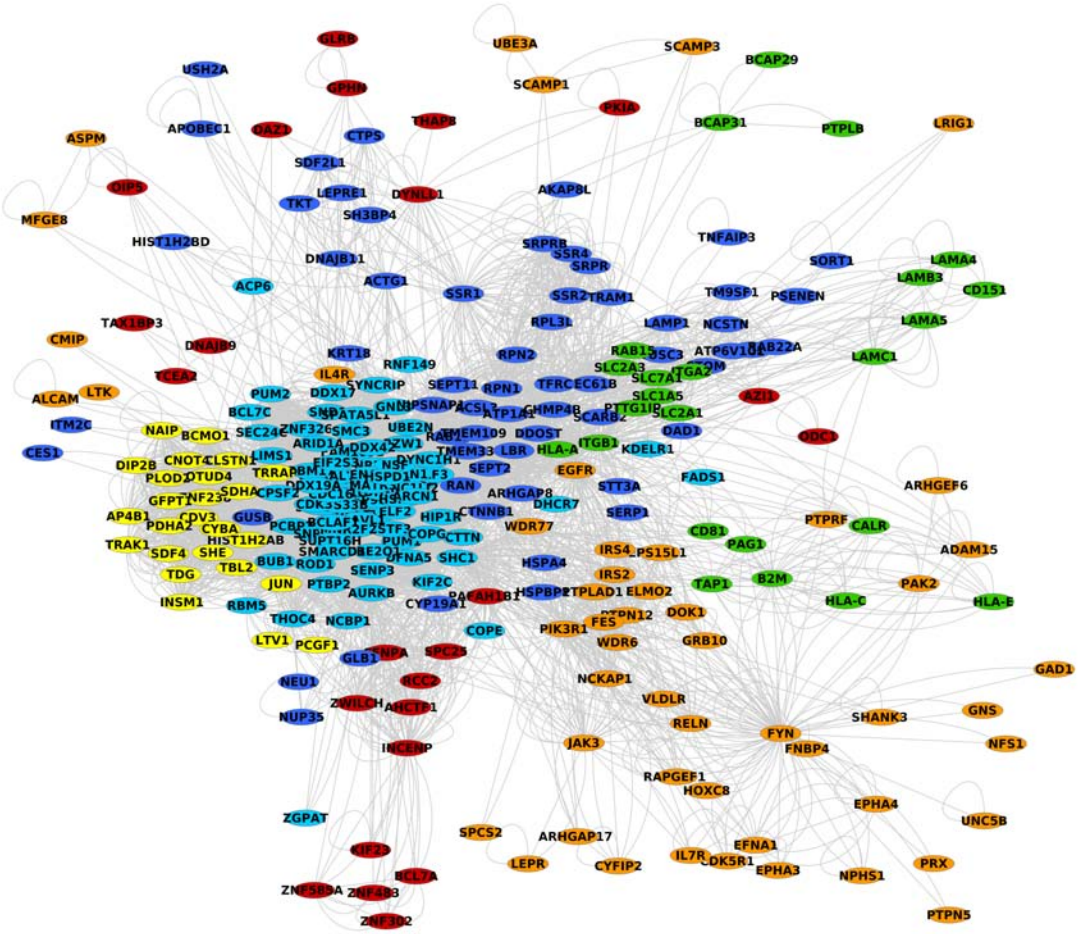


Figure 1. Dengue (A) and West Nile (B) gene networks using protein-protein interactions of differentially expressed genes from gene expression arrays. Module membership is represented by color.

The modules produced by the Dynamic Tree Cut method were compared by both the common genes within modules and by the miRNA regulators of target genes within modules across Dengue and West Nile modules. Analyzing all pairwise comparisons across the viruses produced a total of 114 module comparisons (6 West Nile modules with 19 Dengue modules).

Using the interactome for the reference set, 11 pairwise module comparisons

tested significant and 12 pairwise comparisons were significant using all probes shared between Dengue and West Nile gene expression arrays as the reference set. Restricting the universe to only the differentially expressed genes for Dengue and West Nile, 12 comparisons showed significant overlap. Comparison of the overlap among these three approaches is summarized in Table 6 below.

'GENE UNIVERSE' REFERENCE SET	# significant pairwise module comparisons out of 114
DE genes Differentially expressed genes in Dengue and West Nile networks	12
Shared probes All probes shared between Dengue and West Nile gene expression arrays	12
Interactome Genes in the interactome (Pathway commons interactions)	11
	# significant pairwise module comparisons overlapping
DE genes + Shared probes	12
DE genes + Interactome	11
Shared probes + Interactome	11

Table 6. Comparison of differences in significant module overlap by examining the genes utilized in the individual analyses

Although it was decided to use the Interactome as the gene reference set, additional analyses using the other two reference sets was done for comparison to examine the effect of changing the 'gene universe' on the number of significant module comparisons. Table 7 shows the module overlaps between Dengue and West Nile that had a significant number of genes overlapping using the Interactome as the reference set.

The counts for the contingency tables are shown in bold: the number of common genes (Number common), number of genes in the Dengue module and not found in the West Nile module (denvYes_wnvNo), and number of genes in the West Nile module and not in the Dengue module (wnvYes_denvNo), and the number of genes in neither module (Neither). The last column shows the p-value, adjusted for false discovery rate (FDR).

Denv Module size	Wnv Module Size	Number common	denvYes_wnvNo	wnvYes_denvNo	Neither	p-value
71	**25**	**9**	**62**	**16**	**11184**	1.22E-12
110	**79**	**13**	**97**	**66**	**11095**	2.54E-11
77	**24**	**6**	**71**	**18**	**11176**	3.56E-07
20	**24**	**4**	**16**	**20**	**11231**	2.10E-06
34	**61**	**5**	**29**	**56**	**11181**	2.26E-05
66	**61**	**6**	**60**	**55**	**11150**	2.76E-05
26	**61**	**4**	**22**	**57**	**11188**	0.000186
66	**79**	**5**	**61**	**74**	**11131**	0.001333
32	**52**	**3**	**29**	**49**	**11190**	0.00532
68	**61**	**4**	**64**	**57**	**11146**	0.005696
26	**25**	**2**	**24**	**23**	**11222**	0.016001

Table 7. Modules where comparisons showed significant overlap of genes (FDR adjusted p-value ≤ 0.05)

Using the miRNAs to test for significant overlap, 20 module comparisons were significant (false-discovery rate (FDR) adjusted p-value ≤ 0.05). Table 8 below summarizes the comparisons of miRNA regulators between modules shown to be significant and lists the size of the modules in number of genes. The counts for the contingency tables are shown in bold: the number of common miRNAs (Number common), number of miRNAs in the Dengue module and not found in the West Nile module (denvYes_wnvNo), and number of miRNAs in the West Nile module and not in the Dengue module (wnvYes_denvNo), and the number of miRNAs in neither module (Neither). The last column shows the p-value, adjusted for false discovery rate (FDR). To

calculate the miRNAs not found in either module, the miRNAs mapping to all differentially expressed target genes for both Dengue and West Nile networks was subtracted from the other three count values in the contingency table.

Denv Module size	Wnv Module Size	Number common	denvYes_wnvNo	wnvYes_denvNo	Neither	p-value
77	24	**5**	**3**	**0**	**36**	0.00361
53	24	**4**	**1**	**1**	**38**	0.00842
66	61	**4**	**4**	**0**	**36**	0.01395
77	61	**4**	**4**	**0**	**36**	0.01395
110	79	**4**	**1**	**3**	**36**	0.02424
71	61	**3**	**2**	**1**	**38**	0.02546
110	61	**3**	**2**	**1**	**38**	0.02546
34	79	**3**	**0**	**4**	**37**	0.02546
66	24	**4**	**4**	**1**	**35**	0.02546
66	52	**4**	**4**	**1**	**35**	0.02546
53	61	**3**	**2**	**1**	**38**	0.02546
29	24	**3**	**1**	**2**	**38**	0.02546
72	79	**3**	**0**	**4**	**37**	0.02546
77	52	**4**	**4**	**1**	**35**	0.02546
77	25	**3**	**5**	**0**	**36**	0.03482
71	24	**3**	**2**	**2**	**37**	0.04301
71	52	**3**	**2**	**2**	**37**	0.04301
110	52	**3**	**2**	**2**	**37**	0.04301
53	52	**3**	**2**	**2**	**37**	0.04301
56	61	**2**	**0**	**2**	**40**	0.04301

Table 8. Modules where comparisons showed significant overlap of miRNA regulators (FDR adjusted p-value ≤ 0.05)

Of the thirteen miRNA regulators mapping to differentially expressed genes in Dengue and West Nile networks, eight mapped to genes from the twenty significant module comparisons: hsa-let-7b, hsa-miR-1, hsa-miR-124, hsa-miR-155, hsa-miR-16, hsa-miR-29c, hsa-miR-30, hsa-miR-373. Co-regulation of the miRNA regulators is shown in Table 9 below. Of the eight miRNA regulators, three were common to up-regulated miRNAs in Dengue and one in common with West Nile differentially

expressed miRNAs.

Denv Module size	Wnv Module Size	Hsa-let-7b	Hsa-miR-1	Hsa-miR-124	Hsa-miR-155	Hsa-miR-16	Hsa-miR-29c	Hsa-miR-30	Hsa-miR-373
77	24		x	x		x	x	x	
53	24		x	x		x		x	
66	61		x	x	x	x			
77	61		x	x	x	x			
110	79	x	x	x	x				
71	61		x	x		x			
110	61		x	x	x				
34	79	x	x	x					
66	24		x	x		x		x	
66	52		x	x	x	x			
53	61		x	x		x			
29	24		x			x		x	
72	79	x	x	x					
77	52		x	x	x	x			
77	25					x	x	x	
71	24		x	x		x			
71	52		x	x		x			
110	52		x	x	x				
53	52		x	x		x			
56	61		x			x			

Table 9. Human miRNAs regulating significant module comparisons

Genes from a significant overlap of modules between Dengue and West Nile was visualized in Cytoscape. It shows a Dengue network module consisting of 110 genes overlapped with a West Nile network module of 79 genes (see Figure 2 below). The figure shows genes common between the two modules, and genes specific to either Dengue or West Nile with edges symbolizing protein-protein interactions. From Table 9, these two modules share four miRNAs: hsa-let-7b, hsa-miR-1, hsa-miR-124 and hsa-miR-155.

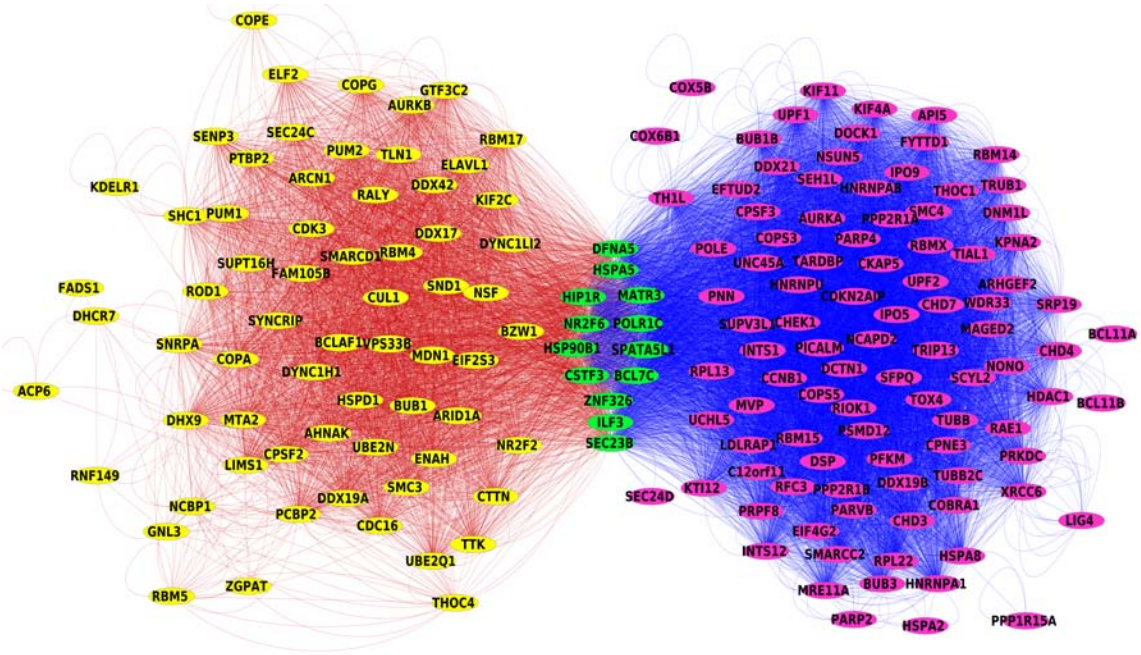


Figure 2. Module for West Nile (yellow genes) overlapped with module for Dengue (pink genes). Green genes represent genes in common to the two modules.

Table A-2 in the Appendix lists significantly enriched GO ontologies in the overlapping genes between Dengue and West Nile gene networks. Twenty-six GO categories were statistically enriched in the overlap. Both membrane-associated and "virus response" pathways were in the top results.

Discussion.

For the first aim, one test out of six was significant: Dengue differentially expressed genes positively associated with one or more differentially expressed miRNAs (Table 2A, p-value = 0.0268). This may be because Dengue had a much larger number of differentially expressed genes that were useable in the analysis (4,396 DE Dengue genes

versus 2,521 DE genes in West Nile). Out of the DE genes, 225 Dengue genes are in TarBase and 120 West Nile genes. Only 7 genes in West Nile were found to be targeted by a differentially expressed miRNA compared with 120 in Dengue. The negative results in West Nile may be support for the hypotheses that West Nile infected cells do not use miRNAs to aid in the response to infection.

Using directionality in gene and miRNA expression, Dengue tested significant for a positive association for up-regulated genes with down-regulated miRNAs (Table 5A, p-value = 6.906e-15). Unfortunately we are limited by the number of up-regulated miRNAs in Tarbase that are usable in the analysis for association with down-regulated genes, given our requirement of experimental validation. In West Nile, the number of up- and down-regulated genes that are in Tarbase is zero and three genes respectively. The counts are too low to use in the statistical test for association. No tests could be done using up-regulated miRNAs because none are in Tarbase for West Nile and only three for Dengue. West Nile also had only two down-regulated miRNAs in Tarbase, therefore we were also limited by not being able to test West Nile for association between up-regulated genes with down-regulated miRNAs. This limitation on target genes and miRNAs experimentally validated does not allow the interpretation of the directionality of expression in infected cells as well as anticipated. The use of predictive target genes may be an alternative to future study. This would depend on the sequence of the miRNA seed sequence only and therefore be solely predictive. Another limitation with that approach is that a large number of genes would be potential targets of a single miRNA and thus may be a confounder in the relationship of expression between target genes and the associated miRNAs.

In the second aim, of the thirteen miRNAs that mapped to all differentially expressed genes in Dengue and West Nile, eight miRNAs were found to regulate the significant module overlaps from aim #2. Hsa-miR-1, hsa-miR-124 and hsa-miR-16 were found to target the most module comparisons with 18, 17 and 14 of them targeted by these three miRNAs. These three miRNAs have not been shown to be associated with viral infection in the literature until now.

For the second aim, several options were available to select a reference set when testing whether a Dengue module significantly overlapped with a module of West Nile. All three options: Interactome genes (those genes in Pathway commons), differentially expressed genes in Dengue and West Nile, and all genes shared on the two viruses' expression arrays were viable in testing the hypothesis. The first option, genes in the Interactome, was chosen because it allows for the best biological interpretation of whether two modules' genes and miRNA regulators are significantly associated compared to all other genes that interact with one another in Pathway Commons and also was chosen because does not limit or bias genes used as a reference set.

When selecting modules to be compared, the methodology used compares all colored modules in Dengue with all colored modules in West Nile. The grey modules (see Appendix: Figure A-1) produced in both viruses were not used in the analysis for comparison because they contained genes that did not interact with any other differentially expressed genes in the set based on protein-protein interactions. This may indicate an additional limitation with regard to interaction annotation.

Table 6 summarizes the number of module comparisons that tested significant out of the total 114 possible comparisons. Although using the Interactome as the 'gene

universe' did not produce the most number of significant comparisons between modules (11 significant comparisons) compared to the other two reference sets (12 significant comparisons in both from Table 6), it contains all significant comparisons that the differentially expressed genes reference set contained. The high overlap between these two options provides even more confirmation that this was a good choice in how to limit the genes to be compared in the analysis. In other words, limiting the reference set to just differentially expressed genes in Dengue and West Nile only eliminated one overlapping module.

Gene Ontology (GO) ontologies were obtained for pathway annotation on the overlap of the Dengue and West Nile networks. Twenty-six GO categories were statistically enriched in the overlap and particularly "virus response" (Table A-1) was of interest and validates the view that the networks represent the response to viral infection in both Flaviviruses.

When examining the module overlap from Figure 2, 45 GO categories were found to be statistically enriched. Several binding GO categories were enriched, and particularly the GO category "virion binding" may be relevant to the infection process in both Dengue and West Nile shown within these two overlapping modules. Both modules contain genes that are statistically enriched for this pathway and show that the cell may be responding to the virus particles being formed by binding to the virion. This evidence to suggest infection response in both Flaviviruses justifies follow up experiments and validation of key genes and miRNA regulators.

CHAPTER 4: Conclusion and Future Work

Summary and Conclusion.

The proposed work to associate miRNAs with changes in viral host response networks was accomplished and shown in the Results section of this thesis. As a result of this work we now know that, based on the data, Dengue gene expression is significantly positively associated with the expression of miRNAs in infected cells versus mock-infected cells. Dengue down-regulated genes are significantly associated with up-regulated miRNAs. This is evidence that the Dengue gene expression and miRNA expression are complementary to one another and suggests that miRNAs and genes are being co-regulated during infection. There is no evidence to suggest that the down-regulated genes are associated with up-regulated miRNAs, but this is because of limiting findings to only confirmed target genes and their miRNAs. Making use of another data source for miRNAs and their target genes other than TarBase such as prediction based databases may be an option for more exploratory avenues. To answer the research question from the first aim, combining two different data types does not predict miRNA-mRNA interactions better than the two independently given the data and the approach. This was an unexpected conclusion but not surprising taking into account the data was restricted to confirmed miRNA-mRNA relationships.

From the second aim, several cellular pathways have been identified based on protein-protein interactions of differentially expressed genes in Dengue and West Nile independently. Of all comparisons, twenty pathway modules' miRNA regulators were found to be significantly associated in the Dengue and West Nile modules. The answer to the second aim's research question: Do the miRNA regulators shared between the West

Nile and Dengue virus regulatory network overlap significantly?, is yes, 8 significant miRNA regulators provide a precise list of potential miRNAs that are responsible for targeting common Flavivirus pathways changing during infection. We expected that miRNAs would significantly overlap between networks but not as defined. These miRNAs provide potential in our own validation and pursuit of a direct approach rather than discovery for miRNA regulators of the viral response in both Flaviviruses.

Implications of Research:

For our collaborators (Dr. Jay Nelson's lab), several forms of experimental data had existed for both West Nile and Dengue. What was not known is what all three of the data types mean together in terms of complementarity and biological meaning. Synchronizing information from miRNA activity, gene expression and RISC-associated genes allows for an exciting and new interpretation of viral infection. Connecting miRNA expression and gene regulation between viruses of the same genus better predicts what regulatory modules are being changed in response to infection. These results provide initial predictions for experimental validation and follow-up. Incorporating the experimental data for viral infection can serve as a model for other related virus families and guide the development of non-viral drugs and gene therapy for treatment of infected individuals in hopes of reducing the number of deaths per year.

References:

(1) Shi P. Molecular Virology and Control of Flaviviruses. : Caister Academic Press; 2012.

(2) CDC Cfdc. Dengue: Clinical Guidance. 2010; Available at: http://www.cdc.gov/dengue/clinicalLab/clinical.html. Accessed June 24, 2012.

(3) Pastorino B, Nougairède A, Wurtz N, Gould E, de Lamballerie X. Role of host cell factors in flavivirus infection: Implications for pathogenesis and development of antiviral drugs. Antiviral Res 2010 9;87(3):281-294.

(4) Chambers TJ, Hahn CS, Galler R, Rice CM. Flavivirus Genome Organization, Expression, and Replication. Annu Rev Microbiol 1990 10/01; 2012/06;44(1):649-688.

(5) Eulalio A, Huntzinger E, Izaurralde E. Getting to the Root of miRNA-Mediated Gene Silencing. Cell 2008 01/11;132(1):9-14.

(6) Gregory RI, Chendrimada TP, Cooch N, Shiekhattar R. Human RISC Couples MicroRNA Biogenesis and Posttranscriptional Gene Silencing. Cell 2005 11/18;123(4):631-640.

(7) Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. Cell 2004 Jan 23;116(2):281-297.

(8) Zambon RA, Vakharia VN, Wu LP. RNAi is an antiviral immune response against a dsRNA virus in Drosophila melanogaster. Cell Microbiol 2006;8(5):880-889.

(9) Zamore PD, Tuschl T, Sharp PA, Bartel DP. RNAi: Double-Stranded RNA Directs the ATP-Dependent Cleavage of mRNA at 21 to 23 Nucleotide Intervals. Cell 2000 3/31;101(1):25-33.

(10) Heiss BL, Maximova OA, Pletnev AG. Insertion of microRNA targets into the flavivirus genome alters its highly neurovirulent phenotype. J Virol 2011 Feb;85(4):1464-1472.

(11) Du P, Kibbe WA, Lin SM. lumi: a pipeline for processing Illumina microarray. Bioinformatics 2008 July 01;24(13):1547-1548.

(12) Smyth G. limma: Linear Models for Microarray Data. 2005:397-420.

(13) Gunaratne P, Creighton C, Watson M, Tennakoon J.
Large-scale integration of MicroRNA and gene expression data for identification of enriched microRNA-mRNA associations in biological systems. Methods Mol Biol 2010;667(6):297-315.

(14) Creighton CJ, Nagaraja AK, Hanash SM, Matzuk MM, Gunaratne PH. A bioinformatics tool for linking gene expression profiling results with public databases of microRNA target predictions. RNA 2008 Nov;14(11):2290-2296.

(15) Nam S, Li M, Choi K, Balch C, Kim S, Nephew KP. MicroRNA and mRNA integrated analysis (MMIA): a web tool for examining biological functions of microRNA expression. Nucleic Acids Res 2009 Jul;37(Web Server issue):W356-62.

(16) Wu X, Watson M. CORNA: testing gene lists for regulation by microRNAs. Oxford Journals 2009;25(6):832.

(17) Papadopoulos GL, Reczko M, Simossis VA, Sethupathy P, Hatzigeorgiou AG. The database of experimentally supported targets: a functional update of TarBase Nucleic Acids Res. 2009 Jan;37(Database issue):D155-8. Epub 2008 Oct 27. 2009 Oct 27 2008;37(Database issue):D155.

(18) Farh KK, Grimson A, Jan C, Lewis BP, Johnston WK, Lim LP, et al. The Widespread Impact of Mammalian MicroRNAs on mRNA Repression and Evolution. Science 2005 12/16;310(5755):1817-1821.

(19) Cerami EG, Gross BE, Demir E, Rodchenkov I, Babur Ã, Anwar N, et al. Pathway Commons, a web resource for biological pathway data. Nucleic Acids Res 2010 11/10.

(20) Dong J, Horvath S. Understanding network concepts in modules. BMC Systems Biology 2007;1(1):24.

(21) Langfelder P, Zhang B, Horvath S. Bioinformatics 2008;24(5):719.

(22) Carlson M, Zhang B, Fang Z, Mischel P, Horvath S, Nelson S. Gene connectivity, function, and sequence conservation: predictions from modular yeast co-expression networks. BMC Genomics 2006;7(1):40.

(23) Ghazalpour A, Doss S, Zhang B, Wang S, Plaisier C, Castellanos R, et al. Integrating Genetic and Network Analysis to Characterize Genes Related to Mouse Weight. PLoS Genet 2006 08/18;2(8):e130.

(24) Maere S, Heymans K, Kuiper M. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics 2005 Aug 15;21(16):3448-3449.

Appendix:

	total	expectation	observation	hypergeometric	fisher	description
GO:0004984	378	71	1	1	2.45E-29	olfactory receptor activity
GO:0004930	760	143	37	1	4.03E-27	G-protein coupled receptor activity
GO:0050896	543	102	20	1	2.56E-23	response to stimulus
GO:0007186	913	172	76	1	1.31E-16	G-protein coupled receptor signaling pathway
GO:0004871	1037	196	101	1	1.14E-13	signal transducer activity
GO:0030968	82	15	48	3.30E-12	1.32E-12	endoplasmic reticulum unfolded protein response
GO:0006987	64	12	39	1.54E-10	7.69E-11	activation of signaling protein activity involved in unfolded protein response
GO:0005604	86	16	43	7.08E-08	3.81E-08	basement membrane
GO:0005783	1002	189	266	4.96E-07	4.57E-07	endoplasmic reticulum
GO:0016021	4599	868	729	1	4.78E-07	integral to membrane
GO:0005764	190	36	69	8.45E-06	5.60E-06	lysosome
GO:0005794	975	184	251	2.26E-05	2.14E-05	Golgi apparatus
GO:0042470	89	17	39	3.42E-05	2.46E-05	melanosome
GO:0005789	640	121	175	3.54E-05	3.72E-05	endoplasmic reticulum membrane

GO:0009615	151	28	52	0.00141	0.001259	response to virus
GO:0008285	361	68	101	0.003968	0.004877	negative regulation of cell proliferation
GO:0000139	427	81	115	0.006358	0.00698	Golgi membrane
GO:0030433	30	6	15	0.018418	0.016952	ER-associated protein catabolic process
GO:0060337	65	12	25	0.024109	0.024918	type I interferon-mediated signaling pathway
GO:0002237	9	2	7	0.027497	0.025279	response to molecule of bacterial origin
GO:0004185	5	1	5	0.028809	0.026703	serine-type carboxypeptidase activity
GO:0005624	564	106	141	0.021841	0.026703	membrane fraction
GO:0006488	32	6	15	0.031621	0.030153	dolichol-linked oligosaccharide biosynthetic process
GO:0043687	169	32	51	0.029333	0.034212	post-translational protein modification
GO:0005788	104	20	35	0.028809	0.034212	endoplasmic reticulum lumen
GO:0018279	87	16	30	0.039108	0.043522	protein N-linked glycosylation via asparagine

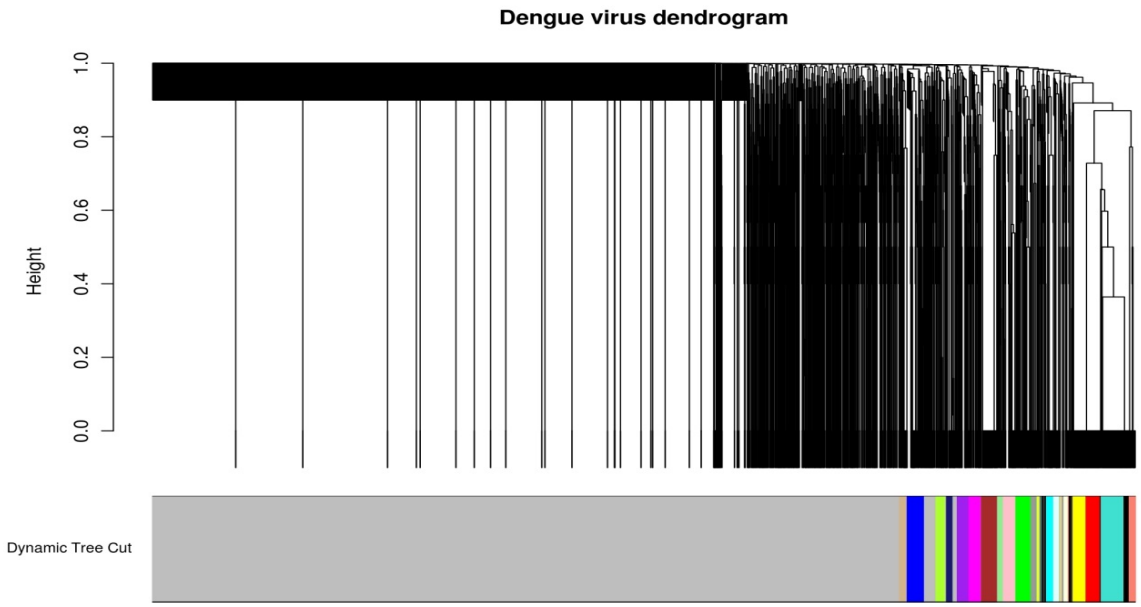
Table A-1. Enriched GO ontologies common between Dengue and West Nile gene networks

GO-ID	p-value	corr p-value	Description
48471	6.78E-05	3.14E-02	perinuclear region of cytoplasm
30433	1.71E-04	3.18E-02	ER-associated protein catabolic process
70013	5.96E-04	3.18E-02	intracellular organelle lumen
5793	6.40E-04	3.18E-02	ER-Golgi intermediate compartment
43233	6.68E-04	3.18E-02	organelle lumen
31247	7.31E-04	3.18E-02	actin rod assembly
60904	7.31E-04	3.18E-02	regulation of protein folding in endoplasmic reticu-lum
43153	7.31E-04	3.18E-02	entrainment of circadian clock by photoperiod
31974	7.49E-04	3.18E-02	membrane-enclosed lumen
16023	9.53E-04	3.18E-02	cytoplasmic membrane-bounded vesicle
16363	9.69E-04	3.18E-02	nuclear matrix
31988	1.06E-03	3.18E-02	membrane-bounded vesicle
31410	1.18E-03	3.18E-02	cytoplasmic vesicle
34399	1.19E-03	3.18E-02	nuclear periphery
31982	1.38E-03	3.18E-02	vesicle
5788	1.43E-03	3.18E-02	endoplasmic reticulum lumen
21577	1.46E-03	3.18E-02	hindbrain structural organization
21589	1.46E-03	3.18E-02	cerebellum structural organization
6987	1.46E-03	3.18E-02	activation of signaling protein activity involved in unfolded protein response
50961	1.46E-03	3.18E-02	detection of temperature stimulus involved in sensory perception
50965	1.46E-03	3.18E-02	detection of temperature stimulus involved in sensory perception of pain
3723	1.51E-03	3.18E-02	RNA binding
48770	1.99E-03	3.38E-02	pigment granule
42470	1.99E-03	3.38E-02	melanosome
51208	2.19E-03	3.38E-02	sequestering of calcium ion
16048	2.19E-03	3.38E-02	detection of temperature stimulus
71318	2.19E-03	3.38E-02	cellular response to ATP
32075	2.19E-03	3.38E-02	positive regulation of nuclease activity
5736	2.19E-03	3.38E-02	DNA-directed RNA polymerase I complex
9648	2.19E-03	3.38E-02	photoperiodism
30662	2.48E-03	3.63E-02	coated vesicle membrane
12505	2.51E-03	3.63E-02	endomembrane system
32069	2.92E-03	3.94E-02	regulation of nuclease activity
42149	2.92E-03	3.94E-02	cellular response to glucose starvation

51082	2.98E-03	3.94E-02	unfolded protein binding
6360	3.65E-03	4.33E-02	transcription from RNA polymerase I promoter
50951	3.65E-03	4.33E-02	sensory perception of temperature stimulus
51787	3.65E-03	4.33E-02	misfolded protein binding
43008	3.65E-03	4.33E-02	ATP-dependent protein binding
3676	4.13E-03	4.60E-02	nucleic acid binding
46790	4.38E-03	4.60E-02	virion binding
4887	4.38E-03	4.60E-02	thyroid hormone receptor activity
9649	4.38E-03	4.60E-02	entrainment of circadian clock
43161	4.47E-03	4.60E-02	proteasomal ubiquitin-dependent protein catabolic process
10498	4.47E-03	4.60E-02	proteasomal protein catabolic process

Table A-2. GO ontologies for Figure 2 overlapping turquoise genes

(A)



(B)

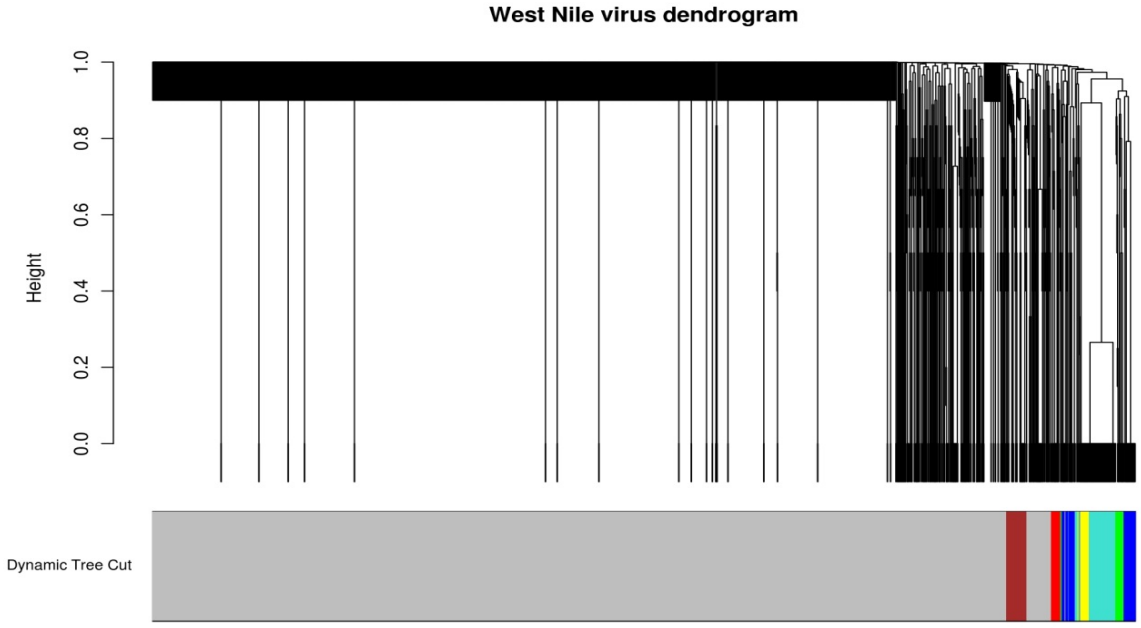


Figure A-1. Dendrogram of Dengue (A) and West Nile (B) virus infections using protein-protein interactions of differentially expressed genes from gene expression arrays and the Dynamic Tree cut method. Modules are shown in colored bands and the grey module represents unassigned genes.

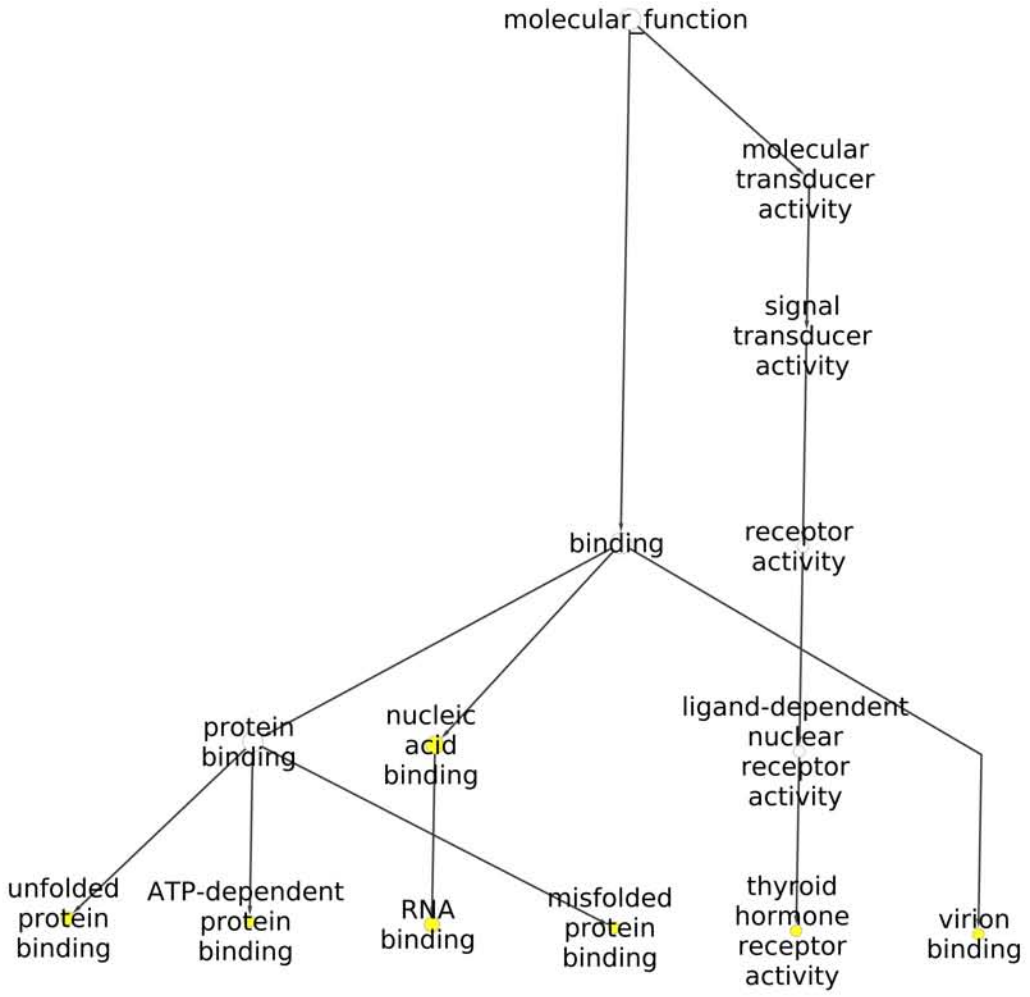


Figure A-2. Graphic showing 14 of the 45 GO biological processes statistically enriched in overlapping genes from Figure 2