1-2021

Computational frameworks for microRNA functional analysis of inter-kingdom and indirect targeting.

Mohammed Sayed
University of Louisville

COMPUTATIONAL FRAMEWORKS FOR MicroRNA FUNCTIONAL ANALYSIS OF INTER-KINGDOM AND INDIRECT TARGETING

By

Mohammed Sayed

B.Sc. Minia University, Egypt, 2009

M.Sc. Alexandria University, Egypt, 2015

A Dissertation

Submitted to the Faculty of the

J.B. Speed School of Engineering of the University of Louisville

in Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

in Computer Science and Engineering

Department of Computer Science and Engineering

University of Louisville

Louisville, Kentucky

May 2021

COMPUTATIONAL FRAMEWORKS FOR MicroRNA FUNCTIONAL ANALYSIS OF INTER-KINGDOM AND INDIRECT TARGETING

By

Mohammed Sayed

B.Sc. Minia University, Egypt, 2009

M.Sc. Alexandria University, Egypt, 2015

A Dissertation Approved on

April 19th, 2021

by the following Committee

Dr. Juw Won Park, Dissertation Chair

Dr. Eric Rouchka

Dr. Olfa Nasraoui

Dr. Nihat Altiparmak

Dr. Huang-Ge Zhang

DEDICATION

This dissertation is dedicated to my parents

ACKNOWLEDGEMENTS

I would like first to thank my advisor Dr. Park for his guidance, patience, and continuous support throughout my Ph.D. I'd also like to thank him for giving me the opportunity to work on interesting projects and encouraging me to pursue my ideas.

I would also like to thank Dr. Rouchka for his valuable feedback about my work. I learned a lot about the concepts of bioinformatics during his classes. I'd also to thank Dr. Nasraoui, Dr. Zhang, and Dr. Altiparmak for being part of my dissertation committee and taking the time to give me suggestions to make it better.

I would like to thank Dr. Julia Chariker for her continuous support and constructive feedback. I'd like to extend my thanks to current and former lab members Sen Yao, Andrey Smelter, Ernur Saka, Kalpani De Silva, Mohamed Chaabane, Aanchal Malhotra, Aryan Neupane, Dr. Jae Hwang, Jonah Daneshmand, and Tae Lim Kook.

ABSTRACT

COMPUTATIONAL FRAMEWORKS FOR MicroRNA FUNCTIONAL ANALYSIS OF INTER-KINGDOM AND INDIRECT TARGETING

Mohammed Sayed

April 19, 2021

Genes are DNA sequences that encode the information needed to synthesize molecules necessary for the function of the cell. Some genes are called protein-coding genes because they have the code required to manufacture proteins. The expression of a certain gene means its product (protein) is produced. Although some genes are not protein-coding, they regulate the gene expression of other protein-coding genes. Of these, microRNAs (miRNAs) are small RNA molecules that inhibit the expression of other genes by binding to their mRNA transcripts. miRNAs have been shown to be linked to several biological processes like development and diseases like cancer.

Recently, researchers have hypothesized that miRNAs are involved in the regulation of the expression of genes from other species. Although tools to predict miRNA target genes are available in the case when miRNAs and target genes belong to the same species, to our knowledge there are no available tools to predict inter-kingdom miRNA target genes (miRNA and target genes belong to two different kingdoms). To address this limitation, we developed an efficient tool to predict potential gut bacterial

genes targeted by miRNAs from edible plants. We successfully predicted ginger miRNAs that target two genes from a gut bacterial strain called *Lactobacillus rhamnosus GG.* To maintain the efficiency of our tool while using a larger number of miRNAs and bacterial strains, we used a hash-table to index the sequences of bacterial genes.

To predict the function of a miRNA, we start by compiling the list of direct target genes (ones with binding sites) and then we search for biological process in which these genes are enriched. This approach does not include other genes affected by the miRNA but do not necessarily have a physical binding site (indirect targets). An example of an indirect target is the gene that doesn't have a binding site and is regulated through other direct targets like transcription factors. To overcome this limitation, we developed *miRinGO* an interactive web application to include these indirect targets in the functional analysis. Our approach showed better performance compared to the existing approach in predicting biological processes known to be targeted by certain miRNAs.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 INTRODUCTION

MicroRNAs (miRNAs) are small non-coding RNA sequences (21-24 nucleotides long) that have a prominent role in gene regulation. MiRNAs mainly act post-transcriptionally by binding to mRNA transcripts and either inhibit translation (protein synthesis) or initiate the degradation of the target transcript. Computational tools have been introduced to predict potential target genes and biological processes affected by a specific miRNA or a set of miRNAs. Despite extensive work in miRNA research, the full spectrum of the role of miRNAs in gene regulation is still to be revealed.

1.1 Motivation

Typically, to predict the function of a miRNA, computational tools follow a two-step process. The first step is to identify the list of potential target genes and then they identify biological processes that are enriched in these target genes. Although many tools are developed to do each of the two steps, they have some limitations. For instance, miRNA target prediction tools are species-specific (i.e. they are used to predict target genes when miRNAs and targets belong to the same species e.g. human). Also, current tools include only target genes with physical binding sites (direct targets) in the enrichment step.

In this dissertation, we discuss the need for computational tools to address inter-kingdom miRNA targeting (part I). We also discuss the motivation to develop new tools for miRNA functional analysis (Part II).

1.1.1 Part I: inter-kingdom miRNA targeting

The overall goal of this project is to determine how edible plants (e.g. ginger and broccoli) affect our gut bacteria and ultimately our health. Gut bacteria play an important role in human health. For example, help digest hard-to-digest dietary fibers, defend the host against harmful microorganisms, and synthesize essential vitamins and amino acids. The imbalance of gut bacteria can lead to multiple diseases including inflammatory bowel diseases (IBD), obesity, liver disease, and cancer.

Studies have shown that food including edible plants can affect the balance of gut bacteria. Edible plant cells use secreted nanoparticles called exosomes to communicate with other cells and tissues [1]. The cargo of exosomes includes different types of molecules: proteins, DNAs, mRNAs, and miRNAs. Researchers hypothesize that exosomes from edible plants may interact with gut bacteria.

MiRNAs play a role in the regulation of target gene expression even across kingdoms [2]. In our case, we are dealing with two different kingdoms (plants and bacteria), and each has its miRNA targeting mechanism. In animals, a limited number of base-pairings (~ 7 bp) is sufficient for miRNA targeting. On the other hand, in plants perfect (or near-perfect) base pairing between miRNA and mRNA sequences is required. Although miRNAs have been shown to target gut bacteria [3], the actual mechanism is still unknown. For the above reasons, new computational tools are needed to identify cross-kingdom miRNA targets.

1.1.2 Part II: human miRNA functional analysis

To computationally predict miRNA-targeted biological processes, typically potential target genes are compiled using one or more miRNA target prediction tools and a standard gene enrichment analysis [4] is used to find potential biological processes enriched in these genes. Although current tools are widely used, they have some limitations. Of these, existing tools consider only direct targets (ones with physical binding sites) of miRNAs but do not consider other targets that do not necessarily have miRNA binding sites [5].

1.2 Dissertation Contributions

To address the aforementioned limitations, we developed the following computational frameworks.

1.2.1 A framework for inter-kingdom miRNA targeting

Given the importance of the miRNA seed region in targeting mechanisms in different kingdoms, animals, plants, and viruses, we developed a seed-based miRNA targeting framework to find potential gut bacterial genes targeted by edible plant miRNAs. In this framework, first gut bacterial genes were searched for potential plant miRNA binding sites. Then, we employed a technique proposed by Murphy et al. [6], initially to predict targets of viral miRNAs, to identify target genes with enriched binding sites. We showed that using our tool we can identify ginger miRNAs that target genes from *Lactobacillus rhamnosus GG* (LGG) bacteria. To validate our results, two miRNAs have been experimentally validated to downregulate and bind to two LGG genes.

1.2.2 A web application for human miRNA functional analysis

We developed, *miRinGO,* an interactive R Shiny application [7] to predict potential biological processes both directly and indirectly targeted by human miRNAs. In addition to a list of top-ranked biological processes, the web application gives the user a visual summary of the results. It is freely available from GitHub at https://github.com/Fadeel/miRinGO

1.3 Dissertation Outline

The remainder of this dissertation is organized as follows. Chapter 2 provides an introduction to molecular biology and miRNA targeting techniques. Chapter 3 introduces a proposed method to predict potential gut bacterial genes targeted by edible plant miRNAs. Chapter 4 introduces the proposed framework to find biological processes indirectly targeted by miRNAs in humans. Chapter 5 provides conclusions and potential future work.

2 BACKGROUND AND LITERATURE REVIEW

2.1 Introduction to molecular biology

2.1.1 Living organisms and cells

It is estimated that there are more than 10 million different living species on Earth [8]. The building block of these organisms is the cell. Although the cell size can vary from 1μm (bacteria) to 100 μm (plants), all cells share mainly two characteristics [9]. First, the ability to take up nutrients from the environment and converting them to other molecules and energy. Second, the ability to make many copies of themselves and passing genetic material to their offspring.

Some organisms are composed of single cells like bacteria but other species like humans can have up to 10^{13} cells. Cells can be categorized into two types: *prokaryotic* and *eukaryotic*. Prokaryotic cells do not have a nucleus and on the other hand, eukaryotic cells have a nucleus where it contains the cell's DNA. Prokaryotes include bacteria and archaea. Animals, plants, and fungi are examples of eukaryotes.

Cells are surrounded by fat-based molecules (phospholipids) forming what is called the cell membrane (or plasma membrane) [10]. Inside the cell membrane, cellular machinery and other structural units exist within a water-based environment called cytoplasm. Cells contain a variety of molecules including nucleic acid, proteins, carbohydrates, and lipids. Nucleic acids and proteins will be discussed in more detail below.

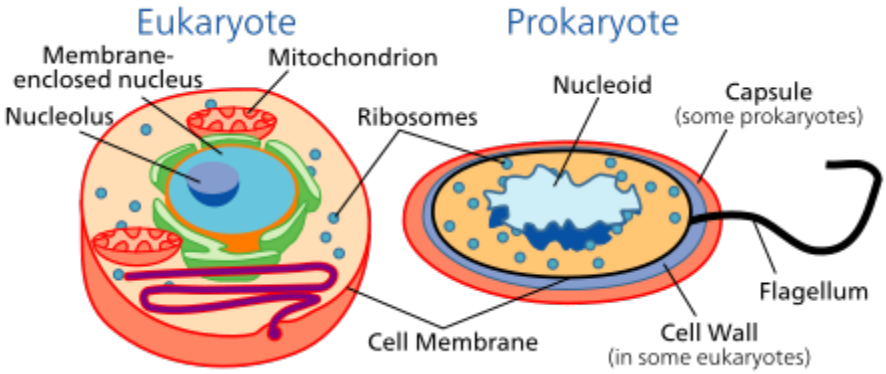

Figure 2-1: Eukaryotic and prokaryotic cells [11]

2.1.2 Nucleic acids

Nucleic acids are the molecules that contain genetic information. There are two types of nucleic acids: deoxyribonucleic acid (DNA) and ribonucleic acid (RNA).

2.1.2.1 DNA

DNA consists of a sequence of smaller molecules called nucleotides [10]. Each nucleotide contains three molecules: a nitrogenous base, a sugar molecule, and a phosphate group. There are four different DNA nucleotides based on the nitrogenous base: adenine (A), thymine (T), cytosine (C), and guanine (G) as shown in Figure 2-2.

Although DNA can be found as single-strand polynucleotides, it is more stable when two strands come together and bases from one strand are bound (paired) to the complementary bases for the other strand via hydrogen bonds as shown in Figure 2-3. Each “A” from one strand always pairs with a “T” from the other strand and each “G” from one strand pairs with a “C” from the other strand.

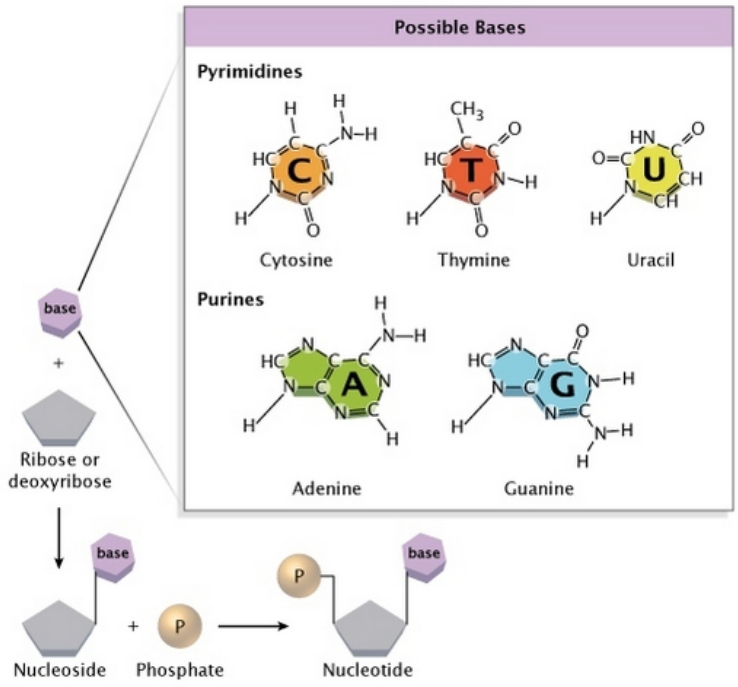

Figure 2-2: chemical structure of nucleotides [12]

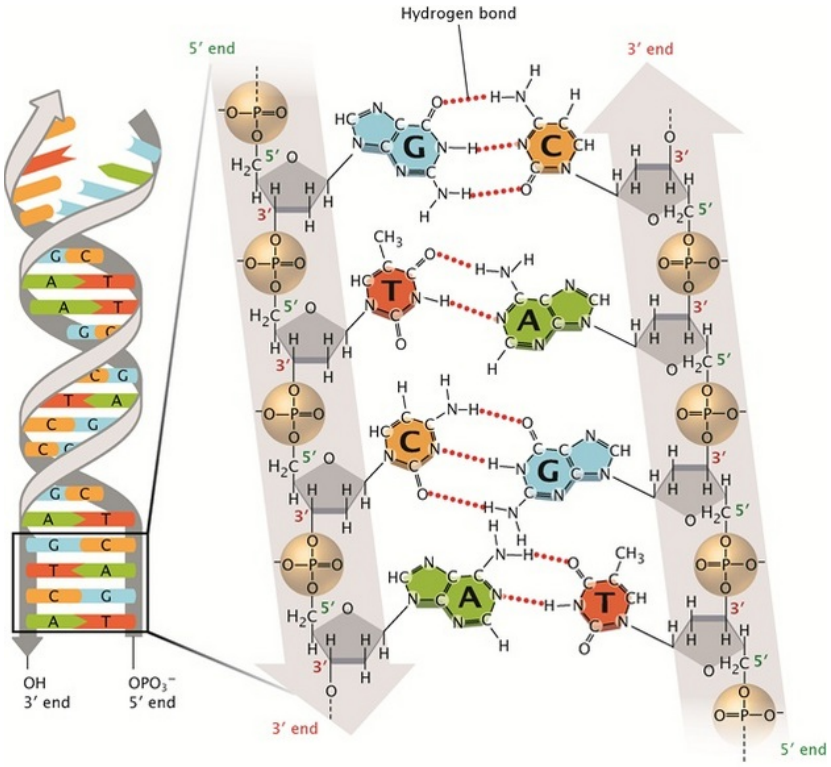

Figure 2-3: DNA base pairing [12]

2.1.3 Proteins

Proteins are large molecules that contribute to a cell's function and structure. For example, structural proteins maintain cell shape and are part of structural elements of connective tissues like cartilage and bone [13]. Other proteins called enzymes work as catalysts for biochemical reactions within the cell. Another category of proteins are attached to the cell membrane and have diverse functionality e.g. transporting molecules from/to the cell, activating an intracellular process upon receiving an extracellular signal or attach the cell to a specific location.

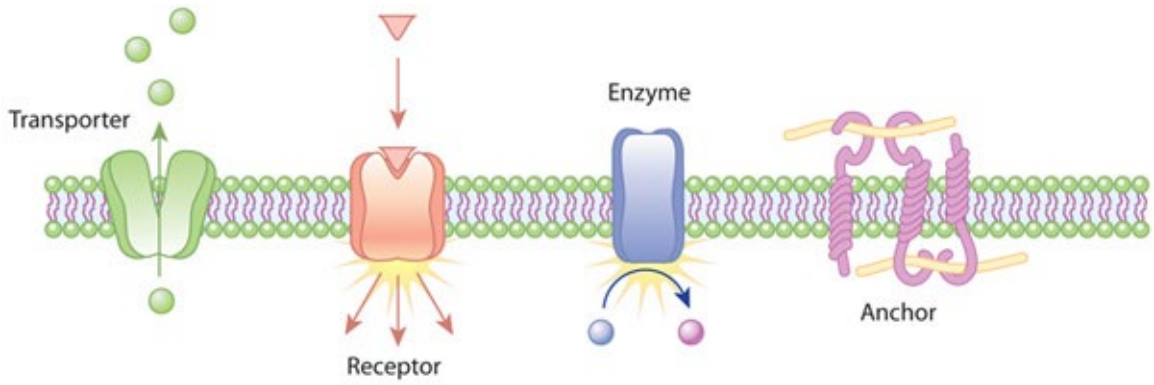

Figure 2-4: Example functions of membrane proteins [10]

Each protein consists of a series of building blocks called amino acids. Each amino acid consists of a central carbon atom, an amino group, a carboxyl group, a hydrogen atom, and a side chain. Multiple amino acids are bound together by peptide bonds to form a long chain of polypeptides or a protein as shown in Figure 2-5. The linear sequence of amino acids is considered the primary structure. Protein's primary structure determines its conformation and ultimately the three-dimensional shape.

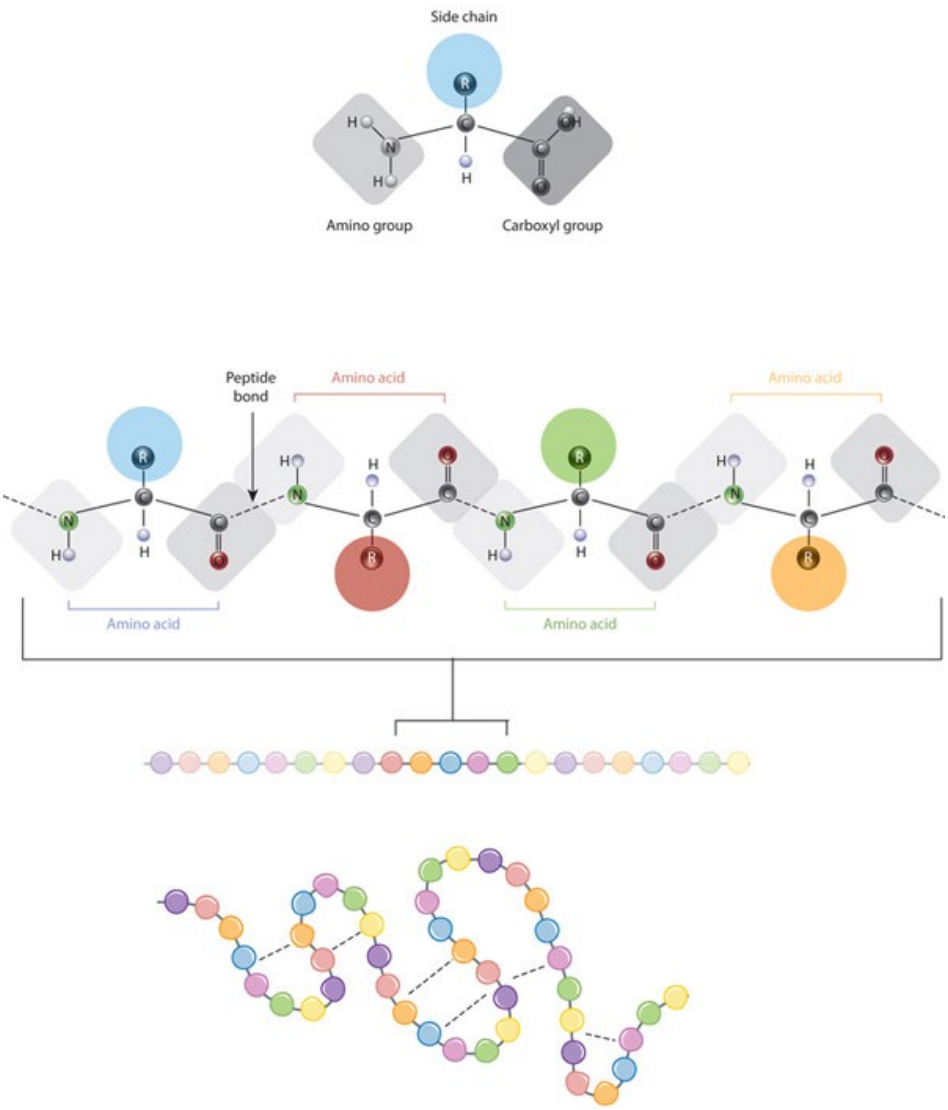

Figure 2-5: Protein building blocks [10]

2.1.4 The central dogma of molecular biology

The term "central dogma" was coined by Francis Crick in 1956 to describe the flow of information (genetic code needed to form a protein) from DNA to protein. It states that

"*once information has got into a protein it can't get out again*" [14].

The information flow was classified into three categories. General transfers occur in most cells and include DNA → DNA, DNA→ RNA, and RNA → protein. Special transfers occur under specific conditions and include the following transfers RNA→ RNA, RNA → DNA, and DNA → protein. The last category includes transfers unlikely to occur (protein → protein, protein → RNA, and protein → DNA).

2.1.5 Genes and gene expression

Genes are DNA sequences that encode the information needed to synthesis a protein. In humans, a child inherits a copy of the same gene from each parent. Genes are expressed when their corresponding proteins are manufactured [15]. Gene expression process goes through two main steps: DNA transcription and translation as shown in Figure 2-6.

DNA transcription is the process in which information required for protein synthesis is transferred to an intermediate molecule called messenger RNA (mRNA). The transcription process is initiated by an enzyme called RNA polymerase where one strand of DNA is used as a template [16].

During translation process, the genetic code in the mRNA molecule is read sequentially to produce a linear chain of amino acids (protein). With help of translation machinery called the ribosome, every three nucleotides in mRNA correspond to one amino acid according to a specific code as shown in Figure 2-7.

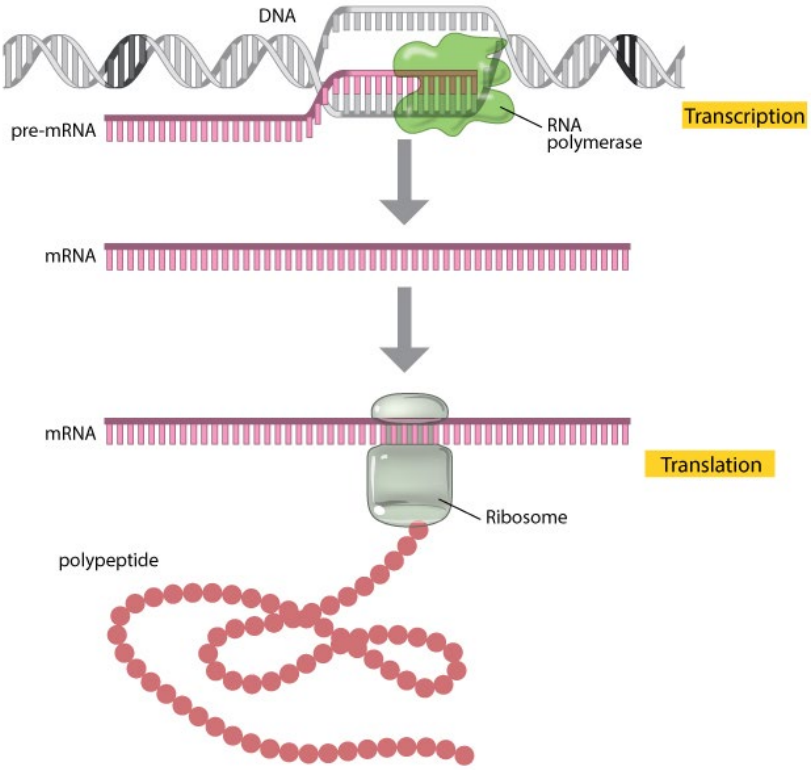

Figure 2-6: Gene expression process [15]

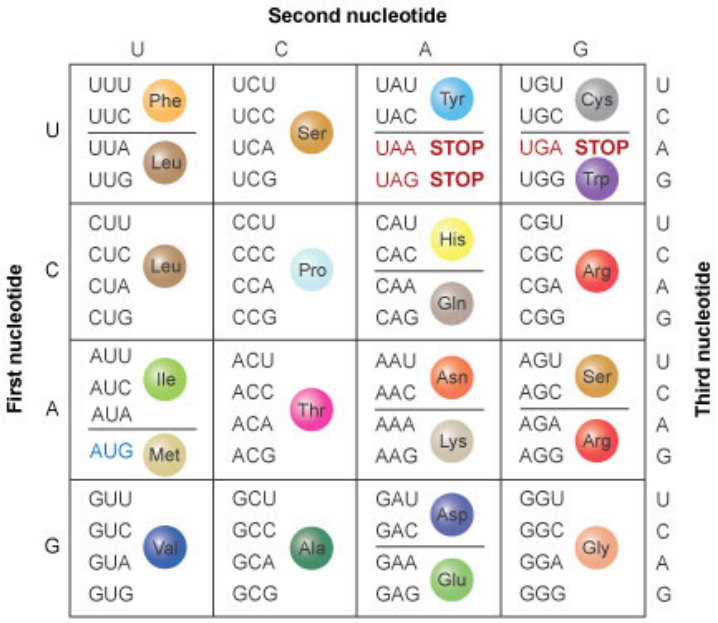

Figure 2-7: Genetic code [15]

2.1.6 Regulation of gene expression

The process of gene expression involves decoding genetic instructions encoded in the DNA to synthesize proteins. The set of proteins expressed in a cell determines the structure and function of this cell. Although all cells in a multicellular organism (e.g. animals, plants, …) have the same copy of the DNA sequence, different sets of genes are expressed in different cell types and some genes are only expressed in a specific cell type, for example, Hemoglobin (which is used to carry oxygen) is expressed specifically in red blood cells. Another protein, tyrosine aminotransferase (which is used to break down tyrosine in food), is specifically expressed in liver cells [8].

The process of gene expression goes through multiple steps from DNA to proteins through RNA, therefore the amount of expression can be controlled at each step. There are multiple levels of regulation as detailed below.

2.1.6.1 Epigenetic alterations

In which genes are switched on or off without genetic changes (changes in DNA sequence) but in response to external factors like development, aging, exercise, and diet [17]. There are two main types of epigenetic changes as shown in Figure 2-8. The first mechanism is called methylation and is facilitated by adding a methyl group to a DNA molecule. If specific regions in the DNA (for example promotor regions) are methylated, this will repress the expression of the nearby gene. The second mechanism which is called histone modification happens when chemical groups are added to histones. Since DNA is wrapped around histones, these modifications can determine which genes are available for expression (unwrapped) or turned off (wrapped).

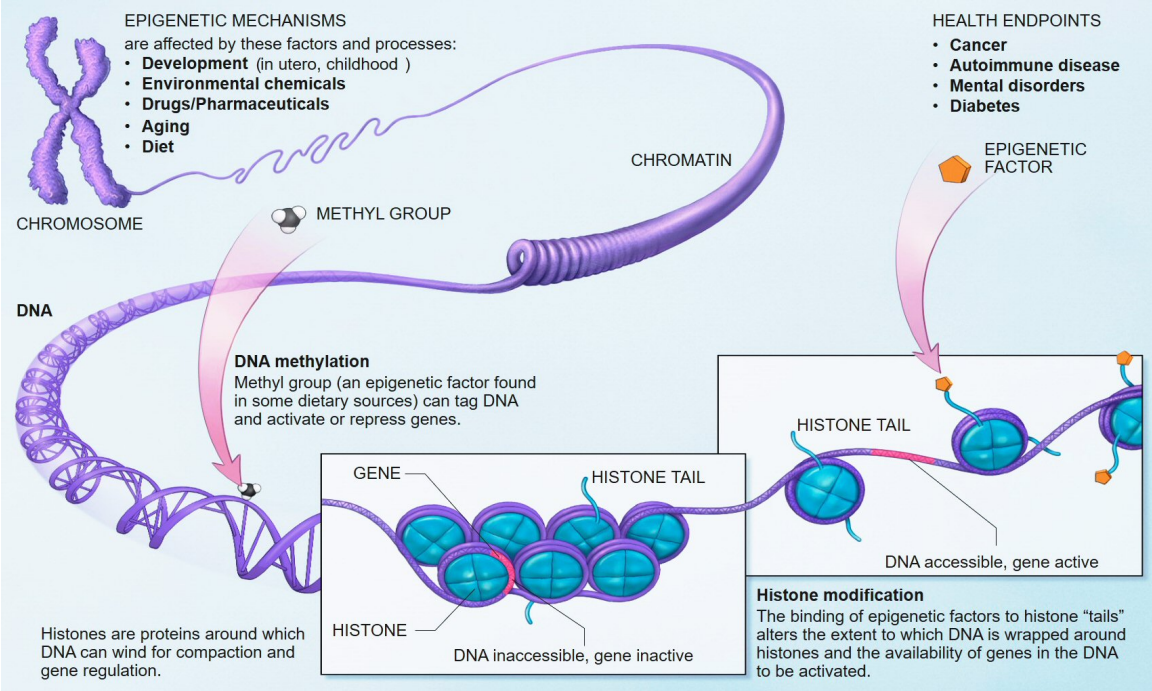

Figure 2-8: Epigenetic mechanisms for gene regulation [18]

2.1.6.2 Transcriptional regulation

In addition to epigenetic mechanisms, a group of protein-coding genes called transcriptional regulators (factors) activate or repress the expression of a gene by binding to a specific DNA sequence usually 5-10 nucleotides near that gene. This binding initiates a sequence of reactions that determines which genes to be transcribed and additionally the rate of transcription [19].

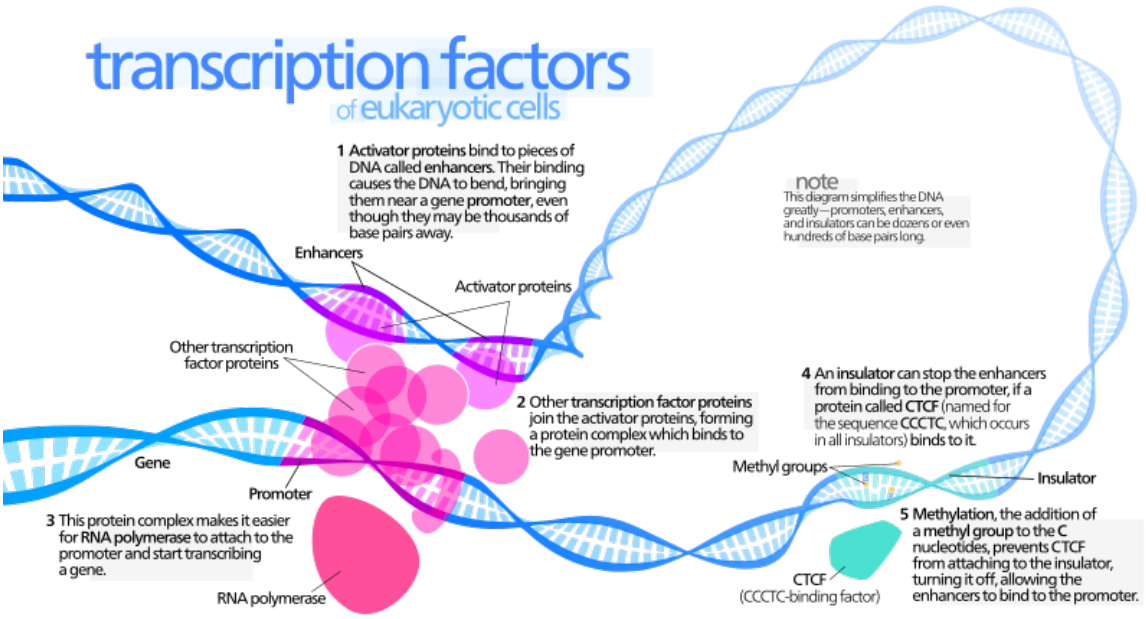

Figure 2-9: Transcription factors initiate the control of gene expression [20]

2.1.6.3 Alternative splicing

Transcription of a gene from a DNA sequence produces a long RNA transcript called precursor mRNA (pre-mRNA) transcript. Pre-mRNA transcripts typically undergo a process called RNA splicing in which non-coding sequences of the gene (introns) are removed and protein-coding sequences (exons) are joined back together to form mature RNA (mRNA) transcript [21] that has the code to synthesis a specific protein. Some genes are alternatively spliced meaning that some exons are either included or excluded in the final mRNA transcript. Alternative splicing (AS) makes it possible for a single gene to produces multiple proteins as shown in Figure 2-10.

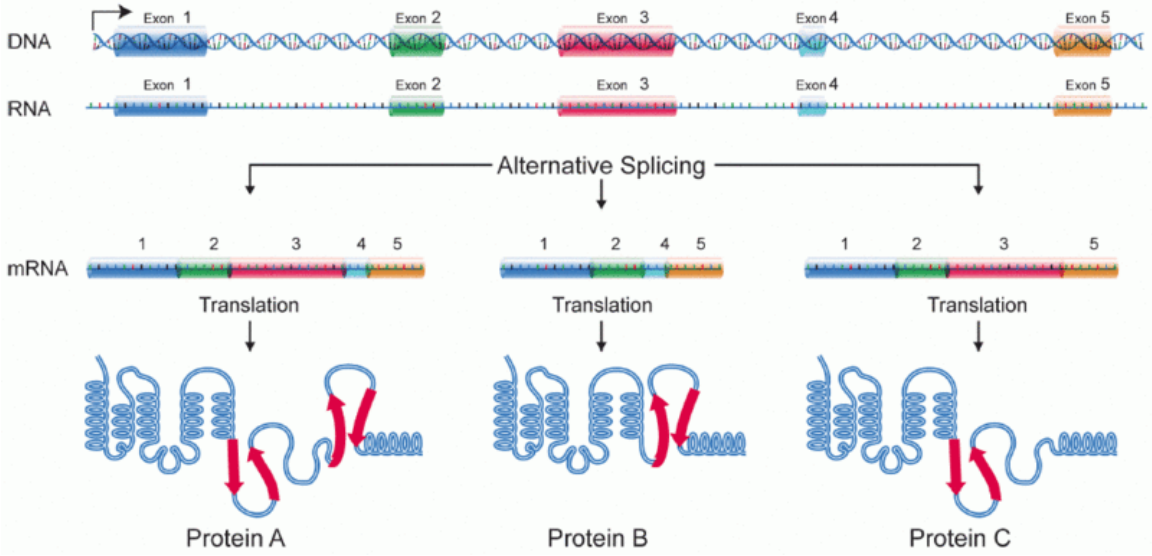

Figure 2-10: An example of an alternatively spliced gene [22]

2.2 MicroRNAs

MicroRNAs (miRNAs) are small non-coding RNA molecules that are ~22 nucleotides and are known to be involved in the regulation of the expression of other protein-coding genes (target genes) and have an essential role in many biological processes including development and diseases like cancer. The first miRNA was discovered in *Caenorhabditis elegans* (*C. elegans*) (the worm) in 1993 [23] [24] by Ambros and Ruvkun groups [25]. This miRNA (*lin-4*) was found to be essential to the normal development of *C. elegans* by regulating the expression of the *lin-14* protein.

2.2.1 Biogenesis of miRNAs

miRNA genes can be found within introns of other protein-coding genes (Mitrons) or independent from other genes (intergenic miRNAs) [25]. Biogenesis of miRNAs starts with transcribing miRNA genes into primary transcripts (pri-miRNAs) using RNA polymerase II/III. Processing of pri-miRNAs to produce mature miRNAs

follows different pathways depending on the locus of the miRNA gene as shown in Figure 2-11.

In the canonical pathway, longer pri-mRNA transcripts (~1K nucleotides) are processed by a microprocessor complex, consists of two proteins (DGCR8 and Drosha) to produce a transcript with a stem-loop structure known as pre-miRNA (~60 nucleotides in animals). After producing pre-miRNAs, they are exported to the cytoplasm where the loop is removed by Dicer protein, and the double-stranded RNA is released. Each strand of the mature miRNA can be loaded into one of the Argonaute (AGO) family of proteins to form a miRNA-induced silencing complex (miRISC).

In the non-canonical pathway, some intronic sequences following the splicing process have the structural features of pre-miRNAs, therefore bypassing the Drosha-mediated processing [26].

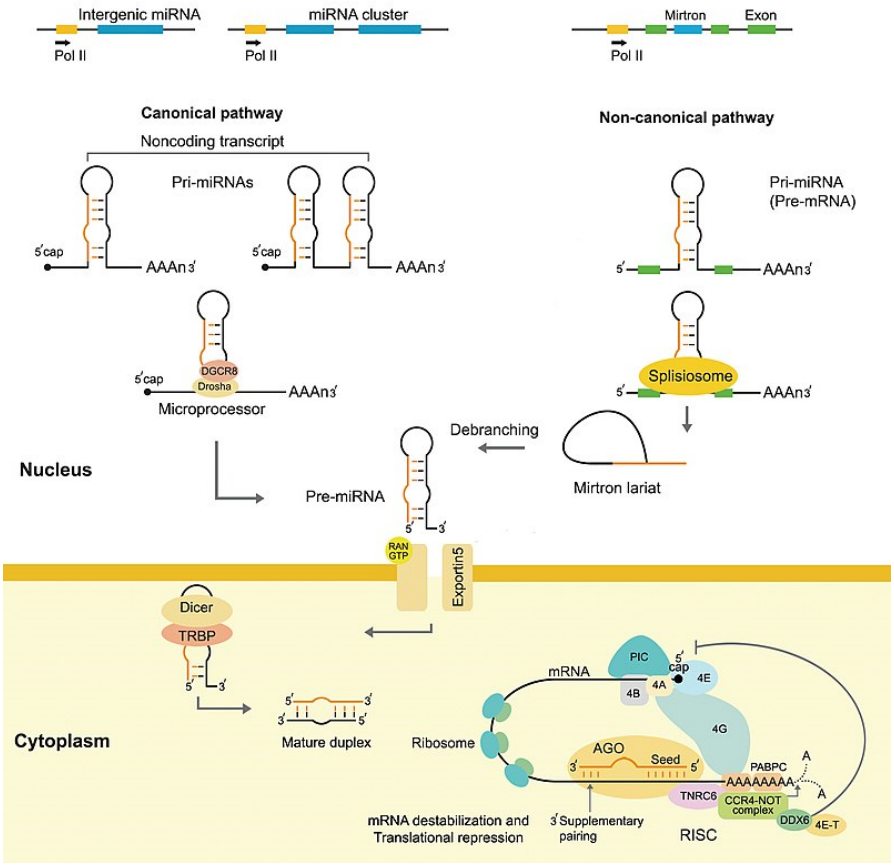

Figure 2-11: Different pathways of animal miRNA biogenesis [27]

2.2.2 MicroRNA targeting mechanism in animals

The following papers in this section cover the process of miRNA-based gene silencing in animals. Such information could enhance the performance of tools developed for the computational prediction of miRNA targets.

Cisse, *et al.* [28] addressed the question of what the minimum number of base-pairings is needed to establish a stable association of two DNA/RNA strands. This could be used to infer some of the rules governing small RNA-based gene silencing, including miRNA-based gene regulation. To answer this question, they synthesized 8 nucleotides (nt) sequences representing the seed of human miR125 with mismatches at the 1st nt

(leaving 7 contiguous nts) and at the 2nd nt (leaving 6 contiguous nts). The association rate with seven contiguous nts was 450 times bigger than the one with 6 nts. This finding supports the hypothesis that at least seven Watson-Crick base pairs in the miRNA seed region are needed for rapid and effective targeting.

Wee, *et al.* [29] provided an insight into how the Argonaute (Ago2) protein in RISC (RNA-induced silencing complex) shapes the miRNA guide. They showed that Ago2 divides a miRNA sequence into five functional domains. These domains are the anchor (1st nt), seed (2nd – 8th nt), central (9th – 12th nt), 3' supplementary (13th – 16th nt), and tail region (17th – 21st nt). Their results showed that the miRNA-guided silencing complex is used for translation inhibition without cleaving target mRNA transcript. On the other hand, small-intervening RNAs (siRNAs) are responsible for mRNA cleavage. Unlike mismatches at the seed ends, mismatches in the center of the seed regions are not favorable for effective gene silencing. Although G-U wobbles in the seed region were considered by existing miRNA target prediction tools, this paper suggests that G-U wobbles should be treated like mismatches and should not be allowed.

Salomon, *et al*. [30] studied the binding properties of guide RNA that are bound to Ago protein within the RISC complex. They found that seed nucleotides [2nd – 5th nt] are involved more at the initial binding to target mRNA and then Ago is conformed to expose nucleotides [6th – 8th nt] and [13th – 16th nt] for more stable binding. Also, Extensive binding on the 3' end could compensate for mismatches in the seed region [4th – 5th nt]. They also suggested that the binding of nucleotides [9th, 10th] of miRNA should be penalized in miRNA target prediction algorithms due to the unfavorable conformation in Argonaute.

Broughton, *et al.* [31] investigated the specificity of seed-based miRNA targeting. To answer this question, they used individual-nucleotide resolution cross-linking immunoprecipitation (iCLIP) to extract miRNA-target chimeras by ligating each miRNA to its target. Since miRNA families have common seed regions but different 3' sequences, they were used to determine if all miRNAs in a family have common targets or not. The results suggest that miRNA family members have shared targets in addition to targets specific to each member. This paper highlights the importance of pairing the miRNA 3' supplementary sequence for enhancing the predictive power of current miRNA target prediction tools.

Chandradoss, *et al.* [32] studied how the Ago protein finds target sites in mRNA. They used a single-molecule FRET (Förster resonance energy transfer) assay to visualize the search process used by Ago2 to find miRNA target sites. They found that the miRNA-bound Ago employs a lateral diffusion mechanism for searching and identifying target sites. First, the sub-seed region [nt 2-4] is used for initial searching and once a complete seed pairing is found [nt 2-8], the binding becomes stable. This model suggests that target regions might be enriched in the trinucleotides [nt 2-4] of the seed region which could be an additional feature to be added to miRNA target prediction tools.

2.2.3 Computational prediction of miRNA targets in animals

Saito *et al.* [33] discussed six main features currently used by miRNA target prediction tools. These features are listed as follows,

1. miRNA-mRNA base pairing at the 5' miRNA seed region [1st to 8th nucleotide]: this feature is considered the most relevant feature used by the silencing complex to find target genes. Stringent seed paring is considered more effective than seed pairing with mismatches.

2. Location of target sites within mRNA transcript: although miRNA-guided silencing complex may target 5'UTR and CDS regions, the 3'UTR contains most of the target sites.

3. Conservation of miRNAs and their targets: miRNA targets tend to be in regions conserved among closely related species. Using conservation information in miRNA target prediction tools could increase specificity. Of course, this increase occurs with a cost to sensitivity.

4. Site accessibility: to enable miRNA-mRNA hybridization, a target site must be easily accessible. Estimation of site accessibility can be achieved by calculating the minimum free energy of an mRNA secondary structure. Although theoretically, an mRNA secondary structure with the minimum free energy is most likely to occur naturally, this might not be the actual structure because RNA secondary structure is dynamic [34]. Also, the computational complexity of these algorithms is $O(n^3)$ [35], where n is the length of the RNA transcript.

5. Multiple target sites Having more than one potential miRNA target site could enhance the role of miRNA as a translation inhibitor as it increases the chances of binding of silencing complex and mRNA transcript.

6. Expression profiles of miRNAs and mRNAs: due to their role in translation inhibition and mRNA degradation, the expression of miRNAs is expected to be negatively correlated with the expression of targeted mRNAs. Although using this feature could help exclude some of the false positives, it could overlook some true miRNA-mRNA pairs that are negatively correlated at the protein level only.

Saito *et al.* [33] provided a useful comparison of approximately 30 different prediction tools based on the presence or absence of these six features. On the other hand, it did not provide a performance analysis (accuracy and computational complexity) for these tools.

Bradley, *et al.* [36] provided a review of two classes of miRNA target prediction tools, namely, classical tools and tools trained by data generated from CLIP (cross-linking and immunoprecipitation) or CLIPL (cross-linking and immunoprecipitation and ligation) high-throughput experiments. This new class of prediction tools (CLIP & CLIPL-based) is an improvement in the identification of transcriptome-wide miRNA targets.

In CLIP-based experiments, RNAs bound to the AGO protein are directly sequenced, but there is still ambiguity linking specific miRNA to a specific mRNA. Although CLIPL-based experiments use a similar process, an additional step is added that ligates a miRNA to its mRNA fragment. By sequencing miRNA-mRNA chimeras, a specific miRNA can be linked to its target site.

Although the high-throughput nature of CLIP&CLIPL experiments makes it suitable for training machine learning-based target prediction algorithms, these methods still suffer from technical difficulties that require well-trained personnel to manage. Another issue with these experiments is that several non-canonical target sites (seed pairing with mismatches and bulges) have been reported, which are thought to be ineffective sites.

Fan, *et al.* [37] provided an overview of 38 sequence-based miRNA target prediction tools and a comparison of seven popular tools based on their predictive performance, ease of use, availability, and impact. The authors compared TargetScan 6.2 [38], PicTar [39] , DIANA-microT-CDS [40], miRanda [41], EIMMo3[42], mirTarget2 v4 [43], miRmap v1.1 [44]. They assessed the predictive performance based on four different datasets, duplex-level (the actual site within mRNA transcript), gene-level, mRNA expression-level, and protein-level.

The results of these assessments suggest that there is no globally best tool. Although TargetScan and miRmap provided high overall predictive performance, PicTar and MirTarget2 showed high specificity. As expected, gene-level prediction performance was better than duplex-level predictions, as gene-level predictions can make use of more information (enrichment of multiple seed pairings). Finally, many (>83%) of non-functional target sites have at least 6bp in their seed-pairings, which makes the miRNA target prediction more challenging.

Oliveira, *et al.* [45] focused on finding the best approach for using miRNA prediction tools. The authors argue that the current trend of using the intersection of targets identified with different prediction tools is not the best way to combine results, as this approach suffers low sensitivity and often misses many true targets. To address this question, they compared the targets generated from four different individual tools (TargetScan [46], miRanda-mirSVR [41], Pita [47], and RNA22 [48]) and the intersection/union of different subsets of these tools. In terms of prediction performance, the authors suggest using the union of target prediction results from Targetscan and miRanda when high specificity is needed and the union of target prediction results from

Targetscan, miRanda, and RNA22 if higher sensitivity is need while preserving good specificity. Although this paper addressed a practical aspect to solve the miRNA target prediction problem, it has some limitations. For its test dataset, true negatives were defined as those that are neither validated (according to mirTarbase database) nor predicted. However, this assumption is not accurate since some true targets might not be validated till this point in time Also, they did not provide a plausible explanation for why this particular combination of tools might have the best predictive performance.

Gumienny, *et al.* [49] introduced MIRZA-G tool for predicting miRNA targets and siRNA off-targets. Since siRNA off-target sites are not likely to be conserved, using conservation-based miRNA target prediction tools is not suitable. In addition to the classical features (site accessibility, position within 3'UTR, flanking G/flanking U content, conservation information), MIRZA-G employs miRNA-target interaction energy as a new feature. A logistic regression model was used to predict the efficacy of the target site based on the previous features.

Agarwal, *et al.* [46] introduced the prediction model used in the latest version of TargetScan (v7). Although TargetScan is considered the leading tool for miRNA target prediction, previous versions were mainly focusing on conserved target sites which reduces its sensitivity and increases the likelihood of missing siRNA off-targets. The new multiple regression model employs 14 features and covers information regarding seed type, context around the target site, 3' pairing, target transcript, and site accessibility.

Since TargetScan considers only canonical seed types (no mismatches or GU wobbles), the authors examined the efficacy of non-canonical target sites which are abundantly discovered in the new CLIP-based experiments. Surprisingly, they found that

these sites are not functional concerning translation inhibition. Although the proposed model surpassed other miRNA target prediction tools, there are still some limitations. First, despite being simple and easy to interpret, the multiple regression model assumes that there is a linear relationship between target site efficacy and each of the input features which may not be the case for some features. Second, regarding test data, only gene expression at the mRNA level was used which may not be sufficient for detecting translation inhibition. Third, only target sites within the 3'UTR are included which could miss other functional targets in the CDS/5'UTR regions.

3 PREDICTION OF POTENTIAL BACTERIAL GENES TARGETED BY PLANT MIRNAS

3.1 Introduction

Previous studies have shown the role of small non-coding RNAs especially miRNAs in the regulation of cross-kingdom gene expression. As described in Liang, *et al.*'s study [50], cross-kingdom regulation occurs when a microRNA and a target gene belong to two different species (e.g., different kingdoms such as animals, plants, bacteria, viruses, etc.).

Although the way cells within the same species communicate is known (e.g. communication between neurons through hormones in humans [51]), the way cells from two different species exchange signals was not discovered until recently.

In this chapter, we propose a computational framework to investigate a new type of cross-kingdom gene regulation by predicting potential gut bacterial genes targeted by edible plant miRNAs (e.g. ginger miRNAs).

3.2 Background

3.2.1 Cross-kingdom gene regulation

Liang, *et al.*'s study [50] provided four different examples of the new role of miRNA in cross-kingdom gene regulation. Firstly, Plant miRNAs were found to exist in human tissue and sera and at relevant concentrations. Of these plant miRNAs, MIR168a was shown to target low-density lipoprotein receptor adaptor protein 1 (*LDLR AP1*) [52].

Also, Human miRNAs were found within the malaria parasite *Plasmodium falciparum* and were shown to affect its growth [53]. On the other hand, Viral miRNA KSHV-miR-K12-11 was found to mimic and share common targets with human miRNA miR-155 [54]. Lastly, Human miRNA miR-122 has a role in stimulating the replication of the hepatitis C virus by stabilizing its viral RNA and prevent its degradation [55] by targeting two 5' UTR sites.

Liu *et al*. [3] provided additional evidence of the role of cross-kingdom miRNA-based gene regulation. They showed that human miRNAs from epithelial and other cells can enter gut bacteria and affect its growth. They also discussed how miRNAs manipulate the composition of the gut microbiota. Interestingly, they mentioned that miRNAs could target ribosomal RNAs, and they could either repress or promote gene expression.

3.3 Methods

3.3.1 Datasets

Top 49 expressed ginger miRNAs (Appendix I) are detected from miRNA-seq data of exosomes-like nanoparticles (SRA accession numbers SRX5085431, SRX5085432, SRX5085433). Reference genome sequences of 8 common gut bacteria (see Table 3-1) were downloaded from NCBI RefSeq database [56] (ftp://ftp.ncbi.nlm.nih.gov/genomes/refseq/bacteria/)

Table 3-1: Common gut bacteria and their RefSeq accession numbers

Bacterial strain	NCBI RefSeq Accession Number
Akkermansia muciniphila	NC_010655.1
Bacteroides fragilis	NC_006347.1
Clostridium perfringens	NC_008261.1

Escherichia coli	NC_000913.3
Enterococcus faecalis	NC_004668.1
Helicobacter pylori	NC_000915.1
Lactobacillus rhamnosus GG	NC_013198.1
Lactobacillus ruminis	NC_015975.1

3.3.2 Overall pipeline

Our pipeline (Figure 3-1) takes as inputs sequences of plant miRNAs and sequences of bacterial genes. Then, for each miRNA-gene pair, we search for hits for both miRNA seed and its reverse complement (This because miRNAs can target both bacterial DNA and mRNA [3]). For fast searching, bacterial gene sequences are indexed into a hash-table as detailed in the next section. Finally, for each miRNA-gene, we compare the actual number of hits with the expected number to find genes with enriched miRNA seed binding sites.

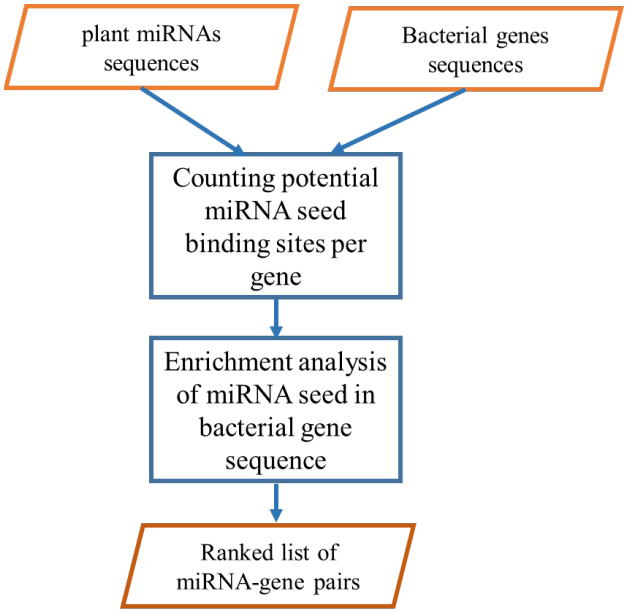

Figure 3-1: Overall pipeline to detect potential bacterial genes targeted by plant miRNAs

3.3.3 Hash table-based miRNA seed binding site detection

Let *T* represents the sequences of bacterial genes, *p* to the pattern (e.g. miRNA seed sequence) we are searching for in *T*, and *H(p)* to be the hash function that converts *p* into an integer number. Our hash table will have *H(k-mer) (*where *k-mer* is a subsequence of length *k* of text *T*) as key and a list of locations of this *p* in *T* as value as shown in Table 3-2

Since our set of keys are static and relatively not too big (for example, the set of all 4^8 possible 8-mers if k = 8), we use a perfect hash function, where each key is mapped to a unique row without collisions. Since DNA/RNA has 4 different nucleotides (A, C, G, T/U), their integer representation can be as follows *I(A) = 0, I(C) = 1, I(G) = 2 and I(T) = 3.* Then the hash value of pattern *p* of length k (k-mer) can be calculated as follows,

$$H(\mathrm{p}) = \sum_{i=0}^{k-1} \big(I(p[i]) \times 4^i \big)$$

For example, $H('ATG') = 2 \times 4^0 + 3 \times 4^1 + 0 \times 4^2 = 14.$

Table 3-2: Target bacterial gene sequences represented as a hash table

Key	Value
…	…
H('ATACACTG')	[<gene1,offset>, <gene656,offset>]
…	…
H('CATCAGGG')	[<gene100,offset>, <gene45,offset>, <gene98,offset>]
…	…

3.3.4 Enrichment analysis

In order to find bacterial genes that are more likely to be targeted, we employ a technique proposed by Murphy et al. [6] initially to predict targets of viral miRNAs. To quantify overrepresentation (enrichment) of a k-mer in a sequence, the actual count C_{actual} is compared to an expected count $C_{expected}$ (how many times we expect to see this k-mer X by chance). $C_{expected}$ is calculated by the following equation,

$$C_{expected} = (L - k + 1) \times P(X)$$

Where *P* is the probability of seeing this k-mer in a target sequence of length *L* given a specific model of this sequence. We use a first-order Markov chain of the target sequence to model it. *P* can be calculated using the following equation,

$$P(X) = p(x_1) \prod_{i=2}^{k} p(x_i \mid x_{i-1})$$

Where x_i is nucleotide at position i of k-mer X, $p(x_i \mid x_{i-1})$ is the probability of having nucleotide x_i after nucleotide x_{i-1} and can be represented as a matrix (Figure 3-2) and $p(x_1)$ is probability of seeing nucleotide x_1.

x_{i-1} \ x_i	A	C	G	T
A	0.2	0.2	0.3	0.3
C	0.4	0.4	0.1	0.1
G	0.1	0.2	0.3	0.4
T	0.1	0.1	0.4	0.4

Figure 3-2: Example of 1st order Markov chain for a DNA sequence

For example, if X = 'ATGC', then the probability of seeing this 4-mer given the 1st order Markov chain model in Figure 3-2 will be calculated as follows,

$$P('ATGC') = p(x_1)[\ p(x_2 \mid x_1) \times p(x_3 \mid x_2) \times p(x_4 \mid x_3)\ ]$$

$$P('ATGC') = p('A')[\ p('T' \mid 'A') \times p('G' \mid 'T') \times p('C' \mid 'G')\ ]$$

$$P('ATGC') = 0.25\ [0.3 \times 0.4 \times 0.2\ ] = 0.006$$

And if we assume that the length of the target sequence (L) is 1000, then the expected count $C_{expected}$ will be 1000× 0.006 which is 6.

Using first-order Markov chain allows us to capture any dinucleotide biases, for example, CG content. Also, given that the average length of a bacterial gene is ~900, there may not be long enough to train higher-order chain models

To quantify the likelihood of one miRNA to target a specific target gene, we calculate P_{value}. Assuming a binomial distribution, P_{value} is the probability of having at least C_{actual} binding sites and is calculated by the following formula,

$$P_{\text{value}} = \sum_{i=C_{actual}}^{L-k+1} \binom{L-k+1}{i} P(X)^i (1-P(X))^{L-k+1-i}$$

3.4 Results

3.4.1 Identification of potential gut bacterial genes targeted by ginger miRNAs

Bacterial mRNAs potentially targeted by ginger miRNAs were identified by enrichment analysis of the reverse complement of the miRNA seed sequence (8mer = nucleotides 1 – 8 from 5' end) in the coding sequence (CDS). The enrichment analysis adopted a framework that utilizes the 1st order Markov model (MM). In this framework, the observed 8mer count in the CDS region of each bacterial mRNA was compared against the background count derived from the 1st order Markov chain model. A p-value was calculated for each miRNA-mRNA pair to estimate the likelihood of having a functional pair. Once all p-values were calculated, the false discovery rate (FDR) was obtained using the Benjamini–Hochberg method [57] for multiple p-value correction.

Table 3-3: Number of potential genes targeted by ginger miRNAs

Bacterial strain	Number of potential target genes	Percentage of total genes (%)	GC content (%)
Akkermansia muciniphila	677	30.0	55.8
Bacteroides fragilis	1088	26.5	43.1
Clostridium perfringens	320	11.2	28.4
Escherichia coli	1168	26.7	50.8
Enterococcus faecalis	866	32.5	37.5
Helicobacter pylori	372	24.5	38.8
Lactobacillus rhamnosus GG	726	28.1	47.0
Lactobacillus ruminis	429	21.3	43.3

Table 3-3 shows that hundreds of bacterial genes are potentially targeted by ginger miRNAs. The percentage of potentials targets varies for different bacterial strains and ranges from 11.2% for *Clostridium perfringens* to 32.5% for *Enterococcus faecalis*. To see if these variations are related to the nucleotide composition of different bacterial genomes, we calculated GC content of each of these genomes. Some bacterial genomes with high GC content like *Akkermansia muciniphila, Escherichia coli, and Lactobacillus rhamnosus GG* tend to have a higher percentage of potential target genes. On the other hand, bacterial genomes with a low percentage of potential target genes like *Clostridium perfringens* has low GC content.

3.4.2 Case study 1: Ginger miRNA ath-miR167a-5p targets LGG *SpaC* gene

Using the top 49 expressed miRNAs from ginger exosomes, we used our tool to find potential bacterial (LGG strain) genes targeted by these miRNAs. We search for potential binding sites for the miRNA seed region (7mer) in both reverse and forward strands (targeting at the DNA level [3]). Of these genes, LGG pilus gene *SpaC* has a role in the colonization of bacteria into host tissues [58]. Table 3-4 shows potential ginger miRNAs targeting the LGG pilus gene. Only six miRNAs have potential binding sites in SpaC coding sequence. Of these, miR-167a-5p has shown significant targeting (p-value < 0.05) with 2 potential seed binding sites.

Teng et al. [59] showed that *SpaC* gene is downregulated at both transcriptional and protein levels when LGG was treated by ginger exosomes. Also, the authors showed that scrambled miR-167a-5p (with mutations at seed region) did not affect the expression of *SpaC*.

Table 3-4: potential ginger miRNAs targeting pilus gene (SpaC)

miRNA ID	seed (7mer)	# of hits	length of transcript	p-value
ath-miR167a-5p	GAAGCTG	2	2688	0.0118
ppt-miR319a	TTGGACT	1	2688	0.1496
mdm-miR535a	GACAACG	1	2688	0.2078
aly-miR396a-3p	TTCAATA	1	2688	0.2473
gma-miR396a-3p	TCAATAA	1	2688	0.2566
aly-miR166a-5p	GAATGTT	1	2688	0.2664

3.4.3 Case study 2: Ginger miRNAs targeting LGG lexA gene

Teng et. al. [59] showed that lexA gene is downregulated at both transcriptional and protein levels when LGG was treated by ginger exosomes. To find potential ginger miRNAs targeting LGG lexA gene, we searched for potential binding sites for all 13 miRNA seed regions with length equals to 8 nucleotides (with different start positions from the 5' end). We found that only miRNA seeds from *miR396* microRNA family have potential binding sites on LGG *lexA* sequence as shown in Table 3-5.

Table 3-5: potential Ginger miRNAs targeting lexA gene

miRNA ID	Seed (8mer)	miRNA seed start site	Target start site	length of transcript
aly-miR396a-5p	TTTCTTGA	10	427	626
gma-miR396e	TTTCTTGA	10	427	626

aly-miR396b-5p	TTTCTTGA	10	427	626
osa-miR396d	TTTCTTGA	10	427	626
gma-miR396h	TTTCTTGA	9	427	626

3.4.4 Computational complexity analysis

To evaluate the computational complexity of the proposed method, we run our program to search for potential binding sites of nucleotides 10 to 17 (8 nucleotides) from the 5'end of the miRNA sequence in different numbers of human gut bacterial genomes (1 to 8 strains) to see how the program scales with more input sequences. We also used different orders of the Markov chain models (0^{th} to 4^{th} order) to measure the effect of model order on the program performance. Running time was measured using cProfile Python package. For each combination of parameters (number of bacterial genomes and model order), we repeated the experiment 3 times and calculated the average run time. Figure 3-3 shows that running time linearly increases with an increased number of bacterial genomes. Using higher-order Markov chain models to model bacterial sequences slightly increased the run time, 4^{th} order models.

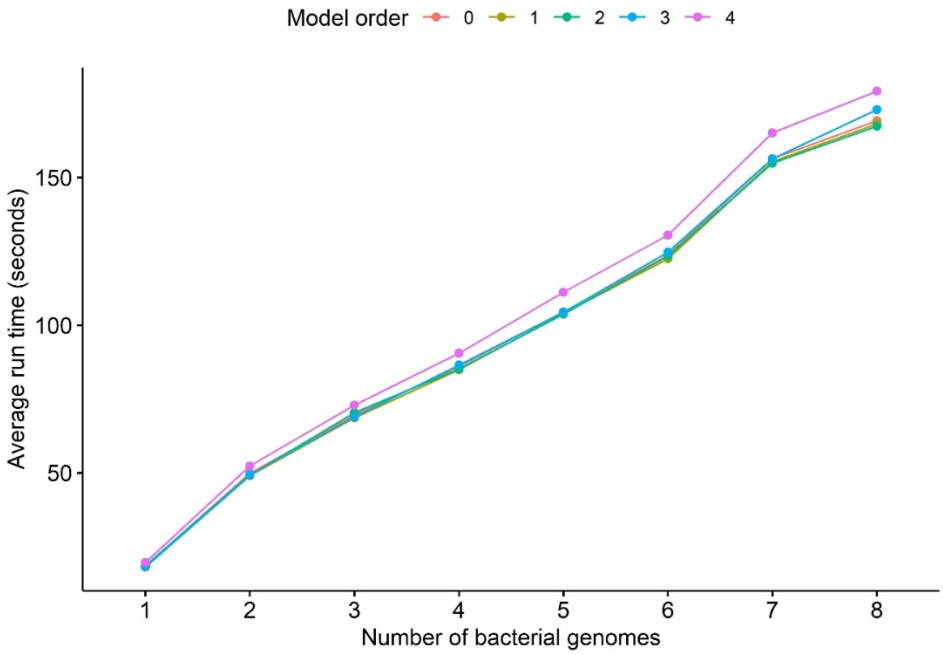

Figure 3-3: Average running time with different number of bacterial genomes and Markov chain model order

We also wanted to measure the effect of the number of input miRNA seeds (query patterns) on the run time of hashing-based pattern searching algorithm. We used different numbers of query patterns (1 to 41 miRNA seeds) and different numbers of target sequences (1 to 6 bacterial genomes). Note that for this experiment, we only included the pattern searching step and excluded other steps (e.g. counting of the number of hits, enrichment analysis using Markov chain models, and writing results to output files). Figure 3-4 shows that the average run time of the hashing-based pattern searching method remains almost constant with increasing the number of miRNAs. On the other hand, the run time of the naïve pattern matching algorithm increases linearly with increasing the number of miRNAs.

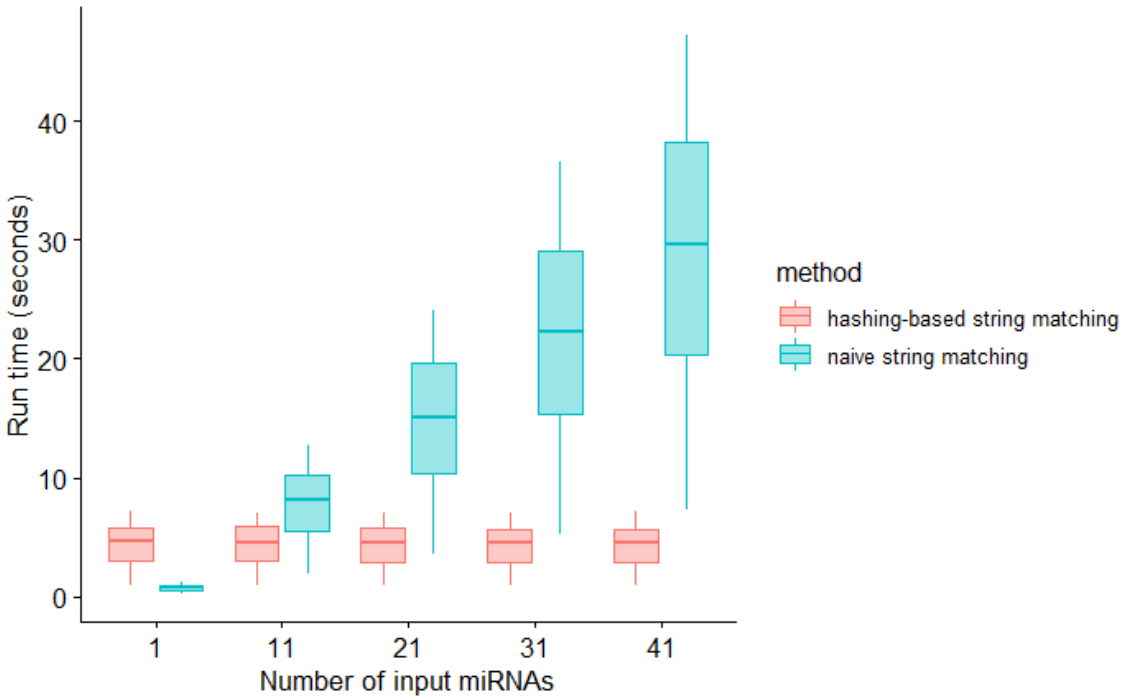

Figure 3-4: hashing-based vs naïve string matching

3.5 Discussion

Our results provide a proof of concept of the use of miRNA seed-based targeting mechanism to predict gut bacterial genes targeted by plant miRNAs. We showed that using our tool we can identify ginger miRNAs experimentally validated to target two LGG genes (*SpaC* and *LexA*). Although, miRNA targeting mechanism is studied in other species in the case of both miRNAs and target genes belong to the same species e.g. plants and animals, to our knowledge there is no known mechanism for cross-species miRNA targeting. Although this framework was used to predict potential mouse bacterial genes targeted by ginger miRNAs, it can be applied to human gut bacteria and miRNAs from any edible plant.

In our framework, we employ a seed-based targeting technique in which we perform an overrepresentation analysis of miRNA seed on the target sequence. This approach was first introduced to predict miRNA targets in humans [60] and also used with viral miRNAs [6]. we search for potential binding sites of miRNA seeds of 7 & 8 nucleotides long. This is based on previous research by Cisse et. al. [28] that suggests at least seven Watson-Crick base pairs in the miRNA seed region are needed for rapid and effective targeting.

We showed that our computational framework can efficiently find potential target genes and the positions of individual binding sites in up to 8 bacterial genomes (more than 23 million nucleotides) in less than 3 minutes. The run time is a linear function of input size, this is because our framework utilizes hash-table and it takes O(n) to build the index of target sequences. But once the index is built, one query takes an O(m) time, where m is the length of the query pattern (miRNA seed in our case), and since m is a constant (usually 7 or 8), the time complexity of one query can be O(1).

Although our proposed framework is promising, it has some limitations. Of these, it only takes into sequence information to predict potential binding sites. Other features like conservation, site accessibility can be utilized to increase the accuracy of our predictions. Also, experimentally validating more targets in bacterial genes is essential to accurately validate our framework and to infer the mechanism by which plant miRNAs target gut bacterial genes.

3.6 Conclusions

In this chapter, we proposed a computational framework to efficiently predict potential bacterial genes targeted by plant miRNA. To decrease the computational complexity of our tool, we employed a hash-table-based index to identify and count potential miRNA binding sites. We employed a Markov chain model to quantify how likely we see these binding sites by chance. Our tool successfully predicted ginger miRNAs that target two experimentally validated LGG target genes. Although our tool can be used to predict potential targeting between other plants and bacterial strains, more wet-lab experiments are needed to infer the inter-kingdom miRNA targeting mechanism.

4 PATHWAY ANALYSIS OF MICRORNA TARGETS IN ANIMALS

4.1 Introduction

Recent studies have suggested that microRNAs (miRNAs) are involved in many diverse biological processes and pathways including normal development and diseases [61]. Animal miRNAs bind to 3'UTR of mRNAs mainly through the short sequence (6-8 NTs) called seed region and act as repressors of target gene expression [62]. Taking into account this short sequence binding, one miRNA can target hundreds or even thousands of genes and subsequently perturb many biological pathways [38].

To computationally predict miRNA-targeted pathways, typically potential target genes are compiled using one or more miRNA target prediction tools and a standard gene enrichment analysis [4] is used to find potential enriched pathways or gene ontology (GO) terms. Although the conventional pipeline is widely used, it has some limitations. Of these, existing tools consider only direct targets (post-transcriptionally regulated) of miRNAs but do not consider indirect targets (transcriptionally regulated) that are not necessarily enriched in miRNA seed-binding sites [5]. Indirect target genes are mainly regulated transcriptionally through transcription factors (TFs) [63, 64]. Furthermore, a study suggested that TFs are preferentially targeted by miRNAs [65].

Figure 4-1 shows a scenario where a biological pathway/process can be missed by classical miRNA pathway analysis tools. Using the classical method, the percentage of

targeted genes is (11%), on the other hand, if we include TF targets (i.e. indirect targets), the percentage of targeted genes will be (67%).

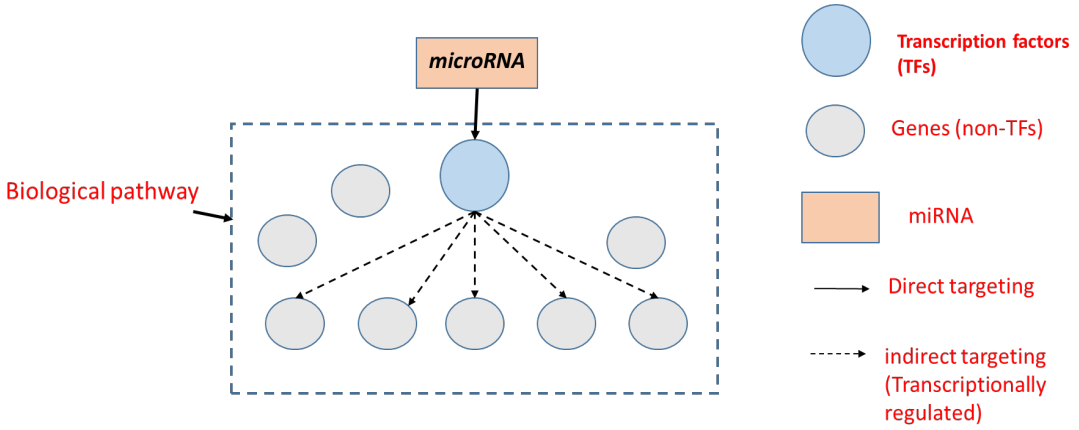

Figure 4-1: miRNAs can indirectly target biological pathways through transcriptions factors

Indirectly regulated genes can be enriched in a specific biological pathway or phenotype. One notable example is the role of miR-200 family and miR-205 in controlling epithelial to mesenchymal transition (EMT) pathway through targeting *ZEB1* and *ZEB2* transcription factors [66]. Other studies have shown how miRNAs regulate cell differentiation through targeting TFs. Of these, Tay et al [67] demonstrated the role of miR-134 in embryonic stem cell differentiation through targeting *Nanog* and *LRH1*. Another study showed that miR-143 and miR-145 can work together to regulate smooth muscle cell differentiation and proliferation through targeting *Klf4* and *ELK1* transcription factors [68].

Several tools and web servers have been developed to predict potential biological pathways targeted by miRNAs [69-73]. While they are all similar in terms of using direct targets only, they use different databases for both miRNA targets and gene ontology annotations [74-78]. miTALOS [70] is the only tool that filters potential targets by

incorporating tissue-specific genes. All tools except for StarBase [71] accept multiple miRNAs as input. A comprehensive comparison of widely used miRNA pathway analysis tools is shown in Table 4-1. In this study, we introduce miRinGO (miRNA indirect target Gene Ontology) that uncovers potential biological pathways affected by indirect targets of human miRNAs especially ones related to cell differentiation and development.

Table 4-1: A comparison of current tools of miRNAs pathway analysis

Features/Tools	mirPath v3.0	StarBase	miTALOS	miRWalk v3.0
Predicted targets databases	TargetScan (v6) / microT-CDS (v5.0)	TargetScan / miRanda/ PITA/ RNA22/ PicTar/…	TargetScan/ miRanda	TargetScan (v7.1) / miRDB
Validated targets databases	TarBase v7.0	CLIP-Seq data	CLIP-Seq data	miRTarbase
Pathways / GO terms databases	KEGG/GO categories	KEGG/GO/ Reactome/ BioCarta	KEGG/ WikiPathways/ Reactome	KEGG/GO/ Reactome
Inclusion of indirect targets?	**No**	**No**	**No**	**No**
Tissue-specific?	No	No	Yes	No
Allows multiple miRNAs?	Yes	No	Yes	Yes

4.2 Methods

4.2.1 Overall pipeline

Our pipeline to predict indirectly targeted biological processes by miRNAs consists of three steps as depicted in Figure 4-2. First, for each miRNA, potential directly-targeted TFs were compiled from the TargetScan database v7.2 [46]. Second, computationally predicted tissue-specific TF-gene associations were collected from the resources' website of (Sonawane et al., 2017) [79]. In the case of multiple input miRNAs, we use the intersection of indirect targets of each miRNA. Third, a hypergeometric test is conducted to find potential targeted biological processes.

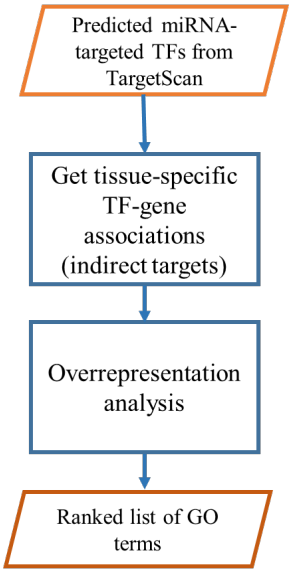

Figure 4-2: Pipeline of our miRNA GO enrichment analysis tool

4.2.2 Input Data

Data used by our tool was compiled from publicly available databases. Putative miRNA targets were downloaded from TargetScan v7.2 [46]. We downloaded the file with all predictions regardless of conservation of miRNA family or miRNA binding sites

and kept high-confidence human miRNA targets (*Cumulative weighted context++ score* < -0.1).

Computationally predicted tissue-specific TF targets were downloaded from the resources website of (Sonawane et al., 2017) [79]. These TF targets were predicted using the PANDA (Passing Attributes between Networks for Data Assimilation) algorithm [80]. PANDA integrates three complimentary sources of information i.e. TF sequences motif data, protein-protein interactions of TFs, and gene co-expression from Genotype-Tissue Expression (GTEx) RNA-Seq data [81]. It contains TF-gene associations from 38 different tissues/tissue locations. We aggregated TF-gene associations from different locations but belong to the same tissue. We had 29 broad tissues after aggregation.

Gene ontology annotations were downloaded using Ensembl Biomart [82] (version GRCh38). GO terms with less than five genes were removed.

4.2.3 Test dataset

To validate our method, we used a 'gold standard' dataset of miRNAs and their experimentally validated functions (GO terms) [83] from ftp://ftp.ebi.ac.uk/pub/databases/GO/goa/HUMAN/goa_human_rna.gaf. We filtered the dataset to include only high confidence annotations (excluded annotations with "*Inferred from Sequence or Structural Similarity*" (ISS), "*Non-traceable Author Statement*" (NAS), and "*Traceable Author Statement*" (TAS) evidence codes). We also removed annotations with no reference article. To keep only relevant annotations, we removed generic GO terms shared by most miRNAs (e.g. "*miRNA mediated inhibition of translation*", "*gene silencing by miRNA*" and "*gene silencing by RNA*"). Cell/tissue ontology was downloaded from http://www.ontobee.org/listTerms/CL?format=tsv . GO terms with less than five genes were

removed. The filtered dataset consists of 335 pairs of miRNAs and their associated GO terms and is available in Supplementary Table S1.

4.3 Results

4.3.1 MicroRNA indirect vs direct targeting

To test the ability of our methodology to predict functions associated with a miRNA, we used a dataset with 335 known miRNA-GO term pairs. All TargetScan-predicted targets were included in this analysis. For each miRNA-GO term pair, resulting GO terms were ranked by the hypergeometric test *p-value* in ascending order, then rank values were converted to a percentile rank by dividing by the total number of GO terms. Finally, we picked the related GO term with the smallest *p-value* (smallest rank value). Known GO terms predicted by the indirect targeting method have a significantly lower (Wilcox signed-rank test, one-sided *p-value* = 0.002417) rank compared to canonical direct targeting as shown in Figure 4-3.

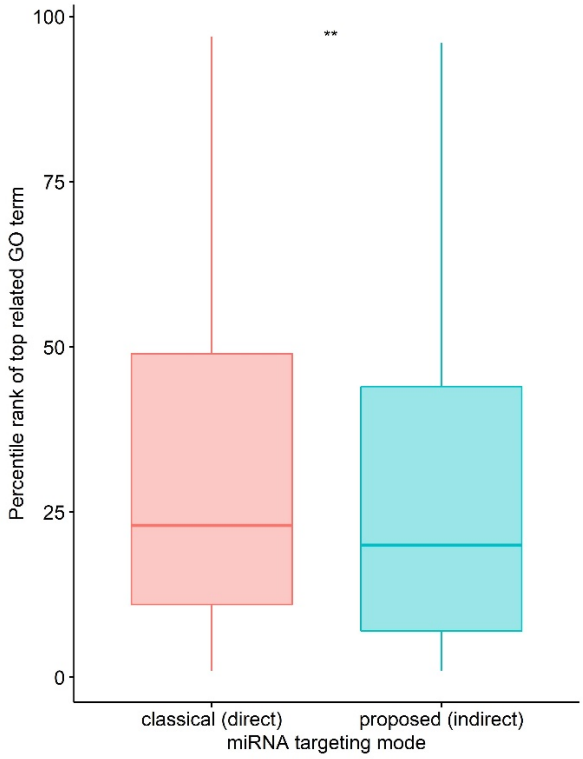

Figure 4-3: Comparison of indirect targeting with direct targeting (** represents *p-value* < *0.01*)

4.3.2 Effect of number of miRNA targets

Since miRNA GO enrichment analysis is affected by targets of the miRNA and TargetScan-predicted targets can have false positives, we investigated the effect of the number of predicted miRNA targets on predicting the known GO terms. We repeated the same analysis but instead of using all predicted targets, we used top (20%, 40%, 60%, 80% and 100%) of potential targets (sorted by TargetScan context++ score [46]). Figure 4-4 shows that in all cases, the average percentile rank of GO terms predicted by IT methodology is lower than DT. Although increasing the number of miRNA targets yielded a lower average rank (better performance), using all of the targets did not give significantly better results compared to using the top 80% of targets and 40% of targets in case of indirect and direct targeting respectively.

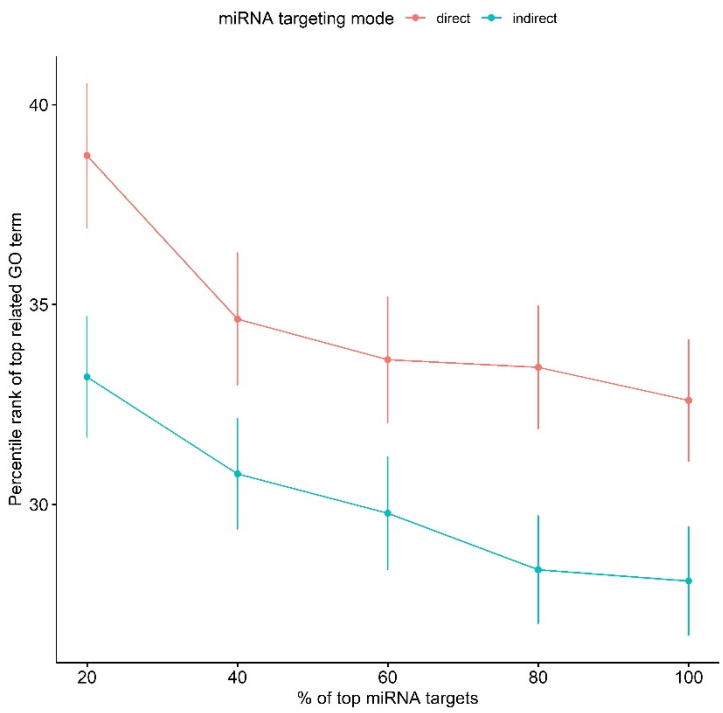

Figure 4-4: Effect of the number of miRNA targets on miRNA GO enrichment analysis. Error bars represent one standard error.

4.3.3 Indirect targeting reveals the role of miRNAs in developmental processes

To investigate biological processes that are more likely to be affected by indirect targeting of miRNAs, we calculated *TF density* per GO term as defined by equation 1.

$$TF\ density = \frac{number\ of\ transcription\ factors\ in\ \ a\ GO\ term}{total\ number\ of\ \ genes\ in\ a\ GO\ term} \quad (1)$$

Table 4-2 shows the top 5 GO BP terms with the highest TF density. All these GO terms are related to the "*developmental process*" and all genes involved are transcription factors.

Table 4-2: Top 5 GO terms with the highest TF density

GO term ID	GO term	# of TFs	# of genes	Parent Process

GO:0001714	endodermal cell fate specification	5	5	developmental process
GO:0003211	cardiac ventricle formation	5	5	developmental process
GO:0003357	noradrenergic neuron differentiation	5	5	developmental process
GO:0021520	spinal cord motor neuron cell fate specification	7	7	developmental process
GO:0021902	commitment of neuronal cell to specific neuron type in forebrain	7	7	developmental process

To see if transcription factors are enriched in development-related GO terms compared to other terms, we divided the GO terms (that have at least one TF) into two groups; one with development-related terms and the second with other terms or processes. We selected development-related terms by searching for GO biological process terms with the following keywords ("*development", "cell fate", "differentiation", "stem cell", "morphogenesis", "cell specification", "formation*"). Figure 4-5 shows that development-related terms (n = 613) tend to have significantly (*p-value* < 2.2e-16, Wilcoxon rank-sum test) higher TF density compared to other terms (n = 1767).

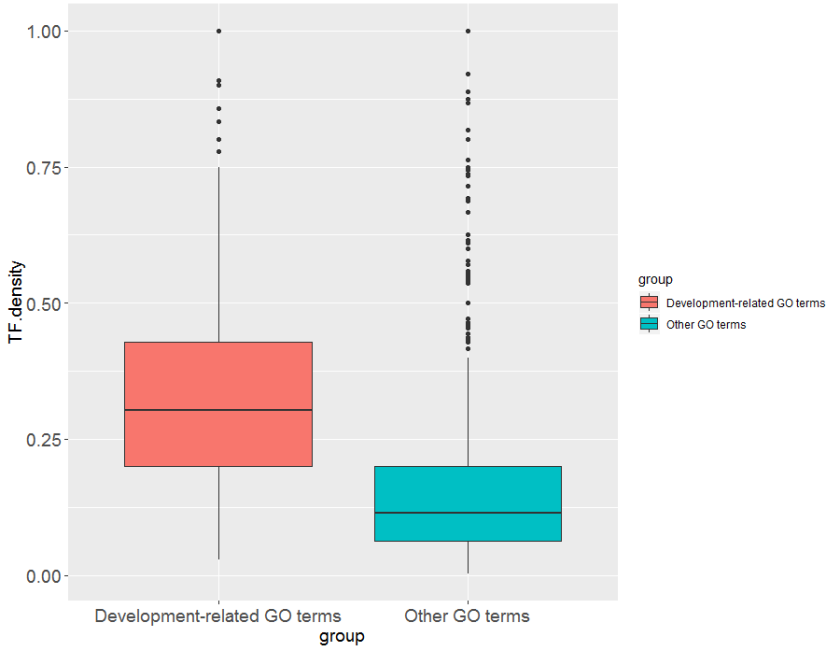

Figure 4-5: Comparison of TF density in development-related GO terms vs. all other terms

4.3.4 Case Study: role of miR-9 in Neurogenesis

To test the ability of our tool to capture relevant targeted development-related GO terms, we picked a miRNA with a known function to be able to compare our predicted GO terms with known ones. Of these miRNAs, miR-9 is a brain-enriched miRNA and has a prominent role in neurogenesis [84-86]. We ran our tool with the following inputs, we selected "brain" as the tissue type, "biological process" as the GO category, "indirect" as the targeting mode, and "100" as the percentage of miRNA targets. Two out of the top 5 GO terms predicted are related to neurogenesis ("*Nervous system development*" and "*brain development*").

To compare our results with existing miRNA pathway analysis tools, we downloaded predicted GO biological process terms for "miR-9" from 4 different web servers (accessed January 15, 2019): mirPath v3.0 [69], StarBase (mirTarPathway

module) v3.0 [71], miTALOS v2 [70] and miRWalk v3.0 [72]. We searched for the highest-ranking GO term related to neurogenesis as shown in Table 4-3. Our tool ranked neurogenesis-related GO terms higher than other tools.

Table 4-3: Comparison of highest-ranking GO terms related to neurogenesis from different miRNA GO enrichment tools

Tool	Highest ranking GO term related to Neurogenesis	Rank
miRinGO	Nervous system development	1
mirPath v3	regulation of neuron maturation	11
miRWalk v3	axonogenesis	13
StarBase v3	Neurogenesis	23
miTALOS v2	N/A	N/A

4.3.5 Multiple miRNAs GO analysis

In all miRNA GO analyses so far, we have used one miRNA as an input. Several studies have shown that miRNAs can work together to regulate certain targets and biological processes [87]. Of these, Gregory et al. [66] showed that the miR-200 family and miR-205 together regulate epithelial to mesenchymal transition (EMT). The miR-200 family consists of miRNAs with two different seed sequences: miR-200a/miR-141 and miR-200b/miR-200c/miR-429. We ran our tool with the following inputs: “kidney” as the tissue type, “biological process” as the GO category, and “indirect” as the targeting mode and “100” as the percentage of miRNA targets. The rank of *“epithelial to mesenchymal transition”* GO term (GO: 0001837) was lower when we used the intersection of indirect

targets of these three miRNAs compared to ranks of GO terms predicted by each miRNA indirect targets as shown in Table 4-4.

Table 4-4: Effect of using multiple miRNAs in capturing EMT-related GO terms

miRNAs	Rank of top GO term related to EMT
miR-200a/miR-141	132
miR-200b/miR-200c/miR-429	147
miR-205-5p	105
All three miRNAs	70

4.3.6 R Shiny Application

For ease of use of our method by researchers, we developed an interactive web application, *miRinGO*, using R Shiny package [7]. It is freely available from GitHub at https://github.com/Fadeel/miRinGO. A screenshot of the application is shown in Figure 4-6.

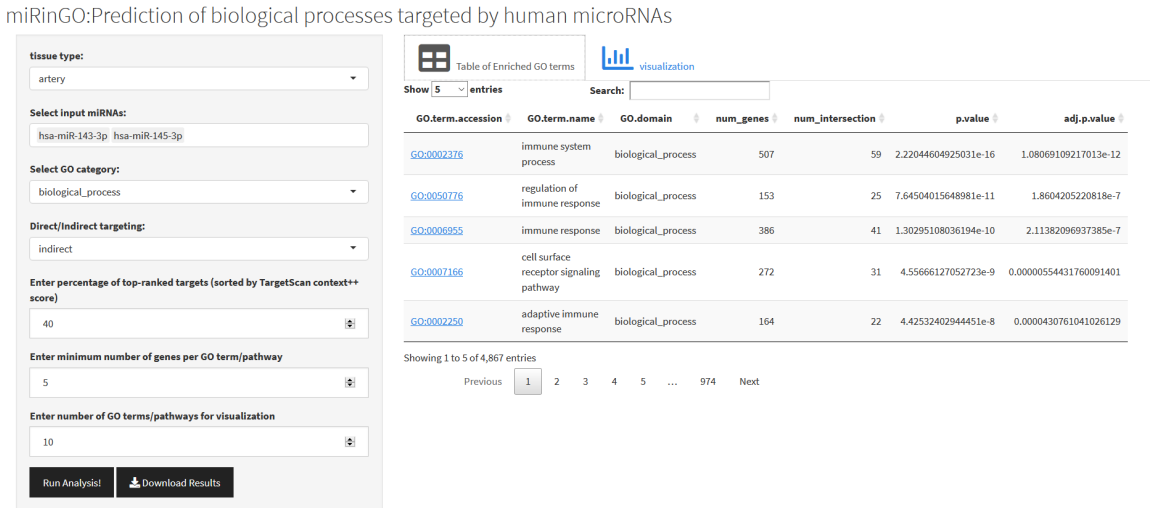

Figure 4-6: A screenshot of the R Shiny application (miRinGO)

The application has two panels, the left one for input data and parameters selection. A description of different input parameters and possible values is detailed in Table 4-5.

Table 4-5: A description of different parameters of the Shiny application

Parameter	Description	Possible values
Tissue type	29 different tissues	*Brain, colon, lung, ..*
input miRNAs	miRBase miRNA ID	*hsa-miR-9-5p*
GO category	Choose the category of GO annotation	*Biological process* OR *cellular component* OR *molecular function*
Targeting mode	Choose mode of miRNA targeting	*Direct* OR *indirect*
Percentage of TargetScan target genes	percentage of top-ranked targeted genes (sorted by TargetScan v7.2 context++ score)	[20% - 100%], step size 20%
minimum number of genes per GO term	A threshold to filter out GO terms with small number of genes	Integer value (default value is 5)
number of GO terms for visualization	Top k enriched GO terms to be visualized	Integer value (default value is 10)

The right panel has two tabs for showing the results of GO enrichment analysis. One tab named “Table of Enriched GO term” to show GO terms more likely to be targeted by selected miRNAs and ranked by the hypergeometric p-value. The table includes information about GO term, number of genes in that term, number of potential miRNA targets that overlap with genes in this term, hypergeometric p-value, and the Benjamini & Hochberg [57] adjusted p-value. To get more details about a specific GO term, GO term accession numbers are linked to their QuickGO [88] webpages.

The other tab named “visualization” provides a visual summary of the enriched GO term. It contains a bar plot with bar height represents the -log10(adjusted p-value) and bar color represents the proportion of potential target genes out of the total number of genes in that GO term. Bar plot of top 15 GO terms indirectly targeted by miR-9-5p in the brain is shown in Figure 4-7.

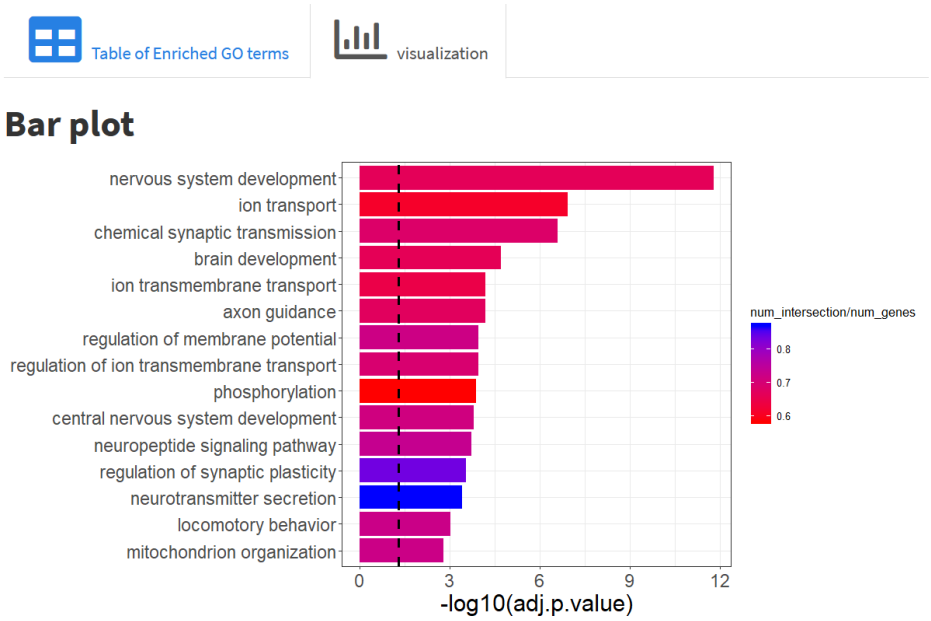

Figure 4-7: Bar plot of top 15 GO terms indirectly targeted by miR-9-5p in brain

miRinGO also provides a visual summary of the top enriched GO terms as WordCloud. To generate the WordCloud, we need to find high-frequency words. First, top enriched GO terms are preprocessed to remove punctuation, numbers, English stop words, and common biological terms (process, cell, miRNA). We used “tm” R package [89] for text preprocessing steps. A snapshot of the WordCloud of top 30 enriched GO terms predicted to be indirectly targeted by miR-16 in the colon is shown in Figure 4-8. We can see that words like mitotic, cycle, segregation, and division are related to cell

cycle. Linsley et. al, had previously shown that overexpression of miR-16 family in HCT116 (human colorectal carcinoma) cell lines regulates cell cycle progression [5].

Figure 4-8: WordCloud of top 30 enriched GO terms predicted to be indirectly targeted by miR-16 in the colon

4.4 Discussion

We propose, miRinGO, a tool that detects biological processes indirectly targeted by miRNAs transcriptionally through transcription factors. Using miRinGO, we can include potential target genes even if there is no physical interaction between miRNA and the regulated genes. To validate this method, we used a dataset of miRNAs and their known targeted GO terms [83]. Although this dataset is considered a significant step towards having a gold standard to validate different miRNA pathway or GO analysis tools, it is still limited to a fraction of human miRNAs and focused more on cardiovascular-related processes. Using this dataset, however, indirect targeting showed better performance in predicting known targeted processes compared to the direct targeting method, even if we use different fractions of input miRNA targets. It is also worth noting that although increasing the number of miRNA targets yielded better

performance, using all of the targets did not give significantly better results compared to using the top 80% of targets and 40% of targets in case of indirect and direct targeting, respectively. This could be because miRNA target prediction tools suffer from having many false positives [90].

Since our method is mainly focused on miRNA-targeted TFs and development-related GO terms or pathways have more TFs than other terms, it is more suitable to use this tool to uncover the tissue-specific roles of miRNAs in development and cell differentiation. Using this method, we predicted biological pathways known to be targeted by miR-9, a miRNA with a known role in neurogenesis. Tan et al. [91] showed that miR-9 regulates neural stem cell differentiation and proliferation by targeting *HES1* transcription factor. Using indirect targeting, three genes related to neuron differentiation (*FEZF2, SOX3,* and *ZHX2*) that are predicted to be targeted by *HES1* (but are not direct targets of miR-9-5p) are now included in GO enrichment analysis as indirect targets of miR-9-5p.

One limitation of our method is that we use two sets of computationally predicted targets: one for miRNA direct targets and the other for tissue-specific TF targets. This might increase the effect of false positives in miRNA GO enrichment analysis. This limitation can be partly alleviated by using only high-confidence miRNA targets (i.e. ones with smaller TargetScan *context*++ score). Although our method outperformed the current miRNA GO analysis method, it is not intended to replace the standard miRNA GO analysis method but on the other hand, to give a different perspective of miRNA roles in regulating biological processes and to uncover ones that are previously

overlooked by current tools, especially ones related to development and cell differentiation.

4.5 Conclusions

Although miRNAs have emerged as important players in gene regulation, the field still lacks computational tools and methods that can predict their functions accurately. We propose a method to predict biological processes that are indirectly targeted by miRNAs transcriptionally through transcription factors. The proposed method provided better performance compared to the existing method. We also developed an interactive web application to make it easier for researchers to investigate the function of miRNA(s) of choice.

5 SUMMARY AND FUTURE WORK

Given the recent advances in DNA/RNA sequencing, new genes/transcripts are being discovered and the need to assign functions to these genes grows. Of these genes, miRNAs are non-protein-coding genes and act as regulators of gene expression of other genes. Recent studies have suggested that miRNAs have a role in many biological processes e.g., cell cycle, proliferation, and development. Moreover, miRNAs may function as tumor suppressors and oncogene. Current experimental techniques for discovering miRNA target genes are either low throughput, costly, or suffer from high false-positive rates. For the above reasons, computational tools are needed to predict targeted genes and ultimately targeted biological processes.

In the first part of this dissertation, we developed a tool to predict gut bacterial genes potentially targeted by miRNAs from edible plants like ginger. To our knowledge, it is the only known tool dedicated to inter-kingdom miRNA targeting. Although we were able to correctly predict ginger miRNAs targeting two bacterial genes, there is still a need in the future for more experimentally validated targets to get a better idea regarding inter-kingdom targeting mechanisms.

Although this tool can find potential target genes of 49 miRNAs in eight different bacterial genomes in less than three minutes, there is still room to improve the scalability of this tool in the future. For instance, we use an integer representation of miRNA seed

sequences (8 nucleotides) which requires 28 bytes in Python 3. This can be reduced by taking advantage of the fact that DNA/RNA sequences have 4 letters, and each nucleotide can be represented by 2 bits. Using this representation, miRNA seed regions can be represented as a vector of 16 bits (2 bytes).

In the second part, we addressed the problem of functional annotation of human miRNAs. Current tools consider only target genes with binding sites. To address this limitation, we developed *miRinGO*, an R Shiny web application that allows researchers to include other indirect target genes e.g., genes regulated through transcription factors. Indirect targeting showed better performance compared to direct targeting in predicting known miRNA-biological process associations. *miRinGO* provides an easy-to-use GUI for researchers with no coding experience to explore potential functions of over 2,000 miRNAs in 38 tissues. Moreover, it provides visual summaries of the results. *miRinGO* is available at https://github.com/Fadeel/miRinGO. Potential future work includes supporting miRNAs from other species e.g. mouse and giving the user the ability to filter target genes based on the existing experimental evidence.

REFERENCES

[1] G. Raposo, and W. Stoorvogel, “Extracellular vesicles: exosomes, microvesicles, and friends,” *J Cell Biol,* vol. 200, no. 4, pp. 373-383, 2013.

[2] H. Zhang, Y. Li, Y. Liu, H. Liu, H. Wang, W. Jin, Y. Zhang, C. Zhang, and D. Xu, “Role of plant MicroRNA in cross-species regulatory networks of humans,” *BMC systems biology,* vol. 10, no. 1, pp. 60, 2016.

[3] S. Liu, A. P. da Cunha, R. M. Rezende, R. Cialic, Z. Wei, L. Bry, L. E. Comstock, R. Gandhi, and H. L. Weiner, “The host shapes the gut microbiota via fecal microRNA,” *Cell host & microbe,* vol. 19, no. 1, pp. 32-43, 2016.

[4] D. W. Huang, B. T. Sherman, and R. A. Lempicki, “Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources,” *Nature protocols,* vol. 4, no. 1, pp. 44, 2008.

[5] P. S. Linsley, J. Schelter, J. Burchard, M. Kibukawa, M. M. Martin, S. R. Bartz, J. M. Johnson, J. M. Cummins, C. K. Raymond, and H. Dai, “Transcripts targeted by the microRNA-16 family cooperatively regulate cell cycle progression,” *Molecular and cellular biology,* vol. 27, no. 6, pp. 2240-2252, 2007.

[6] E. Murphy, J. Vaníček, H. Robins, T. Shenk, and A. J. Levine, “Suppression of immediate-early viral gene expression by herpesvirus-coded microRNAs: implications for latency,” *Proceedings of the National Academy of Sciences,* vol. 105, no. 14, pp. 5453-5458, 2008.

[7] C. J. Chang W, Allaire JJ, Xie Y, McPherson J, "shiny: Web Application Framework for R," *R package* 2019.

[8] B. Alberts, *Molecular Biology of the Cell*: W.W. Norton, 2017.

[9] A. Zien, “A primer on molecular biology,” pp. 3-34, 2004.

[10] C. M. O’Connor, J. U. Adams, and J. J. C. Fairman, MA: NPG Education, “Essentials of cell biology,” vol. 1, pp. 54, 2010.

[11] "Cell Types," https://en.wikipedia.org/wiki/File:Celltypes.svg.

[12] L. J. N. E. Pray, “Discovery of DNA structure and function: Watson and Crick,” vol. 1, no. 1, 2008.

[13] "Protein function," https://www.nature.com/scitable/topicpage/protein-function-14123348/.

[14] F. H. Crick, "On protein synthesis." p. 8.

[15] S. Clancy, and W. Brown, “Translation: DNA to mRNA to,” *Nature Education*, 2008.

[16] L. J. A. i. Hunter, and m. biology, “Molecular biology for computer scientists,” pp. 1-46, 1993.

[17] "What is Epigenetics?," https://www.cdc.gov/genomics/disease/epigenetics.htm.

[18] "Epigenetic mechanisms," https://commons.wikimedia.org/w/index.php?title=File:Epigenetic_mechanisms.jpg&oldid=508703722.
[19] D. S. Latchman, “Transcription factors: an overview,” vol. 29, no. 12, pp. 1305-1312, 1997.
[20] "Transcription Factors," https://en.wikipedia.org/wiki/File:Transcription_Factors.svg.
[21] D. Most, L. Ferguson, and R. A. Harris, "Chapter 6 - Molecular basis of alcoholism," *Handbook of Clinical Neurology*, E. V. Sullivan and A. Pfefferbaum, eds., pp. 89-111: Elsevier, 2014.
[22] "Alternative Splicing," 03/06, 2021; https://en.wikipedia.org/wiki/File:DNA_alternative_splicing.gif.
[23] R. C. Lee, R. L. Feinbaum, and V. J. c. Ambros, “The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14,” vol. 75, no. 5, pp. 843-854, 1993.
[24] B. Wightman, I. Ha, and G. J. C. Ruvkun, “Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans,” vol. 75, no. 5, pp. 855-862, 1993.
[25] J. O'Brien, H. Hayder, Y. Zayed, and C. J. F. i. e. Peng, “Overview of microRNA biogenesis, mechanisms of actions, and circulation,” vol. 9, pp. 402, 2018.
[26] J. G. Ruby, C. H. Jan, and D. P. J. N. Bartel, “Intronic microRNA precursors that bypass Drosha processing,” vol. 448, no. 7149, pp. 83-86, 2007.
[27] "File:MicroRNA biogenesis.jpg," https://commons.wikimedia.org/wiki/File:MicroRNA_biogenesis.jpg.
[28] I. I. Cisse, H. Kim, and T. Ha, “A rule of seven in Watson-Crick base-pairing of mismatched sequences,” *Nature structural & molecular biology,* vol. 19, no. 6, pp. 623-627, 2012.
[29] L. M. Wee, C. F. Flores-Jasso, W. E. Salomon, and P. D. Zamore, “Argonaute divides its RNA guide into domains with distinct functions and RNA-binding properties,” *Cell,* vol. 151, no. 5, pp. 1055-1067, 2012.
[30] W. E. Salomon, S. M. Jolly, M. J. Moore, P. D. Zamore, and V. Serebrov, “Single-molecule imaging reveals that argonaute reshapes the binding properties of its nucleic acid guides,” *Cell,* vol. 162, no. 1, pp. 84-95, 2015.
[31] J. P. Broughton, M. T. Lovci, J. L. Huang, G. W. Yeo, and A. E. Pasquinelli, “Pairing beyond the seed supports microRNA targeting specificity,” *Molecular cell,* vol. 64, no. 2, pp. 320-333, 2016.
[32] S. D. Chandradoss, N. T. Schirle, M. Szczepaniak, I. J. MacRae, and C. Joo, “A dynamic search process underlies microRNA targeting,” *Cell,* vol. 162, no. 1, pp. 96-107, 2015.
[33] T. Saito, and P. Sætrom, “MicroRNAs–targeting and target prediction,” *New biotechnology,* vol. 27, no. 3, pp. 243-249, 2010.
[34] X. Li, H. Kazan, H. D. Lipshitz, and Q. D. Morris, “Finding the target sites of RNA-binding proteins,” *Wiley Interdisciplinary Reviews: RNA,* vol. 5, no. 1, pp. 111-130, 2014.

[35] R. Aguirre-Hernández, H. H. Hoos, and A. Condon, "Computational RNA secondary structure design: empirical complexity and improved methods," *BMC bioinformatics,* vol. 8, no. 1, pp. 34, 2007.
[36] T. Bradley, and S. Moxon, "An Assessment of the Next Generation of Animal miRNA Target Prediction Algorithms," *MicroRNA Detection and Target Identification: Methods and Protocols*, pp. 175-191, 2017.
[37] X. Fan, and L. Kurgan, "Comprehensive overview and assessment of computational prediction of microRNA targets in animals," *Briefings in bioinformatics,* vol. 16, no. 5, pp. 780-794, 2014.
[38] R. C. Friedman, K. K.-H. Farh, C. B. Burge, and D. P. Bartel, "Most mammalian mRNAs are conserved targets of microRNAs," *Genome research,* vol. 19, no. 1, pp. 92-105, 2009.
[39] D. Grün, Y.-L. Wang, D. Langenberger, K. C. Gunsalus, and N. Rajewsky, "microRNA target predictions across seven Drosophila species and comparison to mammalian targets," *PLoS computational biology,* vol. 1, no. 1, pp. e13, 2005.
[40] M. D. Paraskevopoulou, G. Georgakilas, N. Kostoulas, I. S. Vlachos, T. Vergoulis, M. Reczko, C. Filippidis, T. Dalamagas, and A. G. Hatzigeorgiou, "DIANA-microT web server v5. 0: service integration into miRNA functional analysis workflows," *Nucleic acids research,* vol. 41, no. W1, pp. W169-W173, 2013.
[41] D. Betel, A. Koppal, P. Agius, C. Sander, and C. Leslie, "Comprehensive modeling of microRNA targets predicts functional non-conserved and non-canonical sites," *Genome biology,* vol. 11, no. 8, pp. R90, 2010.
[42] D. Gaidatzis, E. van Nimwegen, J. Hausser, and M. Zavolan, "Inference of miRNA targets using evolutionary conservation and pathway analysis," *BMC bioinformatics,* vol. 8, no. 1, pp. 69, 2007.
[43] X. Wang, and I. M. El Naqa, "Prediction of both conserved and nonconserved microRNA targets in animals," *Bioinformatics,* vol. 24, no. 3, pp. 325-332, 2007.
[44] C. E. Vejnar, and E. M. Zdobnov, "MiRmap: comprehensive prediction of microRNA target repression strength," *Nucleic acids research,* vol. 40, no. 22, pp. 11673-11683, 2012.
[45] A. C. Oliveira, L. A. Bovolenta, P. G. Nachtigall, M. E. Herkenhoff, N. Lemke, and D. Pinhal, "Combining results from distinct microRNA target prediction tools enhances the performance of analyses," *Frontiers in genetics,* vol. 8, 2017.
[46] V. Agarwal, G. W. Bell, J.-W. Nam, and D. P. Bartel, "Predicting effective microRNA target sites in mammalian mRNAs," *elife,* vol. 4, pp. e05005, 2015.
[47] M. Kertesz, N. Iovino, U. Unnerstall, U. Gaul, and E. Segal, "The role of site accessibility in microRNA target recognition," *Nature genetics,* vol. 39, no. 10, pp. 1278-1284, 2007.
[48] K. C. Miranda, T. Huynh, Y. Tay, Y.-S. Ang, W.-L. Tam, A. M. Thomson, B. Lim, and I. Rigoutsos, "A pattern-based method for the identification of MicroRNA binding sites and their corresponding heteroduplexes," *Cell,* vol. 126, no. 6, pp. 1203-1217, 2006.
[49] R. Gumienny, and M. Zavolan, "Accurate transcriptome-wide prediction of microRNA targets and small interfering RNA off-targets with MIRZA-G," *Nucleic acids research,* vol. 43, no. 3, pp. 1380-1391, 2015.

[50] H. Liang, K. Zen, J. Zhang, C.-Y. Zhang, and X. Chen, "New roles for microRNAs in cross-species communication," *RNA biology,* vol. 10, no. 3, pp. 367-370, 2013.
[51] W. D. Branton, E. Mayeri, P. Brownell, and S. B. Simon, "Evidence for local hormonal communication between neurones in Aplysia," *Nature,* vol. 274, no. 5666, pp. 70-72, 1978.
[52] L. Zhang, D. Hou, X. Chen, D. Li, L. Zhu, Y. Zhang, J. Li, Z. Bian, X. Liang, and X. Cai, "Exogenous plant MIR168a specifically targets mammalian LDLRAP1: evidence of cross-kingdom regulation by microRNA," *Cell research,* vol. 22, no. 1, pp. 107-126, 2012.
[53] G. LaMonte, N. Philip, J. Reardon, J. R. Lacsina, W. Majoros, L. Chapman, C. D. Thornburg, M. J. Telen, U. Ohler, and C. V. Nicchitta, "Translocation of sickle cell erythrocyte microRNAs into Plasmodium falciparum inhibits parasite translation and contributes to malaria resistance," *Cell host & microbe,* vol. 12, no. 2, pp. 187-199, 2012.
[54] E. Gottwein, N. Mukherjee, C. Sachse, C. Frenzel, W. H. Majoros, J.-T. A. Chi, R. Braich, M. Manoharan, J. Soutschek, and U. Ohler, "A viral microRNA functions as an orthologue of cellular miR-155," *Nature,* vol. 450, no. 7172, pp. 1096-1099, 2007.
[55] C. L. Jopling, M. Yi, A. M. Lancaster, S. M. Lemon, and P. Sarnow, "Modulation of hepatitis C virus RNA abundance by a liver-specific MicroRNA," *science,* vol. 309, no. 5740, pp. 1577-1581, 2005.
[56] K. D. Pruitt, T. Tatusova, and D. R. Maglott, "NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins," *Nucleic acids research,* vol. 35, no. suppl 1, pp. D61-D65, 2007.
[57] Y. Benjamini, and Y. J. J. o. t. R. s. s. s. B. Hochberg, "Controlling the false discovery rate: a practical and powerful approach to multiple testing," vol. 57, no. 1, pp. 289-300, 1995.
[58] R. Lowry, S. Balboa, J. L. Parker, and J. G. Shaw, "Aeromonas flagella and colonisation mechanisms," *Advances in microbial physiology*, pp. 203-256: Elsevier, 2014.
[59] Y. Teng, Y. Ren, M. Sayed, X. Hu, C. Lei, A. Kumar, E. Hutchins, J. Mu, Z. Deng, and C. Luo, "Plant-derived exosomal MicroRNAs shape the gut microbiota," *Cell host & microbe,* vol. 24, no. 5, pp. 637-652. e8, 2018.
[60] H. Robins, and W. H. J. P. o. t. N. A. o. S. Press, "Human microRNAs target a functionally distinct population of genes with AT-rich 3′ UTRs," vol. 102, no. 43, pp. 15557-15562, 2005.
[61] J. A. Vidigal, and A. Ventura, "The biological functions of miRNAs: lessons from in vivo studies," *Trends in cell biology,* vol. 25, no. 3, pp. 137-147, 2015.
[62] Y. Tomari, and P. D. Zamore, "Perspective: machines for RNAi," *Genes Dev,* vol. 19, no. 5, pp. 517-29, Mar 1, 2005.
[63] C. P. Bracken, H. S. Scott, and G. J. Goodall, "A network-biology perspective of microRNA function and dysfunction in cancer," *Nature Reviews Genetics,* vol. 17, no. 12, pp. 719, 2016.
[64] S. J. Gosline, A. M. Gurtan, C. K. JnBaptiste, A. Bosson, P. Milani, S. Dalin, B. J. Matthews, Y. S. Yap, P. A. Sharp, and E. J. C. r. Fraenkel, "Elucidating

microRNA regulatory networks using transcriptional, post-transcriptional, and histone modification measurements," vol. 14, no. 2, pp. 310-319, 2016.
[65] R. Shalgi, D. Lieber, M. Oren, and Y. J. P. c. b. Pilpel, "Global and local architecture of the mammalian microRNA–transcription factor regulatory network," vol. 3, no. 7, pp. e131, 2007.
[66] P. A. Gregory, A. G. Bert, E. L. Paterson, S. C. Barry, A. Tsykin, G. Farshid, M. A. Vadas, Y. Khew-Goodall, and G. J. J. N. c. b. Goodall, "The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1," vol. 10, no. 5, pp. 593-601, 2008.
[67] Y. Tay, J. Zhang, A. M. Thomson, B. Lim, and I. J. N. Rigoutsos, "MicroRNAs to Nanog, Oct4 and Sox2 coding regions modulate embryonic stem cell differentiation," vol. 455, no. 7216, pp. 1124-1128, 2008.
[68] K. R. Cordes, N. T. Sheehy, M. P. White, E. C. Berry, S. U. Morton, A. N. Muth, T.-H. Lee, J. M. Miano, K. N. Ivey, and D. J. N. Srivastava, "miR-145 and miR-143 regulate smooth muscle cell fate and plasticity," vol. 460, no. 7256, pp. 705, 2009.
[69] I. S. Vlachos, K. Zagganas, M. D. Paraskevopoulou, G. Georgakilas, D. Karagkouni, T. Vergoulis, T. Dalamagas, and A. G. Hatzigeorgiou, "DIANA-miRPath v3.0: deciphering microRNA function with experimental support," *Nucleic Acids Res,* vol. 43, no. W1, pp. W460-6, Jul 1, 2015.
[70] M. Preusse, F. J. Theis, and N. S. Mueller, "miTALOS v2: analyzing tissue specific microRNA function," *PLoS One,* vol. 11, no. 3, pp. e0151771, 2016.
[71] J. H. Li, S. Liu, H. Zhou, L. H. Qu, and J. H. Yang, "starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data," *Nucleic Acids Res,* vol. 42, no. Database issue, pp. D92-7, Jan, 2014.
[72] C. Sticht, C. De La Torre, A. Parveen, and N. Gretz, "miRWalk: An online resource for prediction of microRNA binding sites," *PLoS One,* vol. 13, no. 10, pp. e0206239, 2018.
[73] I. Ulitsky, L. C. Laurent, and R. J. N. a. r. Shamir, "Towards computational prediction of microRNA function and activity," vol. 38, no. 15, pp. e160-e160, 2010.
[74] M. Kanehisa, and S. Goto, "KEGG: kyoto encyclopedia of genes and genomes," *Nucleic acids research,* vol. 28, no. 1, pp. 27-30, 2000.
[75] A. Fabregat, S. Jupe, L. Matthews, K. Sidiropoulos, M. Gillespie, P. Garapati, R. Haw, B. Jassal, F. Korninger, and B. May, "The reactome pathway knowledgebase," *Nucleic acids research,* vol. 46, no. D1, pp. D649-D655, 2017.
[76] D. N. Slenter, M. Kutmon, K. Hanspers, A. Riutta, J. Windsor, N. Nunes, J. Mélius, E. Cirillo, S. L. Coort, and D. Digles, "WikiPathways: a multifaceted pathway database bridging metabolomics to other omics research," *Nucleic acids research,* vol. 46, no. D1, pp. D661-D667, 2017.
[77] D. Nishimura, "BioCarta," *Biotech Software & Internet Report: The Computer Software Journal for Scient,* vol. 2, no. 3, pp. 117-120, 2001.
[78] M. Ashburner, C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, and J. T. Eppig, "Gene ontology: tool for the unification of biology," *Nature genetics,* vol. 25, no. 1, pp. 25, 2000.

[79] A. R. Sonawane, J. Platig, M. Fagny, C.-Y. Chen, J. N. Paulson, C. M. Lopes-Ramos, D. L. DeMeo, J. Quackenbush, K. Glass, and M. L. J. C. r. Kuijjer, "Understanding tissue-specific gene regulation," vol. 21, no. 4, pp. 1077-1088, 2017.
[80] K. Glass, C. Huttenhower, J. Quackenbush, and G.-C. J. P. o. Yuan, "Passing messages between biological networks to refine predicted interactions," vol. 8, no. 5, pp. e64832, 2013.
[81] M. Uhlén, L. Fagerberg, B. M. Hallström, C. Lindskog, P. Oksvold, A. Mardinoglu, Å. Sivertsson, C. Kampf, E. Sjöstedt, and A. Asplund, "Tissue-based map of the human proteome," *Science,* vol. 347, no. 6220, pp. 1260419, 2015.
[82] D. R. Zerbino, P. Achuthan, W. Akanni, M. R. Amode, D. Barrell, J. Bhai, K. Billis, C. Cummins, A. Gall, and C. G. Girón, "Ensembl 2018," *Nucleic acids research,* vol. 46, no. D1, pp. D754-D761, 2017.
[83] R. P. Huntley, B. Kramarz, T. Sawford, Z. Umrao, A. Kalea, V. Acquaah, M. J. Martin, M. Mayr, and R. C. Lovering, "Expanding the horizons of microRNA bioinformatics," *RNA,* vol. 24, no. 8, pp. 1005-1017, 2018.
[84] M. Coolen, S. Katz, and L. Bally-Cuif, "miR-9: a versatile regulator of neurogenesis," *Frontiers in cellular neuroscience,* vol. 7, pp. 220, 2013.
[85] R. Madelaine, S. A. Sloan, N. Huber, J. H. Notwell, L. C. Leung, G. Skariah, C. Halluin, S. P. Paşca, G. Bejerano, and M. A. Krasnow, "MicroRNA-9 couples brain neurogenesis and angiogenesis," *Cell reports,* vol. 20, no. 7, pp. 1533-1542, 2017.
[86] K. N. Ivey, and D. Srivastava, "MicroRNAs as regulators of differentiation and cell fate decisions," *Cell stem cell,* vol. 7, no. 1, pp. 36-41, 2010.
[87] C. P. Bracken, H. S. Scott, and G. J. J. N. R. G. Goodall, "A network-biology perspective of microRNA function and dysfunction in cancer," vol. 17, no. 12, pp. 719-732, 2016.
[88] D. Binns, E. Dimmer, R. Huntley, D. Barrell, C. O'donovan, and R. J. B. Apweiler, "QuickGO: a web-based tool for Gene Ontology searching," vol. 25, no. 22, pp. 3045-3046, 2009.
[89] D. Meyer, K. Hornik, and I. J. J. o. s. s. Feinerer, "Text mining infrastructure in R," vol. 25, no. 5, pp. 1-54, 2008.
[90] N. Pinzón, B. Li, L. Martinez, A. Sergeeva, J. Presumey, F. Apparailly, and H. J. G. r. Seitz, "microRNA target prediction programs predict many false positives," vol. 27, no. 2, pp. 234-245, 2017.
[91] S. L. Tan, T. Ohtsuka, A. González, and R. Kageyama, "Micro RNA 9 regulates neural stem cell differentiation by controlling H es1 expression dynamics in the developing brain," *Genes to cells,* vol. 17, no. 12, pp. 952-961, 2012.

APPENDIX I

miRNA ID	mature sequence
gma-miR319a	TTGGACTGAAGGGAGCTCCC
aly-miR319a-3p	TTGGACTGAAGGGAGCTCCCT
ppt-miR319a	CTTGGACTGAAGGGAGCTCC
ppt-miR319c	CTTGGACTGAAGGGAGCTCCC
aly-miR396a-5p	TTCCACAGCTTTCTTGAACTG
gma-miR396e	TTCCACAGCTTTCTTGAACTGT
gma-miR396h	TCCACAGCTTTCTTGAACTG
ptc-miR319e	TTGGACTGAAGGGAGCTCCT
aly-miR159a-3p	TTTGGATTGAAGGGAGCTCTA
mtr-miR319c-3p	TTGGACTGAAGGGAGCTCCCA
aly-miR166a-3p	TCGGACCAGGCTTCATTCCCC
gma-miR166p	TCGGACCAGGCTTCATTCCC
bdi-miR166f	TCTCGGACCAGGCTTCATTCC
gma-miR166m	CGGACCAGGCTTCATTCCCC
gma-miR166u	TCTCGGACCAGGCTTCATTC
bdi-miR166e-3p	CTCGGACCAGGCTTCATTCCC
gma-miR319p	TTTTGGACTGAAGGGAGCTCC

aly-miR396b-5p	TTCCACAGCTTTCTTGAACTT

miRNA ID	mature sequence
gma-miR6300	GTCGTTGTAGTATAGTGG
sbi-miR166k	TCGGACCAGGCTTCATTCCT
osa-miR396d	TCCACAGGCTTTCTTGAACGG
aly-miR168a-5p	TCGCTTGGTGCAGGTCGGGAA
gma-miR168b	TCGCTTGGTGCAGGTCGGG
aly-miR162a-3p	TCGATAAACCTCTGCATCCAG
gma-miR162a	TCGATAAACCTCTGCATCCA
aly-miR858-5p	TTTCGTTGTCTGTTCGACCTT
ath-miR858b	TTCGTTGTCTGTTCGACCTTG
zma-miR396g-3p	GTTCAAGAAAGCTGTGGAAGA
gma-miR4995	AGGCAGTGGCTTGGTTAAGGG
mtr-miR166c	TCGGACCAGGCTTCATTCCTC
aly-miR156a-5p	TGACAGAAGAGAGTGAGCAC
aly-miR396a-3p	GTTCAATAAAGCTGTGGGAAG
ath-miR156j	TGACAGAAGAGAGAGAGCAC
bdi-miR156a	TGACAGAAGAGAGAGAGCACA
gma-miR156f	TTGACAGAAGAGAGAGAGCACA
gma-miR396a-3p	TTCAATAAAGCTGTGGGAAG

miRNA ID	mature sequence
mdm-miR156t	TTGACAGAAGAGAGAGAGCAC
mdm-miR535a	TGACAACGAGAGAGAGCACGC
osa-miR396a-3p	GTTCAATAAAGCTGTGGGAA
ptc-miR156k	TGACAGAAGAGAGGGAGCAC
stu-miR156f-5p	CTGACAGAAGAGAGTGAGCA
aly-miR167a-5p	TGAAGCTGCCAGCATGATCTA
aly-miR167d-5p	TGAAGCTGCCAGCATGATCTGG
bdi-miR398a	TGTGTTCTCAGGTCGCCCCTG
gma-miR167c	TGAAGCTGCCAGCATGATCTG
aly-miR157a-5p	TTGACAGAAGATAGAGAGCAC
aly-miR157d-5p	TGACAGAAGATAGAGAGCAC
aly-miR164a-5p	TGGAGAAGCAGGGCACGTGCA
aly-miR166a-5p	GGAATGTTGTCTGGCTCGAGG

APPENDIX II

Tissue name	miRNA	GO term name	GO ID	reference	Cell IDs	Cell types
artery	hsa-let-7b-5p	positive regulation of angiogenesis	GO:0045766	PMID:28159509	CL:0000071	blood vessel endothelial cell
artery	hsa-let-7e-5p	negative regulation of DNA-binding transcription factor activity	GO:0043433	PMID:30670152	CL:0002618	endothelial cell of umbilical vein
artery	hsa-let-7e-5p	negative regulation of tyrosine phosphorylation of STAT protein	GO:0042532	PMID:30670152	CL:0002618	endothelial cell of umbilical vein
artery	hsa-let-7f-5p	negative regulation of transforming growth factor beta receptor signaling pathway	GO:0030512	PMID:28345812	CL:0002618	endothelial cell of umbilical vein
artery	hsa-let-7g-5p	negative regulation of inflammatory response	GO:0050728	PMID:24291274	CL:0002618	endothelial cell of umbilical vein
artery	hsa-let-7g-5p	negative regulation of pathway-restricted SMAD protein phosphorylation	GO:0060394	PMID:24291274	CL:0002618	endothelial cell of umbilical vein
artery	hsa-let-7g-5p	negative regulation of transforming growth factor beta receptor signaling pathway	GO:0030512	PMID:24291274	CL:0002618	endothelial cell of umbilical vein
artery	hsa-let-7g-5p	positive regulation of angiogenesis	GO:0045766	PMID:24291274	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-1-3p	positive regulation of vascular associated smooth muscle cell apoptotic process	GO:1905461	PMID:26166810	CL:0002539	aortic smooth muscle cell
artery	hsa-miR-101-3p	negative regulation of protein ubiquitination	GO:0031397	PMID:24844779	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-106b-5p	negative regulation of angiogenesis	GO:0016525	PMID:26956882	CL:0000071	blood vessel endothelial cell
artery	hsa-miR-10a-5p	positive regulation of blood vessel endothelial cell proliferation involved in sprouting angiogenesis	GO:1903589	PMID:22955733	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-10a-5p	positive regulation of cell migration involved in sprouting angiogenesis	GO:0090050	PMID:22955733	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-10a-5p	positive regulation of vascular endothelial growth factor receptor signaling pathway	GO:0030949	PMID:22955733	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-10b-5p	positive regulation of blood vessel endothelial cell proliferation involved in sprouting angiogenesis	GO:1903589	PMID:22955733	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-10b-5p	positive regulation of cell migration involved in sprouting angiogenesis	GO:0090050	PMID:22955733	CL:0002618	endothelial cell of umbilical vein

artery	hsa-miR-10b-5p	positive regulation of vascular endothelial growth factor receptor signaling pathway	GO:0030949	PMID:22955733	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-1224-5p	negative regulation of Notch signaling pathway	GO:0045746	PMID:28717225	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-1224-5p	positive regulation of sprouting angiogenesis	GO:1903672	PMID:28717225	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-1224-5p	positive regulation of vascular endothelial growth factor receptor signaling pathway	GO:0030949	PMID:28717225	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-124-3p	cellular response to hypoxia	GO:0071456	PMID:23853098	CL:0002591	smooth muscle cell of the pulmonary artery
artery	hsa-miR-124-3p	negative regulation of protein dephosphorylation	GO:0035308	PMID:23853098	CL:0002591	smooth muscle cell of the pulmonary artery
artery	hsa-miR-125a-5p	negative regulation of angiogenesis	GO:0016525	PMID:25116893	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-125a-5p	negative regulation of DNA-binding transcription factor activity	GO:0043433	PMID:30670152	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-125a-5p	negative regulation of tyrosine phosphorylation of STAT protein	GO:0042532	PMID:30670152	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-125a-5p	positive regulation of endothelial cell apoptotic process	GO:2000353	PMID:25116893,PMID:28467514	CL:0002618,CL:0002618	endothelial cell of umbilical vein,endothelial cell of umbilical vein
artery	hsa-miR-125a-5p	positive regulation of sprouting angiogenesis	GO:1903672	PMID:27252357	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-125b-5p	negative regulation of angiogenesis	GO:0016525	PMID:22391569	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-126-3p	cellular response to hypoxia	GO:0071456	PMID:28578351	CL:2000008	microvascular endothelial cell
artery	hsa-miR-126-3p	negative regulation of inflammatory response	GO:0050728	PMID:28578351	CL:2000008	microvascular endothelial cell
artery	hsa-miR-126-3p	positive regulation of angiogenesis	GO:0045766	PMID:23136161	CL:0002544	aortic endothelial cell
artery	hsa-miR-126-3p	positive regulation of phosphatidylinositol 3-kinase signaling	GO:0014068	PMID:28578351	CL:2000008	microvascular endothelial cell
artery	hsa-miR-126-3p	positive regulation of protein kinase B signaling	GO:0051897	PMID:28578351	CL:2000008	microvascular endothelial cell
artery	hsa-miR-126-3p	positive regulation of sprouting angiogenesis	GO:1903672	PMID:27780851,PMID:28578351	CL:0002618,CL:2000008	endothelial cell of umbilical vein,microvascular endothelial cell
artery	hsa-miR-126-5p	positive regulation of blood vessel endothelial cell proliferation involved in sprouting angiogenesis	GO:1903589	PMID:28124060	CL:0002618	endothelial cell of umbilical vein

artery	hsa-miR-126-5p	positive regulation of cell migration involved in sprouting angiogenesis	GO:0090050	PMID:28124060	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-130a-3p	positive regulation of angiogenesis	GO:0045766	PMID:23136161	CL:0002544	aortic endothelial cell
artery	hsa-miR-130a-3p	positive regulation of vascular endothelial cell proliferation	GO:1905564	PMID:24960162	UBERON:0002012	pulmonary artery
artery	hsa-miR-130a-3p	positive regulation of vascular smooth muscle cell proliferation	GO:1904707	PMID:24960162	UBERON:0002012	pulmonary artery
artery	hsa-miR-132-3p	positive regulation of angiogenesis	GO:0045766	PMID:21868695	CL:0000071	blood vessel endothelial cell
artery	hsa-miR-132-3p	positive regulation of protein kinase B signaling	GO:0051897	PMID:21868695	CL:0000071	blood vessel endothelial cell
artery	hsa-miR-132-5p	cholesterol homeostasis	GO:0042632	PMID:24924687	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-132-5p	fatty acid homeostasis	GO:0055089	PMID:24924687	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-132-5p	positive regulation of endothelial cell apoptotic process	GO:2000353	PMID:24924687	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-133a-3p	cellular response to cytokine stimulus	GO:0071345	PMID:28257760	CL:0000359	vascular associated smooth muscle cell
artery	hsa-miR-133a-3p	negative regulation of low-density lipoprotein particle clearance	GO:0010989	PMID:28257760	CL:0000359	vascular associated smooth muscle cell
artery	hsa-miR-138-5p	negative regulation of G1/S transition of mitotic cell cycle	GO:2000134	PMID:28450935	CL:0002546	embryonic blood vessel endothelial progenitor cell
artery	hsa-miR-138-5p	negative regulation of nitric-oxide synthase activity	GO:0051001	PMID:24244340	CL:2000008	microvascular endothelial cell
artery	hsa-miR-138-5p	negative regulation of p38MAPK cascade	GO:1903753	PMID:28450935	CL:0002546	embryonic blood vessel endothelial progenitor cell
artery	hsa-miR-138-5p	negative regulation of sprouting angiogenesis	GO:1903671	PMID:24244340	CL:2000008	microvascular endothelial cell
artery	hsa-miR-140-5p	cellular response to hypoxia	GO:0071456	PMID:27021683	CL:0002591	smooth muscle cell of the pulmonary artery
artery	hsa-miR-140-5p	positive regulation of BMP signaling pathway	GO:0030513	PMID:27214554	CL:0002591	smooth muscle cell of the pulmonary artery
artery	hsa-miR-140-5p	positive regulation of vascular associated smooth muscle cell apoptotic process	GO:1905461	PMID:27021683	CL:0002591	smooth muscle cell of the pulmonary artery
artery	hsa-miR-143-3p	negative regulation of angiogenesis	GO:0016525	PMID:25801897	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-143-3p	positive regulation of angiogenesis	GO:0045766	PMID:26311719	CL:1001568	pulmonary artery endothelial cell

artery	hsa-miR-145-5p	negative regulation of angiogenesis	GO:0016525	PMID:25801897	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-146a-5p	cellular response to cytokine stimulus	GO:0071345	PMID:25515214	CL:2000044	brain microvascular endothelial cell
artery	hsa-miR-146a-5p	cellular response to glucose stimulus	GO:0071333	PMID:28433754	CL:0002585	retinal blood vessel endothelial cell
artery	hsa-miR-146a-5p	negative regulation of angiogenesis	GO:0016525	PMID:23619365	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-146a-5p	negative regulation of interleukin-6 production	GO:0032715	PMID:28433754	CL:0002585	retinal blood vessel endothelial cell
artery	hsa-miR-146a-5p	negative regulation of NIK/NF-kappaB signaling	GO:1901223	PMID:25515214	CL:2000044	brain microvascular endothelial cell
artery	hsa-miR-146a-5p	negative regulation of tyrosine phosphorylation of STAT protein	GO:0042532	PMID:28433754	CL:0002585	retinal blood vessel endothelial cell
artery	hsa-miR-146a-5p	positive regulation of fibroblast growth factor receptor signaling pathway	GO:0045743	PMID:27121396	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-149-3p	negative regulation of fibroblast growth factor receptor signaling pathway	GO:0040037	PMID:24463821	CL:0002544	aortic endothelial cell
artery	hsa-miR-149-5p	negative regulation of fibroblast growth factor receptor signaling pathway	GO:0040037	PMID:24463821	CL:0002544	aortic endothelial cell
artery	hsa-miR-149-5p	negative regulation of interleukin-6 production	GO:0032715	PMID:24299952	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-152-3p	negative regulation of tumor necrosis factor-mediated signaling pathway	GO:0010804	PMID:24813629	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-155-5p	negative regulation of inflammatory response	GO:0050728	PMID:21310411	CL:0000071	blood vessel endothelial cell
artery	hsa-miR-155-5p	positive regulation of sprouting angiogenesis	GO:1903672	PMID:27731397	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-15a-5p	cellular response to glucose stimulus	GO:0071333	PMID:30365148	CL:0002585	retinal blood vessel endothelial cell
artery	hsa-miR-15a-5p	negative regulation of angiogenesis	GO:0016525	PMID:23867820	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-15a-5p	negative regulation of cell migration involved in sprouting angiogenesis	GO:0090051	PMID:22692216	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-15a-5p	negative regulation of G1/S transition of mitotic cell cycle	GO:2000134	PMID:23867820	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-15a-5p	negative regulation of inflammatory response	GO:0050728	PMID:30365148	CL:0002585	retinal blood vessel endothelial cell
artery	hsa-miR-15a-5p	negative regulation of NF-kappaB transcription factor activity	GO:0032088	PMID:30365148	CL:0002585	retinal blood vessel endothelial cell

artery	hsa-miR-15b-5p	branching involved in blood vessel morphogenesis	GO:0001569	PMID:23688497	UBERON:0007777	umbilical vein endothelium
artery	hsa-miR-15b-5p	negative regulation of angiogenesis	GO:0016525	PMID:27208409	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-16-5p	branching involved in blood vessel morphogenesis	GO:0001569	PMID:23083510	UBERON:0007777	umbilical vein endothelium
artery	hsa-miR-16-5p	cellular response to glucose stimulus	GO:0071333	PMID:30365148	CL:0002585	retinal blood vessel endothelial cell
artery	hsa-miR-16-5p	negative regulation of fibroblast growth factor receptor signaling pathway	GO:0040037	PMID:21885851	CL:0002544	aortic endothelial cell
artery	hsa-miR-16-5p	negative regulation of inflammatory response	GO:0050728	PMID:30365148	CL:0002585	retinal blood vessel endothelial cell
artery	hsa-miR-16-5p	negative regulation of NF-kappaB transcription factor activity	GO:0032088	PMID:30365148	CL:0002585	retinal blood vessel endothelial cell
artery	hsa-miR-16-5p	negative regulation of vascular endothelial growth factor signaling pathway	GO:1900747	PMID:21885851	CL:0002544	aortic endothelial cell
artery	hsa-miR-17-5p	cellular response to hypoxia	GO:0071456	PMID:27640178	CL:0002591	smooth muscle cell of the pulmonary artery
artery	hsa-miR-17-5p	positive regulation of vascular smooth muscle cell proliferation	GO:1904707	PMID:22161164	UBERON:0002012	pulmonary artery
artery	hsa-miR-185-3p	negative regulation of ERK1 and ERK2 cascade	GO:0070373	PMID:28277742	CL:0002539	aortic smooth muscle cell
artery	hsa-miR-185-5p	negative regulation of angiogenesis	GO:0016525	PMID:26694763	CL:2000008	microvascular endothelial cell
artery	hsa-miR-193a-3p	negative regulation of G1/S transition of mitotic cell cycle	GO:2000134	PMID:28276476	CL:0002546	embryonic blood vessel endothelial progenitor cell
artery	hsa-miR-196a-5p	cellular response to vascular endothelial growth factor stimulus	GO:0035924	PMID:22773844	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-196a-5p	negative regulation of cell migration involved in sprouting angiogenesis	GO:0090051	PMID:22773844	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-19b-3p	negative regulation of cell migration involved in sprouting angiogenesis	GO:0090051	PMID:22197821	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-19b-3p	negative regulation of G2/M transition of mitotic cell cycle	GO:0010972	PMID:22197821	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-19b-3p	negative regulation of serine-type endopeptidase activity	GO:1900004	PMID:24998411	CL:0002618	endothelial cell of umbilical vein

artery	hsa-miR-200a-3p	positive regulation of blood vessel endothelial cell migration	GO:0043536	PMID:21698760	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-200b-3p	negative regulation of angiogenesis	GO:0016525	PMID:21081489	CL:2000008	microvascular endothelial cell
artery	hsa-miR-200b-3p	negative regulation of blood vessel endothelial cell migration	GO:0043537	PMID:21081489	CL:2000008	microvascular endothelial cell
artery	hsa-miR-20a-5p	cellular response to vascular endothelial growth factor stimulus	GO:0035924	PMID:22696064	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-20a-5p	negative regulation of cell migration involved in sprouting angiogenesis	GO:0090051	PMID:22696064	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-20a-5p	negative regulation of protein kinase activity	GO:0006469	PMID:25447536	CL:0002591	smooth muscle cell of the pulmonary artery
artery	hsa-miR-20a-5p	positive regulation of angiogenesis	GO:0045766	PMID:28097093	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-20a-5p	positive regulation of vascular smooth muscle cell proliferation	GO:1904707	PMID:22450430	CL:0002591	smooth muscle cell of the pulmonary artery
artery	hsa-miR-20b-5p	cellular response to tumor necrosis factor	GO:0071356	PMID:28595801	CL:2000008	microvascular endothelial cell
artery	hsa-miR-20b-5p	negative regulation of angiogenesis	GO:0016525	PMID:24048733	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-20b-5p	positive regulation of cellular senescence	GO:2000774	PMID:28595801	CL:2000008	microvascular endothelial cell
artery	hsa-miR-21-3p	negative regulation of NF-kappaB transcription factor activity	GO:0032088	PMID:25327529	CL:0002544	aortic endothelial cell
artery	hsa-miR-21-5p	BMP signaling pathway	GO:0030509	PMID:18548003	UBERON:0002012	pulmonary artery
artery	hsa-miR-21-5p	cellular response to lipopolysaccharide	GO:0071222	PMID:29039542	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-21-5p	negative regulation of GTPase activity	GO:0034260	PMID:22158624	CL:0000359	vascular associated smooth muscle cell
artery	hsa-miR-21-5p	positive regulation of angiogenesis	GO:0045766	PMID:27708252	CL:2000008	microvascular endothelial cell
artery	hsa-miR-21-5p	positive regulation of blood vessel endothelial cell proliferation involved in sprouting angiogenesis	GO:1903589	PMID:30106099	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-21-5p	positive regulation of inflammatory response	GO:0050729	PMID:21636785	CL:0000071	blood vessel endothelial cell
artery	hsa-miR-21-5p	positive regulation of vascular associated smooth muscle cell migration	GO:1904754	PMID:20693317,PMID:20693317,PMID:21817107	UBERON:0002012,UBERON:0002	pulmonary artery,pulmonary artery,femoral artery

					012,UBERON:0002060	
artery	hsa-miR-21-5p	positive regulation of vascular endothelial growth factor signaling pathway	GO:1900748	PMID:30106099	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-21-5p	positive regulation of vascular smooth muscle cell proliferation	GO:1904707	PMID:20693317,PMID:21817107	UBERON:0002012,UBERON:0002060	pulmonary artery,femoral artery
artery	hsa-miR-21-5p	regulation of cell shape	GO:0008360	PMID:21817107	CL:0000359	vascular associated smooth muscle cell
artery	hsa-miR-210-3p	positive regulation of angiogenesis	GO:0045766	PMID:18417479	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-210-3p	positive regulation of cell migration	GO:0030335	PMID:23322395	CL:2000008	microvascular endothelial cell
artery	hsa-miR-210-3p	tube formation	GO:0035148	PMID:23322395	CL:2000008	microvascular endothelial cell
artery	hsa-miR-212-3p	negative regulation of angiogenesis	GO:0016525	PMID:25217442	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-212-3p	positive regulation of Notch signaling pathway	GO:0045747	PMID:25217442	CL:0000071	blood vessel endothelial cell
artery	hsa-miR-214-3p	cellular response to hypoxia	GO:0071456	PMID:27144530	CL:0002591	smooth muscle cell of the pulmonary artery
artery	hsa-miR-214-3p	negative regulation of cell migration	GO:0030336	PMID:25656649	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-214-3p	negative regulation of cell proliferation	GO:0008285	PMID:25656649	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-214-3p	negative regulation of vascular associated smooth muscle cell migration	GO:1904753	PMID:27927633	UBERON:0000947	aorta
artery	hsa-miR-214-3p	negative regulation of vascular smooth muscle cell proliferation	GO:1904706	PMID:27927633	UBERON:0000947	aorta
artery	hsa-miR-214-3p	positive regulation of G1/S transition of mitotic cell cycle	GO:1900087	PMID:27144530	CL:0002591	smooth muscle cell of the pulmonary artery
artery	hsa-miR-214-5p	positive regulation of vascular smooth muscle cell proliferation	GO:1904707	PMID:28684904	UBERON:0002012	pulmonary artery
artery	hsa-miR-218-5p	negative regulation of MAP kinase activity	GO:0043407	PMID:21385766	CL:0000071	blood vessel endothelial cell
artery	hsa-miR-22-3p	positive regulation of inflammatory response	GO:0050729	PMID:28112401	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-221-3p	negative regulation of inflammatory response	GO:0050728	PMID:21310411	CL:0000071	blood vessel endothelial cell
artery	hsa-miR-221-3p	negative regulation of sprouting angiogenesis	GO:1903671	PMID:27780851	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-221-3p	platelet-derived growth factor receptor signaling pathway	GO:0048008	PMID:19088079	CL:0002591	smooth muscle cell of the pulmonary artery

artery	hsa-miR-221-3p	positive regulation of blood vessel endothelial cell migration	GO:0043536	PMID:19351599	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-222-3p	negative regulation of inflammatory response	GO:0050728	PMID:21310411	CL:0000071	blood vessel endothelial cell
artery	hsa-miR-223-3p	negative regulation of GTPase activity	GO:0034260	PMID:27121304	CL:0002591	smooth muscle cell of the pulmonary artery
artery	hsa-miR-23a-3p	cellular response to vascular endothelial growth factor stimulus	GO:0035924	PMID:21536891	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-23a-3p	negative regulation of vascular permeability	GO:0043116	PMID:27741223	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-23a-3p	positive regulation of blood vessel endothelial cell proliferation involved in sprouting angiogenesis	GO:1903589	PMID:21536891	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-23a-3p	positive regulation of cell migration involved in sprouting angiogenesis	GO:0090050	PMID:21536891	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-23a-3p	positive regulation of ERK1 and ERK2 cascade	GO:0070374	PMID:21536891	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-23b-3p	cellular response to vascular endothelial growth factor stimulus	GO:0035924	PMID:21536891	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-23b-3p	negative regulation of sprouting angiogenesis	GO:1903671	PMID:27741223	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-23b-3p	positive regulation of blood vessel endothelial cell proliferation involved in sprouting angiogenesis	GO:1903589	PMID:21536891	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-23b-3p	positive regulation of cell migration involved in sprouting angiogenesis	GO:0090050	PMID:21536891	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-23b-3p	positive regulation of ERK1 and ERK2 cascade	GO:0070374	PMID:21536891	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-23b-3p	positive regulation of vascular permeability	GO:0043117	PMID:27741223	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-24-3p	negative regulation of angiogenesis	GO:0016525	PMID:23774796	CL:2000008	microvascular endothelial cell
artery	hsa-miR-24-3p	positive regulation of endothelial cell apoptotic process	GO:2000353	PMID:21788589	CL:0000071	blood vessel endothelial cell
artery	hsa-miR-24-3p	positive regulation of reactive oxygen species biosynthetic process	GO:1903428	PMID:21788589	CL:0000071	blood vessel endothelial cell
artery	hsa-miR-24-3p	positive regulation of vascular associated smooth muscle cell apoptotic process	GO:1905461	PMID:23774796	CL:2000008	microvascular endothelial cell
artery	hsa-miR-26a-5p	negative regulation of BMP signaling pathway	GO:0030514	PMID:24047927	CL:0002618	endothelial cell of umbilical vein

artery	hsa-miR-26a-5p	negative regulation of cell migration involved in sprouting angiogenesis	GO:0090051	PMID:24047927,PMID:28602162	CL:0002618,CL:0002618	endothelial cell of umbilical vein,endothelial cell of umbilical vein
artery	hsa-miR-26a-5p	negative regulation of DNA-binding transcription factor activity	GO:0043433	PMID:24047927	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-26a-5p	negative regulation of G1/S transition of mitotic cell cycle	GO:2000134	PMID:24047927	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-26a-5p	negative regulation of nitric-oxide synthase activity	GO:0051001	PMID:28602162	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-26a-5p	transforming growth factor beta receptor signaling pathway	GO:0007179	PMID:20857419	CL:0002539	aortic smooth muscle cell
artery	hsa-miR-26a-5p	vascular endothelial growth factor signaling pathway	GO:0038084	PMID:28602162	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-27a-3p	cellular response to vascular endothelial growth factor stimulus	GO:0035924	PMID:21536891	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-27a-3p	positive regulation of blood vessel endothelial cell proliferation involved in sprouting angiogenesis	GO:1903589	PMID:21536891	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-27a-3p	positive regulation of cell migration involved in sprouting angiogenesis	GO:0090050	PMID:21536891	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-27a-3p	positive regulation of ERK1 and ERK2 cascade	GO:0070374	PMID:21536891	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-27a-5p	negative regulation of NF-kappaB transcription factor activity	GO:0032088	PMID:25327529	CL:0002544	aortic endothelial cell
artery	hsa-miR-27b-3p	cellular response to vascular endothelial growth factor stimulus	GO:0035924	PMID:21536891	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-27b-3p	positive regulation of blood vessel endothelial cell proliferation involved in sprouting angiogenesis	GO:1903589	PMID:21536891	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-27b-3p	positive regulation of cell migration involved in sprouting angiogenesis	GO:0090050	PMID:21536891	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-27b-3p	positive regulation of ERK1 and ERK2 cascade	GO:0070374	PMID:21536891	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-296-5p	positive regulation of cell migration involved in sprouting angiogenesis	GO:0090050	PMID:18977327	CL:2000044	brain microvascular endothelial cell
artery	hsa-miR-296-5p	positive regulation of vascular endothelial growth factor receptor signaling pathway	GO:0030949	PMID:18977327	CL:2000044	brain microvascular endothelial cell

artery	hsa-miR-29a-3p	negative regulation of angiogenesis	GO:0016525	PMID:28637396	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-29a-3p	positive regulation of angiogenesis	GO:0045766	PMID:23541945	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-29a-3p	positive regulation of G1/S transition of mitotic cell cycle	GO:1900087	PMID:23541945	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-29b-3p	negative regulation of collagen biosynthetic process	GO:0032966	PMID:22269326	CL:0002547	fibroblast of the aortic adventitia
artery	hsa-miR-29c-5p	negative regulation of angiogenesis	GO:0016525	PMID:26175848	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-29c-5p	negative regulation of insulin-like growth factor receptor signaling pathway	GO:0043569	PMID:26045889,PMID:26175848	CL:0002618,CL:0002618	endothelial cell of umbilical vein,endothelial cell of umbilical vein
artery	hsa-miR-30a-3p	positive regulation of angiogenesis	GO:0045766	PMID:23960241	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-30a-3p	transforming growth factor beta receptor signaling pathway	GO:0007179	PMID:23960241	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-30b-5p	negative regulation of angiogenesis	GO:0016525	PMID:28977001	CL:0000071	blood vessel endothelial cell
artery	hsa-miR-30b-5p	negative regulation of endothelial cell apoptotic process	GO:2000352	PMID:27464494	CL:2000018	endothelial cell of coronary artery
artery	hsa-miR-30b-5p	positive regulation of protein phosphorylation	GO:0001934	PMID:28977001	CL:0000071	blood vessel endothelial cell
artery	hsa-miR-30b-5p	positive regulation of sprouting angiogenesis	GO:1903672	PMID:23086751	CL:0000071	blood vessel endothelial cell
artery	hsa-miR-30b-5p	positive regulation of transforming growth factor beta receptor signaling pathway	GO:0030511	PMID:28977001	CL:0000071	blood vessel endothelial cell
artery	hsa-miR-30c-5p	negative regulation of sprouting angiogenesis	GO:1903671	PMID:27780851	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-30e-5p	negative regulation of endothelial cell apoptotic process	GO:2000352	PMID:27464494	CL:2000018	endothelial cell of coronary artery
artery	hsa-miR-30e-5p	negative regulation of sprouting angiogenesis	GO:1903671	PMID:27780851	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-31-5p	positive regulation of angiogenesis	GO:0045766	PMID:28097093	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-31-5p	positive regulation of blood vessel endothelial cell migration	GO:0043536	PMID:26933040	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-31-5p	positive regulation of sprouting angiogenesis	GO:1903672	PMID:26933040	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-329-3p	negative regulation of cell migration involved in sprouting angiogenesis	GO:0090051	PMID:23878390	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-329-3p	negative regulation of vascular endothelial growth factor signaling pathway	GO:1900747	PMID:23878390	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-329-3p	NIK/NF-kappaB signaling	GO:0038061	PMID:23878390	CL:0002618	endothelial cell of umbilical vein

artery	hsa-miR-342-5p	negative regulation of protein kinase B signaling	GO:0051898	PMID:26857067	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-342-5p	negative regulation of transforming growth factor beta receptor signaling pathway	GO:0030512	PMID:26857067	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-342-5p	negative regulation of vascular endothelial growth factor signaling pathway	GO:1900747	PMID:26857067	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-34a-5p	cellular response to hypoxia	GO:0071456	PMID:27302634	CL:0002591	smooth muscle cell of the pulmonary artery
artery	hsa-miR-34a-5p	negative regulation of angiogenesis	GO:0016525	PMID:24048733	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-34a-5p	negative regulation of calcium ion import	GO:0090281	PMID:27302634	CL:0002591	smooth muscle cell of the pulmonary artery
artery	hsa-miR-34a-5p	negative regulation of sprouting angiogenesis	GO:1903671	PMID:23426265	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-34a-5p	negative regulation of vascular associated smooth muscle cell migration	GO:1904753	PMID:26493107	UBERON:0000947	aorta
artery	hsa-miR-34a-5p	negative regulation of vascular smooth muscle cell proliferation	GO:1904706	PMID:24792364,PMID:26493107	CL:0002546,UBERON:0000947	embryonic blood vessel endothelial progenitor cell,aorta
artery	hsa-miR-361-5p	negative regulation of angiogenesis	GO:0016525	PMID:25203061	CL:0000071	blood vessel endothelial cell
artery	hsa-miR-362-3p	negative regulation of G1/S transition of mitotic cell cycle	GO:2000134	PMID:28890348	CL:0000359	vascular associated smooth muscle cell
artery	hsa-miR-377-3p	negative regulation of sprouting angiogenesis	GO:1903671	PMID:25251394	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-424-5p	negative regulation of angiogenesis	GO:0016525	PMID:28566713	CL:0000071	blood vessel endothelial cell
artery	hsa-miR-424-5p	negative regulation of cell migration involved in sprouting angiogenesis	GO:0090051	PMID:28566713	CL:0000071	blood vessel endothelial cell
artery	hsa-miR-424-5p	negative regulation of ERK1 and ERK2 cascade	GO:0070373	PMID:23263626	CL:1001568	pulmonary artery endothelial cell
artery	hsa-miR-424-5p	negative regulation of fibroblast growth factor receptor signaling pathway	GO:0040037	PMID:21885851,PMID:23263626	CL:0002544,CL:1001568	aortic endothelial cell,pulmonary artery endothelial cell
artery	hsa-miR-424-5p	negative regulation of G0 to G1 transition	GO:0070317	PMID:23263626	CL:1001568	pulmonary artery endothelial cell
artery	hsa-miR-424-5p	negative regulation of vascular endothelial growth factor signaling pathway	GO:1900747	PMID:21885851	CL:0002544	aortic endothelial cell
artery	hsa-miR-424-5p	negative regulation of vascular smooth muscle cell proliferation	GO:1904706	PMID:23263626	CL:0002591	smooth muscle cell of the pulmonary artery

artery	hsa-miR-451a	positive regulation of vascular associated smooth muscle cell migration	GO:1904754	PMID:25006399	UBERON:0002012	pulmonary artery
artery	hsa-miR-4632-3p	cellular response to platelet-derived growth factor stimulus	GO:0036120	PMID:28701355	CL:0002591	smooth muscle cell of the pulmonary artery
artery	hsa-miR-483-5p	negative regulation of cell migration involved in sprouting angiogenesis	GO:0090051	PMID:21893058	CL:0000071	blood vessel endothelial cell
artery	hsa-miR-487b-3p	positive regulation of blood vessel endothelial cell proliferation involved in sprouting angiogenesis	GO:1903589	PMID:25660232	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-487b-3p	positive regulation of cell migration involved in sprouting angiogenesis	GO:0090050	PMID:25660232	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-492	negative regulation of angiogenesis	GO:0016525	PMID:23802567	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-495-3p	negative regulation of endothelial cell apoptotic process	GO:2000352	PMID:25466836	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-495-3p	positive regulation of G1/S transition of mitotic cell cycle	GO:1900087	PMID:25466836	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-503-5p	negative regulation of angiogenesis	GO:0016525	PMID:28566713	CL:0000071	blood vessel endothelial cell
artery	hsa-miR-503-5p	negative regulation of cell-substrate adhesion	GO:0010812	PMID:21220732	CL:0002543	vein endothelial cell
artery	hsa-miR-503-5p	negative regulation of cell migration involved in sprouting angiogenesis	GO:0090051	PMID:28566713	CL:0000071	blood vessel endothelial cell
artery	hsa-miR-503-5p	negative regulation of endothelial cell migration	GO:0010596	PMID:21220732	CL:0002543	vein endothelial cell
artery	hsa-miR-503-5p	negative regulation of ERK1 and ERK2 cascade	GO:0070373	PMID:23263626	CL:1001568	pulmonary artery endothelial cell
artery	hsa-miR-503-5p	negative regulation of fibroblast growth factor receptor signaling pathway	GO:0040037	PMID:23263626	CL:1001568	pulmonary artery endothelial cell
artery	hsa-miR-503-5p	negative regulation of G0 to G1 transition	GO:0070317	PMID:23263626	CL:1001568	pulmonary artery endothelial cell
artery	hsa-miR-503-5p	negative regulation of vascular smooth muscle cell proliferation	GO:1904706	PMID:23263626	CL:0002591	smooth muscle cell of the pulmonary artery
artery	hsa-miR-505-3p	negative regulation of angiogenesis	GO:0016525	PMID:25449503	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-638	negative regulation of G1/S transition of mitotic cell cycle	GO:2000134	PMID:23554459	CL:0002539	aortic smooth muscle cell
artery	hsa-miR-638	platelet-derived growth factor receptor signaling pathway	GO:0048008	PMID:23554459	CL:0002539	aortic smooth muscle cell

artery	hsa-miR-665	negative regulation of canonical Wnt signaling pathway	GO:0090090	PMID:29118903	CL:0002539	aortic smooth muscle cell
artery	hsa-miR-7-5p	negative regulation of sprouting angiogenesis	GO:1903671	PMID:27431648	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-92a-3p	cellular response to low-density lipoprotein particle stimulus	GO:0071404	PMID:24255059	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-92a-3p	negative regulation of inflammatory response	GO:0050728	PMID:22267480	CL:1000413	endothelial cell of artery
artery	hsa-miR-92a-3p	negative regulation of nitric oxide biosynthetic process	GO:0045019	PMID:21768538	CL:0000071	blood vessel endothelial cell
artery	hsa-miR-92a-3p	positive regulation of interleukin-6 production	GO:0032755	PMID:24255059	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-92a-3p	positive regulation of monocyte chemotactic protein-1 production	GO:0071639	PMID:24255059	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-92a-3p	positive regulation of sprouting angiogenesis	GO:1903672	PMID:26299712	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-939-5p	negative regulation of angiogenesis	GO:0016525	PMID:28115160	CL:0002618	endothelial cell of umbilical vein
artery	hsa-miR-939-5p	negative regulation of cell-matrix adhesion	GO:0001953	PMID:28115160	CL:0002618	endothelial cell of umbilical vein
blood	hsa-let-7f-5p	negative regulation of interleukin-17 production	GO:0032700	PMID:21508257	CL:0000813	memory T cell
blood	hsa-miR-106a-5p	negative regulation of interleukin-8 secretion	GO:2000483	PMID:26265888	CL:2000001	peripheral blood mononuclear cell
blood	hsa-miR-144-3p	positive regulation of cholesterol storage	GO:0010886	PMID:24733347	CL:0000517	macrophage derived foam cell
blood	hsa-miR-144-3p	positive regulation of interleukin-1 beta secretion	GO:0050718	PMID:24733347	CL:0000517	macrophage derived foam cell
blood	hsa-miR-144-3p	positive regulation of interleukin-6 secretion	GO:2000778	PMID:24733347	CL:0000517	macrophage derived foam cell
blood	hsa-miR-144-3p	positive regulation of tumor necrosis factor secretion	GO:1904469	PMID:24733347	CL:0000517	macrophage derived foam cell
blood	hsa-miR-146a-5p	negative regulation of cholesterol storage	GO:0010887	PMID:21329689	CL:0000517	macrophage derived foam cell
blood	hsa-miR-146a-5p	negative regulation of cytokine production involved in inflammatory response	GO:1900016	PMID:29896267	CL:0000583	alveolar macrophage
blood	hsa-miR-146a-5p	negative regulation of inflammatory response	GO:0050728	PMID:21329689	CL:0000517	macrophage derived foam cell
blood	hsa-miR-146a-5p	negative regulation of interleukin-6 secretion	GO:1900165	PMID:21329689,PMID:29896267	CL:0000517,CL:0000583	macrophage derived foam cell,alveolar macrophage
blood	hsa-miR-146a-5p	negative regulation of interleukin-8 secretion	GO:2000483	PMID:21329689	CL:0000517	macrophage derived foam cell
blood	hsa-miR-146a-5p	negative regulation of toll-like receptor 4 signaling pathway	GO:0034144	PMID:21329689	CL:0000517	macrophage derived foam cell

blood	hsa-miR-155-5p	cellular response to low-density lipoprotein particle stimulus	GO:0071404	PMID:21030878	CL:0000517	macrophage derived foam cell
blood	hsa-miR-155-5p	cholesterol homeostasis	GO:0042632	PMID:21030878	CL:0000517	macrophage derived foam cell
blood	hsa-miR-155-5p	negative regulation of inflammatory response	GO:0050728	PMID:21030878	CL:0000517	macrophage derived foam cell
blood	hsa-miR-155-5p	negative regulation of NF-kappaB transcription factor activity	GO:0032088	PMID:21030878	CL:0000517	macrophage derived foam cell
blood	hsa-miR-155-5p	negative regulation of protein localization to nucleus	GO:1900181	PMID:21030878	CL:0000517	macrophage derived foam cell
blood	hsa-miR-181a-5p	negative regulation of tumor necrosis factor production	GO:0032720	PMID:23516523	CL:0000235	macrophage
blood	hsa-miR-181b-5p	negative regulation of defense response to bacterium	GO:1900425	PMID:25505240	CL:0000235	macrophage
blood	hsa-miR-181b-5p	negative regulation of innate immune response	GO:0045824	PMID:25505240	CL:0000235	macrophage
blood	hsa-miR-181b-5p	negative regulation of phagocytosis	GO:0050765	PMID:25505240	CL:0000235	macrophage
blood	hsa-miR-181b-5p	negative regulation of signal transduction	GO:0009968	PMID:25505240	CL:0000235	macrophage
blood	hsa-miR-181b-5p	positive regulation of inflammatory response	GO:0050729	PMID:25505240	CL:0000235	macrophage
blood	hsa-miR-182-5p	cholesterol homeostasis	GO:0042632	PMID:28855441	CL:0000235	macrophage
blood	hsa-miR-182-5p	positive regulation of cytokine secretion	GO:0050715	PMID:28855441	CL:0000235	macrophage
blood	hsa-miR-182-5p	positive regulation of lipoprotein lipase activity	GO:0051006	PMID:28855441	CL:0000235	macrophage
blood	hsa-miR-182-5p	positive regulation of NIK/NF-kappaB signaling	GO:1901224	PMID:28855441	CL:0000235	macrophage
blood	hsa-miR-183-5p	transforming growth factor beta receptor signaling pathway	GO:0007179	PMID:24586048	CL:0000623	natural killer cell
blood	hsa-miR-19a-3p	positive regulation of B cell receptor signaling pathway	GO:0050861	PMID:26017478	CL:2000001	peripheral blood mononuclear cell
blood	hsa-miR-19b-3p	cholesterol homeostasis	GO:0042632	PMID:25084135,PMID:25765596	CL:0000517,CL:0000517	macrophage derived foam cell,macrophage derived foam cell
blood	hsa-miR-20a-5p	negative regulation of inflammatory response	GO:0050728	PMID:28972028	CL:0000492	CD4-positive helper T cell
blood	hsa-miR-20a-5p	negative regulation of phosphatidylinositol 3-kinase signaling	GO:0014067	PMID:28972028	CL:0000492	CD4-positive helper T cell
blood	hsa-miR-20a-5p	negative regulation of protein kinase B signaling	GO:0051898	PMID:28972028	CL:0000492	CD4-positive helper T cell
blood	hsa-miR-26b-5p	defense response to virus	GO:0051607	PMID:26222045	CL:2000001	peripheral blood mononuclear cell

blood	hsa-miR-26b-5p	negative regulation of defense response to virus	GO:0050687	PMID:26222045	CL:2000001	peripheral blood mononuclear cell
blood	hsa-miR-27a-5p	cellular response to transforming growth factor beta stimulus	GO:0071560	PMID:28791023	CL:0000623	natural killer cell
blood	hsa-miR-302a-3p	cellular response to low-density lipoprotein particle stimulus	GO:0071404	PMID:25524771	CL:0000235	macrophage
blood	hsa-miR-302a-3p	cholesterol homeostasis	GO:0042632	PMID:25524771	CL:0000235	macrophage
blood	hsa-miR-33b-5p	cholesterol homeostasis	GO:0042632	PMID:24931346	CL:0000581	peritoneal macrophage
blood	hsa-miR-361-5p	regulation of inflammatory response	GO:0050727	PMID:28444107	CL:0000517	macrophage derived foam cell
blood	hsa-miR-488-3p	negative regulation of inflammatory response	GO:0050728	PMID:28915828	CL:0000235	macrophage
blood	hsa-miR-488-3p	negative regulation of tumor necrosis factor biosynthetic process	GO:0042536	PMID:28915828	CL:0000235	macrophage
blood	hsa-miR-590-3p	cholesterol homeostasis	GO:0042632	PMID:25149060	CL:0000235	macrophage
blood	hsa-miR-590-3p	negative regulation of inflammatory response	GO:0050728	PMID:25149060	CL:0000235	macrophage
blood	hsa-miR-758-3p	cellular response to low-density lipoprotein particle stimulus	GO:0071404	PMID:21885853	CL:0000235	macrophage
blood	hsa-miR-920	negative regulation of inflammatory response	GO:0050728	PMID:28915828	CL:0000235	macrophage
blood	hsa-miR-920	negative regulation of tumor necrosis factor biosynthetic process	GO:0042536	PMID:28915828	CL:0000235	macrophage
brain	hsa-miR-195-5p	response to ischemia	GO:0002931	PMID:30497184	UBERON:0000955	brain
heart	hsa-let-7b-5p	positive regulation of angiogenesis	GO:0045766	PMID:26296645	UBERON:0002349	myocardium
heart	hsa-miR-1-3p	negative regulation of cardiac muscle hypertrophy	GO:0010614	PMID:19933931	UBERON:0006566	left ventricle myocardium
heart	hsa-miR-1-3p	positive regulation of protein phosphorylation	GO:0001934	PMID:19131648	CL:0000746	cardiac muscle cell
heart	hsa-miR-1-3p	positive regulation of sprouting angiogenesis	GO:1903672	PMID:23625462	CL:0000513	cardiac muscle myoblast
heart	hsa-miR-1-3p	regulation of release of sequestered calcium ion into cytosol by sarcoplasmic reticulum	GO:0010880	PMID:19131648	CL:0000746	cardiac muscle cell
heart	hsa-miR-155-5p	negative regulation of cell migration involved in sprouting angiogenesis	GO:0090051	PMID:28408180	UBERON:0000948	heart
heart	hsa-miR-155-5p	negative regulation of hydrogen peroxide-induced cell death	GO:1903206	PMID:20550618	CL:0010021	cardiac myoblast

heart	hsa-miR-21-5p	regulation of calcium ion transmembrane transport via high voltage-gated calcium channel	GO:1902514	PMID:25107449	CL:0000746	cardiac muscle cell
heart	hsa-miR-223-3p	positive regulation of glucose import	GO:0046326	PMID:20080987	CL:0000746	cardiac muscle cell
heart	hsa-miR-223-3p	positive regulation of protein localization to plasma membrane	GO:1903078	PMID:20080987	CL:2000046	ventricular cardiac muscle cell
kidney	hsa-miR-590-5p	negative regulation of epithelial to mesenchymal transition	GO:0010719	PMID:26459119	CL:1000497	kidney cell
liver	hsa-miR-144-3p	cholesterol homeostasis	GO:0042632	PMID:23519695	CL:0000182	hepatocyte
liver	hsa-miR-148a-3p	cholesterol homeostasis	GO:0042632	PMID:26437365	CL:0000182	hepatocyte
liver	hsa-miR-148a-3p	negative regulation of low-density lipoprotein particle clearance	GO:0010989	PMID:26437365	CL:0000182	hepatocyte
liver	hsa-miR-185-5p	negative regulation of low-density lipoprotein particle clearance	GO:0010989	PMID:26523989	CL:0000182	hepatocyte
liver	hsa-miR-199a-5p	negative regulation of low-density lipoprotein particle clearance	GO:0010989	PMID:26163491	CL:0000182	hepatocyte
liver	hsa-miR-199a-5p	negative regulation of receptor internalization	GO:0002091	PMID:26163491	CL:0000182	hepatocyte
liver	hsa-miR-21-5p	cellular response to virus	GO:0098586	PMID:27571873	CL:0000182	hepatocyte
liver	hsa-miR-27a-3p	negative regulation of low-density lipoprotein particle clearance	GO:0010989	PMID:26318398	CL:0000182	hepatocyte
liver	hsa-miR-27b-3p	cholesterol homeostasis	GO:0042632	PMID:26520906	CL:0000182	hepatocyte
liver	hsa-miR-27b-3p	negative regulation of low-density lipoprotein particle clearance	GO:0010989	PMID:26520906	CL:0000182	hepatocyte
liver	hsa-miR-30c-5p	negative regulation of fatty acid biosynthetic process	GO:0045717	PMID:23749231	UBERON:0002107	liver
liver	hsa-miR-548p	negative regulation of fatty acid biosynthetic process	GO:0045717	PMID:28336556	CL:0000182	hepatocyte
muscle_skeletal	hsa-miR-15b-5p	positive regulation of translation	GO:0045727	PMID:25403480	CL:0000737	striated muscle cell
muscle_skeletal	hsa-miR-16-5p	positive regulation of translation	GO:0045727	PMID:25403480	CL:0000737	striated muscle cell
nerve	hsa-miR-103a-3p	negative regulation of cyclin-dependent protein serine/threonine kinase activity	GO:0045736	PMID:27343180	CL:0000540	neuron
nerve	hsa-miR-103a-3p	negative regulation of peptidyl-threonine phosphorylation	GO:0010801	PMID:27343180	CL:0000540	neuron

nerve	hsa-miR-106b-5p	cellular response to amyloid-beta	GO:1904646	PMID:27520374	CL:0000540	neuron
nerve	hsa-miR-106b-5p	negative regulation of peptidyl-tyrosine phosphorylation	GO:0050732	PMID:27520374	CL:0000540	neuron
nerve	hsa-miR-125b-1-3p	positive regulation of protein phosphorylation	GO:0001934	PMID:28947385	CL:0000540	neuron
nerve	hsa-miR-125b-1-3p	positive regulation of tau-protein kinase activity	GO:1902949	PMID:28947385	CL:0000540	neuron
nerve	hsa-miR-132-3p	negative regulation of nitric-oxide synthase activity	GO:0051001	PMID:28089352	CL:0000540	neuron
nerve	hsa-miR-132-3p	negative regulation of nitric oxide biosynthetic process	GO:0045019	PMID:28089352	CL:0000540	neuron
nerve	hsa-miR-132-3p	negative regulation of peptidyl-serine phosphorylation	GO:0033137	PMID:28089352	CL:0000540	neuron
nerve	hsa-miR-140-5p	cellular response to amyloid-beta	GO:1904646	PMID:29253717	CL:0000540	neuron
nerve	hsa-miR-146a-5p	negative regulation of protein kinase B signaling	GO:0051898	PMID:27241555	CL:0000540	neuron
nerve	hsa-miR-146a-5p	positive regulation of apoptotic process	GO:0043065	PMID:27241555	CL:0000540	neuron
nerve	hsa-miR-15a-5p	negative regulation of cyclin-dependent protein serine/threonine kinase activity	GO:0045736	PMID:27343180	CL:0000540	neuron
nerve	hsa-miR-15a-5p	negative regulation of peptidyl-threonine phosphorylation	GO:0010801	PMID:27343180	CL:0000540	neuron
nerve	hsa-miR-212-3p	negative regulation of nitric-oxide synthase activity	GO:0051001	PMID:28089352	CL:0000540	neuron
nerve	hsa-miR-212-3p	negative regulation of nitric oxide biosynthetic process	GO:0045019	PMID:28089352	CL:0000540	neuron
nerve	hsa-miR-212-3p	negative regulation of peptidyl-serine phosphorylation	GO:0033137	PMID:28089352	CL:0000540	neuron
nerve	hsa-miR-218-5p	negative regulation of collagen biosynthetic process	GO:0032966	PMID:19913496	CL:0000122	stellate neuron
nerve	hsa-miR-29b-3p	negative regulation of collagen biosynthetic process	GO:0032966	PMID:19913496	CL:0000122	stellate neuron
nerve	hsa-miR-98-5p	cellular response to amyloid-beta	GO:1904646	PMID:27541017	CL:0000540	neuron
skin	hsa-miR-181b-5p	negative regulation of toll-like receptor 4 signaling pathway	GO:0034144	PMID:27641447	CL:0000312	keratinocyte

skin	hsa-miR-200b-3p	negative regulation of blood vessel endothelial cell migration	GO:0043537	PMID:22499991	UBERON:0002067	dermis
skin	hsa-miR-203a-3p	negative regulation of cytokine production involved in inflammatory response	GO:1900016	PMID:23608026	CL:0000312	keratinocyte
skin	hsa-miR-203a-3p	negative regulation of interleukin-8 secretion	GO:2000483	PMID:23608026	CL:0000312	keratinocyte

CURRICULUM VITA

Mohammed Sayed
Department of Computer Science and Engineering
University of Louisville
Louisville, KY, 40292
Email: mohammed.sayed@louisville.edu

Education

- Ph.D., Computer Science and Engineering (Expected: May 2021)
  University of Louisville, Louisville, KY
  Advisors: Dr. Juw Won Park
- MSc, Biomedical Image Processing, 2010 - 2015
  Alexandria University, Alexandria, Egypt
- B.Sc., Computer Science and Engineering, 2004-2009
  Minia Universit, Minia, Egypt

Experience

- Graduate teaching assistant, 08/2017 – 04/2021
  Computer Science and Engineering department
  University of Louisville, Louisville, KY
- Bioinformatics scientist intern, 06/2019 – 08/2019
  National Marrow Donor Program
  Minneapolis, MN

Awards

- Grosscurth Fellowship, Speed School of Engineering, University of Louisville, 2015 – 2016

Publications

- **Sayed M**, Park JW. 2020. miRinGO: Prediction of biological processes indirectly targeted by human microRNAs. bioRxiv doi: 10.1101/2020.07.24.220335
- Alugubelly, N., Mohammad, A. N., Edelmann, M. J., Nanduri, B., **Sayed, M.,** Park, J. W., & Carr, R. L. (2019). Proteomic and transcriptional profiling of rat amygdala following social play. Behavioural brain research, 376, 112210.
- Teng Y, Ren Y, **Sayed M**, Hu X, Lei C, Kumar A, Hutchins E, Mu J, Deng Z, Luo C, Sundaram K. Plant-Derived Exosomal MicroRNAs Shape the Gut Microbiota. Cell host & microbe. 2018 Nov 14; 24(5):637-52.
- **Sayed, Mohammed**, Jae Yeon Hwang, and Juw Won Park. Sequence homology in circular RNA detection. Proceedings of the 2018 Conference on Research in Adaptive and Convergent Systems. ACM, 2018.
- **Abdelfadeel, M. A.,** ElShehaby, S., & Abougabal, M. S. (2014, December). Automatic segmentation of left ventricle in cardiac MRI using maximally stable extremal regions. In 2014 Cairo International Biomedical Engineering Conference (CIBEC) (pp. 145-148). IEEE.

.