**Distribution Agreement**

In presenting this thesis as a partial fulfillment of the requirements for a degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis in whole or in part in all forms of media, now or hereafter now, including display on the World Wide Web. I understand that I may select some access restrictions as part of the online submission of this thesis. I retain all ownership rights to the copyright of the thesis. I also retain the right to use in future works (such as articles or books) all or part of this thesis.

Evan Yang March 22nd, 2024

# Transcriptomic Changes in Fontan Surgery Sheep Models and a Novel miRNA-mRNA Interaction Analysis

by

Evan Yang

Dr. Michael Davis
Co-advisor

Dr. Kevin McAlister

Co-advisor

Department of Quantitative Theory and Methods

Dr. Michael Davis

Co-advisor

Dr. Kevin McAlister

Co-advisor

Dr. Adam Glynn

Committee Member

2024

# Transcriptomic Changes in Fontan Surgery Sheep Models and a Novel miRNA-mRNA Interaction Analysis

By

Evan Yang

Dr. Michael Davis

Co-advisor

Dr. Kevin McAlister

Co-advisor

An abstract of a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Science with Honors

Department of Quantitative Theory and Methods

2024

Abstract

Transcriptomic Changes in Fontan Surgery Sheep Models and a Novel miRNA-mRNA Interaction Analysis

By: Evan Yang

Patients with congenital heart defects require a series of three surgeries known as Fontan surgery. While it is lifesaving, both cardiac and system complications can arise because multiple open-heart surgeries are needed. The underlying mechanisms and causes of these complications are not understood.

Extracellular vesicles are miniscule vesicles secreted by cells, and they play an important role in cell-to-cell communication. Circulating RNA in blood and in extracellular vesicles have been shown to have biological significance, such as mediating biological processes. Transcriptomics, or the study of RNA molecules, can reveal the causative role of RNA molecules and identify potential biomarkers for these complications.

To further understand single ventricular flow, researchers at Nationwide Children's Hospital have been working to establish a sheep model of Fontan palliation, as the sheep's cardiovascular system strongly mimics a human's cardiovascular system. Here, they have successfully established a single ventricular flow in a pair of cloned sheep, Justin and Julian. Justin is the sheep that underwent Fontan surgery, and Julian serves as a control model.

To investigate the impact of Fontan surgery, this thesis will investigate the transcriptomic changes in serum and extracellular vesicles within Justin and Julian.

Serum samples have been taken from preoperative and postoperative time points. From these serum samples, extracellular vesicles were isolated. Both serum and extracellular vesicle samples underwent RNA sequencing.

Three bioinformatic pipelines have been established that analyze mRNA profiles and miRNA profiles. The first mRNA pipeline analyzes clustering and differential gene expression within RNA data. The second miRNA pipeline analyzes clustering and differential gene expression within miRNA data. The third miRNA-mRNA correlation analysis uses a Pearson correlation to correlate miRNA and mRNA expression levels.

The mRNA bioinformatic pipeline uncovered transcriptomic changes in response to Fontan surgery. The miRNA bioinformatic pipeline demonstrated the potential of integrating predicted miRNA data for new clustering patterns. The miRNA-mRNA pipeline proposes a novel method to discover miRNA and mRNA interactions within serum. MiRNAs play an inhibitory role in mRNA expression. The miRNA-mRNA pipeline offers a high-throughput way to investigate the mRNA targets of miRNAs in any biological system, with an applicability towards extracellular vesicles.

# Transcriptomic Changes in Fontan Surgery Sheep Models and a Novel miRNA-mRNA Interaction Analysis

By

Evan Yang

Dr. Michael Davis

Co-advisor

Dr. Kevin McAlister

Co-advisor

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Science with Honors

Department of Quantitative Theory and Methods

2024

Table of Contents

# 1. Abstract

Patients with congenital heart defects require a series of three surgeries known as Fontan palliation. After Fontan surgery, single ventricular circulation is established. While Fontan surgery is lifesaving, many complications, including both cardiac and systemic side effects, can arise within these pediatric patients because multiple open-heart surgeries need to be conducted. While the altered hemodynamics of this newly established blood flow is thought to contribute to these complications, the underlying mechanisms and causes of these complications are not understood.

Extracellular vesicles are miniscule vesicles secreted by cells, and they play an important role in cell-to-cell communication. Circulating RNA in blood and in extracellular vesicles have been shown to have biological significance, such as mediating biological processes. Transcriptomics, or the study of RNA molecules, can reveal the causative role of RNA molecules and identify potential biomarkers for these complications.

To further understand single ventricular flow, researchers at Nationwide Children's Hospital have been working to establish a sheep model of Fontan palliation, as the sheep's cardiovascular system strongly mimics a human's cardiovascular system. Here, they have successfully established a single ventricular flow in a pair of cloned sheep, Justin and Julian. Justin is the sheep that underwent Fontan surgery, and Julian serves as a control model.

To investigate the impact of Fontan surgery, this thesis will investigate the transcriptomic changes in serum and extracellular vesicles within Justin and Julian. Serum samples have been taken from preoperative and postoperative (1 week, 1

month, 2 months) time points. From these serum samples, extracellular vesicles were isolated. Both serum and extracellular vesicle samples underwent RNA sequencing.

Three bioinformatic pipelines have been established that analyze total RNA profiles and miRNA profiles. The first total RNA pipeline analyzes clustering and differential gene expression within RNA data. The second miRNA pipeline analyzes clustering and differential gene expression within miRNA data. The third miRNA-mRNA correlation analysis uses a Pearson correlation to correlate miRNA and mRNA expression levels.

As for results, the mRNA bioinformatic pipeline uncovered transcriptomic changes in response to Fontan surgery. The miRNA bioinformatic pipeline demonstrated the potential of integrating predicted miRNA data for new clustering patterns. The miRNA-mRNA bioinformatic pipeline is a newly developed methodology to discover miRNA and mRNA interactions within serum. MiRNAs play a large regulatory and inhibitory role in mRNA expression. The third miRNA-mRNA pipeline offers a high-throughput way to investigate the mRNA targets of miRNAs in any biological system, with an applicability towards extracellular vesicles.

## 2. Introduction

Hypoplastic Left Heart Syndrome is a congenital heart disease that affects roughly 1 in 5,000 newborns in the United States. Patients with Hypoplastic Left Heart Syndrome possess an underdeveloped left ventricle. The left ventricle is responsible for maintaining systemic circulation and pumping blood throughout the body. Babies born with Hypoplastic Left Heart Syndrome must rely on the right ventricle to sustain

circulation, resulting in poor systemic circulation. Without intervention, Hypoplastic Left Heart Syndrome is invariably fatal. The two available surgical interventions are heart transplants or a three-step palliative surgery known as Fontan surgery.

To address congenital heart defects such as Hypoplastic Left Heart Syndrome, pediatric patients undergo a series of three heart surgeries called the Norwood procedure, Glenn Surgery, and Fontan surgery. These three procedures, culminating in Fontan surgery, remain the ultimate therapeutic option for pediatric patients with congenital heart diseases. The end goal of these surgeries is to establish a single ventricular blood flow within the patient. However, because of the altered hemodynamics, both cardiac and systemic complications can arise after Fontan surgery. Cardiac complications include ventricular dysfunction, arrhythmia, exercise intolerance, cyanosis, and elevated pulmonary vascular resistance (Mazza et al., 2021). Systemic complications include renal complications, plastic bronchitis, protein-losing enteropathy, and Fontan Association Liver Disease (Mazza et al., 2021). These complications lower the prognosis for Fontan patients. Moreover, due to the developing nature of the pediatric patient's cardiovascular system, this surgery can lead to developmental complications. As a result, identifying biomarkers and understanding drivers behind these complications are of high clinical significance. Through the early detection of complications through biomarkers, timely intervention and treatment can improve clinical outcomes for Fontan patients.

*Extracellular Vesicles and Serum*

Serum and extracellular vesicles (EVs) contain various proteins, nucleic acids, and biomolecules that are both informative to and causative of different biological processes. Extracellular vesicles are a heterogenous group of nanosized vesicles secreted by all cell types found in all biofluids. EVs play a key role in cell-to-cell communication and contain molecular cargo such as proteins and nucleic acids.

The molecular content in both serum and EVs provide valuable information about the physiological state of an organism. For example, miRNAs have been a topic of interest for their role in regulating mRNA synthesis and expression (Neudecker et al., 2016). The cargo of EVs is reflective of the physiological state of their cell of origin, and the content of EVs can have therapeutic effects. Within the heart, nucleic acids in CPC-EVs possess both pro-angiogenic and pro-migratory properties (Hoffman et al., 2023). Because of the therapeutic and predictive power of RNA in both serum and EVs, the transcriptome serves as a valuable pool of information for clinical medicine.

*Computational Theory of miRNA-mRNA Correlation Analysis*

Originally, this pipeline was developed with the idea of correlating extracellular vesicle miRNA to mRNAs from specific tissues or cells. MiRNAs in extracellular vesicles are delivered to target cells or tissues. Inside the target, these miRNAs perform various silencing or inhibitory functions, oftentimes downregulating the expression of mRNA. By collecting miRNA expression data in extracellular vesicles and mRNA expression data in target tissues or cells, this correlation analysis offers a high throughput way of identifying the mRNA targets of the miRNA content from EVs. The goal is to help

understand potential miRNA-mRNA interactions in understudied animals, such as sheep or dogs.

Extracellular vesicles are known to carry miRNA content, and these EVs deliver their molecular cargo to specific tissues or cells. However, due to a lack of tissue-specific transcriptomic data, this pipeline was developed using serum transcriptomic data.

Circulating nucleic acid in the blood can have various biological origins. Nucleic acids can be released from passive processes, such as cell death or apoptosis, or active processes, such as secretion of nucleic acids from cells (Suraj et al., 2017). Serum miRNA can be transported to target cells, such as through vesicles or Argonaute proteins (O'Brien et al., 2018). In these target cells, miRNA can interact with an RNA-Induced Silencing Complex to silence genes (Pratt and MacRae,2009). As a result, serum miRNA and mRNA levels have the potential to be informative about the inhibitory mRNA targets of miRNAs.

However, with the advent of high-throughput RNA-sequencing, the abundance of information and data can pose challenges. Due to the time-consuming nature of computational analysis, much of this information and data will not be utilized, as the data may contain other nucleic acids and debris that are not of interest. As a result, the aim of this pipeline is to offer a high-throughput method to utilize abundant miRNA and mRNA transcriptomic data to identify potential miRNA-mRNA interactions.

Firstly, our methodology uses normalized miRNA and mRNA data matrices. For dimensionality reduction, the mRNA matrix is then cross-referenced with the mRNA

differential gene expression analysis. The matrix is subsetted to mRNAs with an adjusted p-value of less than 0.1, indicating the gene was likely differentially expressed.

Next, these mRNAs are correlated with known miRNAs using a Pearson correlation. A Pearson correlation assumes a linear relationship. Because regulatory miRNAs can only influence mRNA production in one cell, I am assuming that each miRNA would have a linear, set effect on the amount of mRNA it is able to silence or degrade.

Afterwards, the p-values of each correlation will be calculated. To adjust for multiple hypothesis testing, an FDR/Benjamini-Hochberg correction will be applied (Benjamini et al., 1995). In theory, this methodology should offer a way to correlate miRNA and mRNA levels and to identify miRNA-mRNA interactions in a high-throughput way.

*Sheep Models of Fontan Surgery*

To understand the causes of postoperative complications and to identify biomarkers for early detection, the Davis Lab is conducting two concurrent studies. One study is investigating the transcriptome of pediatric patients who have undergone Fontan. The other study investigates the transcriptome in sheep models of Fontan surgery. Both studies have advantages and disadvantages to their methodology. The study on human patients translates directly into clinical medicine, and the human genome and transcriptome is extensively studied. However, it is limited in several ways, in which the sheep Fontan model can address. For example, collecting samples at set

time points and detailing postoperative complications is more difficult in human patients than within animal models.

Animal models have various benefits that studying human patients lack, such as being able to take more samples across a wide variety of time scales. Furthermore, because animal models are always near the laboratory, it is easier to detail complications and collect clinical measurements.

In 2019, surgical experimentation demonstrated the feasibility of a sheep model of single ventricular circulation (Van Puyvelde et al., 2019). Additionally, collaborators at Nationwide Children's Hospital have been working on developing a sheep model of single-ventricle circulation. The cardiac anatomy of a sheep is similar to the cardiac anatomy of humans, and the rate of development of sheep allows for sheep to serve as a viable animal model. Furthermore, the technical difficulties of Fontan surgery prevents this model from being established in smaller animal models, such as mice or rodents.

This thesis aims to develop and document methods to analyze and use transcriptomic data from these ovine animal models. Furthermore, it aims to present a high-throughput method to computationally identify potential mRNA targets of miRNAs in any biological system. Through the development of these computational methods, a deeper understanding of the transcriptome within any animal model can be gained. Understanding the genes and RNA involved in Fontan surgery can provide new therapeutic targets to relieve side effects, identify biomarkers for early intervention, and offer real-time monitoring of patients.

# 3. Materials and Methods

## Experimental Procedures

*Study Design*

A pair of cloned sheep were created at Nationwide Children's Hospital by collaborators. One cloned sheep, named Justin, underwent single stage Fontan surgery. The other cloned sheep, named Julian, served as a control.

Serum samples were extracted from the pair of cloned sheep before undergoing Fontan surgery. After Justin underwent Fontan surgery, serum samples were collected at identical timepoints from both sheep. The time points were 1 week before surgery, 1 week after surgery, 1 month after surgery, and 2 months after surgery. These serum samples were frozen and sent to the Davis Lab.

Following processing, triplicates EV samples and singular serum samples were collected from each timepoint.

*Serum Total RNA Extraction*

From the serum samples, RNA was isolated and extracted using a Norgen Biotek Plasma/Serum RNA Purification kit. The RNA samples were then frozen in a -80 degrees Celsius freezer. The frozen samples were sent off to the Non Human Primate Genomics Core for total RNA sequencing.

*EV Isolation and Serum Extraction*

From the serum samples, extracellular vesicles were isolated via column chromatography using IZON qEVORIGINAL Gen 2 column kit. These extracellular

vesicles were quantified using Nanosight, where round vesicles were qualitatively observed and the size of the vesicles was plotted.

After Nanosight quantification, RNA from the extracellular vesicles was isolated using Norgen Biotek Exosomal RNA Isolation kit. The RNA samples were frozen in the -80 degree Celsius refrigerator. The frozen samples were sent off to the Non Human Primate Genomics Core for total RNA sequencing.

## Computational Procedures

mRNA Workflow

*mRNA Data Alignment using STAR*

FASTQ files from RNA sequencing were received from the Non Human Primate Genomics Core. First, fastqc was used to conduct quality control on the FASTQ files. Next, the FASTQ files were trimmed using trim-galore.

To extract ovine mRNA data, the sheep reference genome “Genome assembly Oar_v4.0” along with GTF annotation files were downloaded from the NCBI database (National Center for Biotechnology Information, n.d.). Using this data, the sheep reference genome was indexed using STAR (Dobin et al., 2013). The FASTQ files were then aligned and quantified using STAR to extract ovine mRNA counts.

*mRNA Principal Component Analysis*

Principal component analysis is a statistical method that is used for data visualization and dimensionality reduction. The raw mRNA read count data from the triplicate EV samples were summed. The summed EV mRNA counts were merged with

the serum mRNA counts. To filter lowly expressed genes, genes with less than 50% expression rate in all samples were filtered out. Furthermore, the genes were further filtered using a filterByExpr function within the edgeR package (rdrr.io, n.d.). A principal component analysis was conducted on this data <Everything_12_13PC1_vs_PC2.tiff> using R.

The serum and EV mRNA raw read count datasets were then separated. Removal of genes with a less than 50% expression rate along with filtering through filterByExpr were conducted for both the separated serum and EV datasets (rdrr.io, n.d.). Principal component analysis was conducted on both datasets <EV_12_13PC1_vs_PC2.tiff> and <Serum_12_13PC1_vs_PC2.tiff> using R.

*mRNA Heat Maps*

Heat maps are a data visualization tool used to depict gene expression levels across a wide variety of experimental conditions. The summed EV and serum datasets were combined. Lowly expressed genes were filtered out using a 50% expression rate benchmark and through the filterByExpr function. The remaining genes were normalized using the TMM approach. A hierarchical clustering approach was used to cluster samples and genes with similar expression patterns together. To identify clusters, the Euclidean distance was used. The variance of the genes across samples were calculated. The 50 genes with the most variance were selected and plotted in a heat map using R <Everything_Top50_12_13.pdf>.

*mRNA Differential Gene Expression Analysis using limma/voom with dream*

A differential gene expression analysis was conducted using the limma/voom workflow combined with the differential expression for repeated measures (dream) approach in R to identify differentially expressed mRNAs.

Limma, or linear models for microarray and RNA-seq data, utilizes a linear model to fit RNA sequencing data. The assumption of homoscedasticity is central to linear models. Homoscedasticity is a property of a dataset where the variance of the dependent variable is consistent or comparable across different values or conditions of the independent variable. However, RNA sequencing data is heteroscedastic, where the variance changes depending on the experimental condition.

As a result, the voom package, also known as mean-variance modeling at the observational level, within the limma/voom workflow helps transform the data. The voom package utilizes a log counts per million transformation to stabilize the variance across conditions. This makes RNA sequencing data more homoscedastic, enabling linear models to be fit. Finally, the dream package applies various statistical corrections to increase the power and reduce false positives in repeated-measure experiments.

First, the data was split into serum and EV subsets. The datasets were normalized using the trimmed mean of M-values (TMM) approach. Quality controls, such as box plots and merged RNA density distributions, were conducted on the data.

The serum dataset was filtered using a 50% expression rate cutoff along with filterByExpr (rdrr.io, n.d.). Using the limma/voom workflow, along with dream, a differential gene expression analysis was conducted on the serum dataset.

The EV dataset was filtered using a 50% expression rate cutoff along with filterByExpr. Using the limma/voom workflow, along with dream, a differential gene expression analysis was conducted on the EV dataset.

## miRNA Workflow

### *miRNA Data Alignment and Novel miRNA Prediction using miRDeep2*

FASTQ files from the RNA sequencing were received from the sequencing core. First, fastqc was used to conduct quality control on the FASTQ files. Next, the FASTQ files were trimmed using trim-galore. Then, a reference genome was built using bowtie along with “Genome assembly Oar_v4.0”. A reference set of experimentally validated miRNAs was downloaded from miRBase (Kozomara et al., 2019). The FASTQ files were then aligned using the quantifier.pl function to extract miRNA counts from the samples.

To extract putative miRNAs, the miRDeep2.pl function was used to computationally predict novel miRNAs (Friedländer et al., 2012). Novel miRNAs with a significant randfold p-value and a miRDeep2 score greater than 1 were selected. A randfold p-value assesses the validity of an miRNA secondary structure through comparison with randomly generated secondary structures. A miRDeep2 score assesses the confidence that a predicted miRNA is a genuine miRNA, with a higher score meaning higher confidence.

*Principal Component Analysis with Mature and Predicted miRNA*

The serum and EV mature miRNA data were combined. This dataset was filtered using filterbyExpr and normalized using the TMM approach (rdrr.io, n.d.). Box plots and merged RNA density distributions were conducted for quality control. A principal component analysis was conducted on the data <Mature_SerumAndEV_12_15>.

The serum mature miRNA data filtered using filterbyExpr and normalized using the TMM approach (rdrr.io, n.d.). Box plots and merged RNA density distributions were conducted for quality control. A principal component analysis was conducted on the data <Mature_Serum_12_15>.

The EV mature miRNA data was filtered using filterbyExpr and normalized using the TMM approach (rdrr.io, n.d.). Box plots and merged RNA density distributions were conducted for quality control. A principal component analysis was conducted on the data <Mature_EV_12_15>.

Then, both serum and EV mature miRNA raw count datasets were combined with serum and EV novel raw count miRNA datasets. The two EV and serum datasets were filtered using filterbyExpr and normalized using the TMM approach. A principal component analysis was conducted on both datasets to visualize how the data would change, <MatureAndNovel_Serum_12_15PC1_vs_PC2.tiff> and <MatureAndNovel_EV_12_15PC1_vs_PC2.tiff>.

*miRNA Heat Maps*

The mature miRNA data was filtered using filterbyExpr and normalized using the TMM approach (rdrr.io, n.d.). The variance for each gene across samples was

calculated.  A hierarchical clustering approach was used to cluster samples and genes with similar expression patterns together. To identify clusters, the Euclidean distance was used.  The 44 remaining miRNAs were plotted using a heat map <Mature_SerumAndEV_12_17.pdf>.

*miRNA Differential Gene Expression Analysis using limma/voom along with dream*

A differential gene expression analysis was conducted using the limma/voom workflow combined with dream in R to identify differentially expressed mRNAs.

First, the data was split into serum and EV subsets because the expected read counts for serum and EV data is different. Therefore, normalization should be conducted separately. Quality controls, such as box plots and merged RNA density distributions, were conducted on the data.

The serum dataset was filtered using filterByExpr (rdrr.io, n.d.). Using a limma/voom workflow, along with dream, a differential gene expression analysis was conducted on the serum dataset.

The EV dataset was filtered using filterByExpr (rdrr.io, n.d.). Using a limma/voom workflow, along with dream, a differential gene expression analysis was conducted on the EV dataset.

miRNA-mRNA Workflow

*miRNA-mRNA Correlation Analysis*

To get a better understanding of miRNA and mRNA interactions occurring within the data, miRNA and mRNA serum data were integrated together. Because of the higher read count of serum data, this interaction analysis will focus on the serum data. First, the miRNA and mRNA data were both filtered using filterbyExpr separately (rdrr.io, n.d.). A requirement of being expressed in half the samples was assigned to the mRNA data to further filter the mRNA data. Then, to account for different library sizes, the miRNA and mRNA datasets were normalized using a log counts-per-million transformation.

Afterwards, to isolate miRNA-mRNA interactions after the operation, the datasets were filtered for only Justin's post-operation data. A correlation matrix was generated between the miRNA and mRNA data using a Pearson correlation. A Pearson correlation measures the linear correlation between two sets of data.

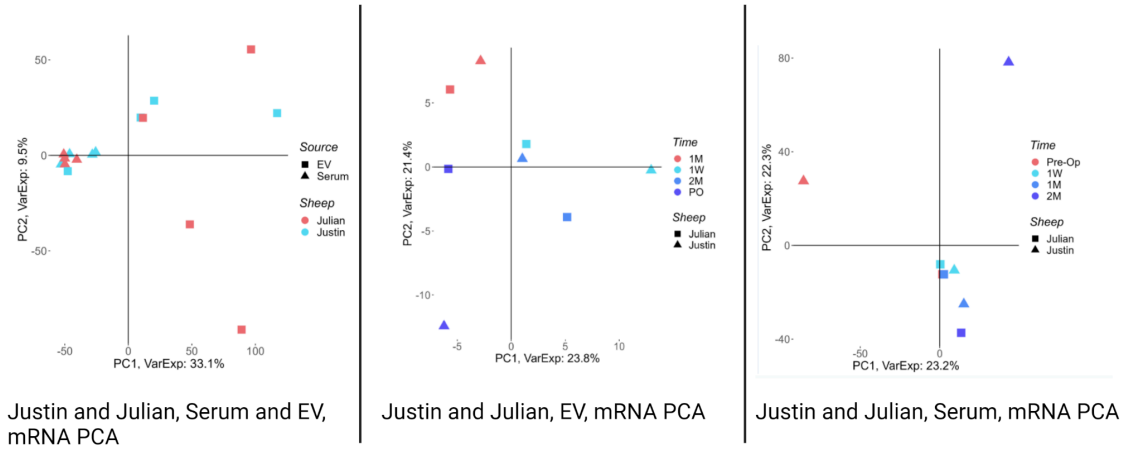
The correlation matrix was cross-referenced with the differential gene expression analysis within Justin's serum data. The correlation matrix was subsetted to mRNAs with an adjusted p-value of less than 0.1 according to the differential gene expression analysis. The correlation matrix was subsetted to mature miRNAs and visualized.

To validate the findings, the p-values for the miRNA-mRNA correlations were calculated. Afterwards, the p-values were corrected using a false discovery rate (FDR) correction.  An FDR/B-H correction is a statistical method that helps adjust and correct p-values in high-throughput experiments (Benjamini et al., 1995). The resulting matrix of p-values were visualized.

# 4. Results and Discussion

mRNA Workflow Results

*Unsupervised Analysis: Principal Component Analysis and Heat Maps*




Justin and Julian, Serum and EV, mRNA PCA

Justin and Julian, EV, mRNA PCA

Justin and Julian, Serum, mRNA PCA

**Figure 1.** Principal Component Analysis of mRNA Data for Justin and Julian

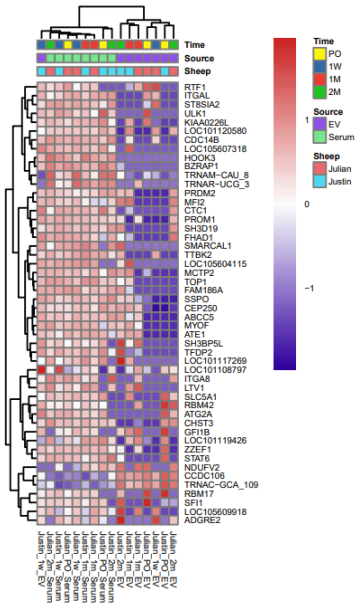



**Figure 2.** Heat Map Clustering Analysis for serum and extracellular vesicle mRNA data from Justin and Julian <Everything_Top50_12_13.pdf>.

Heat maps and principal component analysis were conducted for exploratory analysis. Distinct clustering and separation of serum samples was observed in both PCA and heat maps (Figure 1, Figure 2). Within the extracellular vesicles, clustering was observed based on time points (Figure 1). However, the serum clustered closely based on sheep of origin (Figure 1). The serum data clustered closely for Julian, who did not undergo surgery. On the other hand, serum data for Justin, who underwent surgery, did not cluster closely to Julian's serum data. This demonstrates a change in the mRNA transcriptome resulting from Fontan surgery.

*Supervised Analysis: Differential Gene Expression Analysis*

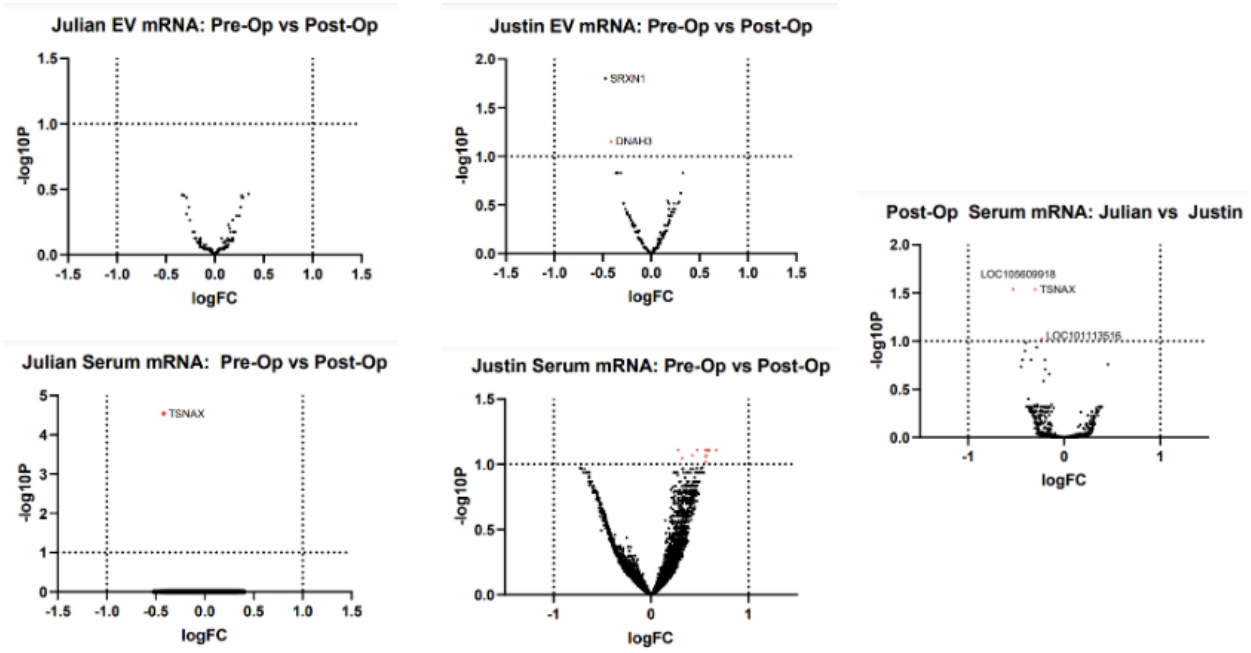



**Figure 3.** Differential Gene Expression Analysis for serum and EV mRNA data from Justin and Julian.

From this differential gene expression analysis, by comparing the serum mRNA results, it is clear to see that Fontan surgery has an impact on the serum transcriptome. Within Julian, there were minimal changes within the serum mRNA. On the other hand, Justin saw vast changes within the serum mRNA. However, the results were not significant, using the standard p-value of 0.05. This is likely because the time points within the samples were extracted too close from the time of surgery. Within other sheep models in the lab, many significant transcriptomic changes have occurred over longer time scales, such as over 6 months to a year. Furthermore, a larger sample size might yield more conclusive results.

## miRNA Workflow Results

*Unsupervised Analysis: Principal Component Analysis and Heat Maps*

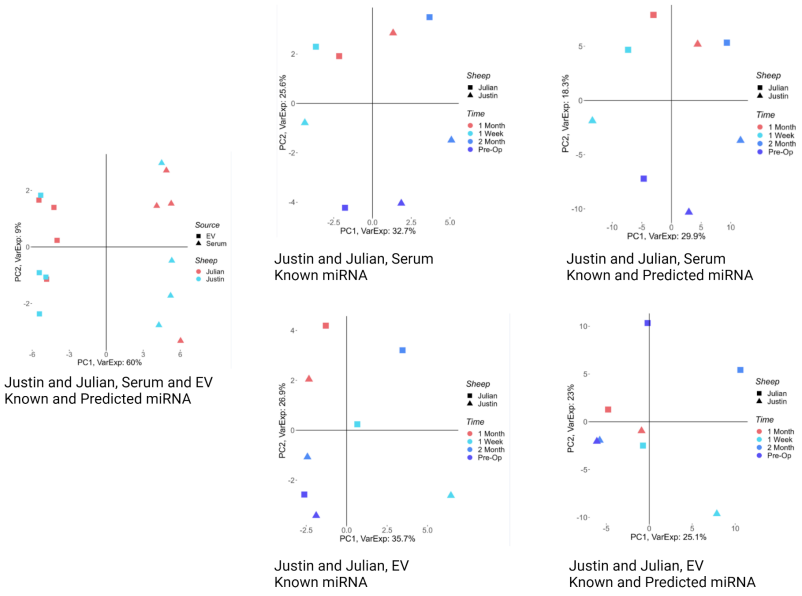



**Figure 4.** Principal Component Analysis for serum and EV miRNA data from Justin and Julian depicting various combinations of validated and predicted miRNA data.

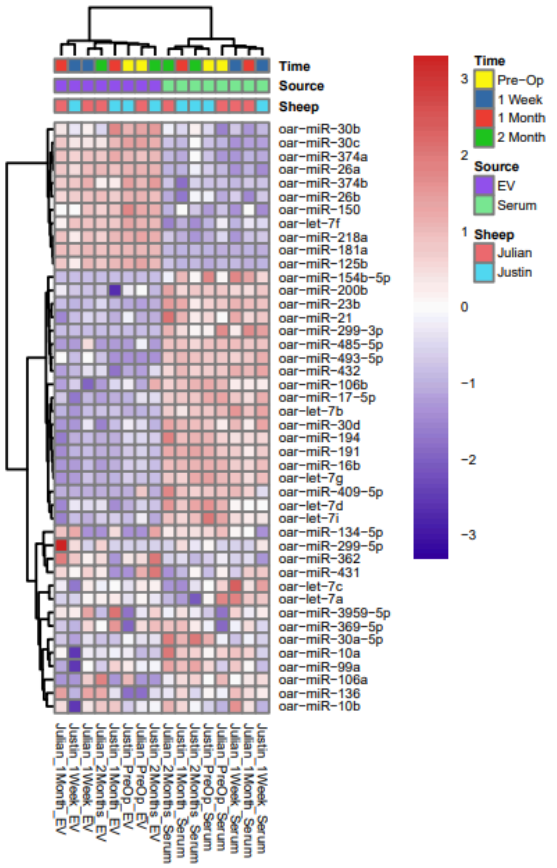



**Figure 5.** Heat Map Clustering Analysis for serum and extracellular vesicle known miRNA data from Justin and Julian <Mature_SerumAndEV_12_17.pdf>.

Heat maps and principal component analysis were conducted for exploratory analysis. Distinct clustering and separation of serum and extracellular vesicle samples was observed in both PCA and heat maps (Figure 4, Figure 5). Integration of predicted miRNA data within the serum samples resulted in different clustering patterns (Figure 4, Figure 5). This supports the idea that the integration of novel, predicted miRNAs can alter results. Because the clustering patterns change drastically when novel, predicted miRNAs are integrated into the analysis, this demonstrates the potential of

undiscovered miRNAs to influence downstream results. This influence of undiscovered miRNAs would be the most prominent in understudied species where the miRNA reference genome is not well documented or deeply understood.

*Supervised Analysis: Differential Gene Expression Analysis*

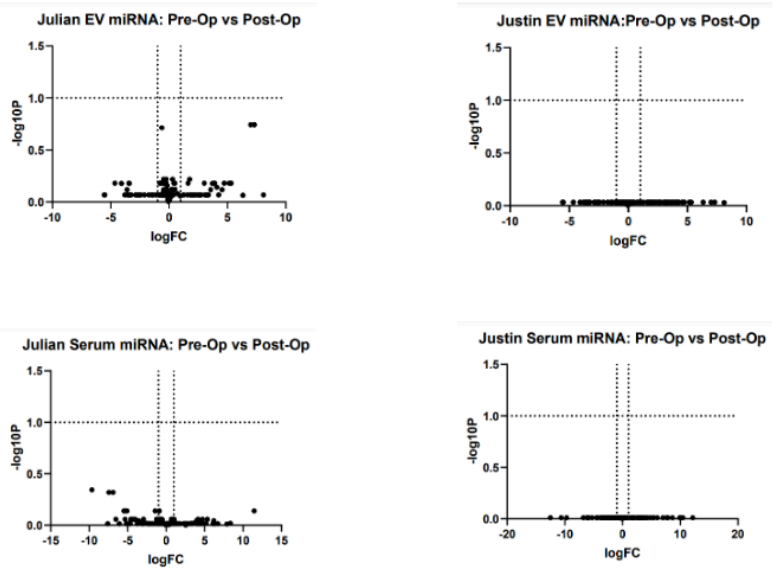



**Figure 6.** Differential Gene Expression Analysis for known, validated miRNA data from Justin and Julian.

This differential gene expression analysis compares the known, validated miRNA levels within both Justin and Julian before and after surgery. The results are not significant.

This lack of significance can be due to various factors. A small sample size might explain why the analysis could not detect significant changes. This pipeline was developed with only 2 pre-operative and 6 post-operative samples, a smaller sample

size for RNA-sequencing experiments. The recommendation for RNA sequencing experimental design is at least 6 samples per condition (Schurch et al., 2016).

Furthermore, the ovine miRNAs are understudied compared to other animal models. In the microRNA database miRBase, there are only 106 validated ovine miRNAs (Kozomara et al., 2019). Comparatively, within miRBase, there are 1917 validated human miRNAs (Kozomara et al., 2019). For further reference, there are currently 1234 validated mouse miRNAs (Kozomara et al., 2019). The relatively low amount of ovine miRNAs could lead to an inability to detect transcriptomic changes among unknown or unvalidated miRNAs.

miRNA-mRNA Correlation Workflow Results

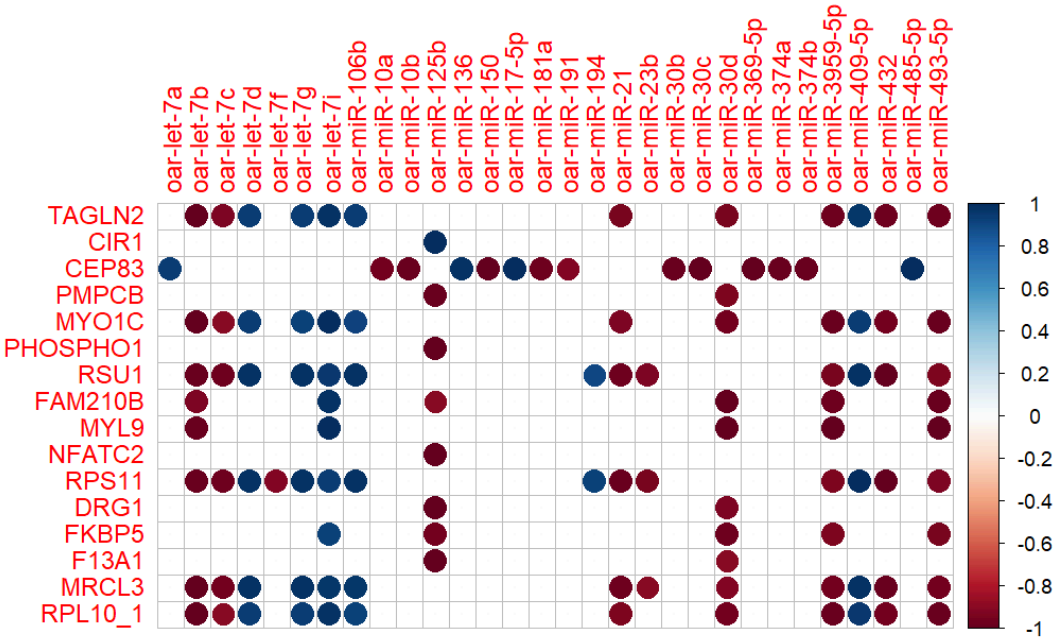



**Figure 7.** A correlation and p-value matrix of known miRNAs along with differentially expressed genes, with correlations with an absolute value below 0.9 removed for clearer visualization.

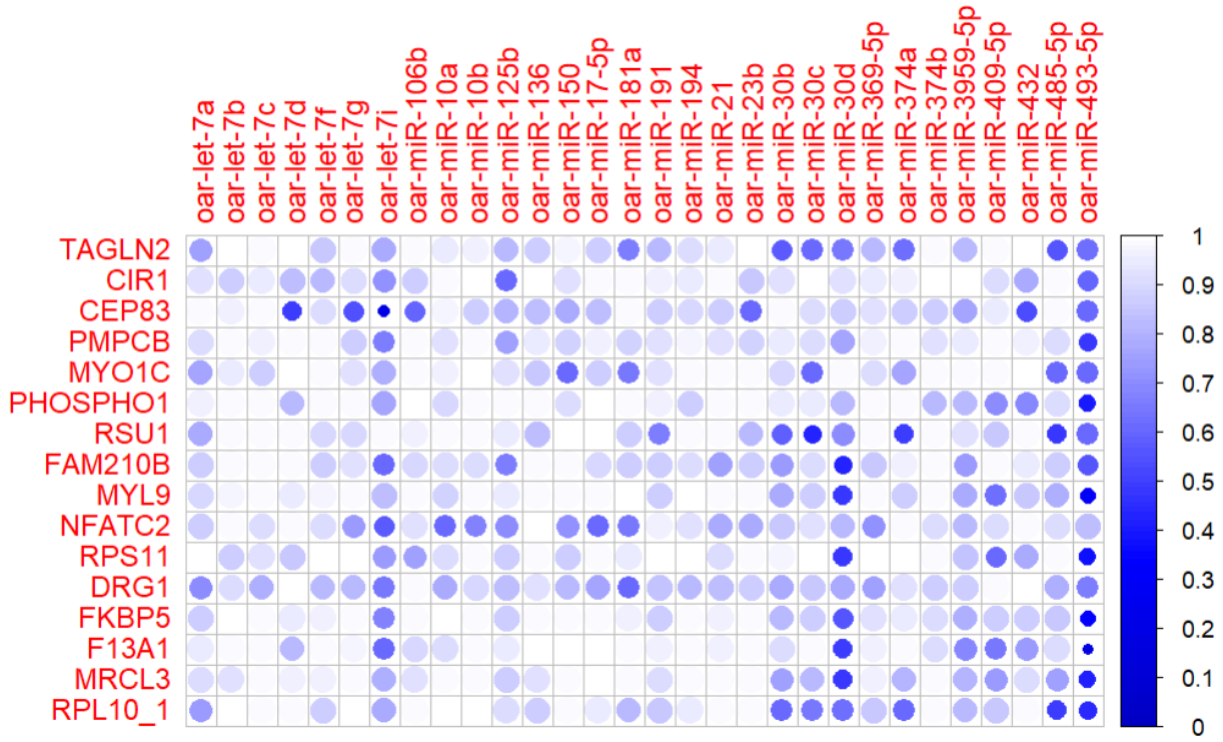



**Figure 8.** A p-value matrix of known miRNAs along with differentially expressed genes.

From this correlation analysis, we can see that oar-miR-30d emerges as a strong signal that is negatively correlated with many of the genes that were differentially expressed in Justin (Figure 7). Furthermore, while the p-values for the correlations of oar-miR-30d are not significant, it does present a lower column of p-values than the other miRNAs (Figure 8). With a larger sample size, future studies within this direction would hopefully increase the significance of this finding.

A literature review reveals that oar-miR-30d improves cardiac function and decreases myocardial fibrosis (Li et al., 2021). Furthermore, many of the mRNAs negatively correlated with miR-30d are associated with cardiac functions or diseases. FAM210B is involved in erythroid, or red blood cell, differentiation (Stelzer et al., 2016). F13A1 encodes a coagulation factor, which assists with clotting (Stelzer et al., 2016). The MYL12A gene, which encodes the MRCL3 proteins, is associated with diseases of the heart, or cardiomyopathies (Stelzer et al., 2016). The role of oar-miR-30d within the cardiovascular system causes it to hold potential as a therapeutic agent.

## 5. Future Directions: miRNA/mRNA Correlation Analysis

The main issue with multiple hypothesis testing is that the likelihood of producing a significant result purely by chance increases as more tests are conducted. Multiple hypothesis testing results in an increased likelihood of committing a Type 1 error, or identifying a false positive.

First, this was addressed by subsetting to purely genes that were differentially expressed. This reduces the amount of hypotheses tested. Furthermore, an FDR correction was applied.

However, another method of dimensionality reduction could be conducted through a weighted gene co-expression network analysis. This reduces the dimensionality of the mRNA data by identifying co-expressed gene networks, or genes that have a similar expression pattern.

Furthermore, new methods of computationally discovering miRNA-mRNA interactions have focused on computationally modeling the feasibility of these interactions (Wong et al., 2020). However, these methods remain computationally time-consuming and heavy. The integration of a correlational approach with a computational modeling approach could provide value by focusing the modeling on predicted interactions.

Transcriptomics allows us to understand what genes and nucleic acids are involved in any biological process. It serves as a valuable tool, with the ability to identify biomarkers, understand causal mechanisms, and provide real-time monitoring. Due to significant decreases in the cost of transcriptomic sequencing, an abundance of biological data has been generated. A high-throughput, unbiased, and untargeted approach towards analyzing this transcriptomic data will discover a wealth of new information.

# Bibliography

Benjamini, Yoav, and Yosef Hochberg. “Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing.” *Journal of the Royal Statistical Society.* Series B (Methodological), vol. 57, no. 1, 1995, pp. 289–300. JSTOR, http://www.jstor.org/stable/2346101. Accessed 8 Apr. 2024.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. “STAR: ultrafast universal RNA-seq aligner.” *Bioinformatics.* 2013 Jan 1;29(1):15-21. doi: 10.1093/bioinformatics/bts635. Epub 2012 Oct 25. PMID: 23104886; PMCID: PMC3530905.

Friedländer, Marc R. , Sebastian D. Mackowiak, Na Li, Wei Chen, Nikolaus Rajewsky, “miRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades”, *Nucleic Acids Research*, Volume 40, Issue 1, 1 January 2012, Pages 37–52, https://doi.org/10.1093/nar/gkr688

Hoffman JR, Park HJ, Bheri S, Platt MO, Hare JM, Kaushal S, Bettencourt JL, Lai D, Slesnick TC, Mahle WT, Davis ME. “Statistical modeling of extracellular vesicle cargo to predict clinical trial outcomes for hypoplastic left heart syndrome.” *iScience.* 2023 Sep 21;26(10):107980. doi: 10.1016/j.isci.2023.107980. PMID: 37868626; PMCID: PMC10589850.

Kozomara A, Birgaoanu M, Griffiths-Jones S. “miRBase: from microRNA sequences to function.” *Nucleic Acids Res.* 2019 Jan 8;47(D1):D155-D162. doi: 10.1093/nar/gky1141. PMID: 30423142; PMCID: PMC6323917.

Li J, Salvador AM, Li G, Valkov N, Ziegler O, Yeri A, Yang Xiao C, Meechoovet B, Alsop E, Rodosthenous RS, Kundu P, Huan T, Levy D, Tigges J, Pico AR, Ghiran I, Silverman MG, Meng X, Kitchen R, Xu J, Van Keuren-Jensen K, Shah R, Xiao J, Das S. “Mir-30d Regulates Cardiac Remodeling by Intracellular and Paracrine Signaling.” *Circ Res.* 2021 Jan 8;128(1):e1-e23. doi: 10.1161/CIRCRESAHA.120.317244. Epub 2020 Oct 22. PMID: 33092465; PMCID: PMC7790887.

Mazza GA, Gribaudo E, Agnoletti G. “The pathophysiology and complications of Fontan circulation.” *Acta Biomed*. 2021 Nov 3;92(5):e2021260. doi: 10.23750/abm.v92i5.10893. PMID: 34738582; PMCID: PMC8689331.

National Center for Biotechnology Information (NCBI)[Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; [1988] – [cited 2024 March 22]. Available from: https://www.ncbi.nlm.nih.gov/

Neudecker V, Brodsky KS, Kreth S, Ginde AA, Eltzschig HK. “Emerging Roles for MicroRNAs in Perioperative Medicine.” *Anesthesiology.* 2016 Feb;124(2):489-506. doi: 10.1097/ALN.0000000000000969. PMID: 26632665; PMCID: PMC4820395.

O'Brien J, Hayder H, Zayed Y, Peng C. “Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation.” *Front Endocrinol (Lausanne)*. 2018 Aug 3;9:402. doi: 10.3389/fendo.2018.00402. PMID: 30123182; PMCID: PMC6085463.

Pratt AJ, MacRae IJ. “The RNA-induced silencing complex: a versatile gene-silencing machine.” *J Biol Chem.* 2009 Jul 3;284(27):17897-901. doi: 10.1074/jbc.R900012200. Epub 2009 Apr 1. PMID: 19342379; PMCID: PMC2709356.

rdrr.io. “filterByExpr: Filter Genes By Expression Level.” rdrr.io, n.d.

Schurch NJ, Schofield P, Gierliński M, Cole C, Sherstnev A, Singh V, Wrobel N, Gharbi K, Simpson GG, Owen-Hughes T, Blaxter M, Barton GJ. “How many biological replicates are needed in an RNA-seq experiment and which differential expression tool should you use? “*RNA.* 2016 Jun;22(6):839-51. doi: 10.1261/rna.053959.115. Epub 2016 Mar 28. Erratum in: RNA. 2016 Oct;22(10 ):1641. PMID: 27022035; PMCID: PMC4878611.

Stelzer G, Rosen N, Plaschkes I, Zimmerman S, Twik M, Fishilevich S, Stein TI, Nudel R, Lieder I, Mazor Y, Kaplan S, Dahary D, Warshawsky D, Guan-Golan Y, Kohn A, Rappaport N, Safran M, Lancet D. “The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses.” *Curr Protoc Bioinformatics.* 2016 Jun 20;54:1.30.1-1.30.33. doi: 10.1002/cpbi.5. PMID: 27322403.

Suraj S, Dhar C, Srivastava S. “Circulating nucleic acids: An analysis of their occurrence in malignancies.” *Biomed Rep.* 2017 Jan;6(1):8-14. doi: 10.3892/br.2016.812. Epub 2016 Nov 16. PMID: 28123700; PMCID: PMC5244781.

Van Puyvelde J, Rega F, Minami T, Claus P, Cools B, Gewillig M, Meyns B. “Creation of the Fontan circulation in sheep: a survival model.” *Interact Cardiovasc Thorac Surg.* 2019 Jul 1;29(1):15-21. doi: 10.1093/icvts/ivz022. PMID: 30789218.

Wong L, You ZH, Guo ZH, Yi HC, Chen ZH, Cao MY. “MIPDH: A Novel Computational Model for Predicting microRNA-mRNA Interactions by DeepWalk on a Heterogeneous Network.” *ACS Omega.* 2020 Jul 9;5(28):17022-17032. doi: 10.1021/acsomega.9b04195. PMID: 32715187; PMCID: PMC7376568.