Discovery and Detection of miRNA Biomarkers for Viral Infection

Submitted in partial fulfilment of the degree requirements for

the degree of

Doctor of Philosophy

in

Chemical Engineering

Virginia Lane DiNenna

M.S., Chemical Engineering, Carnegie Mellon University
M.S., Colloids, Polymers, and Surfaces, Carnegie Mellon University

B.S., Chemical Engineering, University of Florida
B.S. Microbiology and Cell Science, University of Florida
B.M., Music Performance, University of Florida

Carnegie Mellon University
Pittsburgh, PA

December, 2022

Acknowledgements

First, I would like to thank my advisors, Jim Schneider and Todd Przybycien. I am grateful for your guidance through my five years of graduate school and I know that your advice will continue to stick with me during my time in industry.

I would also like to thank the members of the Schneider and Przybycien groups, who I had the privilege to work with over the past five years. Peter Hayes, Kimberly Hui, Shiny Samuel, Bowen Huo, Bruce Yan, Kyle Tynan, Lynn Mao, and Isabella Vendetti, thank you for always being helpful and encouraging.

A huge thank you to my collaborator at Rensselaer Polytechnic Institute, Yang Bai, who spent much time teaching me about next generation sequencing and who always went out of her way to be supportive.

Thank you to my collaborators at University of Pittsburgh, Millipore Sigma, and elsewhere (Anthony St. Leger, Julia Gefter, Kathleen Garwood, Leanne Labriola, Alexei Slesarev, Jaya Onuska, and Afshin Sohrabi) for their support and guidance.

Thank you to my committee members, Anne Robinson, Joel McManus and Elizabeth Wayne, for their honest guidance on my thesis and willingness to answer all of my questions.

Abstract

Despite its rare occurrence, viral infection in pharmaceutical production cultures still poses a great threat to both pharmaceutical companies economically and to the public health. However, current methodologies used to determine viral contamination have many drawbacks in addition to being both time and labor intensive. In this thesis, we make a new connection between pharmaceutical CHO-K1 production cultures and the ability of the miRNA transcriptome to indicate cellular stress – particularly in the case of viral infection. Monitoring of the host cell miRNA transcriptome is a real-time, direct measurement of host cell response to the viral infection and may be utilized in a new technique for fail-fast adventitious agent detection.

Here, we quantified virulence of 5 different viruses (Reo3, MMV, PI2, EMC, and VSV) in CHO-K1 cells. CHO-K1 cells did not show cytopathic effects after being inoculated with PI2, however, the other 4 viruses were used to identify both general and virus-specific biomarkers for infection. Overall, 10 miRNAs (cgr-miR-93-3p, cgr-miR-32-5p, cgr-miR-33, cgr-miR-17-5p, cgr-miR-24-5p, cgr-miR-18a-5p, cgr-miR-450b-5p, cgr-miR-542-3p, cgr-miR-21-3p, and cgr-miR-29b-3p) have shown promise as biomarkers for general viral infection in CHO-K1 cultures compared to healthy, uninfected cells. In addition to passing through the double filtering method of fold change and p-value, these miRNAs showed excellent discrimination ability between infected and uninfected cultures with a receiver operating curve (ROC) analysis.

The concept of host cell miRNA biomarkers was also leveraged to lay the groundwork for a more efficient diagnostic method for ocular surface infection. Using herpes simplex virus type 1 (HSV-1) infection of human corneal epithelial cells (HCLEs), miRNA biomarkers were identified to describe this technology. Overall, the top miRNA biomarker for an early host response to HSV-1 infection in HCLE cells was found to be hsa-miR-378f. Additionally, both hsa-miR-615-3p and hsa-miR-495-3p showed a linear increase in their fold change over time and were good classifiers without a proclivity towards false positives or false negatives.

Finally, we also explored the use of DNA and LNA probes as a more inexpensive option compared to the γPNAA probes previously used with the MTE method. MTE is a quicker and more cost-effective option to replace qPCR and Next Generation Sequencing in quantifying miRNA expression. As a fail-fast method, it could be easily used at numerous points during an upstream pharmaceutical manufacturing process without a large time or capital investment. It was observed that secondary structure of short nucleic acids played a large role in their detection strategy using MTE. Unstructured short nucleic acids could be detected using DNA probes; however, structured targets needed the stronger probe chemistry of LNA to overcome their self-complementarity. The ideal substitution percentage for LNA probes was found to be in the 30% range to optimize binding and prevent aggregation.

Table of Contents

List of Tables

List of Figures

Chapter 1 – miRNA Biomarkers for Viral Infection: Motivation, Background, and Thesis Objectives

1.1 Adventitious Agents in Bioprocessing

The treatment of disease with medications manufactured through bioprocessing has become the bedrock of western medicine. However, the integrity of these bioprocesses and the purity of the resulting drug substances and drug products is essential to our trust of modern-day, biopharmaceuticals. Contamination of a bioprocess or drug product can lead to changes in or create extraneous bioactivity, resulting in disastrous and confusing results for patient outcomes. An adventitious agent, as defined by the World Health Organization, is a contaminating microorganism which has been unintentionally introduced into the manufacturing process for a biological product.[1] This encompasses bacteria, viruses, fungi, parasites, mycoplasmas, and a number of other different types of agents. These agents are mainly introduced into bioprocesses through raw materials, but may also occur out of improper manufacturing procedures or unsterile conditions.

The Federal Drug Administration (FDA) has produced safety guidelines for industry to identify and assure clearance of certain types of adventitious agents.[2] Typical bioprocesses include control systems and processing redundancies to mitigate adventitious agents using what is known as the biosafety triangle. This includes prevention, detection, and removal.[34] Focus on aspects of both upstream and downstream processing are integral in the current approach to prevention of adventitious agents. In upstream processes, much of the attention is on the

sterilization of source materials and the engineering of cell lines for viral resistance.[5] In downstream purification processes, adventitious agents are removed or inactivated using methods that include filtration, chemical inactivation, and chromatography.[6] However, despite the many controls in place, some adventitious agents still slip through the safeguards with disastrous results. One key example of this happened in 2009 when Genzyme shut down a production plant for several weeks following a viral infection, resulting in shortages of Cerezyme and Fabrazyme used to treat Gaucher disease and causing permanent health consequences for some people.[7,8]

1.1.1 Economic Impact of Viral Infection in Production Cultures

In this thesis, viral adventitious agents are the primary focus. Between 1985 and 2020, there were 26 viral infection incidents in bioprocessing facilities reported to the Consortium on Adventitious Agent Contamination in Biomanufacturing (CAACB).[9] Despite their rare occurrence, viral infection in pharmaceutical production cultures still poses a great threat to both pharmaceutical companies economically and to the public health. The economic impacts of viral adventitious agent detection on a production plant are exorbitant. If viral contamination is discovered, the contaminated batch of therapeutics is discarded. Any one-time use manufacturing equipment, including protein A resins costing millions of dollars, must also be thrown away. The production facility is subsequently shut down to undergo an extensive and expensive decontamination, contributing to additional losses in opportunity costs which some estimates have put as high as 1 billion

USD.[10] This opportunity cost includes loss revenue from damage to reputation, medication shortages, and widespread public mistrust.[11]

1.1.2   Detection of Viral Adventitious Agents

The current industry standard for detection of viral adventitious agents, is to perform an in-vivo or an in-vitro virus (IVV) assay[12] at the end point of the upstream cell culturing bioprocess, before harvesting, where a potential virus has had the most time to replicate and the viral titer would be at its highest. However, IVV assays often take weeks to obtain results, by which point a recombinant-biologic product has already spent significant time in the downstream bioprocess. In-vivo assays in particular are inconvenient to industry as the pharmaceutical industry has recently moved to reduce animal testing in general.[12] Additionally, these assays traditionally only test for panels of viruses known to impact humans.

If cell culture parameters show discrepancies and a viral contamination is suspected before the harvesting step of a bioprocess, other methodologies used to determine viral contamination fall into a couple different categories; most notably, those that are used to measure viral infection in cells, and those that are used to analyze for the presence of viral particles.[13] Viral infection in cells is measured primarily through cell growth perturbation assays, such as a viral plaque assay, endpoint dilution assay, or fluorescent focus assay. Each of these has its own drawbacks in addition to being both time and labor intensive.[13,14] For methods which analyze for the presence of viral genetic material as biomarkers of viral contamination, such as Next Generation Sequencing (NGS) or quantitative polymerase chain reaction (qPCR), one shortcoming is that the viral genome(s) to be detected must already be

known. NGS is also very sensitive to the presence of any nucleic acids, including fragments, and therefore has a high likelihood of giving false positives.

Here, the difference between endogenous retrovirus-like particles (RVLP) and non-endogenous adventitious agents should be noted. Cell lines used in biomanufacturing processes produce their own RVLPs, however, these have not been shown to cause infection in other cell line and are partially reduced by clearance procedures in the downstream bioprocess. Conversely, viral adventitious agents are not expected to be present in a cell culture, and are often a concern for infection in other cell lines or organisms. These require a much more stringent testing methods which measure to lower levels of detection. Any viral contamination of this kind would result in the shutdown of the bioprocess for decontamination and the discarding of the biological product.

1.2 Introduction to miRNA biomarkers

The usefulness of microRNAs (miRNAs) as markers for cellular metabolic state and health is just beginning to be understood for applications the field of bioprocessing. Currently, miRNAs can be targeted to engineer hyper productive strains of Chinese Hamster Ovary (CHO) production cultures.[15,16] In this thesis, we draw a new link between miRNA biomarkers and the continuous monitoring of cell health through analysis the miRNA transcriptome for the detection of viral adventitious agents in pharmaceutical cultures.

1.2.1 What is a miRNA?

MicroRNA, also known as miRNA or miR, is a class of small noncoding RNA molecules first discovered in 1993.[17] These molecules are typically 18 to 22 nucleotides in length and are responsible for regulating gene expression at the post-transcriptional level. Each miRNA is capable of binding to hundreds of different types of mRNA targets to repress protein translation.[18] The study of miRNA, having emerged in the past couple decades,[19] has broad implications for an increasing number of fields including cancer research, immunology, and fundamental biology. Many studies have shown how changes in miRNA expression can be used as an indicator for various diseases, cancers, and cellular stresses as miRNA regulation is intimately tied to cell housekeeping processes and health.[20–24]

1.2.2 Canonical Biogenesis

The function of miRNAs is as short, noncoding RNA molecules which repress protein expression at the post-translational level. To recognize the potential miRNAs have as biomarkers of cellular stress, it is imperative to understand how they impact cellular expression and their biogenesis. For mammalian cells, there are a couple different pathways a miRNA can take for biogenesis – one canonical pathway which the majority of miRNAs follow, and a few other non-canonical pathways.[25] The canonical biogenesis pathway, illustrated in Figure 1.1, consists of five steps: transcription, cleavage by DROSHA, nuclear export, cleavage by DICER, and finally strand selection.[26,27] First, the pri-miRNA transcript is generated by RNA polymerase II, although a few are transcribed by RNA

polymerase III.[28] Pri-miRNA stands for primary miRNA and should not be confused with pre-miRNA which stands for precursor miRNA as explained in step 2. The pri-miRNA transcript forms an imperfect hairpin loop with the 3' and 5' ends extending outwards.[29] Second, these ends are cleaved off by DROSHA, a ribonuclease enzyme which is stabilized by DGCR8, an RNA binding protein. This cleavage leaves a protruding single stranded 3' end which is two or three base pairs long.[30] DROSHA is thought to recognize the basal junction as a reference point for the cleavage point of the hairpin, eleven base pairs away. This newly cleaved hairpin is called a pre-miRNA transcript. Third, as transcription takes place in the nucleus, the pre-miRNA transcript is then exported through a nuclear pore out into the cytoplasm by a protein complex of Ran-GTP and Exportin5. Fourth, once in the cytoplasm, the pre-miRNA transcript binds to the RISC complex and is subject to another cleavage by DICER, another ribonuclease. The RISC complex consists of the TRPB, DICER, and AGO proteins. TRPB recognizes the pre-miRNA transcript and keeps DICER from cleaving other substrates. Once bound, DICER cleaves off the hairpin end of the transcript to create a double stranded miRNA with a protruding 5' end. Fifth, AGO selects one of the strands of the double stranded miRNA for degradation. This passenger strand is removed and degraded while the other guide strand, also known as the mature miRNA sequence, stays incorporated in the RISC complex. The seed region of this guiding strand, which is base pairs two through seven on the 5' end, guides the RISC complex towards finding a complementary mRNA sequence.[31] In mammalian cells, this seed region then binds to the 3' UTR of a target mRNA to mark it for degradation by P bodies[32]. As

mRNAs are translated to produce proteins, this degradation process suppresses protein expression at the post transcriptional level.
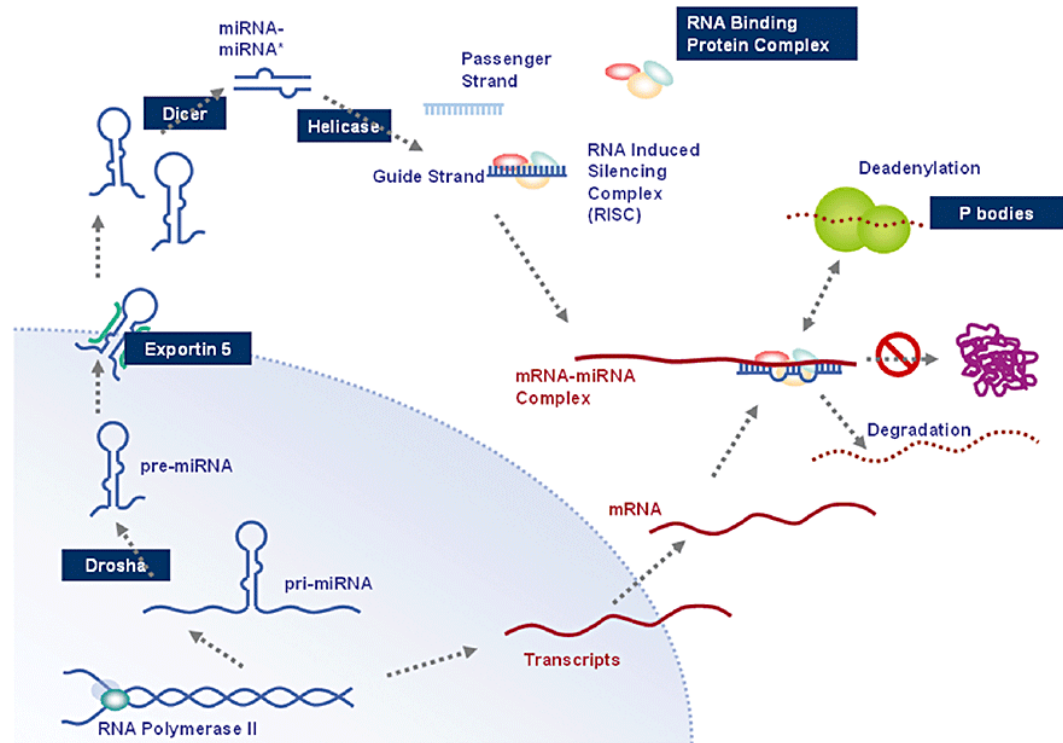
Figure 1.1 Schematic figure from Scaria et al. (2007) of the biogenesis process and mRNA silencing mechanism of action for miRNA.[22]

1.2.3   Non-canonical Biogenesis

Non-canonical pathways of miRNA biogenesis are typically similar to the canonical pathway, but differ in one or two steps. Here, we will divide the non-canonical pathways into two different categories: DROSHA independent pathways and DICER independent pathways. The first DROSHA independent pathway is called the mirtron pathway.[33] In this case, the microRNA comes from an intron, found between two exons. Instead of DROSHA, a spliceosome splices the two exons together and creates a lariat structure with the intron, which then forms the hairpin structure of a pre-miRNA transcript. In other DROSHA independent

7

pathways, a pre-miRNA transcript can also be derived from small nucleolar RNAs (snoRNAs) or short hairpin RNAs (shRNAs).[34] However, much about these pathways is still unknown. In the second category of DICER independent pathways, mir-451 is known to undergo cleavage by AGO2 instead of DICER to form a mature miRNA.

1.2.4 Nomenclature

miRNA biogenesis is the basis for understanding miRNA nomenclature. A miRNA is typically named as following: species of origin – type of sequence –

mmu-miR-1a-1-5p
a   b   c d e   f

Figure 1.2 miRNA nomenclature: a) Species of origin (Mus musculus) b) Sequence type c) miRNA family d) Family member e) Chromosomal location f) Strand of origin.

family/member – chromosomal location – strand of origin. Species of origin is delineated with the first letter of the genus and first two letters of the species. For example, *Homo sapien* would be *hsa* or *Mus musculus* would be *mmu*. For the type of sequence, mir indicates a precursor sequence, miR indicates a mature miRNA, and MIR indicates an entire family. MicroRNA families encompass a group of miRNAs which derive from a similar ancestor. Each miRNA family has a number and subsequent family members are given their own letter (1a, 1b, 1c). Most miRNAs do not need to have their chromosomal location indicated unless the same miRNA arises from more than one location in the genome. In this case, the locations are numbered. Last, the strand of origin is either -3p or -5p, depending on which end of the stem loop sequence the mature miRNA originated from. In another variation of this nomenclature, an * symbol is added to indicate the passenger strand. Ultimately, as an example, the full name of a miRNA will turn out to be as

such: mmu-miR-1a-1-5p. It is worth noting that some miRNAs do not follow this pattern of nomenclature, such as lin-4 and let-7. These are very specific cases in which historical naming took precedent over a new system of nomenclature.

1.3   Why miRNA?

In order to recognize the use of miRNAs as biomarkers, the qualities of an ideal biomarker must first be understood. The first and most obvious requirement of a biomarker is that it shows a statistically significant correlation with the condition or disease it is being used to detect.[35] Furthermore, a biomarker should be specific enough that it rules out other possible conditions which may present similarly. A panel of biomarkers used as a composite indicator or fingerprint may be helpful for increasing the specificity relative to reliance on a single biomarker. The false positive and false negative rate (or sensitivity and specificity) should be known and optimized for the application.

A biomarker must be reproduceable over the intended population of use. Ideally, it should not be overly dependent on random outside forces other than the condition to be detected. Therefore, in screening for a biomarker, factors such as age of population/culture and along with many others must be taken into account and corrected for. Moreover, a biomarker is useless if it cannot be detected. Ease of detection is a necessity for an ideal biomarker.

Finally, biomarkers are not practical if they cannot be accessed. Some biomarkers can only be accessed through invasive means such as biopsies, in which case there is a tradeoff between necessity and the level of invasion. However, an ideal

biomarker would be minimally invasive in correlation with necessity. Additionally, a biomarker must also have enough stability to be detectable between the time it is accessed and subsequently assayed.

Thousands of scientific studies are performed each year to determine miRNA biomarkers for illnesses ranging from infectious diseases to cancer and even to trauma.[36] Given the large amount of time and money invested into these studies, it is apparent that miRNA biomarkers are growing in popularity and becoming fundamental to the scientific literature.[20–24] Below, ten reasons are outlined why miRNAs make ideal biomarkers and are preferable when compared to other possible biological markers.

1.3.1 Response to cellular stresses

Most importantly, it is well established that miRNAs change in response to a variety of cellular stressors. Differential expression of miRNAs has been observed with viral infections, bacterial infections, physical trauma, and cancer to name a few categories. Additionally, these changes have been studied and established for a wide range of organisms: not only the standard human, mouse, and rat specimens, but also for more obscure organisms including crustaceans, tomato plants, and flounder.[37–39]

1.3.2 Either general or specific

MicroRNA biomarkers have the potential to be either general or specific. MicroRNAs help control cellular homeostasis and each one can target hundreds of mRNAs, therefore, a miRNA is more likely to be a marker of general stress than other types of RNA. However, the use of multiple miRNAs as biomarkers can be

used to create a "fingerprint" for a specific cellular stress or disease.[40] A receiver operating curve analysis may be utilized to determine highly specific combinations of biomarkers for differentiation of different cellular stresses.

1.3.3 Homology between species

MicroRNAs are known to be highly conserved and maintain a high degree of homology between species. Often, the discovery of novel miRNAs in one organism can be followed up by an orthologue search for the same miRNAs in other organisms.[41,42] MicroRNAs contribute towards metabolic regulation and maintaining cellular homeostasis, making them a fundamental building block of biological function.[43] This is an advantage for biomarker discovery, as a biomarker for disease in one organism may translate to other organisms.[44]

It should be noted that miRNAs in plants function differently from miRNAs in animal species. Animal miRNAs have a short seed region which binds to a complementary mRNA within the 3' untranslated region to mark it for degradation. Most plant miRNAs bind to sites in coding regions.[45] Plant miRNAs also utilize the 5' end of their sequence to pair as well as the 3' end, which results in plants having fewer targets per miRNA than animals.[46] The animal and plant kingdoms each retain their own conserved miRNAs,[38] but due to the differences in form and function, crossover between the kingdoms should not be expected.

1.3.4 Small number of targets

The number of possible miRNA targets is small compared to the number other RNAs, such as mRNA, that exist in a single organism. The total number of experimentally validated miRNAs in humans is around 2,300.[47] This relatively low

number compared to the hundreds of thousands of mRNAs present allows for a wider array of experimental methods for comprehensive biomarker screening, including microarrays or qPCR.

1.3.5 Ubiquity

MicroRNAs are ubiquitous. From the canonical biogenesis process, most miRNAs begin inside the cell nucleus and end up in the cytoplasm as part of the RISC complex. However, miRNAs have also been found in numerous other places beyond cellular cytoplasm, many of which have yet to be understood. Extracellular miRNAs circulate freely throughout the body in blood serum and other fluids. MicroRNAs are known to be present in at least 12 different body fluids, including tears, saliva, breast milk, and spinal fluid.[48] Exosomal miRNAs contribute to communication between cells through "bioactive vesicles" in which miRNAs are transported.[49] MicroRNAs have also been found present in mitochondria, endomembranes, and other organelles.[50,51]

1.3.6 Accessibility and Abundancy

The ubiquity of miRNA results in their accessibility. Many of the fluids in which miRNAs can be found are able to be obtained noninvasively, such as saliva, ocular fluid, urine, and breast milk. This is extremely important when it comes to using diagnostic biomarkers clinically, as invasive procedures can cause longer recovery times and complications, leading to negative health outcomes for patients and an increased health care burden. Many miRNAs are also present in large enough quantities to be easily measurable. The range of copies of a single miRNA per cell varies from less than 10 to over 30,000.[52]

1.3.7 Direct measurement of host health

Monitoring of the host cell miRNA transcriptome is a real-time, direct measurement of host cell health. Expression of miRNAs provides the benefit of gaining insight into the overall health of an organism or culture and not just the presence of a pathogen. This offers an advantage over using a virus or bacteria as a biomarker for a couple of reasons. First, the transcriptome is likely to change before the pathogen proliferates sufficiently to become detectable. Second, testing for the presence of the pathogen without understanding how it is affecting the host can lead to misdiagnosis or incorrect treatment. Organisms may be carriers of a pathogen without needing treatment, as many pathogens are opportunistic.

Additionally, a direct measurement of host cell health could be an important step into the world of personalized medicine. Measuring host response directly would allow for the tailoring of treatment specifically to that individual and the response to treatment could also be measured directly.

1.3.8 Early Detection

Host cell miRNA transcriptomes have been shown to display differential expression before any cytopathic effects could be visualized or symptoms noticed by a patient.[53,54] This gives the advantage of both early detection and detection of latent issues.[55] Early detection is particularly important in diseases which have a long incubation period or for which early treatment is essential.

1.3.9 Latent Infections

Diagnosis of latent infections presents a challenge due to a lack of symptoms and other visual cytopathic effects. However, transcriptomic analysis is able to identify

issues at the genomics level which cannot be visually observed. As a real-time direct measurement of cell health, miRNAs have been shown to be useful in identifying latent issues which would otherwise be difficult to detect. Most current research of miRNAs for detection of a latent infection is focused on latent tuberculosis.[56,57]

1.3.10 Stability

Stability is an important characteristic of a biological biomarker, and given the short length of these RNA molecules, degradation potential may be thought to be a concern. However, miRNAs are observed to be less susceptible to ribonuclease degradation than mRNA and other RNA classes.[58] It has been shown that miRNA isolations stored at $-80°C$, can remain stable for a minimum of 10 months.[59] In another study, miRNA stability in plasma samples was observed for up to 4 years of freezing time, at which point miRNAs were seen to start degradation at different rates depending on AU content.[60] One miRNA in particular withstood a freezing time of 14 years. Additionally, circulating miRNAs have been shown to withstand up to 8 freeze-thaw cycles and incubation at room temperature for 24 hours with minimal changes.[61] Ultimately, the stability of miRNAs has been well studied and established for numerous types of fluids, including serum, blood, plasma, and even urine.[60–63]

1.4 Challenges of miRNA Biomarkers

Since the discovery of microRNA (miRNA or miR) in 1993,[17] this field of study has burgeoned and numerous applications have been discovered. In the past decade, thousands of papers have been published on miRNA biomarker diagnostics every

year.[36] However, as of 2019 there were only 8 companies with products utilizing miRNA diagnostic biomarkers either available commercially or in development.[36] MicroRNAs are not currently widely used as biomarkers in clinical practice and there are a few key reasons why.

1.4.1 Lack of Standard Procedures

Due to the relative newness of this field, it can be seen in the literature that the identification and utilization of miRNA biomarkers lacks standard protocols. There is a vast array of options for isolation of RNA with different experimental methods, and even more options for how data is analyzed once the RNA is extracted. However, one standard which seems to be consistent is the use of qPCR as the "gold standard" for biomarker validation. Despite this, many different types of statistical tests are still used with different cutoffs for significant p-value and fold change to determine differential expression. More recently, some miRNA studies have also been done using machine learning to predict biomarkers.[55] Due to the vast number of algorithm options, a standard protocol for this analysis would be beneficial. Once standard procedures are developed and the challenges of clinical application are increasingly overcome, miRNA biomarkers could become more widely utilized clinically.

1.4.2 Challenging Detection

Small nucleic acids have traditionally presented challenges for detection due to their short nature. Nevertheless, many different detection techniques have been created to overcome this. Methods such as NGS, microarrays, and qPCR have all been shown to be successful in detection of miRNAs, each with their own

advantages and disadvantages. Although some rare miRNAs have as few as 10 copies per cell,[52] a number of different methods have shown potential to detect miRNAs down to the 1 fM concentration range.[64]

1.4.3 Few Clinical Biomarkers

Finally, though an abundant number of studies have been done on determining miRNA biomarkers, they are not yet widely used for diagnostics either commercially or clinically. A couple of reasons may account for this. First, is the difficulty of translating research into clinical practice. Much upfront capital must be invested in a product or diagnostic to bring it through the development phase to be commercially available, particularly in terms of satisfying regulatory requirements. Additionally, the reproducibility of miRNA biomarker-based assays has been shown to be an issue, showing the need for homogeneous protocols for this relatively new field of research.[65] Finally, the variability of clinical subjects impacts the usefulness of miRNA biomarkers in practice. Life may be unlike the controlled nature of a research lab, and a variety of other environmental factors are likely play a role in cellular expression of miRNAs.

1.5 Thesis Objectives

In this thesis, we focus on the identification and utilization of host cell miRNA biomarkers for viral infection. This concept has significant applications in both bioprocessing and the medical field, which are addressed with the following objectives.

1. **Validation that CHO-K1 cells show changes in miRNA transcriptome with viral infection.** Before a comprehensive search for biomarkers, it is necessary to validate that the CHO-K1 cells found in many pharmaceutical production cultures do change their miRNA transcriptome with viral infection. This first step is necessary groundwork to build the rest of the thesis.

2. **Identification of general and specific biomarkers for viral infection.** In this thesis, we make a new connection between pharmaceutical CHO production cultures and the ability of the miRNA transcriptome to indicate cellular stress – particularly in the case of viral infection. Monitoring of the host cell miRNA transcriptome is a real-time, direct measurement of host cell response to the viral infection. We will compare miRNA transcriptome changes between viral infection caused by double-stranded and single-stranded, enveloped and non-enveloped, and DNA and RNA viruses to gain insight on how changes in miRNA expression are connected with the modes of propagation for different types of viruses. This will lead to the identification of host cell biomarkers which are specific to each virus. The identification of general stress biomarkers gives our method an additional benefit better understanding the overall health of production cultures. This is important as unhealthy production cultures could produce unforeseen consequences, including variability in the critical quality attributes of secreted protein products. In addition to the importance for the pharmaceutical industry and advances in the healthcare field, this work will

contribute to a more comprehensive understanding of the fundamental modes of action of viral infection in CHO-K1 cells.

3. **Analysis of herpes simplex virus (HSV) miRNA biomarkers in Human Corneal Limbal Epithelial (HCLE) cells.** We will take the data analysis methods and experimental protocols developed in the previous chapters and apply them to a new system of HSV infected ocular fluid to identify useful clinical biomarkers. Identifying miRNA biomarkers in this model system will lay the groundwork for the potential use of noninvasive ocular fluid for diagnosis of HSV in human patients.

4. **Examination of data analysis methods.** One of the current challenges of miRNA biomarkers is the lack of standard protocols for identifying and utilizing them. In this thesis, we will analyze and compare the merits of different analysis methods including fold change, P-value, programs to analyze large data sets, receiver operating curves, and machine learning.

5. **Explore the detection of short nucleic acid targets using micelle tagging electrophoresis (MTE) with DNA and LNA probes.** In this thesis, we explore the use of LNA probes as a more inexpensive option compared to the γPNAA probes previously used with the MTE method. MTE is a quicker and more cost-effective option to replace qPCR and Next Generation Sequencing in quantifying miRNA expression and as a fail-fast method, it could be easily used at numerous points during an upstream pharmaceutical manufacturing process (Figure 1.3) without a large time or capital investment. This research is not only important for use in the

pharmaceutical industry, but has far reaching implications for any field that necessitates the detection of miRNA.
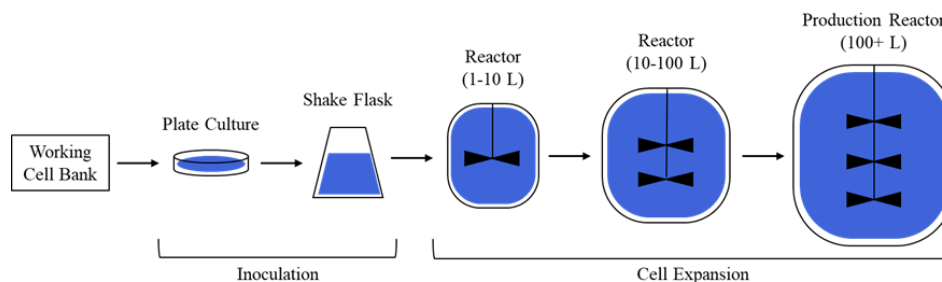
Figure 1.3 Monitoring of miRNA biomarkers may take place at any point in an upstream bioprocess.

1.6 References

1. World Health Organization. Annex 3 Recommendations for the evaluation of animal cell cultures as substrates for the manufacture of biological medicinal products and for the characterization of cell banks. Replacement of Annex 1 of WHO Technical Report Series, No. 878. 89–90 (2013).

2. Food and Drug Administration. Guidance for Industry Q5A Viral Safety Evaluation Derived From Cell Lines of Human or Animal Origin Guidance for Industry. 33 (1998).

3. Morrow, J. K. Advances in Upstream Technologies Reduce Viral-Contamination Risks. *Viral Safety Merck* (2017).

4. Shiratori Masaru and Kiss, R. Risk Mitigation in Preventing Adventitious Agent Contamination of Mammalian Cell Cultures. in *New Bioprocessing Strategies: Development and Manufacturing of Recombinant Antibodies and Proteins* (ed. Kiss Bob and Gottschalk, U. and P. M.) 75–93 (Springer International Publishing, 2018). doi:10.1007/10_2017_38.

5. Ferreira, M. M. Laying the Foundation for Viral Safety: Mitigating the Risk of Viral Contamination in Vaccines, Cell, and Gene Therapies. *Viral Safety Merck* (2017).

6. DePalma, A. Viral Safety in Monoclonal Antibody Manufacturing: Various Technologies to Prevent, Detect, and Remove Virus Contamination. *Viral Safety Merck* (2017).

7. Pollack, A. Genzyme Drug Shortage Leaves Users Feeling Betrayed. *New York Times* (2010).

8. Armstrong, D. Genzyme Shuts Down Plant Tainted by Virus. *The Wall Street Journal* https://www.wsj.com/articles/SB124515915116918837 (2009).

9. Barone, P. W. *et al.* Viral contamination in biologic manufacture and implications for emerging therapies. *Nat Biotechnol* **38**, 563–572 (2020).

10. Westman, D. Cost and impact of a bioburden incident. Preprint at (2017).

11. Phuong, J. M., Penm, J., Chaar, B., Oldfield, L. D. & Moles, R. The impacts of medication shortages on patient outcomes: A scoping review. *PLoS One* **14**, e0215837 (2019).

12. Slawinski, K. Keeping Up with Viral Safety Trends in Bioprocessing: Next Generation Sequencing and Quality by Design. *Viral Safety Merck* (2017).

13. Rodruiges, K. & Hebert, C. Why measurement of viral infectivity matters and how to improve it. https://www.lumacyte.com/the-cyte-blog/2018/2/21/why-measurement-of-viral-infectivity-matters-and-how-to-improve-it (2018).

14. Adair, R. Viral Contaminant Testing in Biopharma Manufacturing. https://www.contractpharma.com/issues/2018-11-01/view_features/viral-contaminant-testing-in-biopharma-manufacturing/ (2018).

15. Schad, M. *et al.* A functional high-content miRNA screen identifies miR-30 family to boost recombinant protein production in CHO cells. *Biotechnol J* **9**, 1279–1292 (2014).

16. Hackl, M. *et al.* Next-generation sequencing of the Chinese hamster ovary microRNA transcriptome: Identification, annotation and profiling of microRNAs as targets for cellular engineering. *J Biotechnol* **153**, 62–75 (2011).

17. Lee, R. C., Feinbaum, R. L. & Ambrost, V. *The C. elegans Heterochronic Gene lin-4 Encodes Small RNAs with Antisense Complementarity to &II-14*. *Cell* vol. 75 (1993).

18. Leva, G. di, Calin, G. A. & Croce, C. M. MicroRNAs: Fundamental facts and involvement in human diseases. *Birth Defects Res C Embryo Today* **78**, 180–189 (2006).

19. Almeida, M. I., Reis, R. M. & Calin, G. A. MicroRNA history: Discovery, recent applications, and next frontiers. *Mutation Research - Fundamental and Molecular Mechanisms of Mutagenesis* **717**, 1–8 (2011).

20. Leung, A. K. L. & Sharp, P. A. MicroRNA Functions in Stress Responses. *Mol Cell* **40**, 205–215 (2010).

21. Lu, L. F. & Liston, A. MicroRNA in the immune system, microRNA as an immune system. *Immunology* **127**, 291–298 (2009).

22. Hariharan, M., Scaria, V., Brahmachari, S. K., Pillai, B. & Maiti, S. Host-virus genome interactions: macro roles for microRNAs. *Cell Microbiol* **9**, 2784–2794 (2007).

23. Skalsky, R. L. & Cullen, B. R. Viruses, microRNAs, and Host Interactions. *Annu Rev Microbiol* **64**, 123–141 (2010).

24. Marquez, R. T. *et al.* Correlation between microRNA expression levels and clinical parameters associated with chronic hepatitis C viral infection in humans. *Laboratory Investigation* **90**, 1727–1736 (2010).

25. Chong, M. M. W. *et al.* Canonical and alternate functions of the microRNA biogenesis machinery (Genes & Development (2010) 24, (1951-1960)). *Genes Dev* **24**, 2228 (2010).

26. Murchison, E. P. & Hannon, G. J. miRNAs on the move: miRNA biogenesis and the RNAi machinery. *Curr Opin Cell Biol* **16**, 223–229 (2004).

27. Siomi, H. & Siomi, M. C. Posttranscriptional Regulation of MicroRNA Biogenesis in Animals. *Mol Cell* **38**, 323–332 (2010).

28. Hahn, S. Structure and mechanism of the RNA Polymerase II transcription machinery. **11**, 394–403 (2005).

29. Krol, J. *et al.* Structural features of microRNA (miRNA) precursors and their relevance to miRNA biogenesis and small interfering RNA/short hairpin RNA design. *Journal of Biological Chemistry* **279**, 42230–42239 (2004).

30. Zeng, Y., Yi, R. & Cullen, B. R. Recognition and cleavage of primary microRNA precursors by the nuclear processing enzyme Drosha. *EMBO Journal* **24**, 138–148 (2005).

31. Lam, J. K. W., Chow, M. Y. T., Zhang, Y. & Leung, S. W. S. siRNA versus miRNA as therapeutics for gene silencing. *Mol Ther Nucleic Acids* **4**, e252 (2015).

32. Nilsen, T. W. Mechanisms of microRNA-mediated gene regulation in animal cells. *Trends in Genetics* **23**, 243–249 (2007).

33. Westholm, J. O. & Lai, E. C. Mirtrons: MicroRNA biogenesis via splicing. *Biochimie* vol. 93 1897–1904 Preprint at https://doi.org/10.1016/j.biochi.2011.06.017 (2011).

34. Abdelfattah, A. M., Park, C. & Choi, M. Y. Update on non-canonical microRNAs. *Biomolecular Concepts* vol. 5 275–287 Preprint at https://doi.org/10.1515/bmc-2014-0012 (2014).

35. Holland, R. L. What makes a good biomarker? *Advances in Precision Medicine* **1**, 66 (2016).

36. Guire, D. v. *How close are miRNAs from clinical practice? A perspective on the diagnostic and therapeutic market*.

37. Zhang, B. C., Zhang, J. & Sun, L. In-depth profiling and analysis of host and viral microRNAs in Japanese flounder (Paralichthys olivaceus) infected with megalocytivirus reveal involvement of microRNAs in host-virus interaction in teleost fish. *BMC Genomics* **15**, (2014).

38. Yin, Z., Li, C., Han, X. & Shen, F. Identification of conserved microRNAs and their target genes in tomato (Lycopersicon esculentum). *Gene* **414**, 60–66 (2008).

39. Huang, T., Xu, D. & Zhang, X. *Characterization of host microRNAs that respond to DNA virus infection in a crustacean*. http://www.biomedcentral.com/1471-2164/13/159 (2012).

40. Wegman, D. W. & Krylov, S. N. Direct miRNA-hybridization assays and their potential in diagnostics. *TrAC - Trends in Analytical Chemistry* **44**, 121–130 (2013).

41. Weber, M. J. New human and mouse microRNA genes found by homology search. *FEBS Journal* **272**, 59–73 (2005).

42. Veeranagouda, Y. *et al.* Identification of MicroRNAs in Macaca fascicularis (Cynomolgus monkey) by homology search and experimental validation by small RNA-seq and RT-qPCR using kidney cortex tissues. *PLoS One* **10**, (2015).

43. Hartig, S. M., Hamilton, M. P., Bader, D. A. & McGuire, S. E. The miRNA Interactome in Metabolic Homeostasis. *Trends in Endocrinology and Metabolism* vol. 26 733–745 Preprint at https://doi.org/10.1016/j.tem.2015.09.006 (2015).

44. Koenig, E. M. *et al.* The beagle dog MicroRNA tissue atlas: Identifying translatable biomarkers of organ toxicity. *BMC Genomics* **17**, (2016).

45. Millar, A. A. & Waterhouse, P. M. Plant and animal microRNAs: Similarities and differences. *Functional and Integrative Genomics* vol. 5 129–135 Preprint at https://doi.org/10.1007/s10142-005-0145-2 (2005).

46. Wang, J., Mei, J. & Ren, G. Plant microRNAs: Biogenesis, homeostasis, and degradation. *Frontiers in Plant Science* vol. 10 Preprint at https://doi.org/10.3389/fpls.2019.00360 (2019).

47. Alles, J. *et al.* An estimate of the total number of true human miRNAs. *Nucleic Acids Res* **47**, 3353–3364 (2019).

48. Weber, J. A. *et al.* The microRNA spectrum in 12 body fluids. *Clin Chem* **56**, 1733–1741 (2010).

49. Hu, G., Drescher, K. M. & Chen, X. M. Exosomal miRNAs: Biological properties and therapeutic potential. *Front Genet* **3**, (2012).

50. Sripada, L., Tomar, D. & Singh, R. Mitochondria: One of the destinations of miRNAs. *Mitochondrion* vol. 12 593–599 Preprint at https://doi.org/10.1016/j.mito.2012.10.009 (2012).

51. Kim, Y. J., Maizel, A. & Chen, X. Traffic into silence: Endomembranes and post-transcriptional RNA silencing. *EMBO Journal* vol. 33 968–980 Preprint at https://doi.org/10.1002/embj.201387262 (2014).

52. Chen, C. *et al.* Real-time quantification of microRNAs by stem-loop RT-PCR. *Nucleic Acids Res* **33**, 1–9 (2005).

53. Farr, R. J. *et al.* Altered microRNA expression in COVID-19 patients enables identification of SARS-CoV-2 infection. *PLoS Pathog* **17**, (2021).

54. Biswas, S., Haleyurgirisetty, M., Lee, S., Hewlett, I. & Devadas, K. Development and validation of plasma miRNA biomarker signature panel for the detection of early HIV-1 infection. *EBioMedicine* **43**, 307–316 (2019).

55. Tribolet, L. *et al.* MicroRNA Biomarkers for Infectious Diseases: From Basic Research to Biosensing. *Frontiers in Microbiology* vol. 11 Preprint at https://doi.org/10.3389/fmicb.2020.01197 (2020).

56. Hashimoto, S. *et al.* Developing a diagnostic method for latent tuberculosis infection using circulating miRNA. *Transl Med Commun* **5**, (2020).

57. Wang, C. *et al.* Comparative mirna expression profiles in individuals with latent and active tuberculosis. *PLoS One* **6**, (2011).

58. Aryani, A. & Denecke, B. In vitro application of ribonucleases: Comparison of the effects on mRNA and miRNA stability. *BMC Res Notes* **8**, (2015).

59. Mraz, M., Malinova, K., Mayer, J. & Pospisilova, S. MicroRNA isolation and stability in stored RNA samples. *Biochemical and Biophysical Research Communications* vol. 390 1–4 Preprint at https://doi.org/10.1016/j.bbrc.2009.09.061 (2009).

60. Balzano, F. *et al.* MiRNA stability in frozen plasma samples. *Molecules* **20**, 19030–19040 (2015).

61. Mitchell, P. S. *et al. Circulating microRNAs as stable blood-based markers for cancer detection. PNAS* (2008).

62. Mall, C., Rocke, D. M., Durbin-Johnson, B. & Weiss, R. H. Stability of miRNA in human urine supports its biomarker potential. *Biomark Med* **7**, 623–631 (2013).

63. Gilad, S. *et al.* Serum microRNAs are promising novel biomarkers. *PLoS One* **3**, (2008).

64. Goldman, J. M., Kim, S., Narburgh, S., Armitage, B. A. & Schneider, J. W. Rapid, multiplexed detection of the let-7 miRNA family using γPNA amphiphiles in micelle-tagging electrophoresis. *Biopolymers* **113**, (2022).

65. Condrat, C. E. *et al.* miRNAs as Biomarkers in Disease: Latest Findings Regarding Their Role in Diagnosis and Prognosis. *Cells* vol. 9 Preprint at https://doi.org/10.3390/cells9020276 (2020).

Chapter 2 – miRNA Changes in CHO-K1 Host Cells with Viral Infection

2.1 Introduction

Chinese hamster ovary (CHO) derived cell lines are widely utilized in the recombinant-biologic industry for the production of numerous types of biological products including enzymes, interferons, and monoclonal antibodies.[1] CHO is used as an ideal host expression system due to its adaptability in manufacturing, history of regulatory approval, and ability to produce complex therapeutics.[2] CHO-K1 specifically, is a subclone of the original CHO cell line and is estimated to be used in the manufacturing of 70% of approved biotech products.[1] In the year 2016, CHO-K1 based bioprocessing systems produced more than $50 billion USD of therapeutic products.[3]

Since 1985, CHO production cultures have accounted for at least 50% of reported major viral contamination events in biomanufacturing, making this cell line an important component of study for adventitious agent detection.[4] For the majority of contamination events, the leading indicator of infection was observed abnormalities in cell culture parameters. For validation, the most commonly used technique for detection of viral adventitious agents was polymerase chain reaction (PCR) using virus-specific probe libraries, followed by in-vitro virus (IVV) testing and electron microscopy.[4] This is problematic, as not all viruses may impact cell culture performance significantly enough to indicate infection. Additionally, the current

techniques used for detection are inefficient, being both time and labor consuming, resulting in dire economic costs in the occurrence of a real contamination event.

The possibility of identifying host cell miRNA biomarkers for viral infection in bioreactors was indicated as a possibility from literature studies which identify miRNA biomarkers correlated with disease states including viral infection, as well as by the extensive overlap between the human and CHO-K1 miRNA transcriptomes. However, changes to the miRNA transcriptome of CHO-K1 cells had not yet been studied under the conditions of viral infection. Therefore, it was necessary to complete a preliminary experimental screen before the commencement of a comprehensive biomarker screen to establish that the miRNA transcriptome in CHO-K1 cells does respond to viral infection. In this work, we utilized a qPCR platform to analyze fold change differences for infected and uninfected total RNA samples. While it is possible to make preliminary observations of potential biomarkers from this work, the primary goal of these preliminary experiments was to only validate that changes in miRNA do occur with viral contamination in CHO-K1 cells. To move forward with a comprehensive biomarker screening, we expected to see more than one miRNA with significant upregulation or downregulation and numerous miRNAs with insignificant expression changes as measured by fold change compared to a healthy, uninfected control.

Despite the fact that miRNA biomarkers have not yet been explored with CHO in regards to utility as an indicator of viral infection, there is still evidence to support the hypothesis that such (a) biomarker(s) exists. Many research studies have shown

25

that the miRNA transcriptome in other organisms changes with viral infection. From tomatoes to crustaceans[5,6] and beyond, this is a well-established phenomenon.[7–14] As miRNAs are highly conserved between species it can be inferred from these pieces of information that CHO will show changes in its miRNA transcriptome upon viral infection. Furthermore, due to this homology, it is possible that viral infection biomarkers may also be shared between some species. CHO-K1 miRNA transcriptome

Although changes to the miRNA transcriptome of CHO-K1 cells had not yet been studied for viral infection, there were previous studies on CHO-K1 transcriptome changes in response to general stress.[15] Johnson et al. showed how libraries of CHO-K1 cultured under differing conditions, including temperature shifts and sodium butyrate treatment, presented changes in miRNA expression.[16] The CHO-K1 miRNA genome has been sequenced by three different groups[16–19] and this data is included in the miRBase database under the listing for *Cricetulus griseus* (Chinese Hamster).[20–25] Currently, there are 245 stem loop and 351 mature miRNA entries for *Cricetulus griseus* in the miRBase repository.[20,24] This would include miRNAs from any CHO cell line or the organism as a whole. Although the total number of miRNAs may seem low, this is typical as there are relatively few miRNAs in any organism compared to other more abundant RNAs such as mRNA. However, the miRBase list is likely not complete as the study of miRNA is a relatively new field and miRNA expression is known to change under different cellular culture conditions. In sequencing CHO-K1 cells under the condition of

viral infection, we expect to that we may identify miRNAs not listed in this database for CHO.

Most of the miRBase entries under *Cricetulus griseus* were compiled from studies utilizing the CHO-K1 cell line specifically. However, the miRBase repository may also contain miRNAs from other CHO cell lines, many of which have different chromosomal rearrangements and possible deletions and mutations.[26] Therefore, it was necessary to determine how many and which of the listed miRNAs were present in specifically in CHO-K1. This was determined by comparing the miRNA sequences from miRBase to a known CHO-K1 genome obtained from chogenome.org.[27] Stem loop sequences were compared using blastn, and mature miRNA sequences were compared using blastn-short on the public server at usegalaxy.org. Human miRNAs listed in miRBase were also blasted against the CHO-K1 genome to determine the extent of homology between the two organisms.
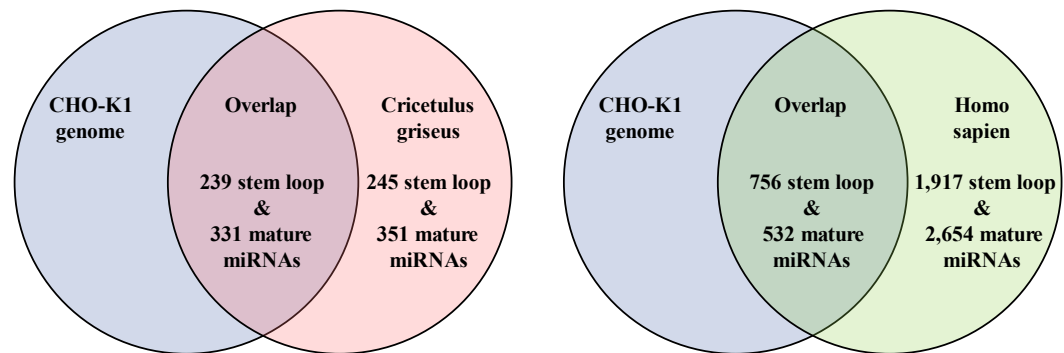
Figure 2.1 Blastn results comparing a) C. griseus and b) H. sapien miRNAs with the CHO-K1 genome.

The Venn diagrams of Blastn results in **Error! Reference source not found.** show the overlap between the CHO-K1 genome and miRNAs from *Cricetulus griseus* and *Homo sapien*. Overlap miRNAs had no mismatches with an expectation value threshold of 10^{-4}. The CHO-K1 genome and *Cricetulus griseus* miRNAs show

almost complete overlap. *Homo sapien* miRNAs show significantly less overlap with the CHO-K1 genome, however the predicted homology is still approximately 40% for stem loop miRNAs and 20% for mature miRNAs. A full list of CHO-K1 mature miRNAs utilized for subsequent experimental work and data analysis may be found in the appendix.

2.2   Literature Search for Potential miRNA Biomarker Targets

After identifying miRNAs present in CHO-K1, a literature search was conducted to put together a list of potential biomarkers which we predicted may be indicators of viral infection. Even though there are no miRNA biomarkers for viral infection in the literature specifically for CHO-K1, a multitude of studies have shown host miRNA changes with viral infection in organisms other than CHO.[5,28–34] A list of potential miRNA targets, shown in Table 2.1, was compiled from studies of miRNA biomarkers for viral infection in those other organisms. The sequences of potential miRNA biomarkers were cross referenced with the sequences CHO-K1 miRNAs to ensure homology.

Potential miRNA Targets				
miR-9	miR-34	miR-145	miR-196	miR-296
miR-17	miR-99	miR-146	miR-213	miR-320
miR-20	miR-100	miR-155	miR-219	miR-351
miR-26a	miR-101	miR-181	miR-221	miR-431
miR-27	miR-106	miR-190	miR-222	miR-448
miR-29a	miR-125b	miR-193	miR-223	let-7b
miR-33				

Table 2.1 List of potential miRNA targets compiled from literature search.

2.3　Materials and Methods

2.3.1　Experimental Design

To determine whether the CHO-K1 miRNA transcriptome changes with viral infection, CHO-K1 cells were infected with Orthoreovirus 3 (Reo3). Reo3 was selected based on its reported occurrence in adventitious bioreactor infections.[4] Four cultures were infected to 10 $TCID_{50}$ with Reo3 and four were left uninfected as healthy controls. Total RNA was extracted on days 3 and 7 post-infection with biological replicates.

2.3.2　Cell Culture

For each sample, 5 mL wells in 6-well plates were seeded with 400,000 CHO-K1 cells. Each well contained 4 mL of F-12K nutrient mixture with 10% fetal bovine serum (FBS) and was cultured overnight at 37°C. Divided into a control group and a viral infection group, the cultures in the viral infection group were infected after 24 hours to 10 $TCID_{50}$. On this day 0, the infection day, culture media was changed for both groups. The virally infected group received a nutrient mixture with 2% FBS while the control group nutrient mixture contained 10% FBS. Higher percentages of FBS are known to reduce cellular uptake of a virus. The CHO-K1 cultures were incubated at 37°C until miRNA was isolated on days 3 and 7 post-infection, with the culture media being changed on day 4 post-infection.

2.3.3　RNA Extraction

The mirVana miRNA isolation kit (ThermoFisher) was used to isolate total RNA, following the protocol for adherent cells.[35] Approximately 10^6 cells for each sample were pelleted and washed in PBS as an input. After harvesting, the cells

29

immediately underwent total RNA extraction. The mirVana miRNA isolation kit was selected as, unlike other RNA extraction kits, it preserves small RNAs. Immediately after the total RNA of each sample was extracted, the sample quality (for organic contaminants) and RNA concentration were evaluated using a NanoDrop instrument.

2.3.4 cDNA Library Synthesis

After validating the quality of the total RNA, the samples were then synthesized into a cDNA library to prepare them for qPCR. Since the qPCR was done using TaqMan probes, it was necessary to modify the mature miRNAs with a 3' Poly(A) tail and a 5' adaptor. This allowed for the amplification of all miRNA in a single reaction. The cDNA synthesis was performed using TaqMan™ Advanced miRNA cDNA Synthesis Kit (ThermoFisher).[36] The cDNA reaction product was stored at -80°C until the following qPCR step was performed.

2.3.5 qPCR

To validate changes in the CHO-K1 miRNA transcriptome under the condition of viral infection, it was the goal to screen a large number of miRNAs, including miRNAs we had identified as potential targets from the literature. Therefore, a commercially available set of plates with preset TaqMan probes for 378 different miRNAs was utilized. These plates, labeled Human A plates, were intended for use with human samples, however many of these miRNAs are predicted to be present in CHO-K1. Experimental work for qPCR was performed by collaborators at Rensselaer Polytechnic Institute (RPI) and these plates were analyzed using the following procedure.

TaqMan Plates were thawed on ice, vortexed, and then centrifuged. A 1 to 10 dilution was prepared using the cDNA samples from the previous step by mixing with 0.1X TE Buffer. In an Eppendorf tube, the following PCR reaction mix was prepared per reaction: 10 μL of 2X TaqMan Master Mix, 1 μL 20X TaqMan miRNA assay, and 4 μL nuclease free water. 15 μL of this mix was transferred to each well in a new 96-well Roche plate and 5 μL of the cDNA sample was added to each well. This Roche plate was sealed, vortexed, centrifuged, and then loaded into the qPCR machine. The qPCR thermal cycling conditions were set to the following: Enzyme activation at 40°C for 20 seconds and then 40 cycles of denaturing at 95°C for 1 second followed by the annealing and extension step at 60°C for 20 seconds. Cycle threshold (C_T) values were recorded for each reaction well to determine concentration levels for each miRNA target.

2.3.6   Data Analysis

Once C_T values were obtained from the qPCR experiments, they were analyzed using the $2^{-\Delta\Delta Ct}$ method to determine fold change (FC).[37,38] Because we had duplicates of each sample, indicated by "$S1$" and "$S2$", our data was analyzed with Eq. 2.1, Eq. 2.2, and Eq. 2.3 using averages, so that $2^{-\Delta\Delta C_T}$ was always equal to 1 for the control group of each sample.

$$\Delta C_{T,target\ miRNA} = C_{T,\ target\ miRNA} - C_{T,\ housekeeping\ miRNA} \qquad Eq.\ 2.1$$

$$\Delta C_{T,cont\ avg} = \frac{(\Delta C_{T,day\ 3,S1} + \Delta C_{T,day\ 3,S2} + \Delta C_{T,day\ 7,S1} + \Delta C_{T,day\ 7,S2})}{4} \qquad Eq.\ 2.2$$

$$\Delta\Delta C_{T,target\ miRNA} = \frac{\Delta C_{T,target\ miRNA}}{\Delta C_{T,\ control\ average\ of\ target\ miRNA}} \qquad Eq.\ 2.3$$

The preset housekeeping gene for the Human A qPCR plates was hsa-miR-16-5p. However, this miRNA showed unexpected variation and other housekeeping genes were identified upon analysis, discussed further in the results section. For each target miRNA, using the $\Delta\Delta C_T$ values from the healthy, uninfected control samples gives both Eq. 2.4 and Eq. 2.5.

$$FC_{target\ miRNA,\ cont} = 2^{-\left(\frac{\Delta\Delta C_{T,day\ 3,S1}+\Delta\Delta C_{T,day\ 3,S2}+\Delta\Delta C_{T,day\ 7,S1}+\Delta\Delta C_{T,day\ 7,S2}}{4}\right)} \qquad Eq.\ 2.4$$

$$FC_{target\ miRNA,\ cont} = 1 \qquad Eq.\ 2.5$$

To determine the fold change in miRNAs from the infected samples on days 3 and 7, Eq. 2.6 and Eq. 2.7 were used with the $\Delta\Delta C_T$ values from infected samples.

$$FC_{target\ miRNA,\ day\ 3} = 2^{-\left(\frac{\Delta\Delta C_{T,day\ 3,S1}+\Delta\Delta C_{T,day\ 3,S2}}{2}\right)} \qquad Eq.\ 2.6$$

$$FC_{target\ miRNA,\ day\ 7} = 2^{-\left(\frac{\Delta\Delta C_{T,day\ 7,S1}+\Delta\Delta C_{T,day\ 7,S2}}{2}\right)} \qquad Eq.\ 2.7$$

2.4 Results and Discussion

Infection of cell cultures was confirmed as cells started to show visual signs of cytopathic effects before their RNA isolation on day 7. Despite appearing infected, cell death had not yet overcome the cultures and there still many live cells available for isolation. Of the human miRNAs screened through qPCR, many did show a significant fold change upon viral infection, indicating that miRNA levels in CHO-K1 do change under conditions of viral infection. Table 2.2 lists the miRNAs which showed significant fold change compared to the healthy control for either day 3 or day 7 post-infection with Reo3, sorted by day 7-fold change. Significant

upregulation was seen most strongly in miR-9-5p, miR-384, miR-409-5p, and miR-208a-3p for both days post-infection. Fewer miRNAs showed downregulation, however, the most notable of these were miR-16-5p, miR-21-5p, and miR-429. Out of 31 miRNAs on the list of potential biomarkers identified from the literature 5 showed significant fold change, including miR-9, miR-20, miR-155, miR-181, and miR-448. Out of 198 miRNAs which yielded qPCR results, 129 did not show any significant fold change. Because of the experimental design which utilized biological duplicates instead of triplicates, the targets identified here would need to undergo to undergo further experimentation for a more rigorous statistical analysis to determine their biological relevance as biomarkers. Nevertheless, this work does show that significant changes in the overall miRNA transcriptome occur with Reo3 infection.

miRNA	Day 3	Day 7	miRNA	Day 3	Day 7
hsa-miR-9-5p	8.21	9.81	hsa-miR-296-5p	-1.45	2.27
hsa-miR-384	5.80	8.14	hsa-miR-653-5p	1.09	2.26
hsa-miR-576-3p	1.64	5.66	hsa-miR-598-3p	-1.58	2.24
hsa-miR-409-5p	2.23	5.17	hsa-miR-450b-3p	2.63	2.20
hsa-miR-2110	-1.13	5.09	hsa-miR-499a-5p	1.15	2.17
hsa-miR-208a-3p	5.60	4.46	hsa-miR-365a-3p	1.16	2.12
hsa-miR-376a-3p	1.66	4.09	hsa-miR-486-5p	1.43	2.11
hsa-miR-330-5p	3.28	4.04	hsa-miR-1-3p	1.57	2.09
hsa-miR-331-5p	1.50	4.02	hsa-miR-152-3p	-1.11	2.07
hsa-miR-203a-3p	4.85	3.77	hsa-miR-193b-3p	1.83	2.06
hsa-miR-509-5p	1.80	3.76	hsa-miR-134-5p	-2.24	1.94
hsa-miR-509-3-5p	2.53	3.57	hsa-miR-485-5p	-2.19	1.67
hsa-miR-660-5p	2.17	3.53	hsa-miR-188-3p	-2.00	1.66
hsa-miR-574-3p	-1.02	3.37	hsa-miR-331-3p	2.03	1.65
hsa-miR-374a-5p	-1.00	3.35	hsa-miR-326	-3.06	1.21
hsa-miR-887-3p	-2.49	3.30	hsa-miR-374b-5p	5.79	1.15
hsa-miR-215-5p	1.81	3.12	hsa-miR-20b-5p	2.38	1.07
hsa-miR-383-5p	1.31	3.06	hsa-miR-106a-5p	2.13	-1.08
hsa-miR-548c-3p	1.43	3.05	hsa-miR-455-3p	-2.98	-1.17
hsa-miR-339-3p	2.66	3.00	hsa-miR-339-5p	-2.16	-1.81
hsa-miR-323b-5p	1.11	2.97	hsa-miR-23b-3p	-2.10	-1.94
hsa-miR-324-3p	2.13	2.87	hsa-miR-429	-2.17	-2.01
hsa-miR-380-3p	1.10	2.83	hsa-miR-20a-5p	1.59	-2.01
hsa-miR-192-5p	4.21	2.80	hsa-miR-542-3p	-1.30	-2.06
hsa-miR-532-3p	1.87	2.78	hsa-miR-93-5p	1.13	-2.07
hsa-miR-139-5p	2.72	2.60	hsa-miR-17-5p	1.15	-2.07
hsa-miR-448	1.16	2.59	hsa-miR-181b-5p	-1.22	-2.07
hsa-miR-449a	-1.12	2.53	hsa-miR-18b-5p	1.01	-2.13
hsa-miR-525-5p	-1.29	2.51	hsa-miR-18a-5p	1.79	-2.17
hsa-miR-153-3p	1.84	2.46	hsa-miR-210-3p	-1.86	-2.24
hsa-miR-525-3p	-1.11	2.44	hsa-miR-155-5p	1.25	-2.27
hsa-miR-502-3p	1.52	2.34	hsa-miR-424-5p	-1.85	-2.36
hsa-miR-219a-5p	1.36	2.30	hsa-miR-21-5p	-1.46	-2.48
hsa-miR-204-5p	-1.97	2.28	hsa-miR-16-5p	1.25	-9.38
hsa-miR-410-3p	1.27	2.28			

Table 2.2 miRNAs which showed significant fold change on either day 3 or day 7 post-infection. FC above 2 is highlighted in green. FC below -2 is highlighted in red.

It should be noted that the preset housekeeping gene on the Human A qPCR plates was intended to be hsa-miR-16-5p. However, since this miRNA showed significant downregulation on day 7, it was necessary to choose a different housekeeping

miRNA. Multiple housekeeping miRNAs were chosen based on a combination of the standard deviation (< 0.7) and the range (< 2) of C_T values for each miRNA. This included miR-25-3p, miR-28-3p, miR-34a-5p, miR-99b-5p, miR-132-3p, miR-137, miR-140-5p, miR-146b-5p, and miR-148b-3p. The C_T values of these miRNAs were averaged and used as $C_{T,\ housekeeping\ miRNA}$ for Eq. 2.1.

2.5 Conclusions

In this work, it was observed that 69 different miRNAs showed differential expression on either day 3 or day 7 post-infection with Reo3, while 129 showed no significant change at all. Therefore, it was validated through qPCR that the host cell miRNA transcriptome in CHO-K1 does show significant changes under the condition of Reo3 viral infection. This step was the necessary groundwork for moving forward with a comprehensive examination of host cell infection biomarkers for multiple types of viruses. Some miRNAs were identified as potential biomarkers for Reo3 infection, including miR-9-5p and miR-384. However, more experimental work with triplicate samples allowing for a more rigorous statistical analysis would be necessary for validation of their usefulness as biomarkers.

2.6 References

1. Walsh, G. Biopharmaceutical Benchmarks 2014. *Nat Biotechnol* **32**, 992–1000 (2014).

2. Xu, X. *et al.* The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line. *Nat Biotechnol* **29**, 735–741 (2011).

3. Mascarenhas, J. *et al.* Novel Genetically Engineered CHO Cells. *Genetic Engineering & Biotechnology News* **36**, 34–35 (2016).

4.   Barone, P. W. *et al.* Viral contamination in biologic manufacture and implications for emerging therapies. *Nat Biotechnol* **38**, 563–572 (2020).

5.   Feng, J., Wang, K., Liu, X., Chen, S. & Chen, J. The quantification of tomato microRNAs response to viral infection by stem-loop real-time RT-PCR. *Gene* **437**, 14–21 (2009).

6.   Huang, T., Xu, D. & Zhang, X. Characterization of host microRNAs that respond to DNA virus infection in a crustacean. *BMC Genomics* **13**, (2012).

7.   Yousefpouran, S. *et al.* The assessment of selected MiRNAs profile in HIV, HBV, HCV, HIV/HCV, HIV/HBV Co-infection and elite controllers for determination of biomarker. *Microb Pathog* **147**, (2020).

8.   Biswas, S., Haleyurgirisetty, M., Lee, S., Hewlett, I. & Devadas, K. Development and validation of plasma miRNA biomarker signature panel for the detection of early HIV-1 infection. *EBioMedicine* **43**, 307–316 (2019).

9.   Visacri, M. B. *et al.* Role of miRNAs as biomarkers of COVID-19: a scoping review of the status and future directions for research in this field. *Biomarkers in Medicine* vol. 15 1785–1795 Preprint at https://doi.org/10.2217/bmm-2021-0348 (2021).

10.   Condrat, C. E. *et al.* miRNAs as Biomarkers in Disease: Latest Findings Regarding Their Role in Diagnosis and Prognosis. *Cells* vol. 9 Preprint at https://doi.org/10.3390/cells9020276 (2020).

11.   Wang, J., Chen, J. & Sen, S. MicroRNA as Biomarkers and Diagnostics. *Journal of Cellular Physiology* vol. 231 25–30 Preprint at https://doi.org/10.1002/jcp.25056 (2016).

12.   Slota, J. A. & Booth, S. A. MicroRNAs in neuroinflammation: Implications in disease pathogenesis, biomarker discovery and therapeutic applications. *Non-coding RNA* vol. 5 Preprint at https://doi.org/10.3390/ncrna5020035 (2019).

13.   Tribolet, L. *et al.* MicroRNA Biomarkers for Infectious Diseases: From Basic Research to Biosensing. *Frontiers in Microbiology* vol. 11 Preprint at https://doi.org/10.3389/fmicb.2020.01197 (2020).

14.   Verma, P., Pandey, R. K., Prajapati, P. & Prajapati, V. K. Circulating microRNAs: Potential and emerging biomarkers for diagnosis of human infectious diseases. *Frontiers in Microbiology* vol. 7 Preprint at https://doi.org/10.3389/fmicb.2016.01274 (2016).

15.   Leung, A. K. L. & Sharp, P. A. MicroRNA Functions in Stress Responses. *Mol Cell* **40**, 205–215 (2010).

16.   Yap, M. *et al.* Conserved MicroRNAs in Chinese hamster ovary cell lines. *Biotechnol Bioeng* **108**, 475–480 (2010).

17.   Hackl, M. *et al.* Computational identification of microRNA gene loci and precursor microRNA sequences in CHO cell lines. *J Biotechnol* **158**, 151–155 (2012).

18.   Hackl, M. *et al.* Next-generation sequencing of the Chinese hamster ovary microRNA transcriptome: Identification, annotation and profiling of microRNAs as targets for cellular engineering. *J Biotechnol* **153**, 62–75 (2011).

19.   Hammond, S., Swanberg, J. C., Polson, S. W. & Lee, K. H. Profiling conserved microRNA expression in recombinant CHO cell lines using illumina sequencing. *Biotechnol Bioeng* **109**, 1371–1375 (2012).

20.   Kozomara, A. & Griffiths-Jones, S. MiRBase: Annotating high confidence microRNAs using deep sequencing data. *Nucleic Acids Res* **42**, 68–73 (2014).

21.   Griffiths-Jones, S. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res* **34**, D140–D144 (2006).

22.   Kozomara, A. & Griffiths-Jones, S. MiRBase: Integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res* **39**, 152–157 (2011).

23.   Griffiths-Jones, S., Saini, H. K., Van Dongen, S. & Enright, A. J. miRBase: Tools for microRNA genomics. *Nucleic Acids Res* **36**, 154–158 (2008).

24.   Kozomara, A., Birgaoanu, M. & Griffiths-Jones, S. MiRBase: From microRNA sequences to function. *Nucleic Acids Res* **47**, D155–D162 (2019).

25.   Griffiths-Jones, S. The microRNA Registry. *Nucleic Acids Res* **32**, 109D – 111 (2004).

26.   Wurm, F. M. CHO quasispecies-Implications for manufacturing processes. *Processes* **1**, 296–311 (2013).

27.   Hammond, S., Kaplarevic, M., Borth, N., Betenbaugh, M. J. & Lee, K. H. Chinese hamster genome database: An online resource for the CHO community at. *Biotechnology and Bioengineering* vol. 109 1353–1356 Preprint at https://doi.org/10.1002/bit.24374 (2012).

28.   Huang, T., Xu, D. & Zhang, X. *Characterization of host microRNAs that respond to DNA virus infection in a crustacean*. http://www.biomedcentral.com/1471-2164/13/159 (2012).

29.   Samir, M., Vaas, L. A. I. & Pessler, F. MicroRNAs in the Host Response to Viral Infections of Veterinary Importance. *Front Vet Sci* **3**, 1–17 (2016).

30.   Guo, Y. E. & Steitz, J. A. Virus Meets Host MicroRNA: the Destroyer, the Booster, the Hijacker. *Mol Cell Biol* **34**, 3780–3787 (2014).

31.   Grassmann, R. & Jeang, K. T. The roles of microRNAs in mammalian virus infection. *Biochim Biophys Acta Gene Regul Mech* **1779**, 706–711 (2008).

32.   Ghosh, Z., Mallick, B. & Chakrabarti, J. Cellular versus viral microRNAs in host-virus interaction. *Nucleic Acids Res* **37**, 1035–1048 (2009).

33.   Scaria, V., Hariharan, M., Maiti, S., Pillai, B. & Brahmachari, S. K. Host-virus interaction: A new role for microRNAs. *Retrovirology* **3**, 1–9 (2006).

34. Hariharan, M., Scaria, V., Brahmachari, S. K., Pillai, B. & Maiti, S. Host-virus genome interactions: macro roles for microRNAs. *Cell Microbiol* **9**, 2784–2794 (2007).

35. *mirVana ™ miRNA Isolation Kit*. (2011).

36. Thermo Fisher Scientific. TaqMan ® Advanced miRNA Assays, USER GUIDE. 0–3 (2016).

37. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2-ΔΔCT method. *Methods* **25**, 402–408 (2001).

38. Schmittgen, T. D. & Livak, K. J. Analyzing real-time PCR data by the comparative CT method. *Nat Protoc* **3**, 1101–1108 (2008).

Chapter 3 – Viral Potency and Infection Progression in CHO-K1 Cells

3.1 Introduction

miRNAs are an integral part of the innate immune response to viruses in mammalian cells. Part of the innate immune response is the production of interferons (IFNs) which either block viral replication or activate an immune response. miRNAs are involved in one of two ways. The IFNs inhibit the expression of cellular miRNAs which promote viral life cycle and replication, or they promote antiviral cellular miRNAs.[1] These miRNAs have a number of different modes by which they accomplish their goal through interactions with their target mRNAs.[2–4] However, viruses can hijack these mechanisms and use them to their advantage. For example, *Herpesvirus saimiri* degrades miR-27, which is known to interact with an mRNA that encodes parts of T-cell receptor signaling pathways, allowing it to evade certain immune responses. Or, using a different mechanism, *Hepatitis C virus* promotes the expression of cellular miR-122, which facilitates its own viral replication.[5] However, these are only two examples and interactions between viruses and host miRNAs are highly complex and somewhat variable.

These interactions are made even more complicated by the fact that some viruses have the ability to produce their own mRNA and miRNAs which shape host-virus interactions at the post-transcriptional level. Viral miRNAs can bind to host cell mRNAs and host cell miRNAs can bind to viral mRNAs in addition to each binding their own miRNAs and mRNAs to control protein expression, which adds another level of complexity to miRNA involvement in immune responses.[1] Here, it is

relevant to note that it would be possible to detect viral miRNAs as biomarkers of a specific viral infection. However, unlike mammalian miRNAs, viral miRNAs have very little homology. For this reason, and because we also seek general biomarkers of infectivity and cellular stress for monitoring of host cell health, it is beneficial to focus on the host miRNA transcriptome response to viral infection instead of the viral genome itself.

In this work, we explore the properties and infection mechanisms of different viruses which infect CHO-K1 and their effects on cellular function for a deeper understanding of why certain miRNAs might show differential expression and why these miRNA biomarkers might differ between infecting viruses. Furthermore, we obtain a measurement of $TCID_{50}$/mL for each virus in the CHO-K1 cell line to allow for comparison between them and discovery of general biomarkers in a subsequent comprehensive miRNA biomarker screening. Although commercial viral solutions have listed $TCID_{50}$/mL values obtained by the manufacturer, these are measured in different cell lines and at different days post-infection. Therefore, it is necessary to have an exact measurement of them specifically in CHO-K1 and at the same day post-infection for comparison between viruses. Experimentally, the goal is first to validate infectivity of each virus in CHO-K1 through visualized cytopathic effects (CPE) and second, to quantify that infectivity using a mathematical model known as the Reed—Muench method.[6]

3.2   Viruses for Infection in CHO-K1

Five different viruses were selected for analysis of miRNA changes with infection in CHO-K1 cells, in consultation with collaborators in the bioprocessing industry.

Of these, 3 had been documented in major viral contamination events for biopharmaceutical manufacturing within the past 20 years.[7] These viruses were selected to span a range of biophysical and lifecycle characteristics to facilitate the identification of both specific and general markers. Characteristics for and classifications of each virus are detailed in Table 3.1. Of these 5 viruses, no research was found in the literature on their interaction with host cell miRNAs or their ability to produce their own viral miRNAs.

Virus	MMV	Reo3	EMC	PI2	VSV
Full Name	Rodent proto-parvovirus 1	Mammalian orthoreovirus 3	Encephalo-myocarditis virus	Parainfluenza virus 2	Vesicular stomatitis virus
Family	Parvoviridae	Reoviridae	Picornaviridae	Paramyxoviridae	Rhabdoviridae
Genus	Protoparvovirus	Orthoreovirus	Cardiovirus	Rubulavirus	Vesiculovirus
Strain		Dearing			
ss/ds	ss	ds	ss	ss	ss
+/-	-	N/A	+	-	-
Envelope	Non-enveloped	Non-enveloped	Non-enveloped	Enveloped	Enveloped
DNA/RNA	DNA	RNA	RNA	RNA	RNA
Size	5-6 kb	1.2-3.9 kb	7.8 kb	15.7 kb	11 kb
Shape	icosahedral	icosahedral	icosahedral	filamentous	bullet shaped

Table 3.1 Classification and biophysical characteristics of viruses for infection in CHO-K1 cultures.[8]

3.2.1 Mammalian Orthoreovirus 3

Mammalian Orthoreovirus 3 (Reo3) is a non-enveloped, double-stranded RNA virus with 10 genome segments. The virion consists of a double layer of protein shells in an icosahedral shape, which tightly pack the genome segments into the viral core.[8] After binding to the cell surface, the outer capsid is degraded and the core is released into the cytoplasm of the cell where its dsRNA segments are simultaneously transcribed to create a + strand RNA copy. This viral mRNA copy is then released from the core and translated. Out of its 10 genome segments, 9 code

for a single protein. The translation of these viral mRNAs greatly inhibits the translation of host cell mRNAs which interferes with normal cellular function. Reo3 can induce both cell-cycle arrest (suspension of cell cycle function) and apoptosis (programmed cell death) in mammalian cells.[8]

Reo3 has been found as an adventitious agent in multiple major contamination events since 1990.[7] The most likely source of Reo3 contamination is through bovine serum which is most commonly used in bioprocesses during the scale-up phase.[9] It is a particularly concerning pathogen, as it is capable of infecting human and primate cell lines and is known to be pathogenic to humans, unlike many of the other adventitious agents capable of infecting CHO-K1 such as MMV.[7] Therefore, effective detection methods for Reo3 are essential to ensure bioprocess safety.

3.2.2 Rodent protoparvovirus 1

Rodent protoparvovirus 1, also known as minute virus of mice or minute mouse virus (MVM or MMV), is a small, non-enveloped, single-stranded DNA virus. It consists of a single, icosahedral-shaped capsid which encapsulates linear viral DNA.[8] Due to their small size, parvoviruses are highly dependent on their host cell functions for replication and are only able to infect cells actively moving through the cell cycle. They enter a host cell through the process of endocytosis where they are uncoated while releasing their DNA into the cell nucleus through nuclear pores. In the nucleus, the viral DNA is first converted into its double stranded form and then the host cell RNA polymerase II transcribes it into viral mRNA. For this process to occur, it is necessary that the host cell be in the cell cycle S-phase.

Nuclear export signal activity allows the release of mature virions from the cell nucleus and then cell lysis results in their rapid release and transmission.[10]

MMV infection is a pervasive risk in pharmaceutical bioprocesses which utilize CHO-K1 cells, with an average of one major contamination event every 5 years.[11] Contamination originates from a number of potential sources, including raw materials utilized in the biomanufacturing process, and is persistent in wild mouse populations making it difficult to control.[7] MMV is highly robust and is known to retain viability for long periods of time in the environment.[10] In some bioprocesses, MMV contamination has been extremely difficult to detect as it did not impact cell viability and was therefore undetectable with in vitro virus assays.[12] Even after detection, viral clearance and validation for MMV is challenging because of its small size.[13] Recently, certain strains of CHO-K1 have been engineered with resistance to MMV by knocking out the host cell receptors to which it binds.[10] Additionally, the use of animal origin-free media has reduced the incidence rate of adventitious MMV contamination; however, this virus still remains a prevailing threat to bioprocesses utilizing CHO-K1 cells.

3.2.3   Encephalomyocarditis virus

Encephalomyocarditis virus (EMC) is a linear single-stranded, positive-sense RNA virus in the picornavirus family.[8] Lacking an envelope, EMC has an icosahedral-shaped capsid which binds to cellular receptors. Once released into the host cell cytoplasm, the viral RNA is translated into proteins by host cellular ribosomes. Picornaviruses are capable of replicating using only their RNA genome, without the need for any of the structural proteins which make up the virion. After

translation, the viral proteins start replication of the viral RNA in association with membrane vesicles induced by the infection.[8] Picornaviruses are known to inhibit the macromolecular functions of its host cell during the course of infection as the infected cell utilizes its machinery primarily for the replication of the virus. Infected cells have greatly reduced vesicular trafficking and synthesis of host nucleic acids and proteins. EMC has been reported to infect CHO-K1 cells and show clear CPE.[13] Itis most likely to be introduced as an adventitious agent in bioprocessing through contamination of raw materials.[14]

3.2.4 Parainfluenza virus 2

Parainfluenza virus 2 (PI2) is a single-stranded, negative-sense RNA virus under the paramyxovirus classification. Myxoviridae is a group of RNA viruses that is split into two different families, ortho- and para-. Paramyxoviruses, very similar to rhabdoviruses, are negative-stranded RNA viruses packaged in a helical nucleocapsid surrounded by a loose-fitting envelope.[8] PI2 enters a host cell through fusion of its envelope proteins with the host plasma membrane. Once inside the cell, as a negative-sense RNA virus, it produces its own RNA polymerase to transcribe a positive-sense mRNA strand, as the host cell lacks this machinery. This mRNA strand then codes for viral proteins using the host cell apparatus. Paramyxovirus envelopes are known to assemble at the plasma membrane. Once the complex between the matrix and envelope proteins and nucleocapsids is formed, the virion is completed by budding out from the cell membrane where it is free to find another host cell.[8]

44

PI2 has not been verified for infection in CHO-K1 cells in the literature. However, other paramyxoviruses, including parainfluenza virus 3 (PI3), have been shown to replicate in CHO-K1. Despite its ability to replicate, PI3 infection in CHO-K1 did not show any CPE, making this adventitious agent difficult to detect.[14] PI3 was implicated in a major viral contamination event in an MRC5 cell line in the 1990s.[7] PI3 and PI2 are most likely to be introduced as adventitious agents into a bioprocess from manufacturing personnel, highlighting the need for sterile procedures.[14] Like Reo3, PI3 is also known to infect human and primate cell lines and to be pathogenic to humans.

3.2.5 Vesicular stomatitis virus

Vesicular stomatitis virus (VSV) is an enveloped, single-stranded, negative-sense RNA virus in the rhabdovirus family. It has a genome made up of a linear RNA molecule which is surrounded by a nucleocapsid supercoil. Its bullet-shaped envelope consists of a single glycoprotein which binds to host cell receptors. Cells infected with VSV often show fusion with their neighbors due to the envelope proteins bound to their membranes. Ultimately, this can create large, multinucleated cells, or syncytia, resulting in tissue destruction or other pathogenic effects.[8] VSV enters a host cell through endocytosis, which releases the viral nucleocapsid into cell cytoplasm. Upon cell entry, the negative-stranded RNA virus is transcribed into a complementary positive-sense mRNA using its own encoded RNA polymerase. VSV encodes 5 different proteins (L, G, M, NS, and N) which are translated throughout the infection cycle.[15] Viruses in the rhabdovirus family are known to be highly virulent, causing rapid changes in host cell metabolism

leading to cell death. Similar to Reo3, VSV obstructs normal cellular function by preventing the synthesis of host cell nucleic acids and proteins.[15]

VSV has been widely studied as a model negative-strand virus and has been reported in the literature to infect CHO-K1 cultures with a rapid onset of CPE.[14] In biomanufacturing processes, VSV is most likely to be introduced as an adventitious agent through raw materials.[14] Due to the highly visible CPE, bioprocesses infected with VSV are obvious. Human infections of VSV are rare and mild, however it is a virus which causes severe effects for cattle.[8]

3.3 Materials and Methods

Each of the viruses in this project employ different mechanisms of action for and have different potencies when infecting CHO-K1 cells. Therefore, it was necessary to obtain an objective measurement of CHO-K1 cytopathic response to infection for each tested virus as opposed to relying on the $TCID_{50}$/mL listed by the manufacturer, measured in different host cell lines and on different days post-infection. We experimentally determined the potency of each virus in our stock suspensions by determining the titer at which 50% of CHO-K1 cultures were killed ($TCID_{50}$). This measurement allowed for subsequent infection studies with different viruses to be conducted at similar infectious doses, facilitating comparison of CHO-K1 transcriptome changes across virus species.

3.3.1 Cell Culture and Infection

Healthy CHO-K1 cells (ATCC) were cultured and plated into 96-well cell culture plates with flat-bottomed wells. Each well was seeded with 40,000 cells in 200 µL

of F-12K nutrient mixture containing 10% fetal bovine serum (FBS). Plates were cultured overnight at 37°C. At > 70% confluence, after approximately 24 hours, the wells were infected with Reo3 (ATCC), MMV (ATCC), PI2 (Genentech), EMC (Genentech), or VSV (ATCC). Virus solutions were created through 12 consecutive 1:10 serial dilutions of the original virus stocks in a F-12K nutrient mixture with 2% FBS. Wells were aspirated, washed with PBS, and 100 µL of virus solution was introduced to each well. Figure 3.1 shows how each column of wells was infected with a consecutive serial dilution of the original virus stock. Cultures were incubated at 37°C.After 2 hours additional F-12K media solution with 2% FBS was added. 96-well plates were observed every 24 hours post-infection for 2 weeks to monitor viral progression. Representative images were taken at each time point using a Keyence BZ-X series microscope using the phase-contrast or brightfield settings at 20x magnification. All cultures were incubated at 37°C until they were discarded. Culture media was replaced with fresh F-12K nutrient mixture containing 2% FBS on days 4 and 8 post-infection to keep the cells from starvation.

Dilution from Stock Solution
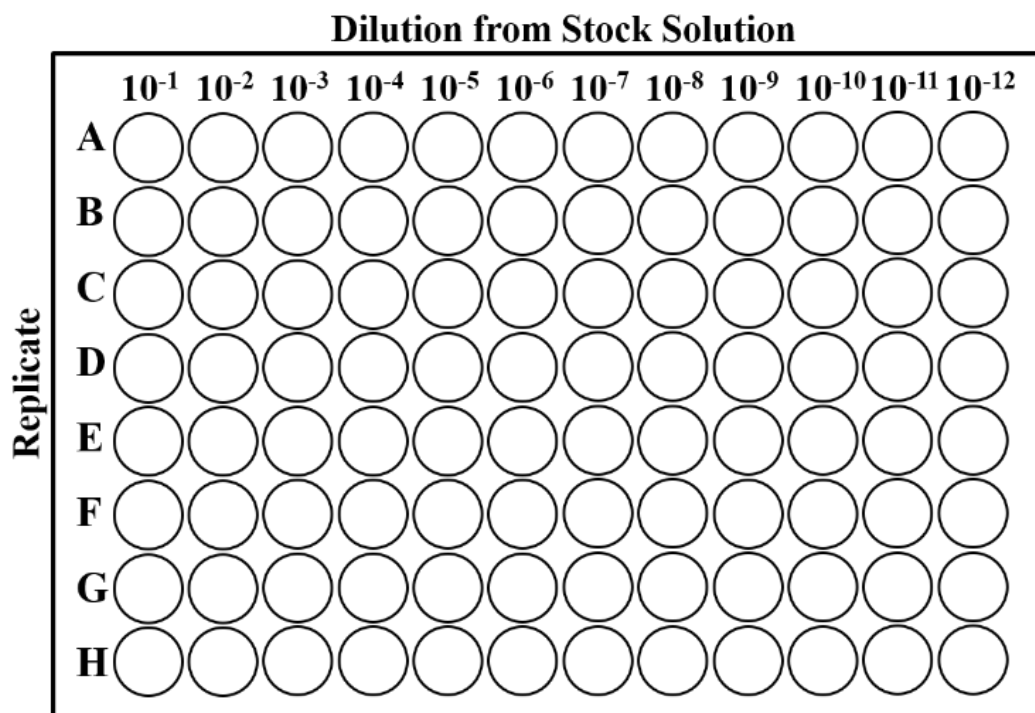
Figure 3.1 Representative 96-well plate for serial dilutions of stock virus solutions to determine TCID$_{50}$/mL values.

3.3.2 Calculation of Relative TCID$_{50}$/mL

Relative TCID$_{50}$/mL was obtained for each virus (Reo3, MVM, EMC, PI2, and VSV) by observing each well in a 96-well plate for cytopathic effects (CPE) every 24 hours after infection. CPE was observed as any significant structural changes in cell morphology caused by the infecting virus compared as to the healthy control cultures. Each column of wells in the plate was a 10-fold serial dilution of virus solution compared to the preceding column as seen in Figure 3.1. When a well showed significant CPE or cell death, it was marked with an X. Relative TCID$_{50}$/mL was calculated on Day 5 post-infection for each virus using the following Reed & Muench method equations where PD is the proportional distance between two dilutions.[6]

$$PD = \frac{(\% \; CPE \; in \; dilution \; above \; 50\%) - (50\%)}{(\% \; CPE \; in \; dilution \; above \; 50\%) - (\% \; CPE \; in \; dilution \; below \; 50\%)} \qquad \text{Eq. 3.1}$$

$$TCID_{50}/mL = \frac{1}{10^{(dilution \; factor \; with \; CPE \; above \; 50\% \; -PD)} * 0.1} \qquad \text{Eq. 3.2}$$

3.4 Results and Discussion

3.4.1 Healthy CHO-K1 Cell Results

Figure 3.2 shows phase-contrast images of the normal cell growth progression of a healthy CHO-K1 culture. Healthy cells reached 70% confluency for infection approximately 24 hours after plating. Uninfected CHO-K1 cells continued to replicate and crowd together over the two weeks of culturing time. Due to this crowding characteristic, cellular morphology became less geometric or round, developing longer structures and forming large swirl patterns instead CHO-K1 cells also showed the ability to grow on top of themselves and form 3-dimensional structures as confluency exceeded 100%.
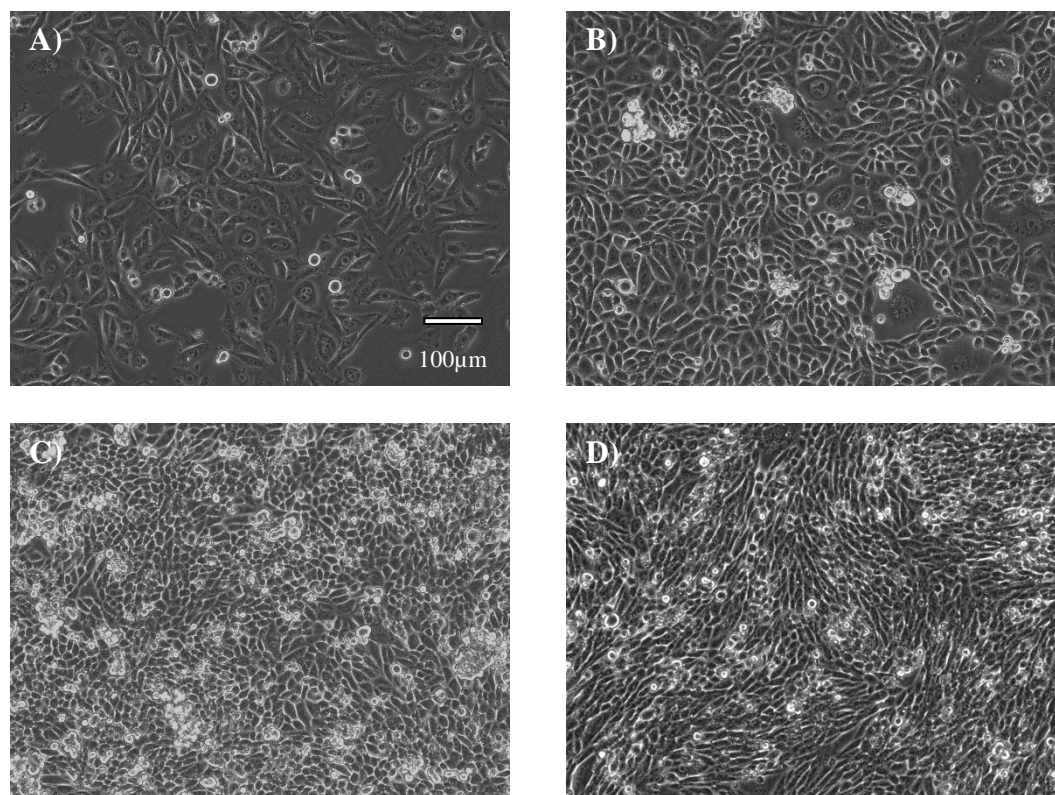
Figure 3.2 Phase contrast images at 20x magnification for the progression of healthy CHO-K1 cultures after A) 24 hours, B) 2 days, C) 4 days, and D) 6 days of culturing time. Scale bar approximate.

3.4.2   Reo3 Results

Reo3 served as the base-case infection to which all other infections were compared.

Figure 3.3 shows the progression of CPE for different dilutions of the stock Reo3

virus. Once CPE was observed, it took multiple days for that culture to completely

die. However, after displaying CPE, cell death was inevitable and CHO-K1 cells

did not show any resistance to the virus. This aligns with the pathogenesis of Reo3,

where the host cell is severely inhibited in translating proteins for its own cellular

function and instead uses its machinery to translate viral mRNAs into viral proteins.

Unable to function properly, cell death is unavoidable.
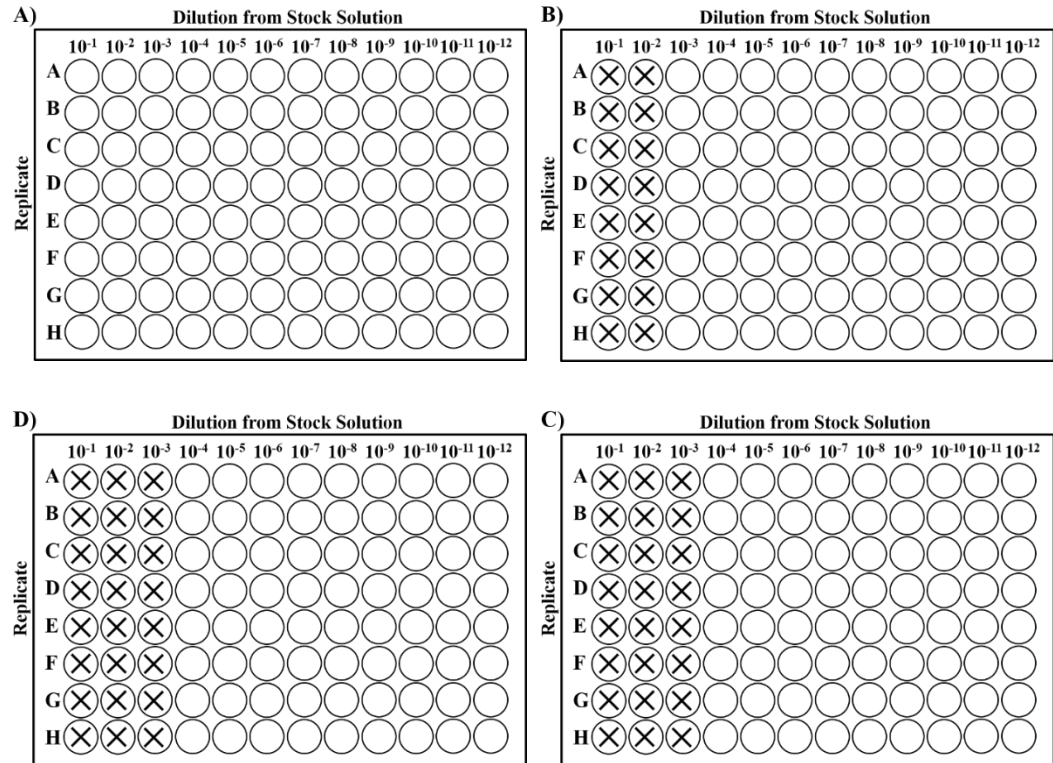
Figure 3.3 Cell death (>50%) progression of Reo3 virus stock dilution infections in CHO-K1 cells on A) day 2, B) day 3, C) day 4, D) day 5 post-infection.

Figure 3.4 shows brightfield images for the progression of Reo3 infection in CHO-K1 cells. Stunted growth compared to healthy CHO-K1 cultures was the first sign of infection. Infected cultures showed the continuous, progressive death until every cell in the culture had died. Cells appeared shriveled-up upon death, consistent with the apoptosis mechanism which Reo3 is known to activate.
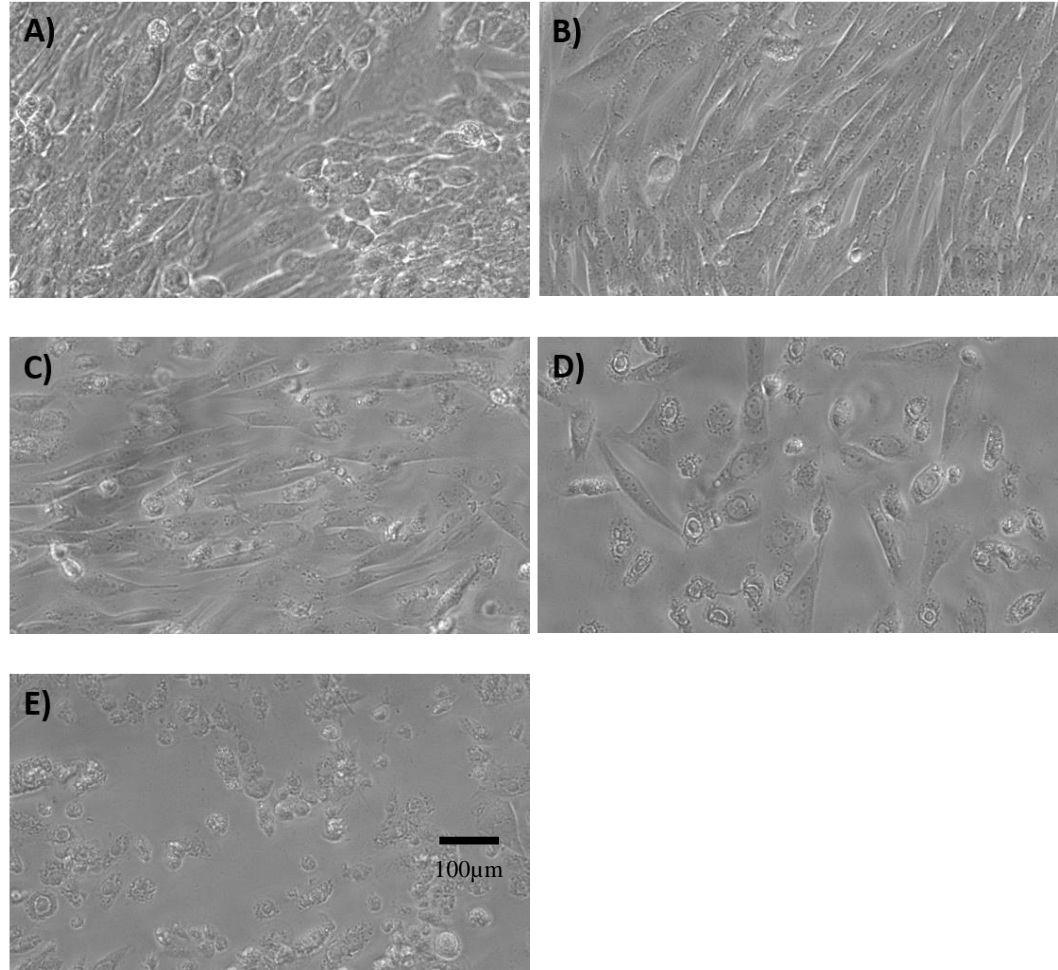
Figure 3.4 Brightfield images at 20x magnification showing the viral progression and morphological changes of Reo3 infection in CHO-K1 cells. A) No CPE, B) stunted culture growth and CPE, C) CPE with <50% cell death, D) CPE with >50% cell death, E) complete cell death. Scale bar approximate.

3.4.3 MMV Results

Figure 3.5 shows the progression of CPE for different dilutions of stock MMV solution. With MMV infection, CHO-K1 cells started showing signs of CPE before Reo3 infection, however, they resisted death longer after showing morphological changes and displayed significant improvement in health after new media was added to the culture. This is consistent with the pathogenesis of MMV, which is completely reliant on functioning host cell machinery to replicate and previous

literature evidence that MMV can be a "silent" infection, not always showing signs of CPE while still being a pervasive adventitious agent.[12]
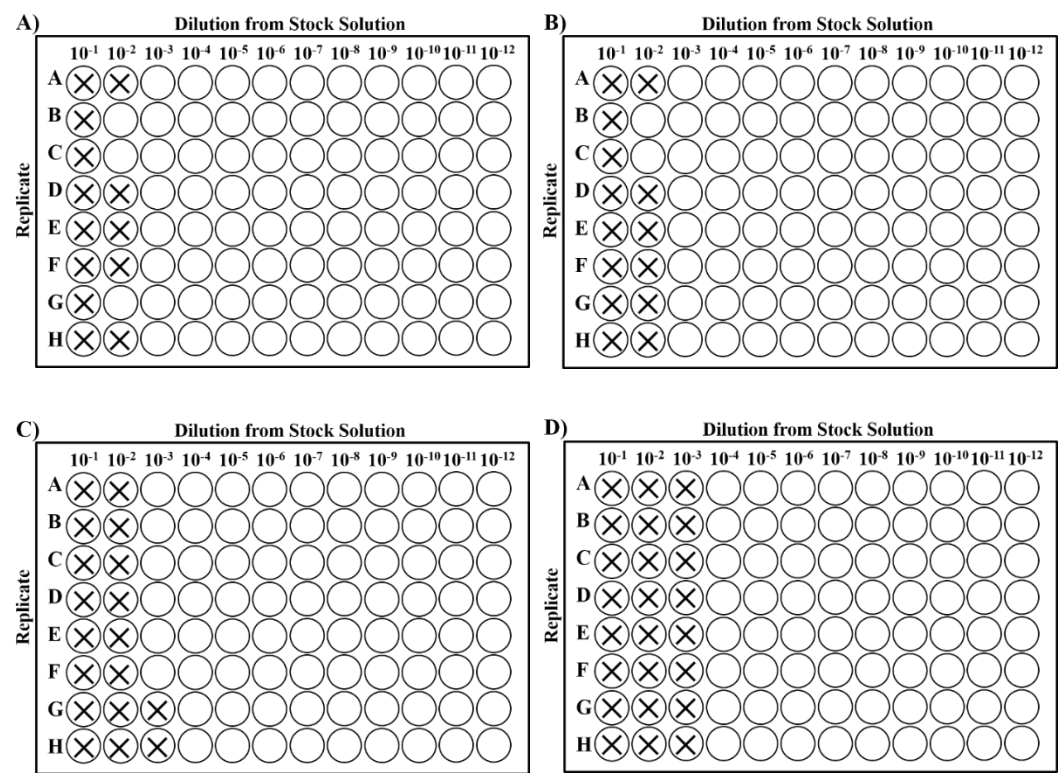
Figure 3.5 Cell death (>50%) progression of MMV virus stock dilution infections in CHO-K1 cells on A) day 2, B) day 3, C) day 4, D) day 5 post-infection.

Figure 3.6 shows brightfield images for the progression of MMV infection in CHO-K1 cells. The first observable evidence of infection was stunted growth compared that of healthy cells. When CPE became apparent, confluence of adherent cells was decreased to approximately 80% and many dead cells were seen floating in solution. Cells which remained adherent showed a robustness to the virus and remained living for over a week before succumbing to death. However, it was clear these cells were still infected as they were not replicating quickly, and their confluence did not increase as it would have in healthy CHO-K1 cultures. Cells

appeared fragmented upon cell death, which is consistent with the cell lysis mechanism that releases MMV virions back into the environment.
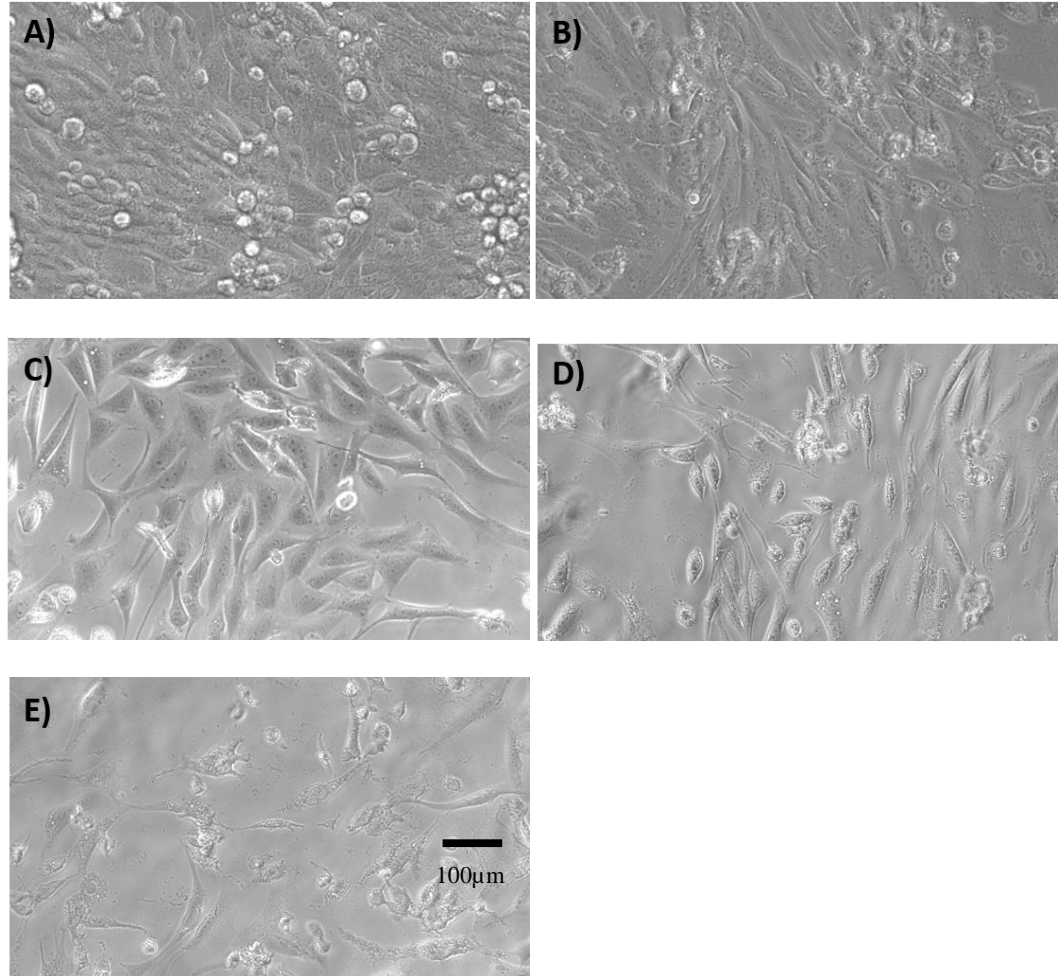
Figure 3.6 Brightfield images at 20x magnification showing the viral progression and morphological changes of MMV infection in CHO-K1 cells. A) No CPE, B) stunted culture growth, C) CPE with <50% cell death, D) CPE with >50% cell death, E) complete cell death. Scale bar approximate.

3.4.4   PI2 Results

PI2 never showed any CPE as far out as 12 days post-infection. Therefore, it was impossible to visually confirm whether the CHO-K1 cells were infected. Figure 3.7 shows brightfield images for a comparison between healthy CHO-K1 cells and PI2 infected cells on the same day. CHO-K1 cells inoculated with PI2 did not show any

stunted growth, morphological changes, or excessive cell death compared to the uninfected control. These results are consistent with literature reports of PI3 not showing CPE in CHO-K1 cells even though it was still replicating the virus.[14]
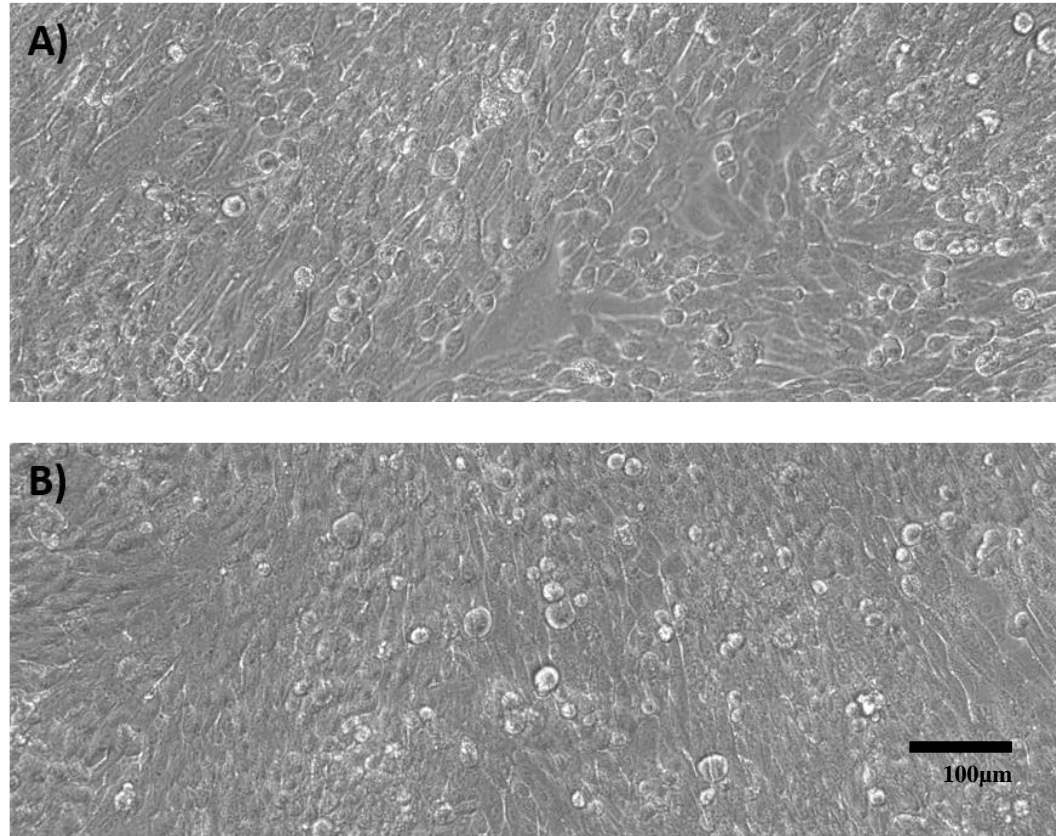
Figure 3.7 Brightfield images at 20x magnification showing A) PI2 viral infection in CHO cells and B) healthy CHO cultures on day 5 post-infection. No CPE was visualized with PI2 infection up to 12 days post-infection. Scale bar approximate.

3.4.5 EMC Results

Figure 3.8 shows the progression of CPE for different dilutions of the stock EMC virus solution. EMC progressed more quickly than both Reo3 and MMV, with the highest viral titer showing complete cell death after only 24 hours post-infection. Viral progression showed a slight delay after day 4, which can likely be attributed

the addition of fresh nutrient media in the wells to keep the cells from dying of starvation.
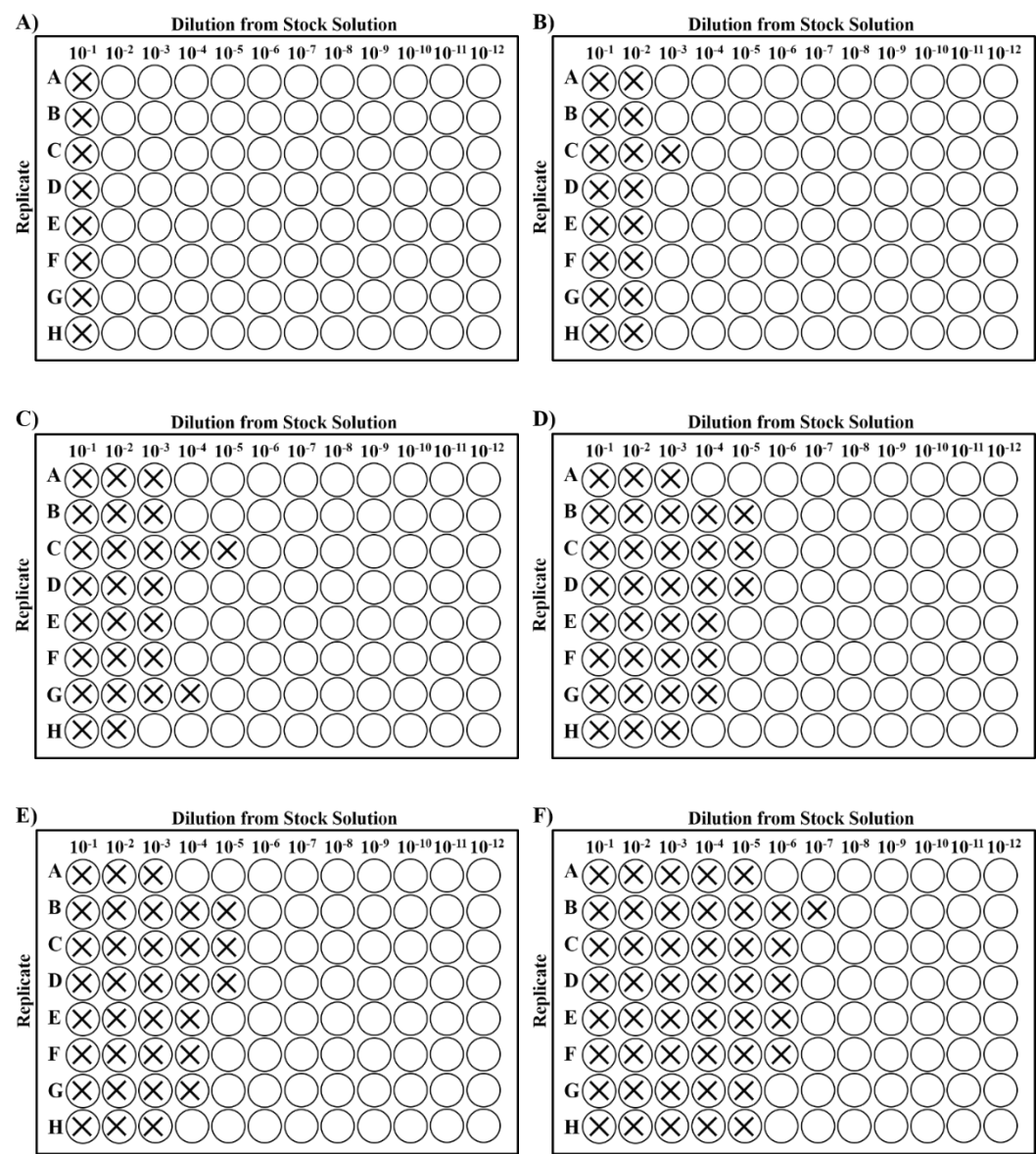
Figure 3.8 Cell death (>50%) progression of EMC virus stock dilution infections in CHO-K1 cells on A) day 1, B) day 2, C) day 3, D) day 4, E) day 5, and F) day 6 post-infection.

Figure 3.9 shows brightfield images for the progression of EMC infection in CHO-K1 cells. Cell cultures did not display any stunted growth compared to healthy CHO-K1, however, cell death happened quickly after infection.. Complete cell

death occurred within one or two days of the first observed CPE. Cells appeared very small and shriveled upon cell death, which is consistent with a viral pathogenesis that inhibits host macromolecular functions.
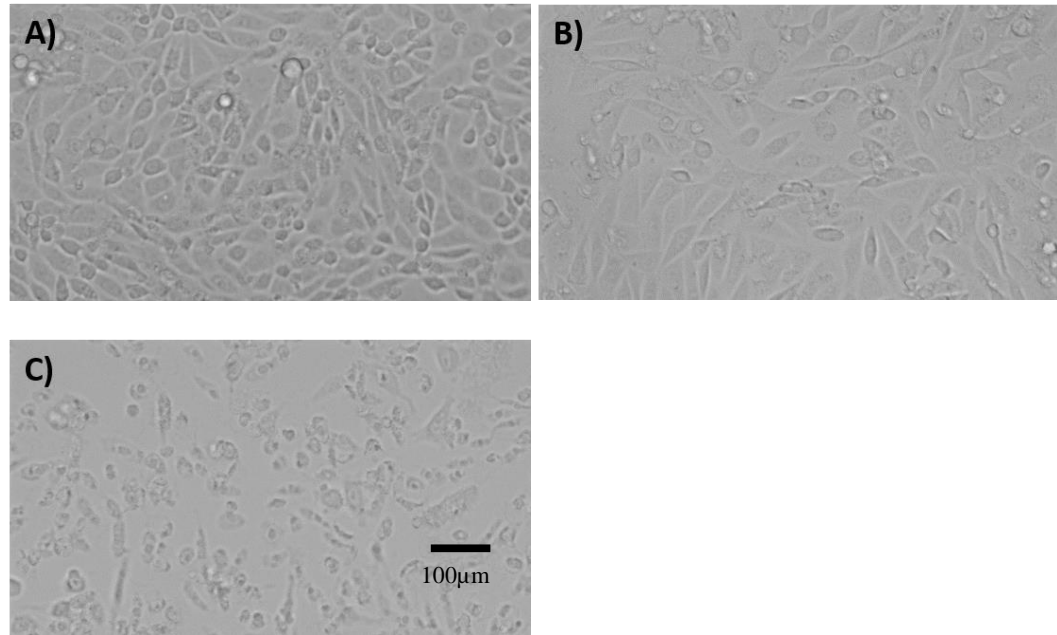
Figure 3.9 Brightfield images at 20x magnification showing the viral progression and morphological changes of EMC infection in CHO-K1 cells. A) No CPE, B) CPE with <50% cell death, C) complete cell death. Scale bar approximate.

3.4.6   VSV Results

VSV infection in CHO-K1 cells was similar to EMC infection, but noticeably more virulent. VSV was the most virulent of the viruses tested for infection of CHO-K1. Figure 3.10 shows the progression of CPE for different dilutions of the stock VSV solution. The 4 most concentrated solutions of VSV showed complete cell death in less than 24 hours post-infection. CHO-K1 cells showed little CPE before from the transition from appearing healthy to complete culture death. This is consistent with previous literature observations on the virulence of rhabdoviruses and the rapidity of CPE onset for VSV in CHO-K1 cells.[14,15]
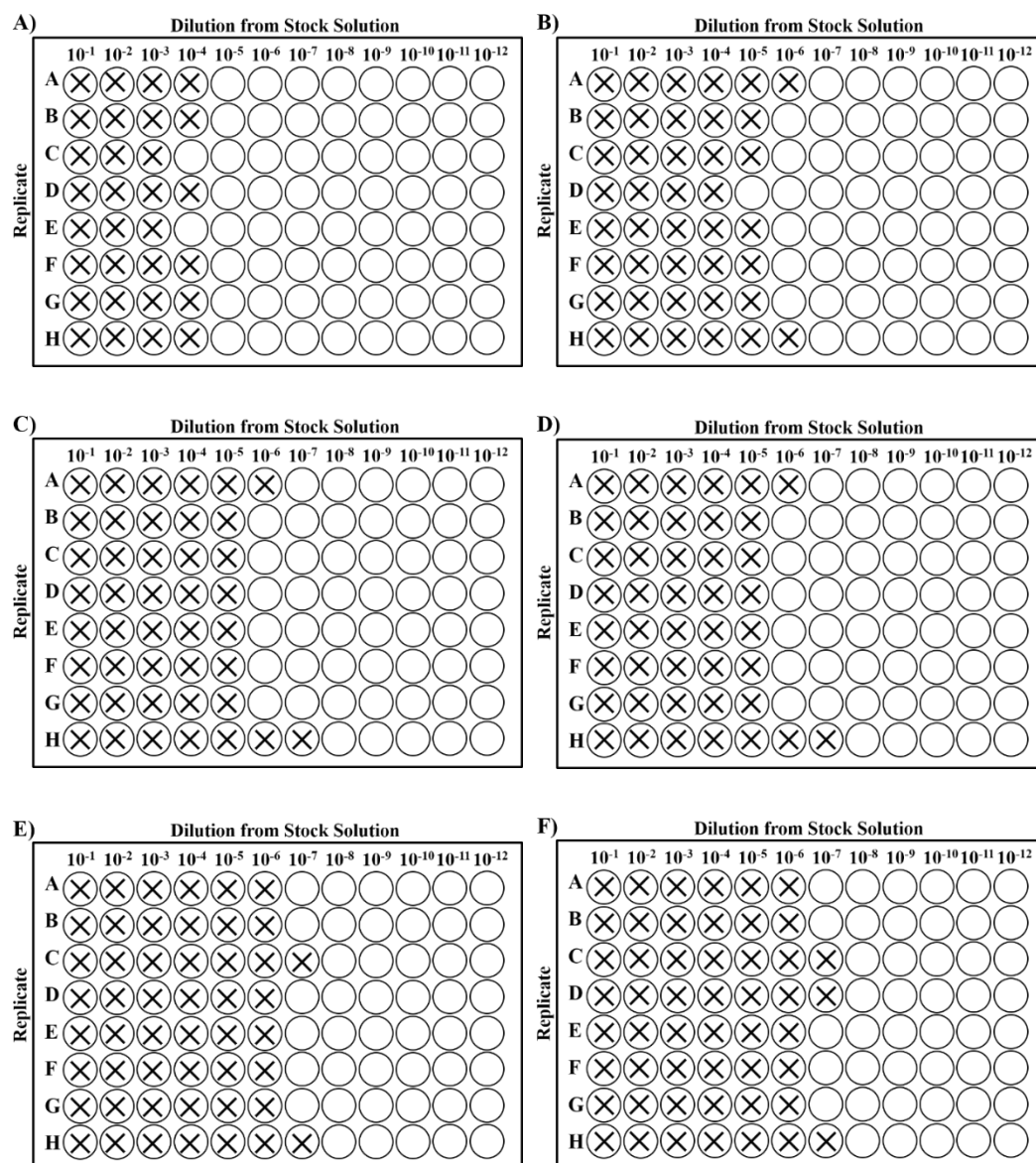
Figure 3.10 Cell death (>50%) progression of VSV virus stock dilution infections in CHO-K1 cells on A) day 1, B) day 2, C) day 3, D) day 4, E) day 5, and F) day 6 post-infection.

Figure 3.11 shows brightfield images for the progression of VSV infection in CHO-K1 cells. At the highest viral titers, dead cells were observed as either being "stuck together" or burst and appeared to create a 3-dimensional plaque. This observation correlates with the literature on the ability of the VSV glycoprotein envelope to cause cell fusion between neighboring cells, leading to their destruction from

58

mechanical stress.[8] Higher titer levels had more glycoprotein envelope available in solution to facilitate this phenomenon. Cell death at the highest titers looked distinctly different from cell death at the lower titers in individual cells were observed as shriveled-up. Similar to Reo3, this shriveled-up morphological change may be a result of the virus preventing synthesis of host cell proteins and nucleic acids, rendering the cell dysfunctional.
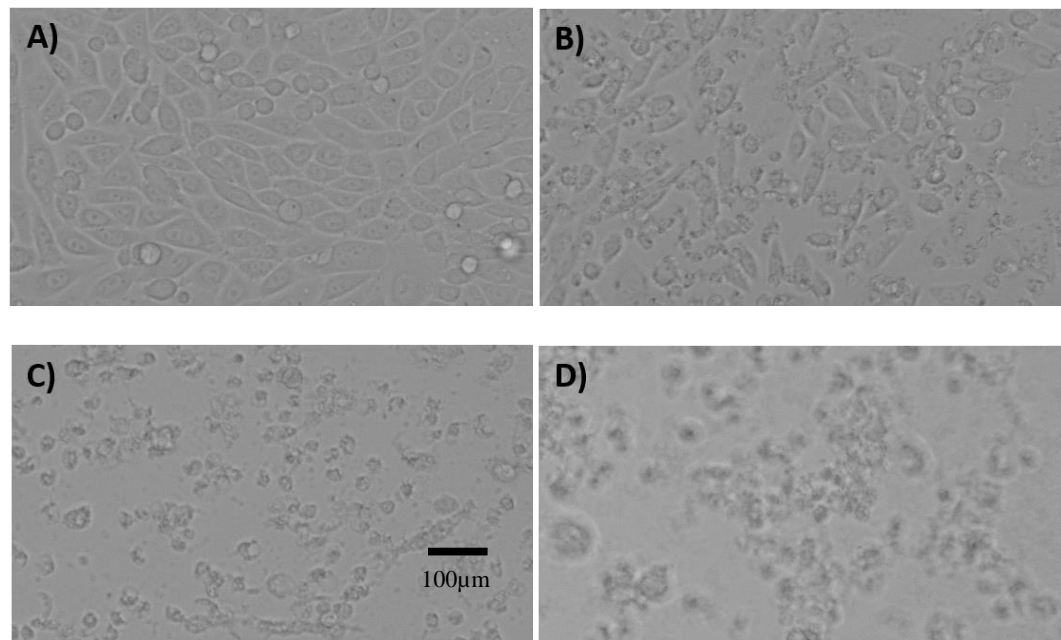
Figure 3.11 Brightfield images at 20x magnification showing the viral progression and morphological changes of VSV infection in CHO-K1 cells. A) No CPE, B) CPE with >50% cell death, C) shriveled morphology of dead cells at lower titer levels, D) clumps of dead cells forming 3D structures for cell death at the 3 highest titer levels. Scale bar approximate.

3.4.7 Comparison of Relative $TCID_{50}$/mL

$TCID_{50}$/mL was calculated on day 5 post-infection using Eq. 3.1 and Eq. 3.2. Table 3.2 contains the relative $TCID_{50}$/mL for each of the virus stock solutions. These measurements, specific to CHO-K1, were used to determine relative viral titers for a comprehensive screening to determine specific and general miRNA biomarkers of viral infection.

59

Virus	$(TCID_{50})/(mL$ ATCC stock solution)
Reo3	$10^{4.5}$
MMV	$10^{4.5}$
EMC	$10^{5.67}$
VSV	$10^{7.67}$

Table 3.2 Relative $TCID_{50}/mL$ measurements of stock viral solutions.

It was observed that Reo3 and MMV have approximately the same virulence for infection in CHO-K1 cells. EMC was one order of magnitude more virulent than MMV and Reo3. VSV was the most virulent in CHO-K1 at two orders of magnitude higher than EMC. A relative $TCID_{50}/mL$ was unable to be calculated for PI2 due to its lack of CPE and cell death compared to healthy control cultures.

3.5 Conclusions

In conclusion, relative $TCID_{50}/mL$ values were obtained for stock solutions of four different viruses in CHO-K1 cells. This necessary step has laid the groundwork for the experimental design and experimental work portions of a comprehensive miRNA biomarker screening. Although viral infectivity of PI2 in CHO-K1 cells could not be confirmed using this method, future work may include the quantification of viral particles produced by CHO-K1 cells inoculated with PI2. This quantification PI2 infectivity in CHO-K1 by a method not reliant on visualization of CPE would validate whether PI2 was a silent infector of CHO-K1, similar to PI3. While each of the 5 viruses tested (Reo3, MMV, PI2, EMC, and VSV) displayed different levels of virulence and cellular morphological variations, many of these changes were observed to be strongly correlated with the pathogenesis and mechanisms of the infecting virus. Due to the calculation of

relative $TCID_{50}/mL$ for each of the stock solutions, it is now possible to compare miRNA transcriptome responses between these infections. This comparison will give a deeper insight into the transcriptomic host cell changes which are impacted by each of the viral mechanisms.

3.6 References

1. David, M. Interferons and MicroRNAs. *Journal of Interferon & Cytokine Research* **30**, 825–828 (2010).

2. Leung, A. K. L. & Sharp, P. A. MicroRNA Functions in Stress Responses. *Mol Cell* **40**, 205–215 (2010).

3. Hariharan, M., Scaria, V., Brahmachari, S. K., Pillai, B. & Maiti, S. Host-virus genome interactions: macro roles for microRNAs. *Cell Microbiol* **9**, 2784–2794 (2007).

4. Scaria, V., Hariharan, M., Maiti, S., Pillai, B. & Brahmachari, S. K. Host-virus interaction: A new role for microRNAs. *Retrovirology* **3**, 1–9 (2006).

5. Guo, Y. E. & Steitz, J. A. Virus Meets Host MicroRNA: the Destroyer, the Booster, the Hijacker. *Mol Cell Biol* **34**, 3780–3787 (2014).

6. Reed, L. J. & Muench, H. A Simple Method of Estimating Fifty Percent Endpoints. *Am J Epidemiol* **27**, 493–497 (1938).

7. Barone, P. W. *et al.* Viral contamination in biologic manufacture and implications for emerging therapies. *Nat Biotechnol* **38**, 563–572 (2020).

8. Acheson, N. H. *Fundamentals of molecular virology*. (John Wiley & Sons, 2011).

9. Nims, R. W. Detection of adventitious viruses in biologicals–a rare occurrence. *Dev Biol (Basel)* **123**, 153—64; discussion 183—97 (2006).

10. Mascarenhas, J. X. *et al.* Genetic engineering of CHO cells for viral resistance to minute virus of mice; Genetic engineering of CHO cells for viral resistance to minute virus of mice. *Biotechnol. Bioeng* **114**, 576–588 (2017).

11. Mascarenhas, J. *et al.* Novel Genetically Engineered CHO Cells. *Genetic Engineering & Biotechnology News* **36**, 34–35 (2016).

12. Moody, M., Alves, W., Varghese, J. & Khan, F. Mouse Minute Virus (MMV) Contamination—A Case Study: Detection, Root Cause Determination, and Corrective Actions. *PDA J Pharm Sci Technol* **65**, 580 (2011).

13. Wickramasinghe, S. R., Han, B., Carlson, J. O. & Powers, S. M. Clearance of minute virus of mice by flocculation and microfiltration. *Biotechnol Bioeng* **86**, 612–621 (2004).

14. Berting, A., Farcet, M. R. & Kreil, T. R. Virus susceptibility of Chinese hamster ovary (CHO) cells and detection of viral contaminations by adventitious agent testing. *Biotechnol Bioeng* **106**, 598–607 (2010).

15. Wagner, R. R. *The Rhabdoviruses*. (Springer Science, 1987).

Chapter 4 – Identification of General Stress Biomarkers for Viral Infection in CHO-K1

4.1 Introduction

MicroRNAs (miRNAs) are short non-coding RNA molecules that regulate gene expression and contribute to metabolic regulation and maintenance of cellular homeostasis, making them a fundamental building block of biological function.[1–3] These short RNA molecules are known to be highly conserved and mediate how cells respond to environmental changes, broadly defined as cellular stress.[4] In this work, we aim to identify one or more general miRNA biomarkers for the cellular stress imposed by viral infection in CHO-K1 cells. MicroRNAs are well studied as biomarkers for various conditions[5–8] and often it is possible to identify multiple miRNAs as a biomarker "fingerprint" for a specific cellular stress or disease state.[9] MicroRNAs are an ideal candidate in the search for a general biomarker for a number of reasons. First, since there are relatively few miRNAs and each miRNA can target hundreds of mRNAs, a miRNA is more likely to be a marker of a general infection response than other types of RNA. Second, the highly conserved nature of miRNAs and their role in maintaining cellular homeostasis makes them more likely than other proteins or nucleic acids to be a marker of general stress. Finally, the discovery of novel miRNAs in one organism is often followed up by an orthologue search for the same miRNAs in other organisms.[10,11] As such, if certain miRNAs or group of miRNAs is (are) found to be (a) biomarker(s) of general viral infection in CHO-K1 cultures, it is possible this could translate to detection of viral infections in other organisms, including humans.[12]

The determination of general miRNA biomarkers for viral infection has implications in both the healthcare field for disease diagnostics and in the biomanufacturing industry for the detection of adventitious virus contamination events. Here, we focus on biomanufacturing applications and the identification of a general miRNA biomarker panel for viral infection in the bioprocessing cell line CHO-K1. In this work, we seek to identify a panel with a minimum of 3 miRNAs which both show differential expression in all 4 viruses compared to a healthy control and act as good classifiers for infection as analyzed from a receiver operating curve analysis. Other literature studies have found that the number of 3 miRNAs is ideal for optimization of utilizing miRNAs as biomarkers for classification purposes.[6] Ultimately, such a panel would lay the groundwork for implementation of this technology into the bioprocessing industry.

4.2   Applications of a miRNA Biomarker Panel for Fail-Fast Viral Adventitious Agent Detection

The CHO-K1 host cell line is widely used in the recombinant-biologic industry for the production of numerous biological products including enzymes, interferons, and monoclonal antibodies.[13] The purity of these products and assured clearance of viral adventitious agents is an ongoing concern for this field.[14] This work will contribute to the development of a fail-fast adventitious agent detection method, that could be easily implemented at numerous points during the upstream manufacturing process without a large time or capital investment. A fail-fast method for adventitious agent detection utilizes the concept of quick, repeated testing to understand the proper time point to cut losses, as opposed to a single,

longer and more accurate test. For detection of viral adventitious agents, an ideal fail-fast test would have a negligible false positive rate and a reasonable false negative rate. With repeated testing early in an upstream bioprocess, despite the high false negative rate, these parameters would allow for detection of an adventitious agent earlier that the current industry standard, (no virus testing except the FDA-mandated CPE assay) which is highly accurate, but time consuming and implemented late in the biomanufacturing process.

In addition to the time and economic benefits, this general biomarker panel would encompass detection of infection by unknown viruses and viruses that may not show visual cytopathic effects in the host cell line. There is an important difference between monitoring the presence of viral particles, as many viral detection methods do, and quantifying the level of infection in cells by monitoring the host cell transcriptome. This distinction is critical as not all viruses are capable of infecting all cells, and in cases where they can, the infection follows a time course where not all host cells are infected at a given time. Monitoring of the host cell miRNA transcriptome is a real-time, direct measurement of host cell response to the viral infection and would create new possibilities for the field of adventitious agent detection.

4.3   Materials and Methods

4.3.1   Experimental Design

Four viruses known to infect CHO-K1 cells via different mechanisms[15] were chosen to screen for the presence of a common miRNA biomarker. These included mammalian orthoreovirus 3 (Reo3), rodent protoparvovirus 1 (MMV),

encephalomyocarditis virus (EMC), and vesicular stomatitis virus (VSV). Relative viral titers for each viral stock solution had previously been determined in CHO-K1 cultures. This enabled experimental viral titers to be chosen for the inoculum based on the virulence of each virus for a proper comparison of the type and extent of the cellular response to each virus. Reo3 and MMV were evaluated at 0.1, 1, and 100 $TCID_{50}$. Due to the high virulence of the EMC and VSV viruses, they could not be experimentally evaluated at 0.1 $TCID_{50}$, or even 1 $TCID_{50}$ for VSV, due to the high number of serial dilutions which would have been necessary to establish those infectivities.

Consistent with the goals of a fail-fast adventitious agent detection method, we desired to seek a host miRNA biomarker response for an early infection time point, before the cells started to show cytopathic effects. Days 1, 2, and 3 post-infection were evaluated at each examined inoculum titer, resulting in a total of 84 total samples which included a healthy control for each time point. The number of RNA isolation samples for each condition (virus, viral titer, and day post-infection) evaluated through next generation sequencing (NGS) is listed in Table 4.1. Each condition was assessed in triplicate to ensure a rigorous statistical analysis.

	Day 1	Day 2	Day 3
Healthy Control	3	3	3
Reo3 0.1 TCID$_{50}$	3	3	3
Reo3 1 TCID$_{50}$	3	3	3
Reo3 100 TCID$_{50}$	3	3	3
MMV 0.1 TCID$_{50}$	3	3	3
MMV 1 TCID$_{50}$	3	3	3
MMV 100 TCID$_{50}$	3	3	3
EMC 1 TCID$_{50}$	3	3	3
VSV 100 TCID$_{50}$	3	3	3

Table 4.1 Number of samples processed using next generation sequencing (NGS) for each virus, viral titer, and day post-infection.

4.3.2   Cell Culture

Healthy CHO-K1 cells (ATCC) were cultured and plated into 6-well plates. Each well was seeded with 400,000 cells in 4 mL of F-12K nutrient mixture containing 10% fetal bovine serum (FBS). Plates were cultured overnight at 37°C. At > 70% confluence, at approximately 24 hours, the wells were infected with 2 mL of a virus solution. Virus solutions were created by diluting the original virus stocks to the desired titer (0.1, 1, or 100 TCID$_{50}$) with a F-12K nutrient mixture containing 2% FBS. Wells were aspirated, washed with PBS, and 2 mL of virus solution was introduced to each well. Cultures were incubated at 37°C. After 2 hours, 2 mL of additional F-12K media solution containing 2% FBS was added. Total RNA was isolated on days 1, 2, and 3 post-infection. All cultures were incubated at 37°C until RNA isolation.

4.3.3   Total RNA Isolation

Total RNA was isolated by organic extraction using the mirVana™ miRNA Isolation Kit, with phenol.[16] First, the cells were trypsinized, washed, and then re-suspended in PBS. Approximately 10^6 cells for each sample were re-pelleted, and

the PBS was aspirated. The cells were lysed by adding 300 μL of lysis/binding solution and vortexing. 30 μL of miRNA homogenate additive was added to the solution and the samples were left on ice for 10 minutes. Next, the 300 μL of acid-phenol: chloroform was added to each sample for the organic extraction of total RNA. Samples were vortexed and the organic and aqueous phases were allowed to settle by centrifuging the samples for 5 minutes at maximum speed. 300 μL of the aqueous phase was pipetted out from each sample into a new 2 mL Eppendorf tube containing 375 μL of 100% ACS grade ethanol. This solution was transferred to a glass-fiber filter cartridge in a new clean tube and centrifuged. The samples were washed with 100% ACS grade ethanol three times eluted into a clean 2 mL Eppendorf tube. Immediately after the total RNA of each sample was extracted, the sample quality and RNA concentration were evaluated using a NanoDrop instrument to check the RNA concentration and determine the extent of organic contaminants.

4.3.4   Next Generation Sequencing

Experimental protocols for next generation sequencing (NGS) were performed by collaborators at Rensselaer Polytechnic Institute. To ensure sample quality, each sample was analyzed using a Bioanalyzer instrument. The samples libraries were prepared for NGS using the Qiagen miRNA library prep, and these were sequenced using Illumina NextSeq 550 and 2000 sequencers. Quality control (QC) analysis of the data was performed using fastqc and multiqc.[17,18] Trimmomatic 038[19] was utilized to clean up the reads and then Bowtie2[20] was used for read mapping to the CHO-K1 reference genome obtained from CHOgenome.org.[21] Using the

miRDeep2 pipeline[22] and the hairpin and mature sequences obtained from miRBase, miRNA was quantified allowing for 1 mismatch and only including reads at least 18 nucleotides in length. Statistical analysis between groups was evaluated using deseq2[23] on the public usegalaxy.org server.[24]

4.3.5   Receiver Operating Curve Analysis

Normalized count files from the NGS analysis were used to create a receiver operating curve (ROC) for binary classification to quantify each miRNA's ability to discriminate between infected and uninfected samples. In this binary system, virally infected samples were classified as positive, and uninfected samples were classified as negative. True positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) were calculated over the range of normalized counts and sensitivity (S_n) and specificity (S_p) were derived from this classification system at each point using Eq. 4.1 and Eq. 4.2.

$$S_n = \frac{TP}{TP + FN} \qquad \text{\textit{Eq. 4.1}}$$

$$S_P = \frac{TN}{TN + FP} \qquad \text{\textit{Eq. 4.2}}$$

Sensitivity (S_n) was plotted against ($1 - S_p$) and an area under the curve (AUC) was calculated for each miRNA. A classifier with an AUC of 0.5 shows no ability to discriminate; our minimum AUC cutoff point for identification of a potentially useful miRNA biomarker was chosen to be 0.75. All calculations were performed and plotted using MATLAB[25] and a script may be found in the Appendix.

4.4 Results and Discussion

4.4.1 General Biomarkers for Viral Infection

First, fold change (FC) and p-values were calculated by combining all titers and time points for each virus and comparing them against the combined healthy control samples from all days. Differentially expressed miRNAs were determined using the double filtering parameters of $p\ value \leq 0.05$ and $|FC| \geq 2$.[26] Figure 4.1 shows the resulting volcano plot with differentially expressed genes marked in the upper right and upper left sectors. MMV and EMC viruses were seen to show the most differentially expressed miRNAs for days 1, 2, and 3 post-infection. Based on the reduced pathogenesis and virulence of Reo3 compared to other viruses in CHO-K1, it was expected that Reo3 would show fewer differentially expressed miRNAs at these earlier time points and this is visualized in the volcano plot.
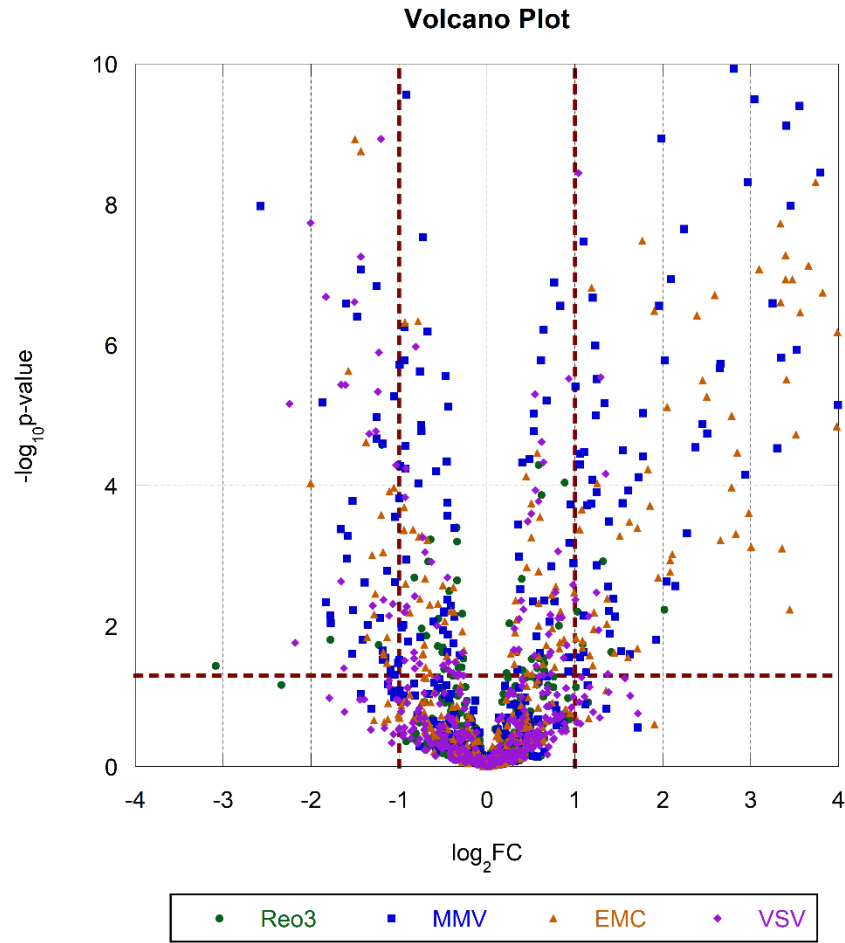
Figure 4.1 Volcano plot for CHO-K1 cells infected with Reo3, MMV, EMC, or VSV compared to healthy control on days 1, 2, and 3 post-infection. A $-log_{10}(pvalue) \geq 1.30$ and $|log_2 FC| \geq 1$ are considered significant, as seen in the top right and left segments of the plot.

Fold change (FC) and p-values were also calculated by comparing each viral condition to the uninfected control at the same time point. Using the same double filtering parameters, Table 4.2 shows miRNAs that were differentially expressed for a minimum of 10 different viral conditions (infecting virus, day post-infection, and viral titer). MicroRNAs are listed in order of how many conditions showed differential expression. Significant fold change ($|\log_2 FC| \geq 1$) is highlighted in red for downregulation and in green for upregulation.

Virus	Reo3									MMV									EMC			VSV		
Day post-infection	Day 1			Day 2			Day 3			Day 1			Day 2			Day 3			Day 1	Day 2	Day 3	Day 1	Day 2	Day 3
TCID$_{50}$	0.1	1	100	0.1	1	100	0.1	1	100	0.1	1	100	0.1	1	100	0.1	1	100	1	1	1	100	100	100
cgr-miR-33	0.0	-0.3	0.0	-0.5	-0.4	-0.1	-0.8	-1.1	-1.3	-2.3	-2.4	-2.1	-2.4	-1.1	-1.5	-1.1	-1.3	-1.2	-1.2	-1.9	-1.6	-1.7	-1.2	-2.6
cgr-miR-15a-3p	0.0	-0.2	0.0	-0.8	-0.8	-0.6	-0.8	-1.1	-1.2	-1.8	-1.7	-1.5	-1.9	-0.7	-1.2	6.1	6.1	6.1	-1.1	-1.2	6.0	-1.4	-1.1	-2.1
cgr-miR-17-5p	-0.1	-0.2	-0.2	-0.3	-0.4	0.0	-0.7	-1.0	-1.2	-1.8	-1.6	-1.7	-1.6	-0.9	-1.5	-2.0	-2.1	-1.6	-1.1	-1.8	-1.6	-1.3	-1.1	-1.6
cgr-miR-93-3p	0.2	0.2	0.1	0.5	0.6	-0.2	0.6	0.7	1.0	1.0	1.0	1.2	2.0	1.1	1.6	8.5	8.5	8.6	1.0	1.7	8.6	0.7	1.3	1.1
cgr-miR-32-5p	-0.1	-0.2	0.0	-0.8	-0.7	-0.5	-0.8	-1.1	-1.4	-2.9	-2.3	-2.0	-2.4	-0.8	-1.1	-3.5	-3.7	-3.2	-0.8	-1.2	-3.4	-1.7	-1.1	-3.5
cgr-miR-497-3p	0.0	-0.2	-0.1	-0.6	-0.7	0.0	-0.4	-0.8	-0.8	-2.4	-2.3	-2.6	-2.4	-1.3	-1.3	3.5	3.5	3.3	-1.6	-1.7	3.3	-1.4	-1.0	-2.2
cgr-miR-615-3p	0.2	0.2	0.1	1.0	1.4	1.0	0.7	0.6	0.9	2.0	1.3	1.5	2.4	0.7	1.6	-7.9	-8.0	-8.0	0.5	1.0	-7.9	0.8	1.1	2.5
cgr-miR-24-5p	-0.1	-0.2	-0.1	-0.3	-0.3	0.0	-0.4	0.2	-0.6	-2.3	-1.9	-2.0	-1.9	-1.0	-1.9	-8.3	-6.2	-8.4	-0.8	-2.0	-6.6	-1.4	-0.9	-2.3
cgr-miR-34c-3p	0.0	0.0	0.0	0.5	0.3	0.1	0.1	0.5	0.8	1.2	1.2	1.2	2.1	1.1	1.3	7.1	7.1	7.2	1.2	1.7	7.3	0.8	1.1	-0.1
cgr-miR-92a-3p	0.0	0.2	0.0	0.6	0.6	0.3	0.7	0.9	1.6	1.4	1.2	1.2	2.6	1.1	1.6	-3.1	-3.0	-3.1	0.9	1.6	-2.8	0.7	1.0	0.4
cgr-miR-130a-5p	0.0	-0.1	0.0	-0.3	-0.4	0.1	-1.0	-1.1	-1.7	-1.5	-1.4	-1.3	-1.3	-0.9	-1.3	7.1	7.2	7.2	-0.6	-1.3	7.0	-0.9	-0.7	-1.7
cgr-miR-142-3p	-0.1	-0.1	0.0	-0.6	-0.6	-0.6	-0.6	-0.7	-1.1	-1.5	-1.4	-1.4	-2.1	-0.8	-1.3	-3.2	-3.0	-3.3	-0.7	-1.3	-3.3	-0.8	-0.9	-1.7
cgr-miR-16	0.0	-0.2	0.1	-0.7	-0.5	-0.2	-0.7	-0.7	-1.0	-1.4	-1.4	-1.2	-1.9	-0.7	-1.1	10.6	10.6	10.7	-1.0	-1.4	10.8	-1.0	-0.9	-1.5
cgr-miR-18a-5p	-0.2	-0.2	-0.1	-0.5	-0.5	-0.3	-0.6	-0.9	-1.0	-1.5	-1.3	-1.3	-1.5	-0.5	-0.9	-6.7	-7.0	-6.3	-0.7	-1.1	-6.1	-1.0	-0.8	-1.7
cgr-miR-137-5p	0.1	0.2	0.0	0.3	0.3	0.2	0.4	0.2	0.1	1.4	0.9	1.2	2.2	1.6	1.6	9.5	9.5	9.6	0.8	2.2	9.6	0.8	1.3	0.7
cgr-miR-181b-5p	-0.1	0.2	0.0	0.5	0.3	0.0	0.0	0.7	1.0	1.0	1.0	0.9	2.0	1.2	1.3	-4.4	-4.2	-4.2	1.1	1.8	-4.2	0.6	1.0	-1.2
cgr-miR-374-3p	0.0	-0.1	0.0	-0.3	-0.3	0.0	-0.7	-0.9	-1.0	-1.4	-1.4	-1.3	-1.3	-0.4	-0.9	0.8	0.7	1.2	-1.1	-1.0	1.0	-0.8	-1.0	-1.7
cgr-miR-423-3p	0.0	0.1	0.0	0.7	0.9	0.7	0.6	0.9	1.4	1.9	1.3	1.5	2.7	0.9	1.9	-1.0	-1.1	-0.9	1.2	2.0	-1.1	0.9	1.2	0.6
cgr-miR-425-5p	-0.1	0.2	0.0	0.3	0.1	-0.2	0.1	0.2	0.5	1.1	1.0	1.0	1.9	1.1	1.3	-4.0	-3.8	-4.0	1.0	1.7	-3.9	0.6	1.0	-0.1
cgr-miR-450b-5p	-0.3	-0.2	-0.2	-0.1	-0.4	-0.2	-2.8	-2.9	-3.2	-1.3	-1.2	-1.1	-0.5	0.4	0.4	-2.5	-2.5	-2.6	-0.6	0.2	-2.5	-0.8	-0.2	-4.3
cgr-miR-542-3p	-0.2	-0.2	-0.2	-0.6	-0.5	-0.3	-1.2	-1.0	-1.4	-0.6	-0.6	-0.6	-1.4	-0.7	-1.0	-6.2	-6.3	-6.3	-0.3	-1.0	-6.2	-0.2	-0.7	-1.0
cgr-let-7c-2	0.1	-0.1	0.0	-0.5	-0.3	0.0	-0.2	-0.3	-0.4	-1.3	-1.3	-1.0	-1.7	-0.7	-1.0	4.9	4.9	4.7	-1.0	-1.1	4.8	-0.7	-0.7	-0.7
cgr-miR-101b-3p	-0.1	-0.2	-0.1	-0.6	-0.5	-0.2	-0.1	-0.2	-0.3	-1.1	-1.1	-1.0	-1.6	-0.7	-1.1	-12.1	-11.7	-12.1	-0.8	-1.2	-12.3	-0.6	-0.8	-0.6
cgr-miR-101b-3p_(2)	-0.1	-0.2	-0.1	-0.6	-0.5	-0.2	-0.1	-0.2	-0.3	-1.1	-1.1	-1.0	-1.6	-0.7	-1.1	-13.2	-12.9	-12.4	-0.8	-1.2	-12.5	-0.6	-0.8	-0.6
cgr-miR-106b-5p	-0.1	-0.1	-0.1	-0.4	-0.4	-0.1	-0.2	-0.5	-0.7	-2.2	-1.8	-1.9	-1.9	-0.4	-1.1	-1.0	-1.1	-0.9	-0.9	-1.2	-0.8	-1.4	-0.6	-1.9
cgr-miR-1306-5p	0.1	0.2	0.0	0.5	0.6	0.3	0.0	0.3	0.5	1.7	1.2	1.6	1.8	0.5	1.1	2.4	2.8	2.9	0.6	0.9	3.1	0.8	0.6	1.9
cgr-miR-142-5p	0.0	-0.3	0.0	-0.5	-0.3	0.1	-0.2	-0.2	-0.5	-1.3	-1.4	-1.6	-2.0	-1.0	-1.0	2.2	2.3	2.1	-0.9	-1.4	2.0	-0.9	-0.8	-0.9
cgr-miR-181d-3p	0.1	-0.1	0.0	1.0	0.8	0.7	0.1	0.2	0.6	1.4	1.1	1.0	2.5	0.9	1.7	6.2	6.2	6.7	0.2	1.2	6.8	0.3	0.6	0.9
cgr-miR-196a-3p	0.1	-0.1	0.1	-0.5	-0.5	-0.5	0.0	-0.2	-0.3	-1.2	-1.1	-1.1	-1.5	-0.3	-0.3	5.7	5.7	5.8	-1.0	-0.8	5.9	-0.9	-0.7	-1.1
cgr-miR-19a	0.0	-0.2	-0.1	-0.6	-0.4	-0.1	-0.1	-0.4	-0.5	-1.9	-1.7	-1.7	-2.0	-0.8	-1.0	-0.2	-0.3	-0.3	-1.2	-1.5	-0.5	-1.3	-1.0	-1.2
cgr-miR-21-3p	-0.3	-0.5	-0.3	-0.2	-0.1	0.0	-0.4	-0.3	-1.1	-1.5	-1.4	-1.3	-1.3	-0.7	-0.6	-10.7	-11.8	-11.7	-0.9	-0.5	-11.1	-0.8	-0.4	-1.4
cgr-miR-29b-3p	0.0	-0.2	-0.1	-0.6	-0.5	-0.2	-0.8	-1.0	-1.4	-1.3	-1.1	-1.0	-2.2	-1.2	-1.6	-0.6	-0.4	-0.5	-0.4	-1.6	-0.7	-0.7	-1.1	-1.1
cgr-miR-301a-3p	0.0	-0.1	0.0	-0.6	-0.4	-0.1	-0.5	-0.5	-0.8	-1.3	-1.1	-1.0	-1.9	-1.0	-1.1	-6.0	-5.9	-5.7	-0.8	-1.5	-5.8	-0.8	-0.9	-0.9
cgr-miR-32-3p	0.1	-0.2	0.1	0.0	-0.1	0.4	-0.6	-0.9	-1.1	-1.2	-1.4	-1.2	-1.6	-0.3	-0.9	2.6	2.7	2.7	-0.7	-0.9	2.9	-0.9	-0.8	-1.5
cgr-miR-652-5p	-0.2	-0.2	-0.2	-0.4	-0.4	-0.1	-0.9	-1.1	-1.3	-1.5	-1.4	-1.4	-1.2	-0.3	-0.6	1.0	0.8	1.6	-1.0	-0.7	1.4	-1.0	-0.8	-1.6
cgr-miR-671-3p	0.1	0.0	0.0	0.1	0.8	0.6	-0.2	-0.1	0.5	1.5	1.0	0.9	2.1	0.8	1.4	5.1	5.3	5.3	0.2	1.2	5.4	0.6	0.9	0.5

Table 4.2 log_2 FC of miRNAs which were differentially expressed for a minimum of 10 conditions (infecting virus, day post-infection, and viral titer). $|log_2 FC| \geq 1$ is significant. Upregulation is highlighted in green. Downregulation is highlighted in red.

As visualized in the volcano plot, MMV and EMC infected cultures showed the strongest host cell miRNA response. Reo3 displayed almost no differentially expressed miRNAs on days 1 and 2 post-infection, however some miRNAs which showed a response on day 3 are similar to those affected by the other viruses. Of the 36 top miRNAs in Table 4.2, 10 showed significant fold change in all 4 viruses with the fold change moving in one direction (either all upregulated or all downregulated). These miRNAs (miR-93-3p, miR-32-5p, miR-33, miR-17-5p, miR-24-5p, miR-18a-5p, miR-450b-5p, miR-542-3p, miR-21-3p, and miR-29b-3p) may be ideal candidates as general biomarkers for viral infection in CHO-K1.

Figure 4.2 shows a dot plot for these 10 miRNAs with each dot represent a different

sample condition (infecting virus, day post-infection, and viral titer).
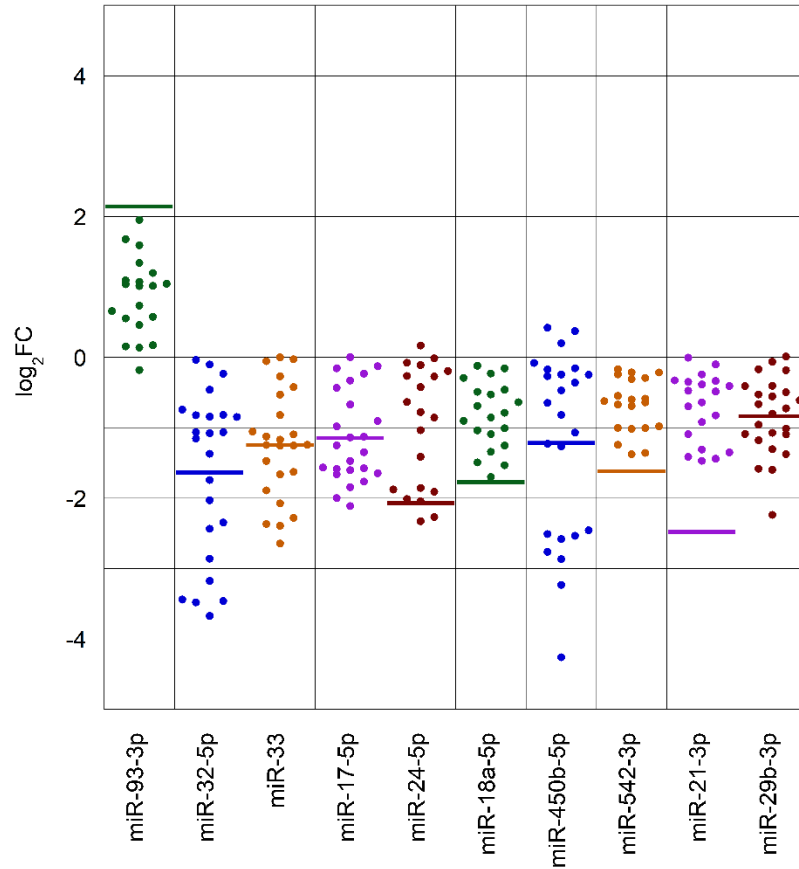
Figure 4.2 Dot plot and $log_2 FC$ averages for top miRNA targets in all infecting viruses (Reo3, MMV, EMC, and VSV). Each dot is a different viral condition (infecting virus, day post-infection, and viral titer). $|log_2 FC| \geq 1$ is significant. Some outliers are not visible on the plot due to their extremity.

It may be observed that the majority of potential general miRNA biomarkers are

downregulated and the only upregulated miRNA is miR-93-3p. We hypothesize

that this is explained by the fact that an infecting virus will take over host cell

pathways and mechanisms, preventing the cell from producing its own nucleic acids and proteins.

4.4.2 Receiver Operating Curve Analysis

In order to assess the classifying capability of each of these double filtered miRNAs, a receiver operating curve (ROC) analysis was utilized. Area under the curve (AUC) was calculated for each miRNA for days 1, 2, and 3 post-infection separately and also for all days combined. These AUC values are inclusive of all viruses tested and are tabulated in Table 4.3.

miRNA	Overall AUC	AUC, Day 1	AUC, Day 2	AUC, Day 3
miR-29b-3p	0.926	0.819	.931	1.00
miR-542-3p	0.923	.917	1.00	1.00
miR-33	0.902	0.903	0.889	0.988
miR-21-3p	0.901	1.00	0.792	0.988
miR-17-5p	0.889	0.917	0.833	1.00
miR-32-5p	0.887	0.882	0.972	0.994
miR-18a-5p	0.872	0.917	0.792	0.988
miR-450b-5p	0.864	0.958	0.444	1.00
miR-93-3p	0.827	0.917	0.903	0.963
miR-24-5p	0.797	0.875	0.806	0.667

Table 4.3 Calculated area under the curve (AUC) values for potential general biomarker miRNAs. AUCs are included for days 1, 2, and 3 post-infection, and an overall AUC for all days combined. An AUC of 0.5 is considered non-discriminant, 1.0 is perfect discrimination.

An AUC of 0.5 is non-discriminant and an AUC of 1 shows perfect discrimination capability. These 10 miRNAs all exhibited AUC values above 0.75 with 4 of them having values upwards of 0.9, indicating that these miRNAs have good discrimination capability between infected and uninfected cultures. Figure 4.3 shows the ROC plots for 8 of these miRNAs for all days post-infection combined.
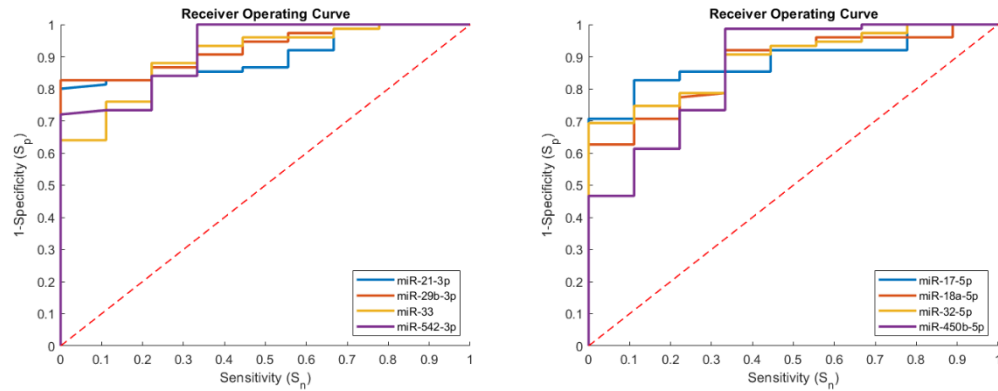
Figure 4.3 Receiver operating curves for top miRNA biomarkers of general viral infection in CHO-K1 cells for days 1, 2, and 3 post-infection combined.

4.5 Conclusions

Overall, we have successfully identified 10 differentially expressed miRNAs (miR-93-3p, miR-32-5p, miR-33, miR-17-5p, miR-24-5p, miR-18a-5p, miR-450b-5p, miR-542-3p, miR-21-3p, and miR-29b-3p) with promise as biomarkers for general viral infection in CHO-K1 cultures compared to healthy, uninfected cells. With Reo3, a less potent virus, these biomarker miRNAs displayed differential expression at a later infection time point (day 3) compared to the other 3 viruses (MMV, EMC, and VSV) for which the miRNAs reacted on days 1 and 2. This is unsurprising as EMC and VSV are known to be virulent in CHO-K1 and take over a culture quickly, causing mass cell death.[15] Additionally, in previous work, MMV-infected cultures were seen to show cytopathic effects (CPE) before those infected with Reo3.

In addition to passing through the double filtering method of fold change and p-value, these miRNAs also showed excellent discrimination ability between infected and uninfected cultures with an ROC analysis. Four miRNAs in particular (miR-29b-3p, miR-542-3p, miR-33, and miR-21-3p) displayed excellent discrimination

75

with an AUC above 0.9. These 4 miRNAs should form the basis for a fail-fast biomarker panel designed to detect viral infection in bioprocesses which utilize CHO-K1 cultures. Future work will involve the development of rapid detection techniques for these miRNAs and testing aimed at their implementation in upstream bioprocesses.

4.6 References

1. Hartig, S. M., Hamilton, M. P., Bader, D. A. & McGuire, S. E. The miRNA Interactome in Metabolic Homeostasis. *Trends in Endocrinology and Metabolism* vol. 26 733–745 Preprint at https://doi.org/10.1016/j.tem.2015.09.006 (2015).

2. Leva, G. di, Calin, G. A. & Croce, C. M. MicroRNAs: Fundamental facts and involvement in human diseases. *Birth Defects Res C Embryo Today* **78**, 180–189 (2006).

3. Leung, A. K. L. & Sharp, P. A. microRNAs: A Safeguard against Turmoil? *Cell* **130**, 581–585 (2007).

4. Leung, A. K. L. & Sharp, P. A. MicroRNA Functions in Stress Responses. *Mol Cell* **40**, 205–215 (2010).

5. Wang, J., Chen, J. & Sen, S. MicroRNA as Biomarkers and Diagnostics. *Journal of Cellular Physiology* vol. 231 25–30 Preprint at https://doi.org/10.1002/jcp.25056 (2016).

6. Tribolet, L. *et al.* MicroRNA Biomarkers for Infectious Diseases: From Basic Research to Biosensing. *Frontiers in Microbiology* vol. 11 Preprint at https://doi.org/10.3389/fmicb.2020.01197 (2020).

7. Verma, P., Pandey, R. K., Prajapati, P. & Prajapati, V. K. Circulating microRNAs: Potential and emerging biomarkers for diagnosis of human infectious diseases. *Frontiers in Microbiology* vol. 7 Preprint at https://doi.org/10.3389/fmicb.2016.01274 (2016).

8. Condrat, C. E. *et al.* miRNAs as Biomarkers in Disease: Latest Findings Regarding Their Role in Diagnosis and Prognosis. *Cells* vol. 9 Preprint at https://doi.org/10.3390/cells9020276 (2020).

9. Wegman, D. W. & Krylov, S. N. Direct miRNA-hybridization assays and their potential in diagnostics. *TrAC - Trends in Analytical Chemistry* **44**, 121–130 (2013).

10. Weber, M. J. New human and mouse microRNA genes found by homology search. *FEBS Journal* **272**, 59–73 (2005).

11. Veeranagouda, Y. *et al.* Identification of MicroRNAs in Macaca fascicularis (Cynomolgus monkey) by homology search and experimental validation by small RNA-seq and RT-qPCR using kidney cortex tissues. *PLoS One* **10**, (2015).

12. Koenig, E. M. *et al.* The beagle dog MicroRNA tissue atlas: Identifying translatable biomarkers of organ toxicity. *BMC Genomics* **17**, (2016).

13. Walsh, G. Biopharmaceutical Benchmarks 2014. *Nat Biotechnol* **32**, 992–1000 (2014).

14. Barone, P. W. *et al.* Viral contamination in biologic manufacture and implications for emerging therapies. *Nat Biotechnol* **38**, 563–572 (2020).

15. Berting, A., Farcet, M. R. & Kreil, T. R. Virus susceptibility of Chinese hamster ovary (CHO) cells and detection of viral contaminations by adventitious agent testing. *Biotechnol Bioeng* **106**, 598–607 (2010).

16. *mirVana* ^{TM} *miRNA Isolation Kit*. (2011).

17. Andrews, S. FastQC: a quality control tool for high througgput sequence data. Preprint at (2010).

18. Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048 (2016).

19. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

20. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357–359 (2012).

21. Hammond, S., Kaplarevic, M., Borth, N., Betenbaugh, M. J. & Lee, K. H. Chinese hamster genome database: An online resource for the CHO community at. *Biotechnology and Bioengineering* vol. 109 1353–1356 Preprint at https://doi.org/10.1002/bit.24374 (2012).

22. Friedländer, M. R., MacKowiak, S. D., Li, N., Chen, W. & Rajewsky, N. MiRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res* **40**, 37–52 (2012).

23. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, (2014).

24. Afgan, E. *et al.* The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic Acids Res* **46**, W537–W544 (2018).

25. MathWorks, I. *MATLAB : the language of technical computing : computation, visualization, programming : installation guide for UNIX version 5*. (Natwick : Math Works Inc., 1996., 1996).

26. Li, W. Volcano plots in analyzing differential expressions with mRNA microarrays. *J Bioinform Comput Biol* **10**, 1–24 (2012).

Chapter 5 – Specific miRNA Biomarker Discovery for Four Virus Infections in CHO-K1

5.1   Introduction

Many studies have shown how miRNA expression is deeply involved in and can be used as an indicator for various diseases, cancers, and cellular stresses.[1–5] In this thesis, we make a new connection between pharmaceutical CHO-K1 production cultures and the ability of the miRNA transcriptome to indicate cellular stress – particularly in the case of viral infection. In previous work, we have identified a panel of general biomarkers for infection in CHO-K1 cells. Here, we analyze 4 viruses separately to compare miRNA transcriptome changes between viral infection caused by double-stranded and single-stranded, enveloped and non-enveloped, and DNA and RNA viruses. Our goal is to identify a minimum of 1-2 miRNA biomarkers specific to each virus to supplement the general biomarker panel previously developed. This work will contribute to a more comprehensive understanding of the fundamental modes of action of viral infection in CHO-K1 cells by analyzing virus-specific biomarkers. Additionally, this work may also give insight on how changes in miRNA expression are connected with the modes of propagation for these different types of viruses and contribute to understanding general cell health even beyond viral infection.

5.2 Materials and Methods

5.2.1 Experimental Design

Four different viruses were analyzed for the changes their infections caused in the miRNA transcriptome of CHO-K1 cells. These four viruses included mammalian orthoreovirus 3 (Reo3), rodent protoparvovirus 1 (MMV), parainfluenza virus 2 (PI2), and encephalomyocarditis virus (EMC). The CHO-K1 miRNA transcriptome was analyzed at later time points (days 3 and 7) after some cells had started showing cytopathic effects (CPE) to ensure infection and that the cellular pathways had time to respond and fully react. The number of RNA isolation samples and each condition (virus and day post-infection) are listed in Table 5.1. The samples from EMC infection on day 7 post-infection were unable to be isolated, as all the cells in these cultures had died from the virus. Each condition was assessed in triplicate to ensure a rigorous statistical analysis.

	Day 3	Day 7
Healthy Control	3	3
Reo3 10 TCID$_{50}$	3	3
MMV 10 TCID$_{50}$	3	3
PI2 10 TCID$_{50}$	3	3
EMC 10 TCID$_{50}$	3	X

Table 5.1 Number of samples processed using Affymetrix and next generation sequencing (NGS) for each virus, viral titer, and day post-infection.

5.2.2 Cell Culture

Healthy CHO-K1 cells (ATCC) were cultured and plated into 6-well plates. Each well was seeded with 400,000 cells in 4 mL of F-12K nutrient mixture containing 10% fetal bovine serum (FBS). Plates were cultured overnight at 37°C. At > 70% confluence, after approximately 24 hours, the wells were infected with 2 mL of a

virus solution. Virus solutions were diluted to 10 $TCID_{50}$, from the manufacturer's listed stock solution $TCID_{50}$ value, with a F-12K nutrient mixture containing 2% FBS. Wells were aspirated, washed with PBS, and 2 mL of virus solution was introduced to each well. Cultures were incubated at 37°C. After 2 hours, 2 mL of additional F-12K media solution containing 2% FBS was added. Total RNA was isolated on days 3 and 7 post-infection. Fresh nutrient media was added on day 4 post-infection to keep the cells from starvation. Triplicates of healthy control cultures and virally infected cultures for each virus were harvested on days 3 and 7 post-infection. All cultures were incubated at 37°C until RNA isolation. Total RNA was isolated by organic extraction using the mirVana™ miRNA Isolation Kit with phenol.[6] After isolation, sample quality was assessed for RNA concentration and organic contaminants using a NanoDrop instrument. Total RNA isolates were stored at $-80°C$ until their analysis with Affymetrix GeneChip.

5.2.3 Affymetrix

Total RNA isolations were analyzed at the Pitt Genomics Core using an Affymetrix miRNA GeneChip 4.0 array. At the genomics core, the miRNA isolations underwent Poly(A) tailing and biotin labeling for detection with the GeneChip arrays. These GeneChips contained 30,424 miRNA probe sets for miRNAs from 203 different organisms (including the Chinese hamster, Cricetulus griseus) and were considered a good tool for a broad analysis of miRNA transcriptomes. Affymetrix data were analyzed with the assistance of Transcriptome Analysis Console (TAC) software from ThermoFisher Scientific.[7]

5.2.4 Next Generation Sequencing

Samples analyzed using Affymetrix were also assessed using next generation sequencing (NGS). To ensure sample quality, each total RNA sample was analyzed using a Bioanalyzer instrument. The samples libraries were prepared for NGS using the Qiagen miRNA library prep, and these were sequenced using Illumina NextSeq 550 and 2000 sequencers. QC analysis of the data was performed using fastqc and multiqc.[8,9] Trimmomatic 038[10] was utilized to clean up the reads and then Bowtie2[11] was used for read mapping to the CHO-K1 reference genome obtained from CHOgenome.org.[12] With the miRDeep2 pipeline[13] and the hairpin and mature sequences obtained from miRBase, miRNA was quantified allowing for 1 mismatch and only including reads at least 18 nucleotides in length. Statistical analysis between groups was evaluated using deseq2 on the public usegalaxy.org server.

5.2.5 Receiver Operating Curve Analysis

The normalized count files from both Affymetrix and NGS data were used to create a receiver operating curve (ROC) curve for binary classification to quantify each miRNA's ability to discriminate between infected and uninfected cultures. In this binary system, virally infected samples were classified as positive, and uninfected samples were classified as negative. True positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) were calculated over the range of normalized counts and sensitivity (S_n) and specificity (S_p) were derived from this classification system at each point using Eq. 4.1 and Eq. 4.2.

$$S_n = \frac{TP}{TP + FN}$$ *Eq. 5.1*

$$S_P = \frac{TN}{TN + FP}$$ *Eq. 5.2*

Sensitivity (S_n) was plotted against ($1 - S_p$) and an area under the curve (AUC) was calculated for each miRNA. A classifier with an AUC of 0.5 shows no ability to discriminate, therefore, our minimum AUC cutoff point for miRNA biomarkers was chosen to be 0.70.

5.3   Results and Discussion

5.3.1   Reo3 Specific Biomarkers

Figure 5.1 shows a volcano plot comparing Affymetrix and NGS data from days 3 and 7 post-infection with Reo3 infection. It can be observed that far more miRNAs were differentially expressed using double filtering criteria from the Affymetrix data than from the NGS data. However, this is the result of insignificant p-values in the NGS data set, not fold change. By day 7, Reo3 infection had taken over the whole cell culture for samples that were to be isolated and CHO-K1 cells showed visual changes to their cellular morphology (cytopathic effects).
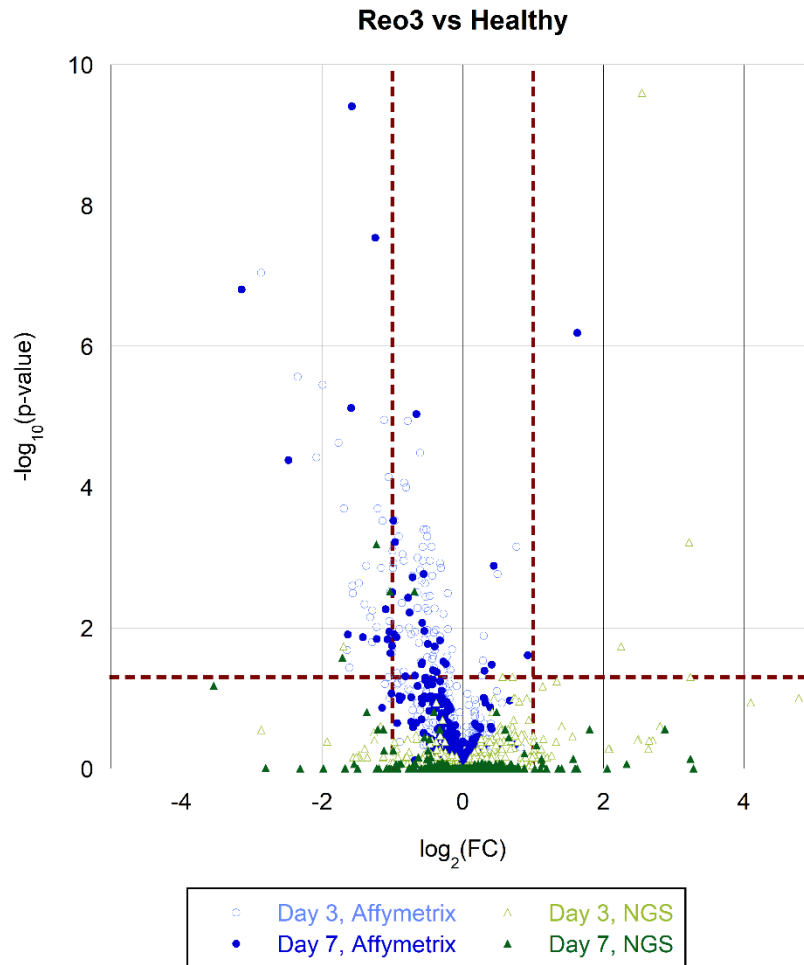
Figure 5.1 Volcano plot of Reo3 infection for Affymetrix and NGS data days 3 and 7 (late infection).

Figure 5.2 shows a volcano plot from early stage Reo3 infection (days 1, 2, and 3 post infection). It is seen that most double filtered miRNAs are from day 3 post-infection, with a few from day 2 and none from day 1, inclusive of all titer values. The strongest miRNA transcriptome response is seen from day 3 post-infection at the highest titer (100 TCID50).
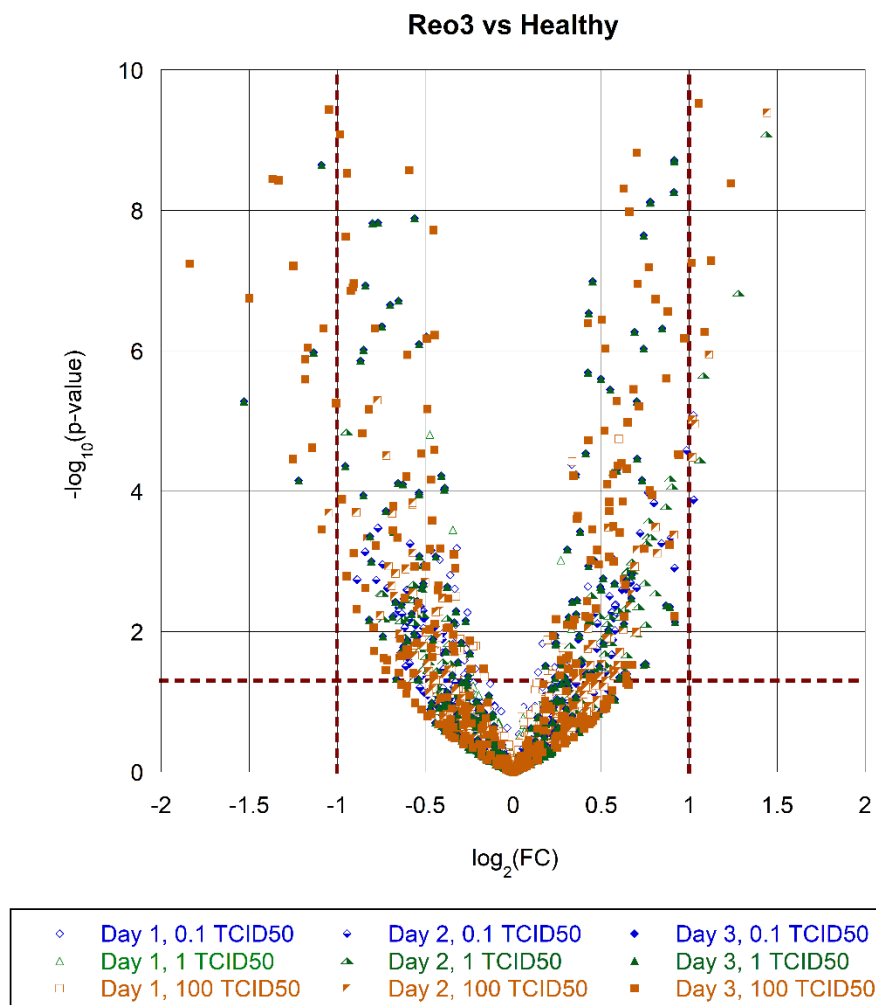
Figure 5.2 Volcano plot of Reo3 infection for 3 titer levels (0.1, 1, and 100 TCID$_{50}$) on days 1, 2, and 3 post-infection (early infection).

Significant miRNAs for NGS (early infection and late infection) and Affymetrix data were filtered out and their $\log_2 FC$ values across all three experiments are tabulated in Table 5.2. Since this work is focused on identification of miRNA biomarkers specific to Reo3 infection, any miRNAs which were identified as general biomarkers were filtered out. Significant miRNAs from Affymetrix data were filtered to include those based on those showed differential expression on both days 3 and 7 post-infection. Since Reo3 viral infection is less virulent than other

viruses tested, significant miRNAs from NGS data early-stage infection, were filtered to include those which showed double filtering for at least 2 out of the 3 time points. Additionally, miRNAs from late-stage infection NGS data were included in Table 5.2 if they showed differential expression for either time point.

miRNA	NGS, early infection (log2FC)			Affymetrix (log2FC)		NGS, late infection (log2FC)	
	Day 1	Day 2	Day 3	Day 3	Day 7	Day 3	Day 7
cgr-miR-125a-3p	-0.07	0.34	0.05	-1.14	-1.01	-0.03	0.95
cgr-miR-222-5p	-0.10	-0.08	0.46	-1.29	-1.04	-0.02	-0.63
cgr-miR-409-5p	0.16	-0.13	-0.17	-1.04	-1.10	1.14	-0.49
cgr-miR-1306-5p	0.24	0.73	0.50	-1.40	-1.22	0.98	-0.76
cgr-miR-210-3p	0.09	-0.09	-0.85	-1.12	-1.58	-0.50	-1.22
cgr-miR-199b	0.08	0.00	-0.89	-1.48	-1.63	0.13	-0.45
cgr-miR-210-5p	0.12	0.44	-1.96	-2.87	-3.14	-1.47	-3.54
cgr-miR-1249	-0.38	1.31	1.10			0.33	-1.19
cgr-miR-1285	0.63	1.18	1.75	0.77	0.14	3.24	1.19
cgr-miR-194-2	-0.32	1.02	1.24	-0.51	-0.43	-1.03	-0.32
cgr-miR-143	1.20	0.07	1.12	-0.07	0.04	2.50	0.01
cgr-miR-222-3p	0.08	0.32	0.22	-0.03	-1.24	0.54	-1.03
cgr-miR-409-3p	-0.12	-0.01	-1.22	-0.64	-2.48	0.01	-1.71
cgr-miR-2424	0.45	1.31	-0.86	0.82	-0.16	2.25	1.80
cgr-miR-3068	0.08	0.31	-0.26			2.55	0.45
cgr-miR-1949	-0.02	0.29	0.76	-0.93	0.92	3.22	0.05

Table 5.2 log$_2$FC values of miRNAs which were significant across Affymetrix and NGS data (early and late infection) for Reo3 infection in CHO-K1. Upregulated log$_2$FC>1 values are highlighted in green. Downregulated log$_2$FC<-1 values are highlighted in red.

From Table 5.2, the miRNAs which showed consist upregulation or downregulation between at least two of the three experiments were miR-210-3p, miR-210-5p, miR-1285, miR-143, miR-2424, miR-222-3p, and miR-409-3p. For these, a ROC analysis was used to calculate AUC values (Table 5.3) for each miRNA to determine their classifier capabilities.

Target miRNA	NGS, early infection	Affymetrix, late infection	NGS, late infection
cgr-miR-210-3p	0.650	1.000	1.000
cgr-miR-210-5p	0.634	1.000	0.972
cgr-miR-1285	0.864	0.792	0.681
cgr-miR-143	0.831	0.556	0.556
cgr-miR-2424	0.597	0.611	0.653
cgr-miR-222-3p	0.617	0.889	0.639
cgr-miR-409-3p	0.593	0.833	0.708

Table 5.3 AUC values for specific biomarker miRNAs in Reo3 infection for NGS (early and late infection) and Affymetrix data. An AUC of 0.5 is non-discriminant, an AUC of 1 shows perfect discrimination ability.

The NGS data set for early infection is much larger than those of Affymetrix and NGS of late infection samples. Therefore, an ROC plot and its corresponding AUC value is likely to be the most reliable for the early infection NGS data. However, given that the data sets measure transcriptome response for different stages of infection, it is not surprising that variation between the results would occur which may be the outcome of actual biological changes in the miRNA transcriptome over the course of infection. The ROC plots associated with Table 5.3 are shown in Figure 5.3. For the specific miRNA biomarker targets of Reo3 infection in CHO-K1 cells, miR-210-3p and miR-210-5p showed perfect discrimination ability for later-stage infection. Overall, across all time points, miR-1285 had the best discrimination ability and the least likelihood of giving false positives.
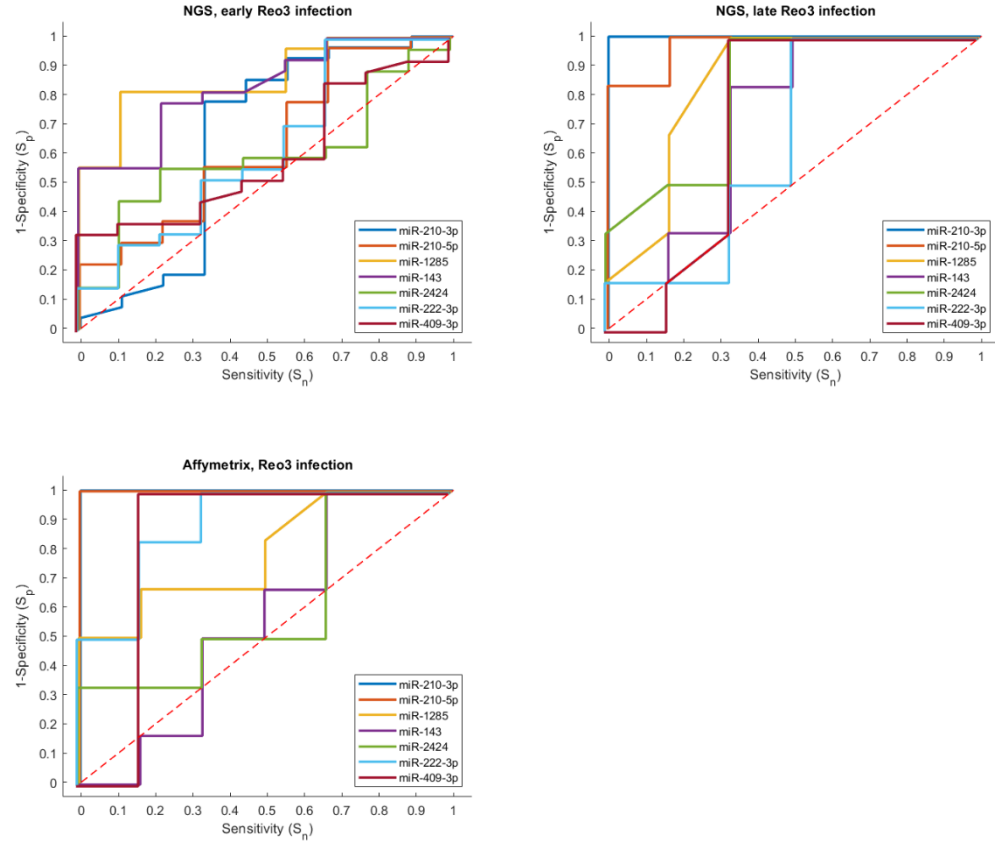
Figure 5.3 ROC plots associated with AUC values in Table 5.3.

5.3.2   MMV Specific Biomarkers

Figure 5.4 shows a volcano plot comparing Affymetrix and NGS data from days 3

and 7 post-infection with MMV infection. As with results from the Reo3 infection

in CHO-K1 cells, it can be observed that far more miRNAs are double filtered from

the Affymetrix data than from the NGS data. Again, this seems to be the result of

insignificant p-values in the NGS data set, not fold change. Double filtered

miRNAs from the Affymetrix data are almost completely downregulated, with only

3 miRNAs from day 7 passing double filtering parameters for upregulation and day

3 post-infection showing a stronger transcriptome response than day 7.

87
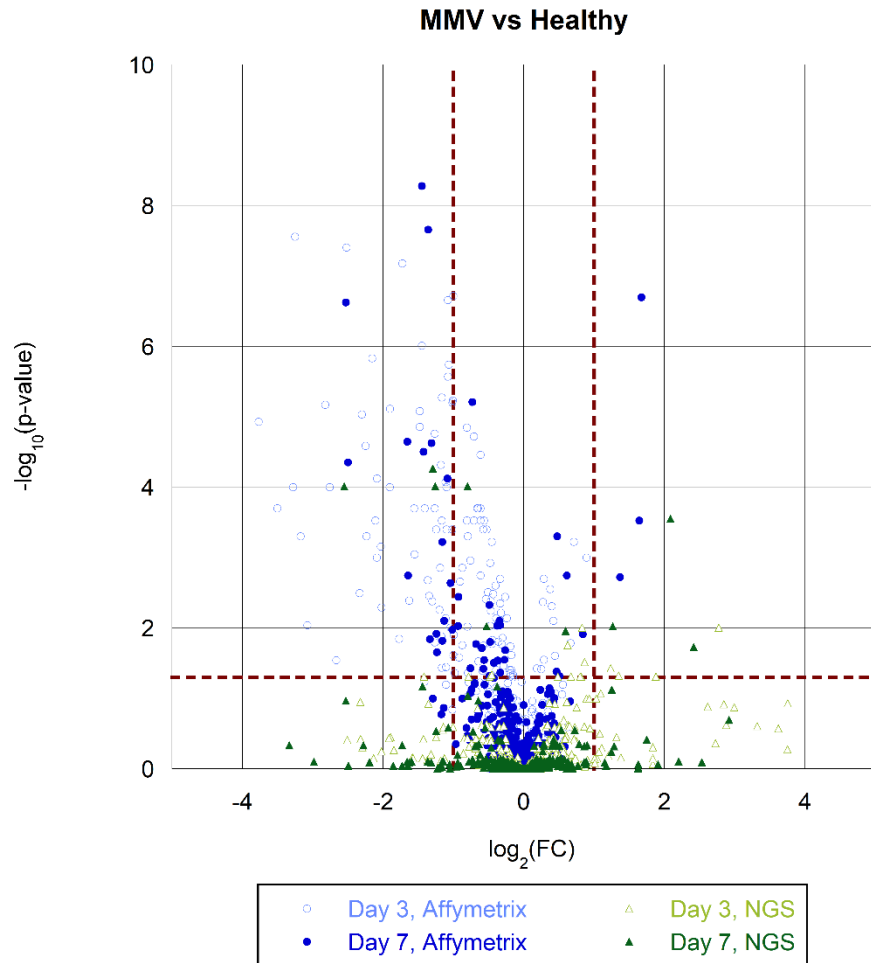
Figure 5.4 Volcano plot of MMV infection for Affymetrix and NGS data days 3 and 7 (late infection).

Figure 5.5 shows the volcano plot from early stage MMV infection (days 1, 2, and 3 post infection) at three different titer levels (0.1, 1, and 100 $TCID_{50}$). Fold change (FC) and p-value data for day 3 across all titers were extremely significant for numerous miRNAs and many of them had to be cut off to display the volcano plot. Viral titer experiments had previously shown that MMV infection displays cytopathic effects before Reo3 infection, which could account for the variation between this data and milder Reo3 results. Again, the strongest transcriptome response between days 1, 2, and 3 post-infection was observed on day 3.
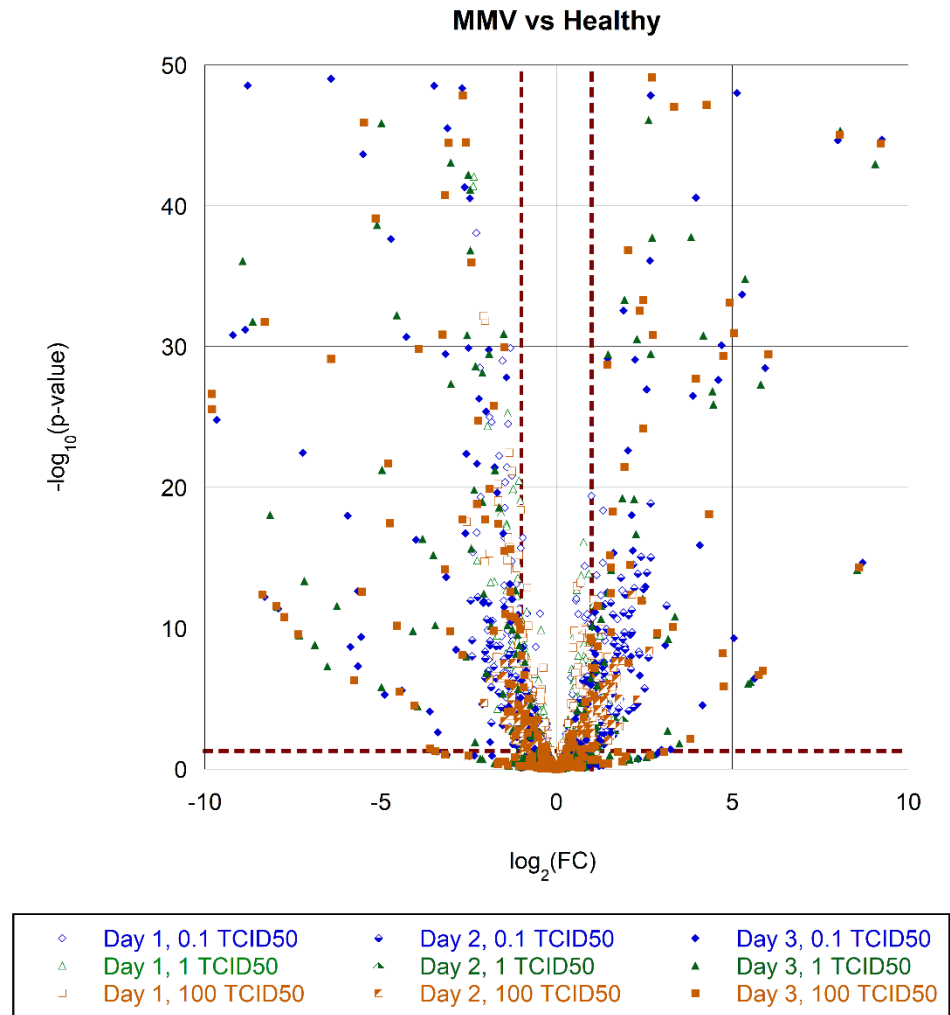
Figure 5.5 Volcano plot of MMV infection for 3 titer levels (0.1, 1, and 100 TCID$_{50}$) on days 1, 2, and 3 post-infection (early infection).. Some extreme data points are cut off.

Significant miRNAs for NGS (early infection and late infection) and Affymetrix data were filtered and their $\log_2 FC$ values across all three experiments are tabulated in Table 5.4. Since this work is focused on identification of miRNA biomarkers specific to MMV infection, any miRNAs which were identified as general biomarkers were filtered out. Significant miRNAs from Affymetrix data were filtered to include those based on those showed differential expression on both

89

days 3 and 7 post-infection. Due to the strong transcriptomic response of CHO-K1 to MMV, NGS data from early infection was filtered to include miRNAs which were double filtered for all three time points and were consistent in being either upregulated or downregulated across samples. MicroRNAs from late-stage infection NGS data were included if they showed differential expression for either time point.

miRNA	NGS, early infection (log2FC)			Affymetrix (log2FC)		NGS, late infection (log2FC)	
	Day 1	Day 2	Day 3	Day 3	Day 7	Day 3	Day 7
cgr-miR-146b-3p	1.40	0.63	8.04	-1.11	-1.13	0.25	0.42
cgr-miR-210-3p	0.14	-0.71	-10.79	-1.72	-1.66	-0.79	-1.26
cgr-miR-210-5p	0.73	0.65	-5.11	-3.25	-2.53	-2.30	-2.53
cgr-miR-222-5p	0.17	0.38	5.90	-1.24	-1.04	0.53	-0.28
cgr-miR-409-3p	1.01	0.54	-2.65	-1.00	-2.50	-0.69	-2.55
cgr-miR-409-5p	0.94	0.45	1.09	-1.01	-1.01	0.48	-0.78
cgr-miR-101b-3p	-1.12	-1.05	-11.37	-3.77	-1.29	-0.47	0.05
cgr-miR-125b	1.21	1.16	8.70	-0.24	-0.19	-0.03	-0.28
cgr-miR-1306-5p	1.64	1.15	2.77	-0.98	-0.59	1.01	-1.73
cgr-miR-137-5p	1.29	1.60	9.25	-0.26	0.84	2.73	0.16
cgr-miR-142-3p	-1.38	-1.20	-3.06	-0.03	0.04	-0.66	-0.25
cgr-miR-181d-3p	1.23	1.71	6.28	-0.18	0.19	1.11	-0.45
cgr-miR-193b-5p	1.02	1.38	1.58	-0.11	0.04	-0.05	1.17
cgr-miR-301a-3p	-1.17	-1.22	-5.66	-2.50	-0.93	-1	-0.23
cgr-miR-499-5p	-1.16	-1.02	-5.75	-0.23	-0.20	-0.26	-0.19
cgr-miR-671-3p	1.26	1.39	5.24	-0.75	0.41	0.91	-1.36
cgr-miR-141	-0.90	-0.60	-0.09	0.08	0.10	0.60	2.09
cgr-miR-222-3p	0.84	1.14	-1.79	0.01	-1.36	0.57	-1.29
cgr-miR-2424	-0.20	0.43	3.27	-1.26	0.21	1.88	2.42
cgr-miR-3068	0.02	0.08	-0.36			2.82	1.27

Table 5.4 log$_2$FC values of miRNAs which were significant across Affymetrix and NGS data (early and late infection) for MMV infection in CHO-K1. Upregulated log$_2$FC>1 values are highlighted in green. Downregulated log$_2$FC<-1 values are highlighted in red.

From Table 5.4, the miRNAs which showed consist upregulation or downregulation between at least two of the three experiments were miR-210-3p, miR-210-5p, miR-409-3p, miR-101b-3p, miR-1306-5p, miR-181d-3p, miR-301a-

3p, and miR-222-3p. A ROC analysis was used to calculate AUC values (Table 5.5) for each of the miRNAs to determine their classifier capabilities.

Target miRNA	NGS, early infection	Affymetrix, late infection	NGS, late infection
cgr-miR-210-3p	0.755	1.000	1.000
cgr-miR-210-5p	0.574	1.000	0.611
cgr-miR-222-3p	0.642	0.833	0.639
cgr-miR-409-3p	0.570	0.972	1.000
cgr-miR-101b-3p	1.000	0.944	0.583
cgr-miR-1306-5p	0.994	0.944	0.542
cgr-miR-181d-3p	0.992	0.708	0.500
cgr-miR-301a-3p	1.000	1.000	0.694

Table 5.5 AUC values for specific biomarker miRNAs in MMV infection for NGS (early and late infection) and Affymetrix data. An AUC of 0.5 is non-discriminant, an AUC of 1 shows perfect discrimination ability.

ROC plots associated with AUC values are shown in Figure 5.6. Out of 8 potential biomarker targets, miR-210-3p and miR-301a-3p showed the best discrimination ability. It may be beneficial for application purposes to avoid miR-409-3p, miR-222-3p, miR1306-5p, miR-101b-3p, and miR-181d-3p as biomarkers for MMV infection, as they are likely to result in a greater number false positives or false negatives compared to other miRNA biomarkers. The remaining miRNA, miR-210-5p, may be utilized as a biomarker, but has significantly less discrimination ability that miR-210-3p and miR-301a-3p.
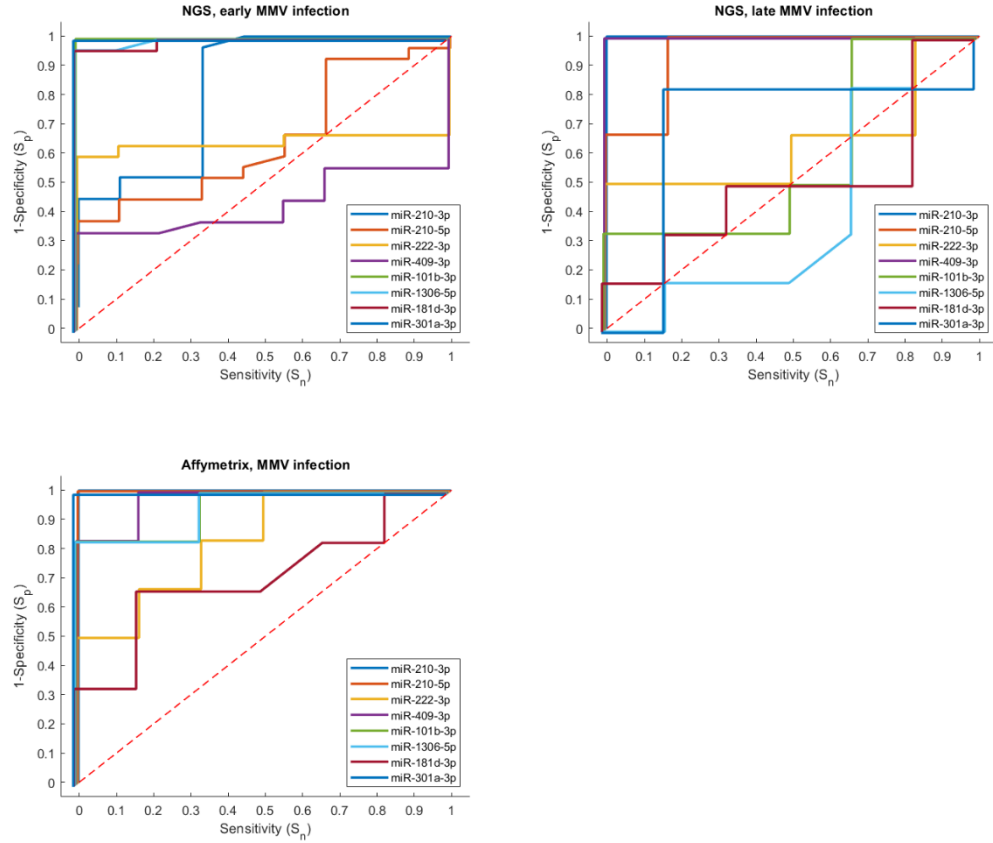
Figure 5.6 ROC plots associated with AUC values in Table 5.5.

5.3.3 EMC Specific Biomarkers

Figure 5.7 shows a volcano plot comparing Affymetrix and NGS data from days 3 and 7 post-infection with EMC infection in CHO-K1 cells. Unlike with the Reo3 and MMV NGS results, far more miRNAs passed the double filtering parameters of significance. An interesting observation is that the Affymetrix results seem to be mostly downregulated, while the NGS results have more upregulated targets. None of the NGS results for day 3 showed differential expression of any miRNA. For both Affymetrix and NGS, the miRNA transcriptome shows the strongest response on day 7 with EMC infection, as opposed to day 3 with MMV infection.
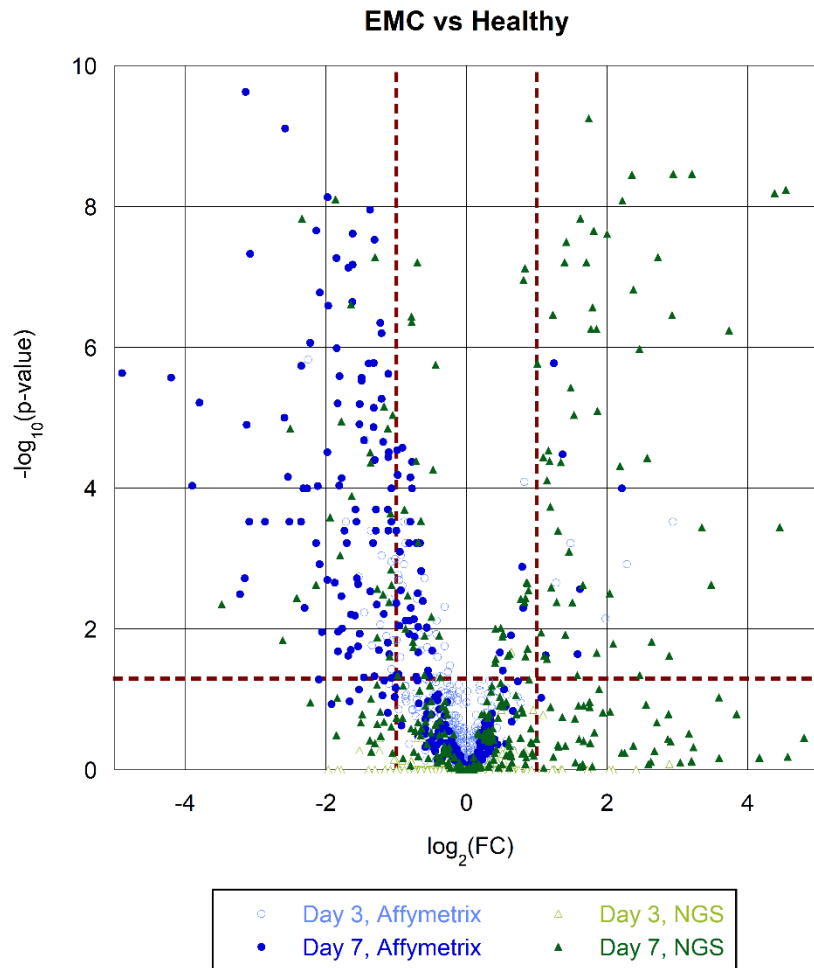
Figure 5.7 Volcano plot of EMC infection for Affymetrix and NGS data days 3 and 7 (late infection).

Figure 5.8 shows the volcano plot from the early stages of EMC infection (days 1, 2, and 3 post infection) at 1 $TCID_{50}$. Similar to early stage MMV infection, fold change and p-value data for some miRNAs on day 3 was extremely significant and many of them had to be cut off on the volcano plot. This volcano plot seems to be in contradiction with Figure 5.7 and the experiments which analyzed later-stage EMC infection in CHO-K1 cells in which the transcriptome did not show a strong response for day 3. However, the titer levels between the two experiments are not comparable, and may have led to the observed discrepancy. Late-stage infection

experiments were performed before virulence and $TCID_{50}$ were properly measured in the CHO-K1 cell line and the titer which was used to infect CHO-K1 cells for the first volcano plot may have been effectively much lower than indicated by the manufacturer's instructions. Additionally, it is possible that cycles of freeze-thawing with our virus stock may have made the particular solution of EMC used for late-stage infection experiments less virulent. As with Reo3 and MMV, the strongest transcriptome response between days 1, 2, and 3 post-infection for EMC was observed on day 3.
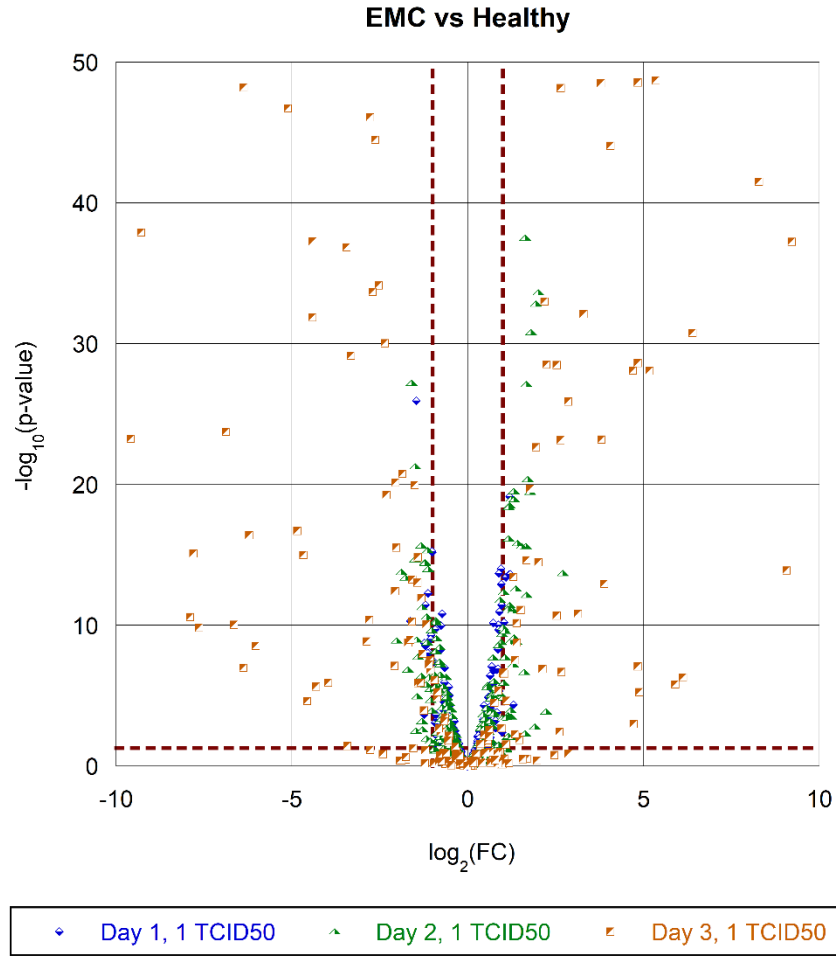
Figure 5.8 Volcano plot of EMC infection for 1 TCID$_{50}$ on days 1, 2, and 3 post-infection (early infection).

Significant miRNAs for NGS (early infection) and Affymetrix data were filtered and their $\log_2 FC$ values across all three experiments are tabulated in Table 5.6. Again, any miRNAs which were identified as general biomarkers were filtered out. Significant miRNAs from Affymetrix data were filtered to include those based on those showed differential expression on both days 3 and 7 post-infection. NGS data from early infection was filtered to include miRNAs which were double filtered for all three time points.

miRNA	NGS, early infection (log2FC)			Affymetrix (log2FC)		NGS, late infection (log2FC)	
	Day 1	Day 2	Day 3	Day 3	Day 7	Day 3	Day 7
cgr-miR-652-5p	-1.03	-0.64	0.62	-1.07	-1.24	-0.56	-2.41
cgr-miR-1306-5p	0.58	0.78	2.51	-1.53	-1.54	1.34	2.09
cgr-miR-505-5p	0.76	0.28	3.86	-1.14	-2.09	-0.53	1.15
cgr-miR-222-5p	0.14	0.18	4.91	-1.04	-2.14	0.36	-0.30
cgr-miR-598	0.06	0.15	1.09	-1.71	-2.32	0.18	-1.07
cgr-miR-210-5p	0.46	0.68	-2.91	-2.25	-2.54	-1.39	1.37
cgr-miR-497-5p	-0.35	-0.28	-0.70	-1.45	-3.09	-0.20	-3.48
cgr-let-7g-3p	-1.20	-0.85	5.99	2.93	-3.15	-0.54	-0.10
cgr-miR-199b	-0.65	-0.81	-0.96	-1.06	-3.79	-0.01	-1.36
cgr-miR-15a-3p	-1.04	-1.02	4.84	-0.06	-0.04	-0.65	-2.21
cgr-miR-181a-5p	1.12	1.28	-1.52	-0.04	-1.07	0.31	-0.28
cgr-miR-181b-5p	1.03	1.69	-2.15	0.06	-0.74	0.12	1.34
cgr-miR-34c-3p	1.07	1.49	6.15	-0.16	-1.11	0.43	0.02

Table 5.6 log$_2$FC values of miRNAs which were significant across Affymetrix and NGS data (early and late infection) for EMC infection in CHO-K1. Upregulated log$_2$FC>1 values are highlighted in green. Downregulated log$_2$FC<-1 values are highlighted in red.

As seen in Table 5.6, the miRNAs which showed consist upregulation or downregulation between at least two of the three experiments were miR-652-5p, miR-21p-5p, miR-497-5p, and miR-199b. A ROC analysis was used to calculate AUC values (Table 5.7) for each of these miRNAs determine their classifier capabilities.

Target miRNA	NGS, early infection	Affymetrix, late infection	NGS, late infection
cgr-miR-652-5p	0.782	1.000	0.681
cgr-miR-21-5p	0.751	0.611	0.667
cgr-miR-497-5p	0.665	1.000	0.625
cgr-miR-199b	0.646	0.917	0.611

Table 5.7 AUC values for specific biomarker miRNAs in EMC infection for NGS (early and late infection) and Affymetrix data. An AUC of 0.5 is non-discriminant, an AUC of 1 shows perfect discrimination ability.

Figure 5.9 shows the ROC plots from which the AUC values in Table 5.7 were calculated. Of these 4 miRNA biomarkers, none shows good discrimination ability across all stages of EMC infection. Almost all of the miRNAs show likelihood

towards giving false positives, and which one changes depending on the method of detection. Overall, miR-652-5p has the most potential as a specific biomarker for EMC infection in CHO-K1 cells, despite its proclivity towards false positives in the late infection NGS data.
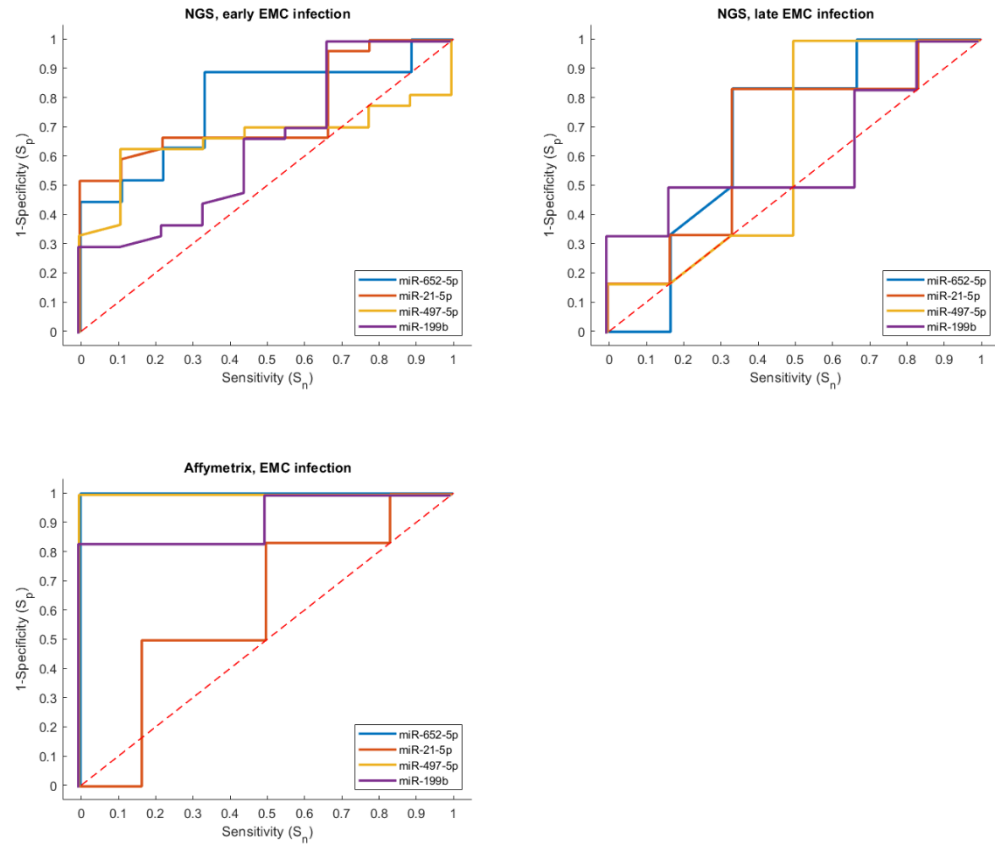
Figure 5.9 ROC plots associated with AUC values in Table 5.7.

5.3.4   VSV Specific Biomarkers

VSV was not analyzed for late-stage miRNA biomarkers through NGS or Affymetrix, due to its virulence in CHO-K1. Cell cultures would not remain viable through 7 days post-infection at any titer level for this virus. NGS data from days 1, 2, and 3 post-infection was used to analyzed biomarkers specific to this virus.

Figure 5.10 shows a volcano plot of the NGS data, in which miRNAs on day 3 show the strongest response to infection.
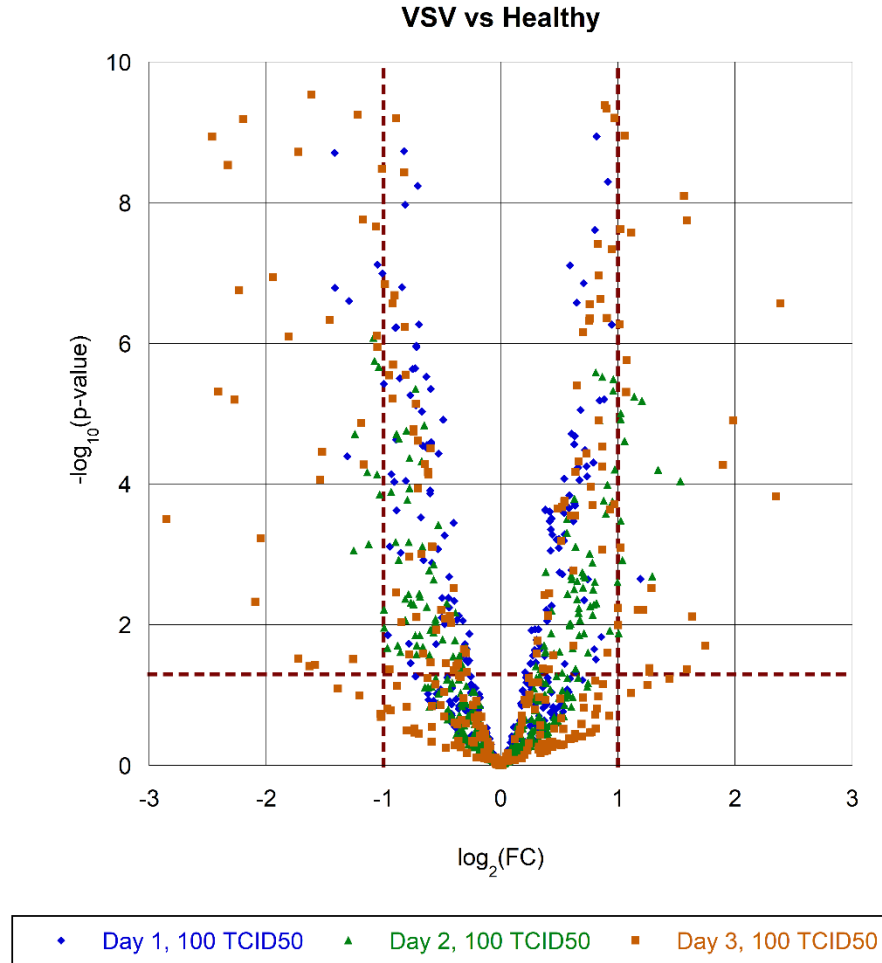
Figure 5.10 Volcano plot of VSV infection for 1 TCID$_{50}$ on days 1, 2, and 3 post-infection (early infection). Some values with extremely high -log$_{10}$p-value are cut off from top of the plot.

The $\log_2 FC$ values for miRNAs which showed differential expression on at least two days post-infection with VSV are listed in Table 5.8. Of these, miR-15a-3p, miR-17-5p, miR-33, and miR-497-3p are the most significant, passing the double filtering method for all three time points. Since miR-17-5p and miR-33 were

previously implicated as biomarkers for general infection in CHO-K1 cells, this indicates that miR-15a-3p and miR-197-3p may be specific to VSV infection.

miRNA	Day 1, log$_2$FC	Day 2, log$_2$FC	Day 3, log$_2$FC
cgr-miR-15a-3p	-1.34	-1.04	-2.13
cgr-miR-17-5p	-1.25	-1.21	-1.68
cgr-miR-33	-1.49	-1.11	-2.88
cgr-miR-497-3p	-1.32	-1.04	-2.33
cgr-miR-29b-3p	-0.70	-1.02	-1.20
cgr-miR-301b	-0.81	-1.34	-1.03
cgr-miR-325-5p	-0.48	-1.10	-2.51
cgr-let-7i	-1.23	-0.92	-1.46
cgr-miR-106b-5p	-1.28	-0.83	-2.13
cgr-miR-18b	-1.24	-0.70	-2.80
cgr-miR-193a	-1.24	0.02	-4.18
cgr-miR-19a	-1.23	-0.97	-1.33
cgr-miR-24-5p	-1.18	-0.92	-2.46
cgr-miR-32-5p	-1.58	-0.94	-3.73
cgr-miR-450a	-1.49	-0.83	-2.72

Table 5.8 log$_2$FC of differentially expressed miRNAs on days 1, 2, and 3 post-infection with VSV.

It can be observed that the most significant double filtered miRNAs in VSV infection are downregulated. This is consistent with the known pathogenesis of VSV, as it will hijack host cell mechanisms making it impossible for the cell to produce its own proteins and nucleic acids.[14]
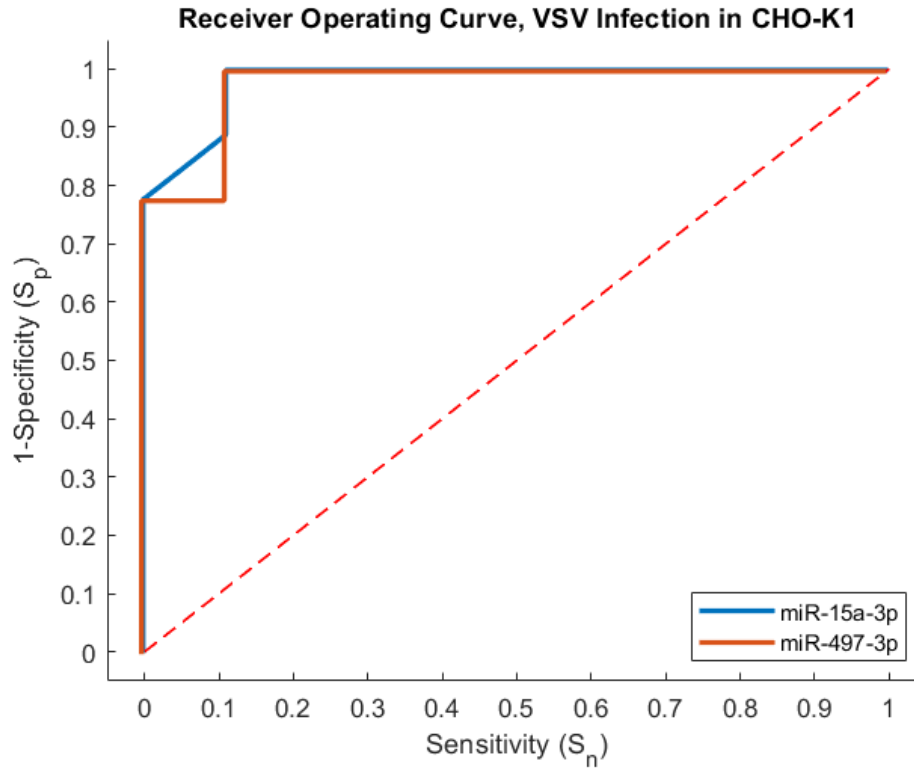
Figure 5.11 Receiver Operating Curve for miR-15a-3p and miR-197-3p from VSV infection in CHO-K1 cells. AUC values are 0.996 and 0.975 respectively.

For the 2 miRNAs which were predicted to be specific biomarkers for VSV infection in CHO-K1 cells, an ROC analysis was performed to determine their ability as classifiers for infection. Both miR-15a-3p and miR-197-3p had AUC values upwards of 0.9, which were 0.996 and 0.975 respectively, indicating their strong ability to act as biomarkers for VSV infection.

5.3.5   Analysis of Infectivity in CHO-K1 Cultures Inoculated with PI2

PI2-inoculated CHO-K1 cells did not show any CPE or cell death compared to a healthy control culture. Therefore, in the absence of further experimentation, it was impossible to tell whether PI2 is capable of infecting CHO-K1 as a silent infection or not. Here, we analyzed the miRNA transcriptome for days 3 and 7 post-

inoculation with PI2 to compare changes with the identified general and specific viral biomarkers in other infections. Figure 5.12 shows a volcano plot of the Affymetrix and NGS results for miRNA transcriptome changes in CHO-K1 on days 3 and 7 post-inoculation.
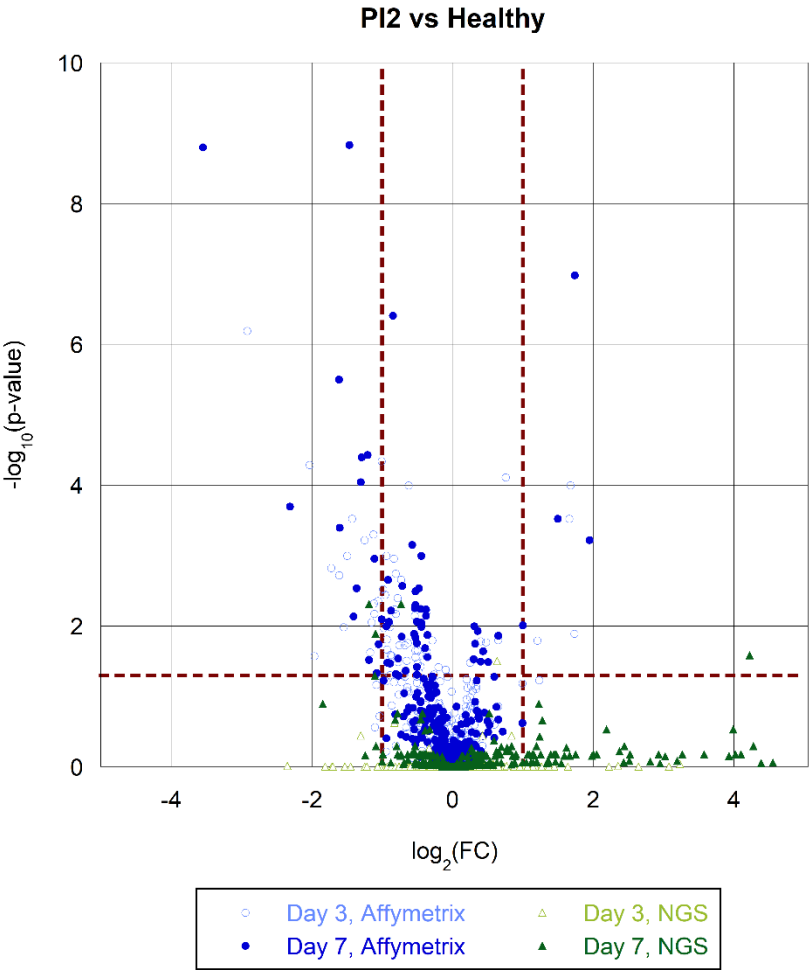
Figure 5.12 Volcano plot of PI2 inoculated CHO-K1 cultures analyzed with Affymetrix and NGS data for days 3 and 7 (late infection).

The $\log_2 FC$ values for differentially expressed miRNAs found with NGS and Affymetrix on days 3 and 7 post-infection are tabulated in Table 5.9. Out of these, miR-21-3p and miR-542-3p had been previously identified as general miRNA

biomarkers for multiple viral infections in CHO-K1 cells. Additionally, miR-210-3p, miR-210-5p, miR-222-5p, miR-222-3p, and miR-1973 were previously observed as miRNA biomarkers specific viruses (Reo3 and MMV).

A)

miRNA	Day 3, log$_2$FC	Day 7, log$_2$FC
cgr-miR-210-3p	-1.01	-1.61
cgr-miR-210-5p	-2.92	-3.55
cgr-miR-222-5p	-1.15	-1.60
cgr-miR-1306-5p	-1.61	-1.05
cgr-miR-21-3p	-0.01	-1.21
cgr-miR-23a-5p	-0.11	-1.30
cgr-miR-27a-5p	-0.73	-1.29
cgr-miR-125a-3p	-0.83	-1.01
cgr-miR-181c-3p	-0.94	-1.11
cgr-miR-200c	-0.19	1.00
cgr-miR-221-5p	-0.62	-1.81
cgr-miR-222-3p	0.12	-1.46
cgr-miR-409-3p	-0.77	-2.31
cgr-miR-409-5p	-0.89	-1.36
cgr-miR-505-5p	-0.06	-1.41
cgr-miR-542-3p	0.04	-1.19
cgr-miR-542-5p	-0.18	-1.08
cgr-miR-1949	-0.03	1.95
cgr-miR-1956	-0.08	1.50
cgr-miR-1973	0.76	1.74

B)

miRNA	Day 3, log$_2$FC	Day 7, log$_2$FC
cgr-miR-1973	3.24	4.23
cgr-miR-210-3p	-0.82	-1.09
cgr-miR-222-3p	0.10	-1.19

Table 5.9 log$_2$FC of differentially expressed miRNAs on days 3 and 7 post-inoculation with PI2 for A) Affymetrix and B) NGS analysis methods.

As with results for previous viral infections, NGS results showed markedly fewer double filtered targets than Affymetrix results. However, all of the differentially expressed miRNAs identified with NGS are consistent with those of Affymetrix. Ultimately, these results taken with the fact that PI3 is known to silently infect CHO-K1 cells, indicates that PI2 may also produce a silent infection in this cell line. However, in the absence of further validation, this is speculation.

5.4 Conclusions

In this work, we successfully identified multiple host cell miRNA biomarkers specific to each viral infection (Reo3, MMV, EMC, and VSV) in CHO-K11. Ultimately, these biomarkers may be used to support a panel of general biomarkers for viral infection in CHO-K1 production cultures in order to obtain more information about a specific infecting virus. Two miRNAs differentially expressed in Reo3 infection, miR-210-3p and miR-210-5p, were shown to have perfect discrimination ability between infected and uninfected Reo3 cultures for later-stage infection. Overall, across all time points, miR-1285 had the best discrimination ability and the least likelihood of giving false positives for Reo3 infection. Specifically for MMV infection, miR-210-3p and miR-301a-3p were identified as differentially expressed biomarkers with excellent discrimination capability. A couple miRNAs including miR-210-3p and miR-210-5p miR-222-5p, miR-222-3p, and miR-1973 appeared as significant biomarkers for multiple viral infections. These miRNAs had not been identified previously as general biomarkers from the early NGS data. Therefore, it is possible that they are general biomarkers of later stage infection, or respond to the pathogenesis of more than one but not all of the tested viruses. One specific biomarker was identified for EMC infection, miR-652-5p, as it showed the best discrimination ability and was the least likely to give false positives. Two miRNAs, miR-15a-3p and miR-497-3p, were identified as specific miRNA biomarkers for VSV infection. The miRNA transcriptome of CHO-K1 cells inoculated with PI2 has been shown to change in ways which are consistent

with viral infection, however, this must be validated with future work to come to a final conclusion.

5.5   References

1.   Leung, A. K. L. & Sharp, P. A. MicroRNA Functions in Stress Responses. *Mol Cell* **40**, 205–215 (2010).

2.   Lu, L. F. & Liston, A. MicroRNA in the immune system, microRNA as an immune system. *Immunology* **127**, 291–298 (2009).

3.   Hariharan, M., Scaria, V., Brahmachari, S. K., Pillai, B. & Maiti, S. Host-virus genome interactions: macro roles for microRNAs. *Cell Microbiol* **9**, 2784–2794 (2007).

4.   Skalsky, R. L. & Cullen, B. R. Viruses, microRNAs, and Host Interactions. *Annu Rev Microbiol* **64**, 123–141 (2010).

5.   Marquez, R. T. *et al.* Correlation between microRNA expression levels and clinical parameters associated with chronic hepatitis C viral infection in humans. *Laboratory Investigation* **90**, 1727–1736 (2010).

6.   *mirVana ᵀᴹ miRNA Isolation Kit*. (2011).

7.   ThermoFisher Scientific. Transcriptome Analysis Console (TAC) Software Data Sheet. (2017).

8.   Andrews, S. FastQC: a quality control tool for high througgput sequence data. Preprint at (2010).

9.   Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048 (2016).

10. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

11. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357–359 (2012).

12. Hammond, S., Kaplarevic, M., Borth, N., Betenbaugh, M. J. & Lee, K. H. Chinese hamster genome database: An online resource for the CHO community at. *Biotechnology and Bioengineering* vol. 109 1353–1356 Preprint at https://doi.org/10.1002/bit.24374 (2012).

13. Friedländer, M. R., MacKowiak, S. D., Li, N., Chen, W. & Rajewsky, N. MiRDeep2 accurately identifies known and hundreds of novel microRNA genes in seven animal clades. *Nucleic Acids Res* **40**, 37–52 (2012).

14. Wagner, R. R. *The Rhabdoviruses*. (Springer Science, 1987).

Chapter 6 – Identification of miRNA Biomarkers for HSV Infection in Ocular Fluid

6.1 Introduction

Due to the continual exposure to the environment, the ocular surface is susceptible to many types of infections ranging from viral, bacterial, and fungal infections. Each poses the same threat – loss of visual acuity – but each has a different treatment regimen. A complicating factor towards treating infectious ocular surface disease, is the lag time between sampling and definitive diagnosis. During this time, an individual may not be aware of their infection and will be more likely to engage in behaviors that lead to further transmission of a contagious pathogen. Additionally, samples are generally obtained by swabbing, which can be uncomfortable and improperly administered leading to false negative results. Here, we identify a new application for miRNA biomarker technology which may use ocular fluids for sampling, which will provide a less invasive strategy for diagnoses as compared to current standards of care.

Corneal herpes simplex virus type 1 (HSV-1) infection is a leading cause of blindness due to infection in the United States. HSV is a double-stranded DNA virus with a linear genome. This genome is enclosed in a double layered capsid of icosahedral shape surrounded by a lipid bilayer envelope including more than 10 different glycoproteins.[1] These glycoproteins bind to the host cell surface within a fusion process that releases the nucleocapsid into the cytoplasm of the cell. Nucleocapsids travel through the cell to the host nucleus where they insert their

DNA through nuclear pores which is then transcribed by cellular polymerase II. One of the key features of all herpesviruses is their ability to establish latency in their host. In the nucleus of a host cell neuron, the viral DNA may become circularized, which represses transcription and replication.[1] Despite the sub-clinical nature of primary HSV-1 infection, HSV-1 gains access to corneal sensory nerves and enters latency in the trigeminal ganglion (TG), where it will remain indefinitely. During times of stress, immunosuppression, or exposure to ultra violet (UV) light, HSV-1 can exit latency, produce live virions, and travel back to the ocular surface. Deposition of live virus in the cornea results in an inflammatory immune response that can disrupt the optically clear cornea resulting in loss of visual acuity. Repeated reactivation events can lead to permanent corneal damage and necessitate a corneal transplant to resolve disease. While corneal transplants are relatively well-tolerated, prior HSV-1 infection dramatically increases the odds that corneal grafts will eventually be rejected. Early intervention after viral reactivation with anti-virals followed by corticosteroids limits excessive inflammatory responses and preserves corneal integrity. Therefore, a diagnostic that can reveal viral reactivation from latency prior to or immediately after symptom onset would prevent or delay the need for future corneal transplants.

This type of disease course makes it an ideal target in developing technologies aimed at identifying microRNA (miRNA) signatures for disease diagnoses. Specifically, samples can be taken longitudinally before a reactivation event, during a reactivation event, and after a reactivation event, so that the kinetics of miRNA expression during a viral infection can be identified. MicroRNAs are short,

noncoding, single-stranded segments of 18-22 nucleotides in length that regulate up to 60% of protein encoding genes.[2] During viral infection, miRNAs are key mediators of the host response by regulating proteins involved in innate and adaptive immune pathways.[3] Host cell miRNAs have been shown to display differential expression before clinical signs of viral infection are evident.[4] Additionally, circulating host miRNAs are known to show changes in differential expression in as little as 1 hour after a precipitating event.[5] A main challenge when sampling ocular fluid is the low volume of tear film and resulting low copy numbers of viral RNA present, which has often led to low sensitivities when using PCR to assay viral RNA directly. However, miRNAs are a much more abundant biomarker that is present in all mammalian cells.

Most infection studies compare the extracellular expression of miRNA (ex-miRNA) profiles of infected patients with healthy controls. Here, we reveal a plan to leverage the early host response to microbial infection to more efficiently diagnose ocular surface infection. Specifically, we lay the groundwork to pursue a new concept in diagnostics which uses host cell miRNA as biomarkers for detection of infection. In this work, we describe this technology using herpes simplex virus type 1 (HSV-1) infection of human corneal epithelial cells (HCLEs). In previous research by Munson et al., a particular viral miRNA (miR-92944) showed differential expression which plateaued in as little as 8 hours during the lytic phase of HSV-1 infection.[6] However, in order to overcome rapid detection challenges associated with low concentrations of viral particles, this work will focus on the host cell miRNA response in RNA samples isolated from HCLE cells infected with

two different strains (KOS and RE) of herpes simplex virus (HSV). We seek to identify a minimum of 3 miRNAs which both show differential expression and act as good classifiers of infection from a receiver operating curve analysis to lay the groundwork for implementation of this technology.

6.2   Materials and Methods

6.2.1   Experimental Design

Multiplicity of infection (MOI) is quantified by the ratio of viral particles to susceptible cells.[7] Three different viral inoculation conditions were tested in this work including HSV-KOS at 1 MOI, HSV-RE at 1 MOI, and HSV-RE at 0.1 MOI against an uninfected control. Different miRNAs can peak at diverse time points over a 24-hour period,[5] showing the need to analyze multiple time points for a complete picture of the miRNA transcriptome changes. Each of these conditions was analyzed at three different time points including 2 hours, 8 hours, and 24 hours post-infection. Triplicate biological samples were created to support statistical analysis. Targets analyzed included 827 different human miRNAs in addition to 5 non-mammalian spike-in probes and 5 internal reference controls.

6.2.2   Cell Culture

Human corneal limbal epithelial cells (HCLE) (ATCC, Manassas, VA), were routinely cultured in keratinocyte serum free media (K-SFM, ref 10724-011, Gibco) and passaged twice a week. Cells were seeded on six-well plates at a density of 1.2 x 10^6 cells per well. At confluent density (>70%), wells were infected with HSV-KOS and HSV-1 RE viruses. Nine wells were infected with HSV-KOS to a final concentration of 1.0 MOI. Another eighteen wells were infected with HSV-1

RE, nine with a final MOI concentration of 1.0 and nine with a final MOI concentration of 0.1. Nine control wells were not infected with either virus. The cells from three wells of each condition were lysed at each time point of two, eight, and twenty-four hours after infection.

6.2.3 RNA Extraction

RNA was extracted using a RNAzol®RT (cat number 190) kit from Molecular Research Center, Inc. (Cincinnati, OH) according to instructions from the manufacturer.[8] At each time point post-infection, the media was removed from three wells of each condition and 0.5 ml of RNAsol RT was added to each well. The solution was passed through the pipette several times to ensure lysis. Next, 0.2 ml of nuclease-free water was added for DNA/protein precipitation and samples were centrifuged for 15 minutes at 12,000 g. Large RNAs were precipitated from the supernatant with 0.2 ml of 75% ethanol. Small RNAs were precipitated from the post-large RNA precipitation supernatant with 0.8 volumes of isopropanol. The final product was washed twice with 0.2 ml of 70% isopronanol and resuspended in nuclease-free water. Samples were analyzed using a Nanodrop instrument to ensure that minimum concentration and purity requirements were met for NanoString processing.

6.2.4 Nanostring Processing and Data Analysis

RNA samples were shipped to NanoString (Seattle, WA) for processing with an nCounter® instrument.[9] NanoString technology measures nucleic acid content directly without amplification. The Human v3 miRNA CodeSet Kit was chosen with over 800 miRNA probes in addition to controls. The resulting .RCC files were

imported into nSolver 4.0 software for analysis.[10] Samples passed all quality control parameters. In order to normalize raw counts, the geometric mean of the negative controls was used as a background subtraction. Samples were further normalized using the geometric mean of the top expressed miRNAs with an average count over 100. A 2-tailed t-test was utilized to calculate the t-statistic and p-value which was then adjusted for false discovery rate (FDR). Fold Change ratios were calculated by comparing infected samples vs. healthy samples from the same time point. Fold changes (FC) and FDR p-values were transformed to $\log_2(FC)$ and $-\log_{10}(p\ value)$ respectively for visualization using a volcano plot.[11] Volcano plots were generated using Kaleidagraph software.[12] Dot plots were generated using Tableau software.

6.2.5 Receiver Operating Curve Analysis

The normalized count files generated using nSolver software were used to create a receiver operating curve (ROC) for binary classification to quantify each miRNA's ability to discriminate between infected and uninfected cultures. In this binary system, virally infected samples were classified as positive, and uninfected samples were classified as negative. True positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) were calculated over the range of normalized counts and sensitivity (S_n) and specificity (S_p) were derived from this classification system at each point using Eq. 4.1 and Eq. 4.2.

$$S_n = \frac{TP}{TP + FN}$$

Eq. 6.1

$$S_P = \frac{TN}{TN + FP}$$

Eq. 6.2

Sensitivity (S_n) was plotted against ($1 - S_p$) and an area under the curve (AUC) was calculated for each miRNA. A classifier with an AUC of 0.5 shows no ability to discriminate, therefore, our minimum AUC cutoff point for miRNA biomarkers was chosen to be 0.75. All calculations were performed and ROC plots were generated using MATLAB software.[13]

6.3   Results and Discussion

6.3.1   Differentially Expressed miRNAs

Fold change (FC) and p-values were calculated by comparing each viral condition to the uninfected control at the same time point. Differentially expressed miRNAs were determined using the double filtering parameters of $p\ value \leq 0.05$ and $|FC| \geq 2$.[11] Figure 6.1 shows the resulting volcano plot with differentially expressed genes marked in the upper right and upper left sectors. Out of the 827 miRNAs screened, 27 showed differential expression for at least 2 viral inoculation conditions and incubation times. These miRNAs along with their fold changes are listed in Table 6.1.
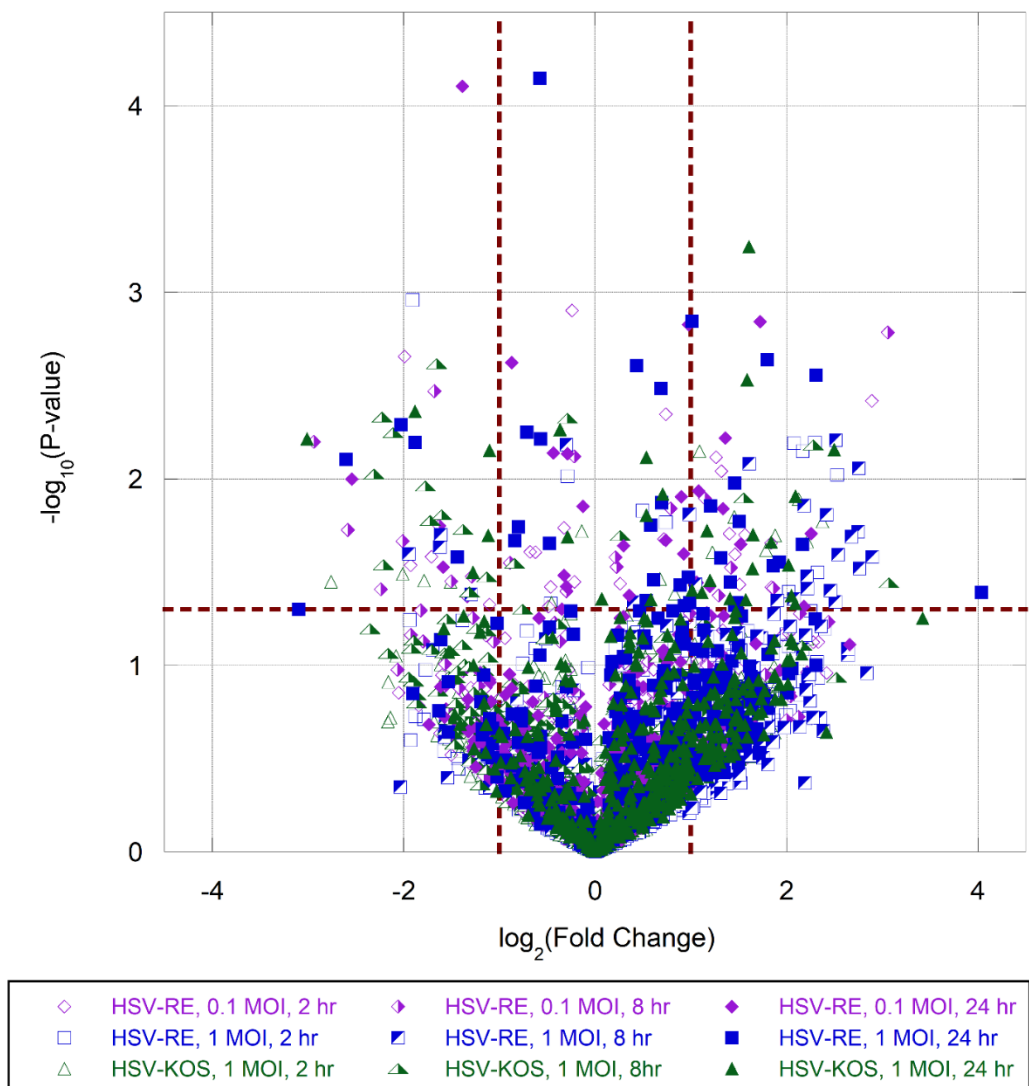
Figure 6.1 Volcano plot showing $log_2 FC$ and $-log_{10} Pvalue$ for the 9 different viral conditions compared with a healthy control from the same time point. Differentially expressed miRNAs, double filtered with a $FC \leq -2$ or $FC \geq 2$ and a $P\ value \leq .05$, are present in the top right and top left sections of the plot.

miRNA	2 hours post-infection			8 hours post-infection			24 hours post-infection		
	KOS 1 MOI	RE 0.1 MOI	RE 1 MOI	KOS 1 MOI	RE 0.1 MOI	RE 1 MOI	KOS 1 MOI	RE 0.1 MOI	RE 1 MOI
hsa-miR-378f	1.03	7.43	2.01	2.65	2.31	2.18	3.68	2.57	2.83
hsa-miR-612	-2	-3.8	-3.67	-1.07	-3.07	2.6	1.28	2.87	2.06
hsa-miR-767-5p	1.03	2.11	2.24	-1.51	1.71	2.15	5.65	3.31	3.23
hsa-miR-1286	1.16	1.25	1.83	-3.65	2.22	3.06	1.78	2.06	2.69
hsa-miR-615-3p	3.69	1.74	2.08	1.75	-1.03	1.81	3.58	1.43	4.96
hsa-miR-6724-5p	-2.85	1.22	1.45	1.95	1.55	3.08	-2.59	-1.72	-6.07
hsa-miR-381-3p	2.04	1.2	-1.72	-4.63	-3.53	-3.85	1.72	-1.08	1.31
hsa-miR-150-5p	1.05	1.64	1	-3.42	-2.83	-3.08	1.59	1.61	2.06
hsa-miR-138-5p	-2.47	1.23	-2	-4.67	-7.65	-2.91	1.55	1.58	1.67
hsa-miR-4284	-1.6	-1.44	-1.19	8.4	2.27	7.46	1.09	1.28	2.07
hsa-miR-412-3p	-1.5	-1.02	2.63	2.94	1.02	2.82	1.85	2.3	2
hsa-miR-548m	-1.1	1.1	-1.08	1.06	2.4	2.66	3.13	2.25	1.76
hsa-miR-1268a	1.12	-1.42	-1.32	2.13	3.64	2.53	1.7	1.6	2.19
hsa-miR-199a-5p	1.01	1.52	1.61	-2.19	-2.85	1.36	2.05	1.3	3.64
hsa-miR-125a-5p	-1.09	-1.06	-1.11	-1.07	1.13	-1.09	-3.68	-2.61	-4.08
hsa-miR-302d-3p	-1.13	-2.15	-1.35	-1.05	1.07	1.38	2.3	1.02	2.49
hsa-miR-503-5p	-1.02	-1.3	-1.11	-1.17	-1.08	-1.34	-8.04	-5.82	-8.56
hsa-miR-345-3p	4.65	3.59	4.77	-1.22	-1.36	1.65	1.41	-1.43	3.4
hsa-miR-495-3p	1.36	1.53	1.05	-1.69	1.35	2.01	2.17	1.42	2.75
hsa-miR-563	1.73	1.18	2.87	1.12	4.37	5.36	-1.22	-3	-1.35
hsa-miR-525-3p	-1.17	-1.31	2.05	-3.04	-3.09	-1.76	1.28	-1.08	1.54
hsa-miR-937-3p	3.45	2.76	1.62	-1.33	1.14	1.64	-2	-1.89	-1.26
hsa-miR-345-5p	-6.77	-2.6	-2.1	-1.26	-4.71	-2.15	-1.33	-1.09	-1.19
hsa-miR-944	1.94	1.14	1.91	-1.01	1.55	5.72	3.01	-1.08	1.31
hsa-miR-640	1.83	1.4	-1.08	-3.15	-3.2	1.33	1.25	-1.31	1
hsa-miR-6503-5p	1.73	2.85	1.62	1.12	2.68	1.3	-1.04	-1.26	1.08
hsa-miR-532-5p	-1.17	1.07	1.48	-2.48	-2.52	-1.02	1.36	-1.08	1.31

Table 6.1 Fold changes of miRNAs which showed differential expression for more than one viral condition. Significantly upregulated miRNAs are highlighted in green. Significantly downregulated miRNAs are highlighted in red.

Differential expression data appeared to be grouped according to time point post-infection. The different strains and titers of HSV-1 acted similarly in their transcriptome changes. When analyzed by time point, differentially expressed miRNAs from this data fell into three different categories: miRNAs with fold change which moved in a linear direction as time progressed, miRNAs which peaked at 2 or 8 hours and then became insignificant, and miRNAs which showed significant up and down regulation at different time points. The behavior of miRNAs in which fold change moved in a linear direction with differential expression at the 24-hour post-infection time point can be seen in Figure 6.2. Out of these 10 miRNAs, 8 show upregulation and 2 show downregulation. Due to their

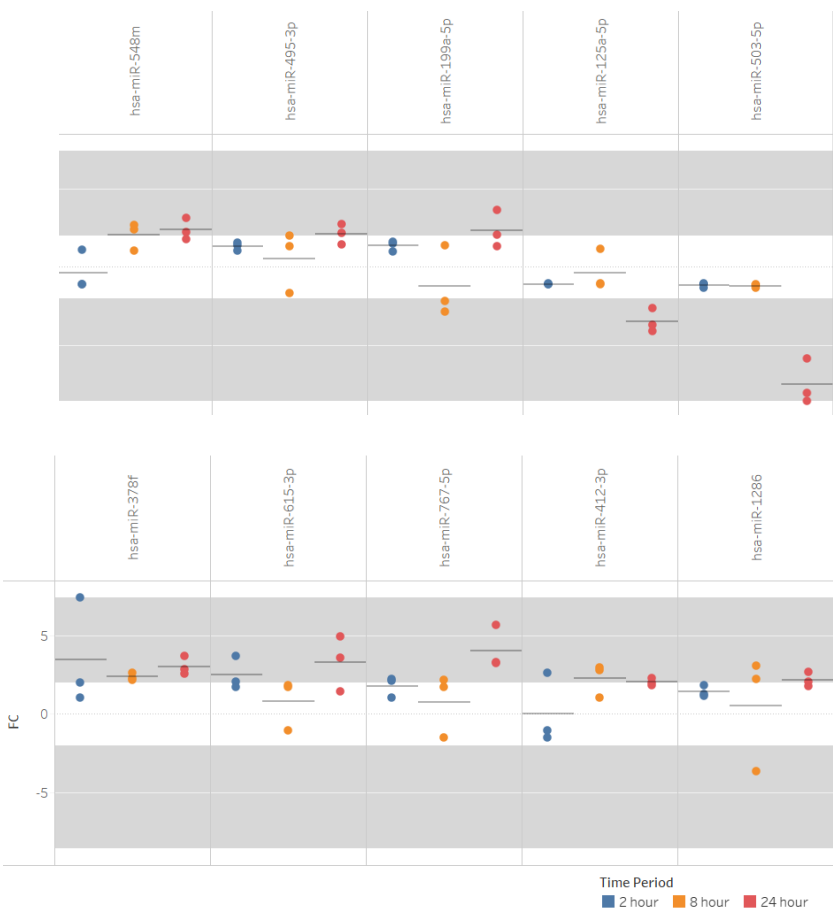linear movement, miRNAs in this category would be the most likely biomarker candidates for clinical use.

Figure 6.2 Double filtered miRNAs for which fold change moved in a linear direction with differential expression at the 24-hour post-infection time point.

MicroRNAs which peaked in their expression at either 2- or 8-hours post-infection are shown in Figure 6.3. Due to the nature of diagnostics in which the time of infection is unknown, these miRNAs are not likely to be useful as biomarkers for clinical use.
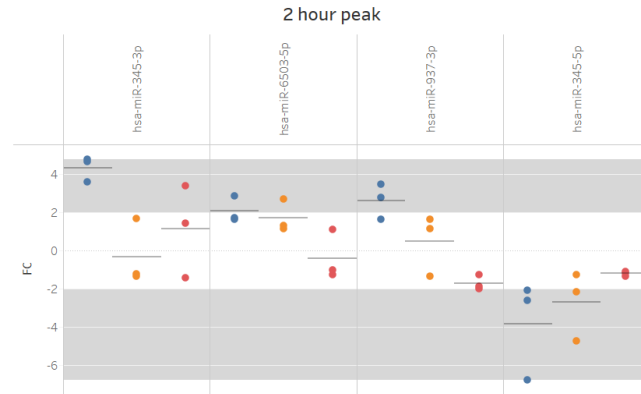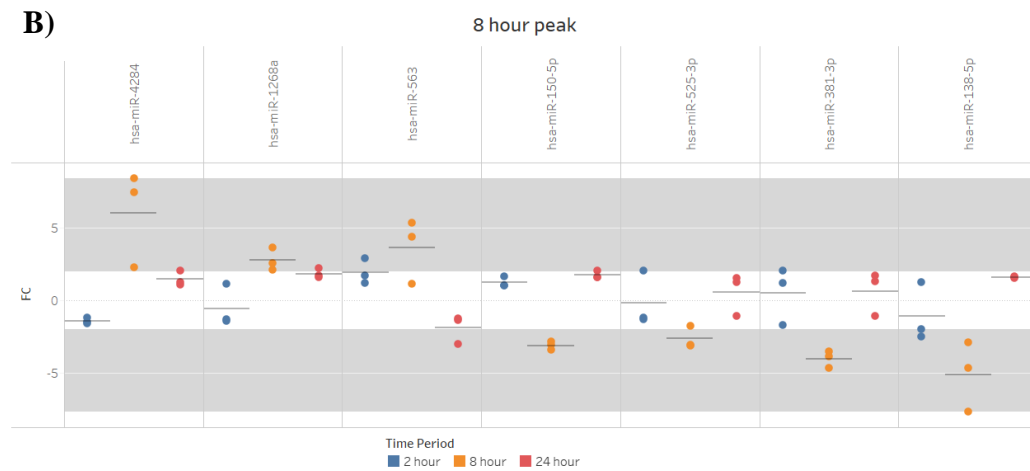
Figure 6.3 Double filtered miRNAs that showed a peak for either A) 2 hours or B) 8 hours post-infection and were insignificant for the 24-hour post-infection time point.

Two miRNAs showed extreme variation in their differential expression, being both significantly upregulated and downregulated at different time points. These are shown in Figure 6.4.
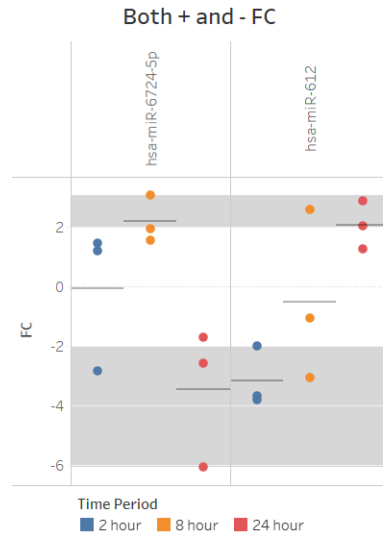
Figure 6.4 miRNAs which showed both upregulated and downregulated fold change at different time points.

6.3.2 Receiver Operating Curve Analysis

In addition to analyzing differential expression through the double filtering method of fold change and p-value, each miRNA was also analyzed for its capability as a classifier of infection through a receiver operating curve analysis. Area under the curve values for the top 27 double filtered targets are listed in Table 6.2.

miRNA	AUC
hsa-miR-378f	0.9218
hsa-miR-615-3p	0.7881
hsa-miR-345-5p	0.7613
hsa-miR-548m	0.7407
hsa-miR-944	0.7284
hsa-miR-6503-5p	0.7202
hsa-miR-1286	0.7181
hsa-miR-1268a	0.714
hsa-miR-345-3p	0.714
hsa-miR-767-5p	0.7099
hsa-miR-495-3p	0.7037
hsa-miR-412-3p	0.6996
hsa-miR-4284	0.6852
hsa-miR-138-5p	0.6626
hsa-miR-563	0.642
hsa-miR-503-5p	0.6132
hsa-miR-125a-5p	0.6008
hsa-miR-199a-5p	0.5885
hsa-miR-937-3p	0.5885
hsa-miR-381-3p	0.5782
hsa-miR-640	0.5679
hsa-miR-6724-5p	0.5556
hsa-miR-302d-3p	0.5391
hsa-miR-532-5p	0.5391
hsa-miR-525-3p	0.537
hsa-miR-150-5p	0.5288
hsa-miR-612	0.5226

Table 6.2 AUC values for top 27 double filtered miRNAs

Of these, 6 miRNAs with an AUC of greater than 0.07 which showed a linear progression of fold change (miR-378f, miR-615-3p, miR-548m, miR-495-3p, miR-767-5p, and miR-1286) are illustrated in Figure 6.5. miR-378f and miR-615-3p showed the best discrimination ability with the least chance of false positives. miR-548m and miR-767-5p may be avoided as miRNA biomarkers due to their higher change of having false positives, which can be seen in the ROC plot where they cross the dashed line. Additionally, miR-1286 is likely to give false negative results.

The last miRNA, miR-495-3p, shows decent, but not excellent ability to discriminate with an AUC of 0.70.
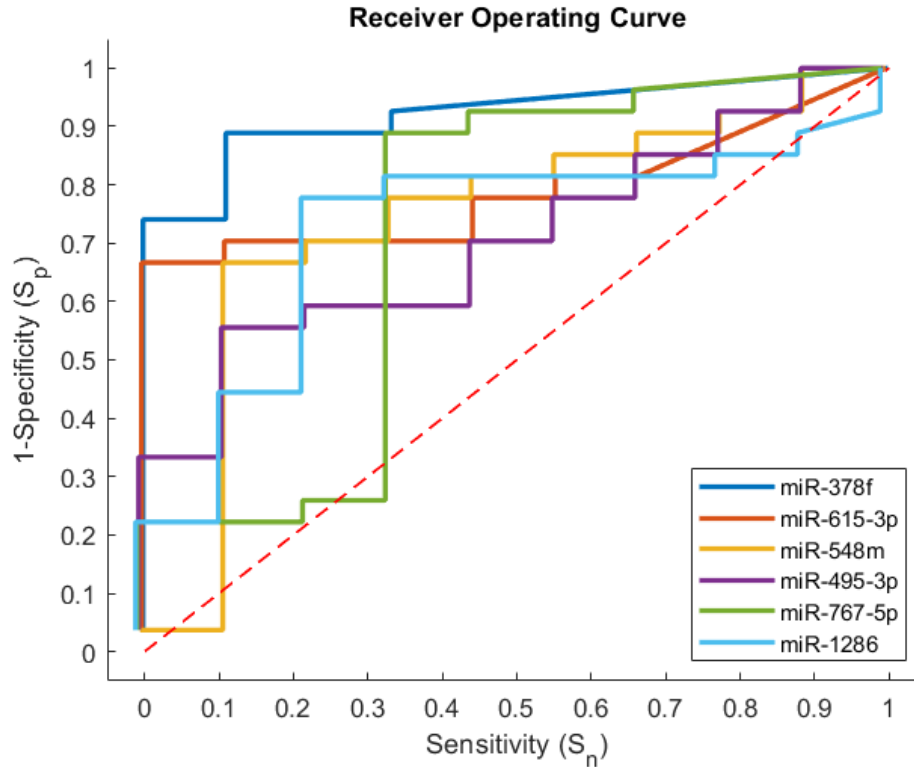

Figure 6.5 Receiver operating curves for top miRNA biomarker targets which have an AUC greater than 0.70. The red dashed line is an AUC=0.5 with no discrimination capability.

6.3.3 Biological Role of Target miRNAs

To analyze potential pathways of differentially expressed miRNAs, a web-based tool named MicroRNA ENrichment TURned NETwork (MIENTURNET) was utilized.[14] MIENTURNET compiles data from miRTarBase (experimentally validated miRNA-target interactions)[15] and TargetScan (computationally predicted miRNA targets)[16,17] to determine which genes were targeted by miRNAs for a miRNA-target enrichment analysis. From the 10 miRNAs for which fold change moved in a linear direction as time progressed (miR-378f, miR-615-3p, miR767-

5p, miR-412-3p, miR-1286, miR-548m, miR-495-3p, miR-199a-5p, miR-125a-5p, and miR-503-5p), 24 experimentally validated target genes were found by miRTarBase which showed interaction with a minimum of 3 miRNAs.

Target genes of HSV biomarker miRNAs					
MCL1	CDH2	COL4A1	EIF5A2	CSNK2A1	RCC2
VEGFA	RRP1B	NIP7	LONRF2	PTBP1	LCOR
HNRNPA1	CLTC	AKT1	TPI1	TNPO2	AGO2
DLST	CERK	ATXN7L3	CDCA4	E2F3	CDKN1A

Table 6.3 Experimentally validated target genes of differentially expressed miRNAs which showed linear fold change over time. Gene targets were generated from miRTarBase software. [15]

Listed in Table 6.3, these 24 target genes were analyzed for their roles in human cellular pathways using the reactome pathway knowledgebase at reactome.org.[18] The most significant pathway, with 6 (AKT1, HNRNPA1, CLTC, PTBP1, COL4A1, and VEGFA) of the 24 input target genes was found to be signaling by receptor tyrosine kinases (RTKs). Involved in signal transduction, RTKs are a class of surface cell proteins which may alter gene expression or cellular metabolism. The Interleukin-4 (IL-4) and Interleukin-13 (IL-13) signaling pathway was also found to be significant, including 4 (AKT1, MCL1, CDKN1A, and VEGFA) input target genes. These two cytokines (IL-4 and IL-13) are known to be highly involved in host immune response.

6.4 Conclusions

In conclusion, we identified 3 miRNAs as top biomarkers for HSV-1 infection in HCLE cells. Overall, the top miRNA biomarker for an early host response to HSV-1 infection in HCLE cells was found to be miR-378f. This miRNA showed differential expression at all 3 early time points post-infection with both strains of HSV-1 (RE and KOS), and was found to display excellent discrimination ability

between infected and uninfected cultures using an ROC analysis. Additionally, both miR-615-3p and miR-495-3p showed a linear increase in their fold change over time and were good classifiers without a proclivity towards false positives or false negatives. These biomarkers lay the groundwork to pursue a new diagnostic which will utilize host miRNA as biomarkers for detection of infection. Future work will include further validation to incorporate these biomarkers into a more efficient test for the diagnosis of ocular surface infection.

6.5 References

1. Acheson, N. H. *Fundamentals of molecular virology*. (John Wiley & Sons, 2011).

2. Friedman, R. C., Farh, K. K. H., Burge, C. B. & Bartel, D. P. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res* **19**, 92–105 (2009).

3. Skalsky, R. L. & Cullen, B. R. Viruses, microRNAs, and Host Interactions. *Annu Rev Microbiol* **64**, 123–141 (2010).

4. Basagoudanavar, S. H. *et al.* Host serum microRNA profiling during the early stage of foot-and-mouth disease virus infection. *Arch Virol* **163**, 2055–2063 (2018).

5. Schulte, C. *et al.* Comparative analysis of circulating noncoding rnas versus protein biomarkers in the detection of myocardial injury. *Circ Res* **125**, 328–340 (2019).

6. Munson, D. J. & Burch, A. D. A novel miRNA produced during lytic HSV-1 infection is important for efficient replication in tissue culture. *Arch Virol* **157**, 1677–1688 (2012).

7. Mistry, B. A., D'Orsogna, M. R. & Chou, T. The Effects of Statistical Multiplicity of Infection on Virus Quantification and Infectivity Assays. *Biophys J* **114**, 2974–2985 (2018).

8. Sigma-Aldrich. RNAzol RT Product Information. Preprint at https://www.sigmaaldrich.com/deepweb/assets/sigmaaldrich/product/documents/381/652/r4533bul.pdf (2012).

9. *nCounter® Expression CodeSet Design Manual*. www.nanostring.com (2011).

10. *Analysis Software User Manual nSolver 4.0*. www.nanostring.com (2018).

11. Li, W. Volcano plots in analyzing differential expressions with mRNA microarrays. *J Bioinform Comput Biol* **10**, 1–24 (2012).

12. *KaleidaGraph Manual Version 4.0*. https://www.synergy.com/documentation/.

13. MathWorks, I. *MATLAB : the language of technical computing : computation, visualization, programming : installation guide for UNIX version 5*. (Natwick : Math Works Inc., 1996., 1996).

14. Licursi, V., Conte, F., Fiscon, G. & Paci, P. MIENTURNET: An interactive web tool for microRNA-target enrichment and network-based analysis. *BMC Bioinformatics* **20**, (2019).

15. Huang, H. Y. *et al.* MiRTarBase 2020: Updates to the experimentally validated microRNA-target interaction database. *Nucleic Acids Res* **48**, D148–D154 (2020).

16. Friedman, R. C., Farh, K. K. H., Burge, C. B. & Bartel, D. P. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res* **19**, 92–105 (2009).

17. McGeary, S. E. *et al.* The biochemical basis of microRNA targeting efficacy. *Science (1979)* **366**, (2019).

18. Gillespie, M. *et al.* The reactome pathway knowledgebase 2022. *Nucleic Acids Res* **50**, D687–D692 (2022).

Chapter 7 – LNA and DNA probes for the Detection of Short Nucleotides using Micelle Tagging Electrophoresis

7.1 Introduction

MicroRNAs (miRNAs) are small noncoding RNA molecules typically 18 to 22 nucleotides in length that are responsible for regulating gene expression at the post-transcriptional level. Each miRNA can bind to hundreds of mRNA targets, making this class of molecules an important area of study for understanding fundamental biological functions.[1] In the past decade, miRNAs have become an increasingly important field of study as biomarkers for a range of cancers, diseases, and other various cellular stresses.[2–6] However, while thousands of studies are published each year to analyze miRNA biomarkers, the number of clinically or commercially usable products which utilize miRNA biomarkers remains less than 10.[7] Although there are numerous contributing factors to this discrepancy, the rapid, cost-effective detection of short nucleotides remains a key challenge. Currently, the most-used method for detection of miRNAs is northern blotting, which is both time and labor intensive.[8] However, research into the use of capillary electrophoresis (CE) for detection of these small nucleic acids has recently become an active field of study. An important advantage of CE is that it allows for the potential use of samples from crude cell lysates and eliminates the extra step of having to isolate the RNA before detection.[9] Additionally, these CE methods detect miRNA directly, requiring no chemical or enzymatic modification, offering advantages over methods such as qRT-PCR or Next Generation Sequencing (NGS).[10]

One CE method called direct quantitative analysis of multiple miRNAs (DQAMmiR) uses complementary DNA probes with peptide drag tags to detect miRNA.[11] Here, a mobility shift is created by using a single-stranded DNA-binding protein (SSB) to bind unbound probes and electrophoretic ally sequester them from bound probes. Micelle-tagging electrophoresis (MTE) is a method which offers a much wider dynamic range and circumvents difficulties encountered when working with proteins, such as surface adsorption, deactivation, and the like. Recently, Goldman et al. have demonstrated how micelle-tagging electrophoresis (MTE) can be utilized with γPNA amphiphile (γPNAA) probes for detection of the let-7 miRNA family.[12] MTE utilizes nonionic surfactant micelles which bind to n-alkane groups at the end of a fluorescently labeled nucleic acid probe. This probe is then detected using rapid, gel-free DNA electrophoresis.[13,14] Using a probe sandwich structure, where a γPNAA with strong binding chemistry binds to a small part of the short nucleotide, allows for weaker DNA chemistry in the second probe.[13] Additionally, different lengths of dsDNA drag tags may then be utilized to create mobility shifts between the 6 members of the let-7 miRNA family.[12] While this method has proven to be effective, there are currently no commercial sources for customizable γPNAs. Synthesis of PNA is challenging and typically limited to sequences less than 15 bp in length in addition to being a time-consuming process.[15] Unfortunately, these supply chain issues greatly decrease the applicability of γPNAs in clinical or commercial uses for detection of miRNAs. Here, we address this challenge by analyzing the applications of both DNA and LNA short nucleic acid probes for use with MTE.

DNA probes are widely available and are relatively inexpensive compared to other options. However, they have a lower binding affinity and for thermodynamic reasons are unusable as probes in a sandwich probe design for miRNA detection, which demands tight binding over a 10 nt sequence. Because MTE multiplex detection utilizes a sandwich design in which varying lengths of the second probe are used to create mobility shifts for different targets, DNA probes would not be useful for detection of multiple targets at once. Locked Nucleic Acids, or LNAs, are more expensive than DNA nucleotides, but are also available commercially for custom sequences. LNAs are nucleic acid analogues where a methylene bridge connects the 4'-carbon with the 2'-oxygen group. Thermodynamically, they have an extremely high binding affinity to and stability with their complementary target, making them a possible substitute for γPNAA probes.[16] LNA probes have been shown to increase detection sensitivity compared to DNA probes with northern blotting experiments for miRNA detection.[17]

In this work, we will test the usefulness and limits of DNA probes and LNA probes for use with MTE. LNA probes have not yet been successfully utilized with MTE. First, we will address the complications of LNA aggregation in probe alkylation and in capillary electrophoresis. Second, we seek to successfully implement the use of LNA probes with MTE. And third, if successful, we will suggest means to improve their implementation. This will be accomplished through the optimization of LNA substitutions percentages in probes and of capillary electrophoresis parameters.

7.2   Materials and Methods

7.2.1   Reagents

Reagents for alkylation and purification, including hexadecyltrimethylammonium bromide (CTAB), (4-dimethylamino) pyridine (DMAP), triphenylphosphine (TPP), 2,2'-dipyridyl disulfide (DPDS), triethylamine (TEA), lithium perchlorate (LiClO4), dimethyl sulfoxide (DMSO), and 4Å molecular sieves were purchased from MilliporeSigma (Burlington, MA). Acetone and acetonitrile were purchased from VWR (Radnor, PA). Glacial acetic acid, UltraPure™ DNase/RNase-Free Distilled Water, RNase AWAY™ Surface Decontaminant, and BODIPY™ FL C16 (4,4-difluoro-5,7-dimethyl-4-bora-3a,4a-diaza-s-indacene-3-hexadecanoic   Acid) were purchased from Thermo Fisher Scientific (Pittsburgh, PA).

For capillary electrophoresis, Triton X-100 surfactant and tris-borate-EDTA (10X TBE) buffer were purchased from MilliporeSigma. POP-6™ Polymer was purchased from Thermo Fisher Scientific. Capillaries were purchased from Polymicro Technologies (Phoenix, AZ). Nucleic acids were purchased from Integrated DNA Technologies (Coralville, IA) using the standard desalting purification. Probe sequences were ordered with a 5' amine modification and designed to optimize hybridization thermodynamics as discussed in the probe design section. Nucleic acids arrived dried and were resuspended in nuclease-free water.

7.2.2   Probe alkylation

For alkylation, DMSO was dried overnight using 4Å molecular sieves. After drying, solutions of 0.5 M TPP, 0.5 M DPDS, and 1.0 M DMAP were made using

DMSO as the solvent. For each sample, reagents for a reaction solution were added to a 1.5 mL microcentrifuge tube in the following order: 5 mM BODIPY solution (20 µL), pure TEA (1 µL), TPP solution (5 µL), DPDS solution (5 µL), and DEA solution (5 µL). The microcentrifuge tube was covered with foil and vortexed at room temperature for 20 minutes. In a separate microcentrifuge tube, 2mM amine modified DNA probe (5 µL) was added to 0.15 M CTAB in nuclease-free water. The microcentrifuge tube was vortexed and then vacuum centrifuged for 30 minutes. Reaction solution (36 µL) was then added to the dried DNA. The microcentrifuge tube was covered in foil, vortexed, and shaken for 6 hours. After shaking, 1 mL of 2.0 w/v% LiClO4 in acetone was added and the solution was vortexed and then centrifuged for 20 minutes at 12,000 RPM. The supernatant was aspirated, the DNA pellet was rinsed twice with acetone (1 mL), and a vacuum centrifuge was used to evaporate off any remaining acetone. Finally, the DNA pellet was resuspended in 0.1M TEAA (300 µL) and purified using HPLC.

7.2.3 Probe purification

A 0.1 M triethylammonium acetate (TEAA) solution was made by adding TEA (12.6 mL) to filtered DI water (900 mL). This buffer was titrated using glacial acetic acid to a pH between 6.8 and 7.2. An Alliance HPLC from Waters Corporation (Milford, MA) or an HPLC from Shimadzu (Kyoto, Japan) was used for purification of DNA and LNA probes. The HPLC column, a Symmetry™ C4 5µm column with a silica resin, was washed using a 100% filtered HPLC grade acetonitrile and then water until pressure stabilized. Samples were separated using

a linear gradient of 0.1 M TEAA buffer to acetonitrile. After purification, DNA probes were lyophilized using a Labconco lyophilizer at 0.0 mbar and 20°C.

7.2.4 Hybridization

Lyophilized DNA probes and dried nucleic acids from IDT were resuspended in nuclease-free water. Concentrations of nucleic acids were measured using a NanoDropTM 2000 Spectrophotometer and 50 nM or 100nM nucleic solutions in 1X TBE for probes and targets were created. Hybridization of probes to targets was achieved through mixing these solutions in a 1:1 ratio unless otherwise noted. For experiments including an annealing step, solutions of probe and target were heated to 95°C for 5 minutes and then either put on ice to cool quickly (fast cool) or allowed to return to room temperature over an hour (slow cool).

7.2.5 Capillary electrophoresis

Capillary electrophoresis experiments were run using a P/ACE MDQ Beckman Coulter instrument (Brea, CA) with laser-induced fluorescence (LIF) and data was collected using 32 Karat software.[18] A fused-silica capillary with 50 µm internal diameter (ID), 20 cm length to detector, and 30 cm total length was used. Electroosmotic flow (EOF) was suppressed by rinsing the capillary with a solution of 10% POP-6 in 1X TBE initially and after every second run. A 24 mM solution of Triton X-100 in 1X TBE was utilized as the running buffer and was used to rinse out the capillary before each separation. Samples were injected hydrodynamically at 0.5 psi for 5 seconds and the applied voltage for electrophoretic separation under reverse polarity was 20 kV. Separations were performed at 22°C unless otherwise noted.

7.3 DNA Probes

7.3.1 DNA Probe Design

Mobilies of micellated dsDNA and ssDNA for short sequences (18-24 bp) were estimated from the literature[19] to determine the elution times of the probe and probe-target complex to ensure that they would be different enough to show two distinct peaks in a detection format. We concluded that a single, fully complementary DNA probe would be appropriate and such probes were designed for detection of short nucleic acid targets.
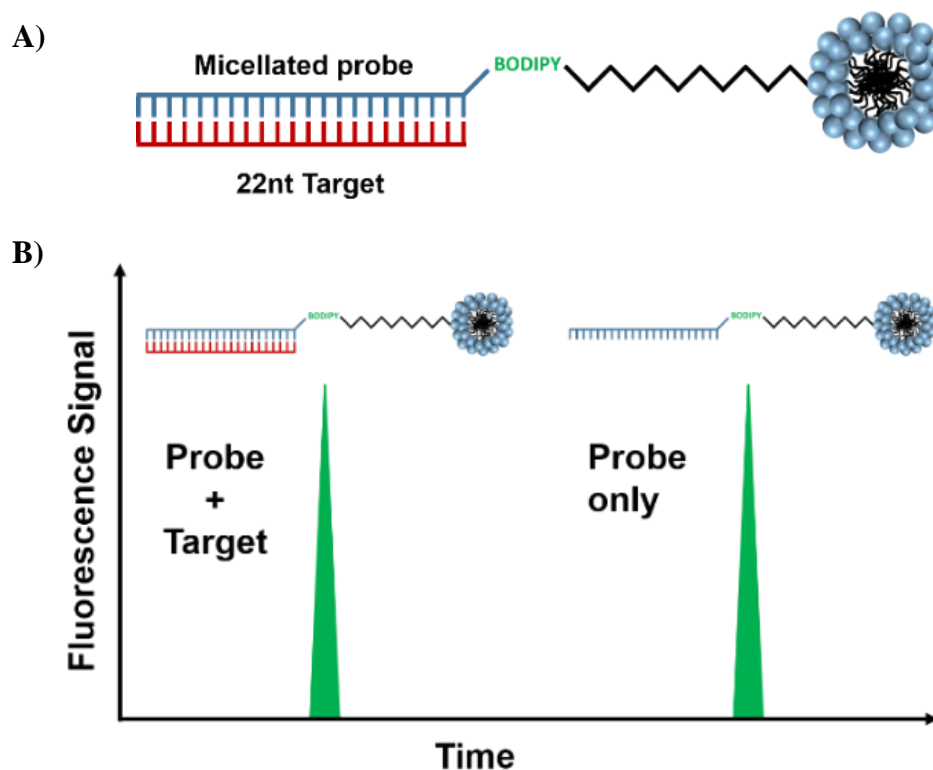

Figure 7.1 A) Diagram of micellated DNA probe hybridized to a short nucleic acid target with Bodipy FL C16 fluorophore. B) Electropherogram detection schematic for single DNA probe hybridized to short nucleic acid target, similar to previous work by the Schneider group.[14]

Figure 7.1 shows a diagram of the fully complementary DNA probe binding to its short target and a schematic for separation and detection of this probe-target

128

complex, similar to previous work from the Schneider group.[14] Bodipy FL C16 with an emission wavelength between 505 and 512 nm was chosen as the fluorophore for detection.

Due their potential as biomarkers for viral infection in CHO-K1 cells, sequences for cgr-miR-21-3p and cgr-miR-1260 were selected as targets. Matching DNA and RNA sequences were both studied as targets to compare detection of different types of oligonucleotides. Probe and target sequences along with their calculated melting temperatures (T_m) are shown below in Table 7.1. A second probe for cgr-miR-21-3p was designed to be 4 bp shorter, with an overhang of the target sequence, to reduce self-complementarity of the probe sequence while maintaining a high enough T_m for binding.

	cgr-miR-1260	cgr-miR-21-3p (probe #1)	cgr-miR-21-3p (probe #2)
Probe (5'-3')	TGGTGGCAGCGGTGGGAT	GACAGCCCATCGACTGCTGTTG	GACAGCCCATCGACTGCT
DNA Target (3'-5')	ACCACCGTCGCCACCCTA	CTGTCGGGTAGCTGACGACAAC	CTGTCGGGTAGCTGACGACAAC
RNA Target (3'-5')	ACCACCGUCGCCACCCUA	CUGUCGGGUAGCUGACGACAAC	CUGUCGGGUAGCUGACGACAAC
DNA target T_m (°C)	61.3	59.4	56.1
RNA target T_m (°C)	59.2	73.1	70.4

Table 7.1 Micellated DNA probe and target sequences with associated melting temperatures estimated using the Integrated DNA Technologies OligoAnalyzer tool.

7.3.2 Predicted Mobilities

Estimation of elution times was an important step to ensure that separation between the probe and probe-target complex was possible. These calculations utilized the following equations and parameters. For ssDNA, the effective friction coefficient

for Triton X-100 micelles (α) was calculated using Eq. 7.1 from the mobilities of free solution and micellated DNA (μ^0_{aDNA} and μ^0_{mDNA} respectively) and the length of the oligonucleotide (L).

$$\alpha = L \left(\frac{\mu^0_{aDNA}}{\mu^0_{mDNA}} \right) - L \qquad \textit{Eq. 7.1}$$

The radius of gyration (R_g) of probe-target complex was calculated using the Kratky-Porod equation (Eq. 7.2) and then used to calculate the hydrodynamic radius (R_H) with the Kirkwood-Riseman approximation (Eq. 7.3).[20,21] Monomer length (b) and persistence length (p) were estimated from the literature with M being the number of monomers in the sequence.[19]

$$R_g^2 \cong \frac{bMp}{3} \left[1 - 3\left(\frac{p}{bM}\right) + 6\left(\frac{p}{bM}\right)^2 - 6\left(\frac{p}{bM}\right)^3 \left(1 - e^{-\frac{b\alpha}{p}}\right) \right] \qquad \textit{Eq. 7.2}$$

$$R_H \cong \frac{2}{3} R_g \qquad \textit{Eq. 7.3}$$

With the approximation that the hydrodynamic radius of dsDNA and ssDNA are similar (Eq. 7.4), α may then be calculated for dsDNA using a rearranged form of the Kratky-Porod equation (Eq. 7.5).

$$R_{H,ssDNA} \approx R_{H,dsDNA} \qquad \textit{Eq. 7.4}$$

$$\alpha \cong \frac{-p}{b} * ln\left(\frac{\left[R_g^2 \left(\frac{3}{bMp}\right) - 1 + 3\left(\frac{p}{bM}\right) - 6\left(\frac{p}{bM}\right)^2 \right]}{6\left(\frac{p}{bM}\right)^3} + 1 \right) \qquad \textit{Eq. 7.5}$$

Mobility of micellated dsDNA was solved for using a rearranged version of Eq. 7.1 seen in Eq. 7.6.

$$\mu_{mDNA}^0 = L\left(\frac{\mu_{aDNA}^0}{\alpha + L}\right)$$

Estimated elution times were calculated using both Eq. 7.7 and Eq. 7.8 with the experimental conditions for the capillary electrophoresis run. These are tabulated in Table 7.2 where V is the applied voltage, E is the electric current, l_t is the total capillary length, and l_D is the capillary length to the detector.

$$E = \frac{V}{l_t}$$

$$t_{eution} = -\frac{l_d}{E\mu}$$

Parameter	
V	$20\ kV$
E	$667\ \dfrac{V}{cm}$
l_d	$20\ cm$
l_t	$30\ cm$

Table 7.2 Capillary electrophoresis run parameters.

Estimated elution times and parameters for double-stranded and single-stranded DNAs are tabulated in Table 7.3.[19,22,23] Separation of electropherogram peaks was estimated to be around 3 minutes.

Parameter	miR-1260 probe	miR-21-3p probe #1	miR-1260 probe + target	miR-21-3p probe #1 + target
μ_{aDNA}^0 $(10^{-4}\frac{cm^2}{Vs})$	−3.01	−3.01	−3.1	−3.1
μ_{mDNA}^0 $(10^{-4}\frac{cm^2}{Vs})$	−0.74	−0.82	−1.55	−1.55
L $(bases)$	18	22	18	22
α $(bases)$	55	59	18	22
M $(monomers)$	18	22	18	22
p (nm)	3.3	3.3	50	50
b (nm)	.43	.43	.34	.34
R_g^2 (nm^2)	2.95	4.46	2.96	4.47
R_g (nm)	1.72	2.11	1.72	2.11
R_H (nm)	1.15	1.41	1.15	1.41
$t(\mu_{aDNA}^0)$ (min)	1.7	1.7	1.6	1.6
$t(\mu_{mDNA}^0)$ (min)	6.75	6.1	3.2	3.2

Table 7.3 Parameters for estimation of CE elution times.[19,22,23]

7.3.3   DNA Probe Results

The electropherograms of the DNA probes bound to DNA analogues of cgr-miR-21-3p and cgr-miR-1260 are shown in Figure 7.2. DNA targets were hybridized to probes in a 1:1 molar ratio. The target-bound cgr-miR-1260 probe eluted before the unhybridized probe and showed a well-resolved separation of over 1 minute with almost complete hybridization. This contrasted greatly with the results of the cgr-miR-21-3p probes which showed almost no hybridization to their DNA target. Two very short peaks were observed around 4 minutes where the probe-target complexes would be expected to elute, however, there was very little binding. The difference in hybridization between the cgr-miR-1260 and cgr-miR-

21-3p probes correlated with the predicted secondary structure of the short nucleic acid targets. Discussed further in the next section, cgr-miR-1260 is an unstructured miRNA while cgr-miR-21-3p folds over onto itself with 6 complementary base pairs. The self-complementarity of the cgr-miR-21-3p probe was decreased by reducing probe #2 by 4 nt to give it a much lower probability of forming secondary structures. However, this did not increase binding compared to probe #1, indicating that the secondary structure of the target plays a key role in the ability of a DNA-based probe to hybridize.
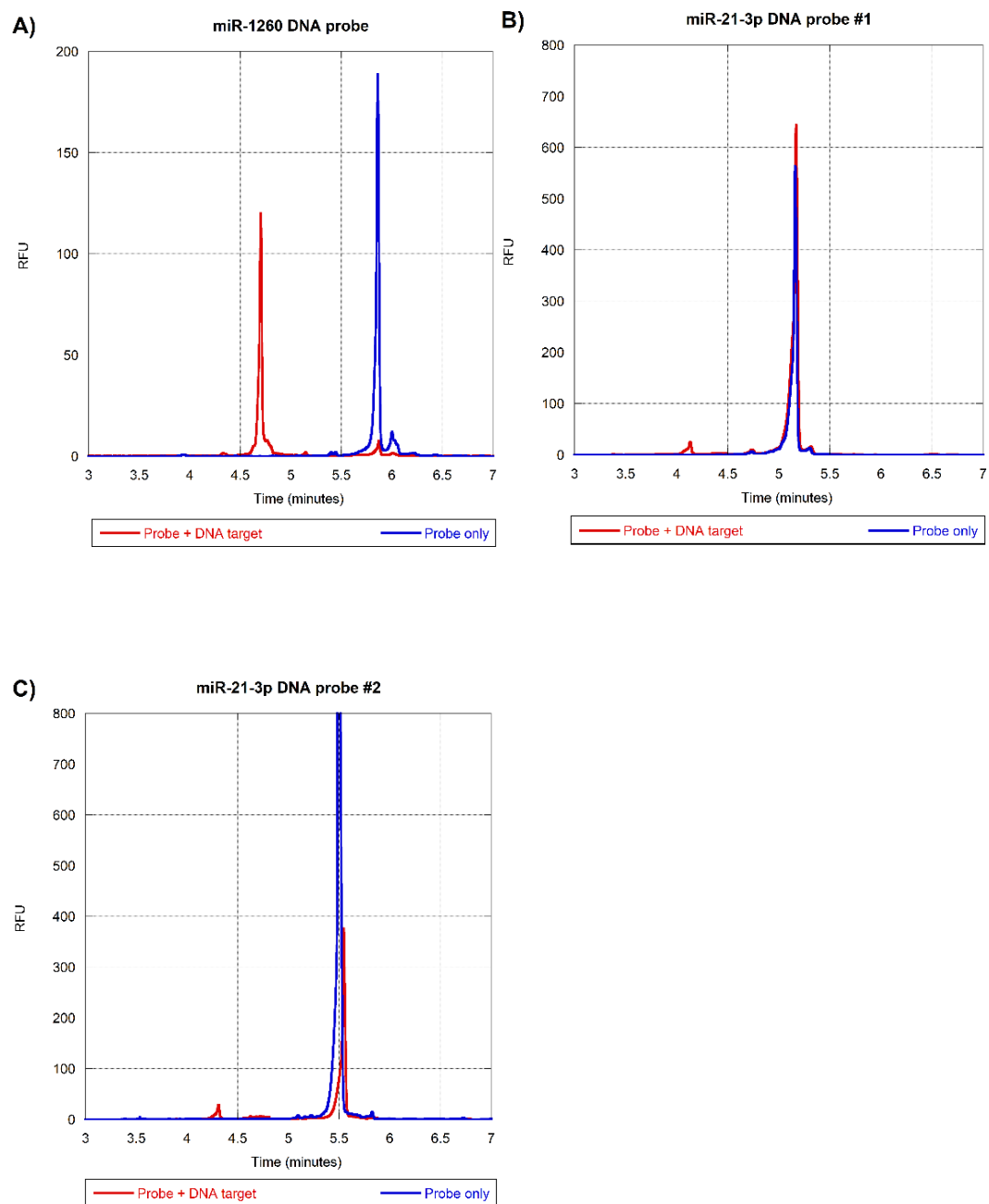
Figure 7.2 A) Separation of cgr-miR-1260 DNA probe and probe hybridized to DNA target at 22°C using MTE. B) cgr-miR-21-3p DNA probe #1 hybridized to DNA target. C) cgr-miR-21-3p DNA probe #2 (shortened) hybridized to DNA target.

134

These results indicate that a DNA probe would be sufficient for detection of a single, unstructured short nucleic acid target with MTE. Separation of cgr-miR-1260 was achieved with a CE run time of less than 6 minutes, allowing for a quick detection of DNA targets without structure. Using Eq. 7.9, mobilities were calculated using elution times of probe and probe-target complexes and are tabulated below in Table 7.4. Here, the applied voltage (V) was $20\ kV$, the total capillary length (l_t) was $30\ cm$, and the length to detector (l_d) was $20\ cm$.

$$\mu = -\frac{l_t l_d}{Vt} \qquad \qquad \textit{Eq. 7.9}$$

Nucleic acid	Experimental elution time (min)	Experimental mobility ($\frac{cm^2}{Vs}$)
miR-21-3p probe #1	5.2	$-0.96 * 10^{-4}$
miR-21-3p probe #1 + DNA target	4.1	$-1.22 * 10^{-4}$
miR-21-3p probe #2	5.5	$-0.91 * 10^{-4}$
miR-21-3p probe #2 + DNA target	4.3	$-1.16 * 10^{-4}$
miR-1260 probe	5.8	$-0.86 * 10^{-4}$
miR-1260 probe + DNA target	4.7	$-1.06 * 10^{-4}$

Table 7.4 Experimental elution times and mobilities of micellated DNA probes and DNA probe-target hybridization complexes.

7.3.4 Optimization of MTE Run Temperature

To account for the competition between the folding of miRNA and probe binding, it was necessary to optimize the run temperature for MTE experiments. Based on the melting temperatures of the hybridized DNA probes, the running temperature was kept below $60°C$. From Figure 7.3, it was observed that a capillary temperature up to $40°C$ did not significantly change detection of a short unstructured DNA target. Temperatures above $22°C$ showed complete binding of probe to target,

whereas the $22°C$ run still displayed a short peak where the unbound probe would elute. Therefore, the ideal operating temperature was estimated to be around $30°C$.
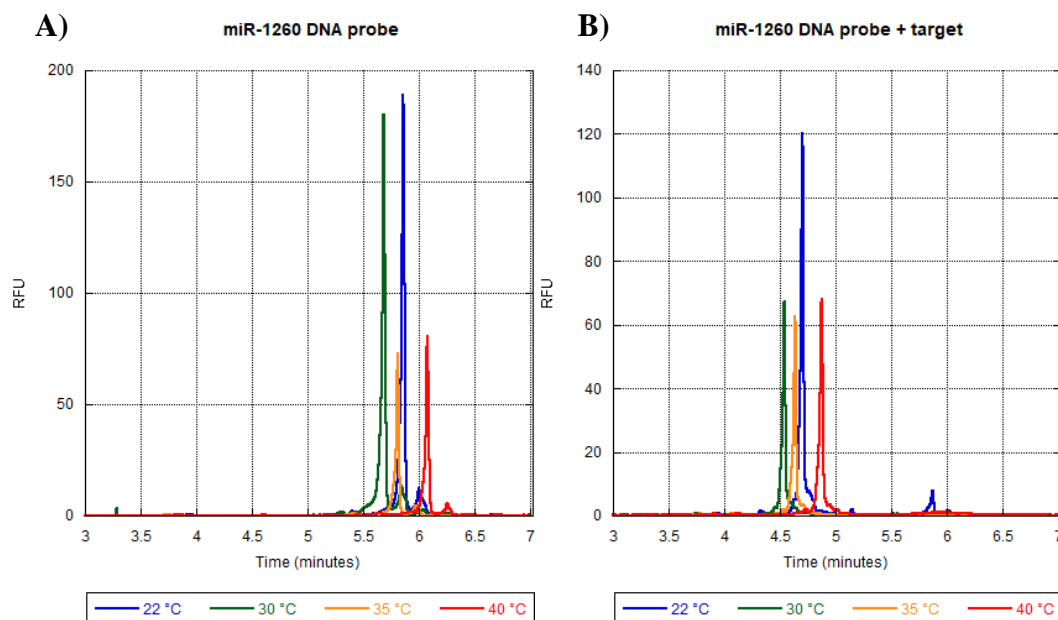
Figure 7.3 Effects of CE run temperature on DNA probe binding to unstructured target. Complete hybridization and separation are observed from $22°C$ to $40°C$.

MTE experiments were also done to increase the capillary temperature with the structured cgr-miR-21-3p DNA target up to $50°C$. However, these higher temperatures produced inconsistent results and did not increase peak resolution or hybridization of the probe to target.

7.4   Secondary Structure of Small Nucleic Acids

The sequences of cgr-miR-21-3p and cgr-miR-1260 were chosen for their biological relevance as biomarkers, however, MTE experiments revealed the need for varying detection strategies. We hypothesize that the difference in hybridization between cgr-miR-1260 and cgr-miR-21-3p targets is due to their differing secondary folding structures. Even though cgr-miR-21-3p is a short 22 nucleotide

sequence, it is self-complementary and folds into a secondary hairpin structure while cgr-miR-1260 does not. Figure 7.4 shows the minimum free energy (MFE) structures predicted for the cgr-miR-21-3p RNA sequence and its DNA analogue.
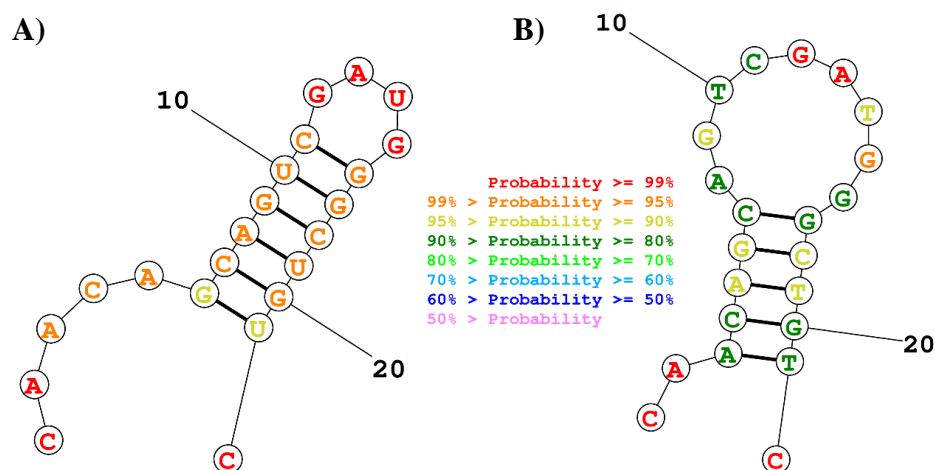
Figure 7.4 Predicted secondary folding structures and binding probabilities of A) cgr-miR-21-3p RNA sequence and B) cgr-miR-21-3p DNA analogue sequence generated using RNAstructure software. [24–26]

Melting temperature (T_m) and free energy change (ΔG) were calculated using UNAFold software and are tabulated in Table 7.5.[27] Folding structures were generated through the RNAstructure software package.[24–26] Interestingly, the cgr-miR-21-3p DNA and RNA sequences fold into different structures with different numbers of base pairs, as opposed to what might be expected. Upon further observation, this is due to G·U wobble in the RNA base pairing. The extra base pair in the RNA structure, as well as the fact that RNA forms stronger bonds than DNA,[28] contributes to a significantly higher melting temperature. Given that RNA hairpin stability, if an approach to nucleic acid detection fails to detect DNA with a hairpin, it will also fail to overcome the secondary structure formation thermodynamics for the analogue RNA sequence.

	miR-21-3p RNA sequence	miR-21-3p DNA analogue
T_m (°C)	78.8	55.1
ΔG ($kcal \cdot mol^{-1}$)	−8.11	− 3.22

Table 7.5 T_m and ΔG of predicted cgr-miR-21-3p secondary folding structures estimated using UNAFold software. [27]

7.5 LNA Probes

7.5.1 LNA Probe Design

Hybridization with a structured target is not a problem which can be fixed purely through better design of DNA probes, but would require the utilization of different nucleic acid backbone chemistry capable of strand invasion. Locked Nucleic Acid (LNA) is a nucleic acid analogue which has been proven to show strong hybridization affinity towards both RNA and DNA. Due to its superior binding properties, invasion into structured regions is possible.[29] LNA has an A-like character similar to RNA and is often substituted into DNA or RNA nucleotide sequences.[30] However, due to this strong binding affinity, self-complementarity should be avoided in LNA-substituted sequences and long sequences often have problems with aggregation. To eliminate probe self-complementarity and decrease the probability of aggregation, a sandwich probe design, similar to that used in Goldman et al. was utilized.[14,31] Figure 7.5 shows a schematic of the sandwich probe complex for binding to a 22-nucleotide target. A short probe with LNA substitution attaches to a micelle using Bodipy FL C16 as the fluorophore.
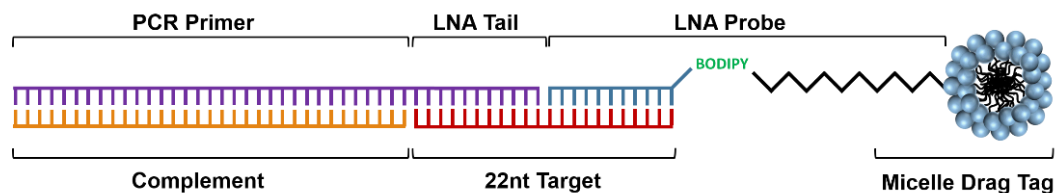
Figure 7.5 Diagram of sandwich design probes binding to a short 22 nt target. Probes which bind to the target utilize LNA substitution with strong binding chemistry for invasion of a structured target. Length of PCR primer and complement probes may be varied to create a mobility shift for detection of multiple targets.

The PCR primer probe is not utilized for amplification here but is a design mechanism for future work. Adding LNA substitution to both the micellated probe and the PCR primer probe allowed them both to have a high enough melting temperature to bind despite their short length. The complement strand and PCR primer with LNA tail form a double-stranded DNA tail which can be tailored in length to control the mobility shift. Probe sequences are listed in Table 7.6.

Probe	Sequence (5'-3')
DNA target	CAACAGCAGTCGATGGGCTGTC
RNA target	CAACAGCAGUCGAUGGGCUGUC
Micellated LNA probe *	GACAGCCCATC
PCR Primer probe	GA**C**TGC**T**GTT**G**CGCAATTGTTATCAGCTATGCGCCGACCAGAAC
Complement	GTTCTGGTCGGCGCATAGCTGATAACAATTGAG

*Table 7.6 Sandwich complex probe sequences for cgr-miR-21-3p target. LNA substitutions are underlined in red. *Four different LNA substitution percentages were designed for the micellated LNA probe and are found in Table 7.7.*

Table 7.7 lists the LNA substitution sequences of the micellated probes along with their T_m to a complementary DNA target. Probes were optimized to space out LNA substitutions while maximizing T_m for binding strength, which was estimated by using nearest neighbor thermodynamic calculations with constants from the literature for DNA, RNA, and LNA substitutions.[32–40] Previous work has

139

demonstrated difficulty with the aggregation of LNA probes at higher LNA percentage substitutions. Therefore, a modified 45% LNA probe was created with a PEG linker ($MW = 212.1 \frac{g}{mol}$) on the 5' end of the sequence to increase solubility.

	18% LNA Probe	27% LNA Probe	45% LNA Probe	45% LNA w/ PEG
LNA Probe (5'-3')	GA**CA**GC**C**CATC	GA**CA**GC**CC**ATC	G**ACA**GC**CCA**TC	G**ACA**GC**CCA**T**C**-PEG
DNA target T_m (°C)	34.8	52.4	57.6	57.6

Table 7.7 Four different substitution percentage designs for cgr-miR-21-3p micellated LNA probe and their associated melting temperatures estimated using IDT OligoAnalyzer tool. LNA substitutions are underlined in red.

7.5.2 DNA – LNA Probe Results

The electropherograms of the LNA sandwich probes hybridized in a 1:1 ratio to a DNA analogue of cgr-miR-21-3p are shown in Figure 7.6. These experiments included no annealing steps for increasing hybridization. The probe with 18% LNA substitution did not show any binding to the target DNA while the 27% and 45% probes demonstrated imperfect hybridization with a well-defined mobility shift of approximately 2.5 minutes. Insight into the optimization of LNA substitution percentage for strand invasion can be obtained from these plots. The 45% LNA probe noticeably precipitated during the alkylation process, resulting in low yields. All probes and targets were diluted to a concentration of 100nM for MTE experiments, however, the 45% LNA probe electropherogram displays a much lower peak when compared to the probes of lower percent LNA substitutions. This indicates that beyond the alkylation process, the higher LNA substitution still presents issues with aggregation under CE conditions, which will cause

inconsistent outcomes for MTE experiments. A 45% LNA probe with a PEG linker on the 5' end for solubility is shown here to fix this aggregation problem. The modified 45% LNA probe was fully soluble during the alkylation process and has a large well-defined peak on the electropherogram. However, the soluble 45% LNA probe did not show increased hybridization of a structured DNA target when compared to the 27% LNA probe. Therefore, the optimal LNA substitution percentage for invasion must be above 18% and below 45%, and is predicted to be around the 27% mark.
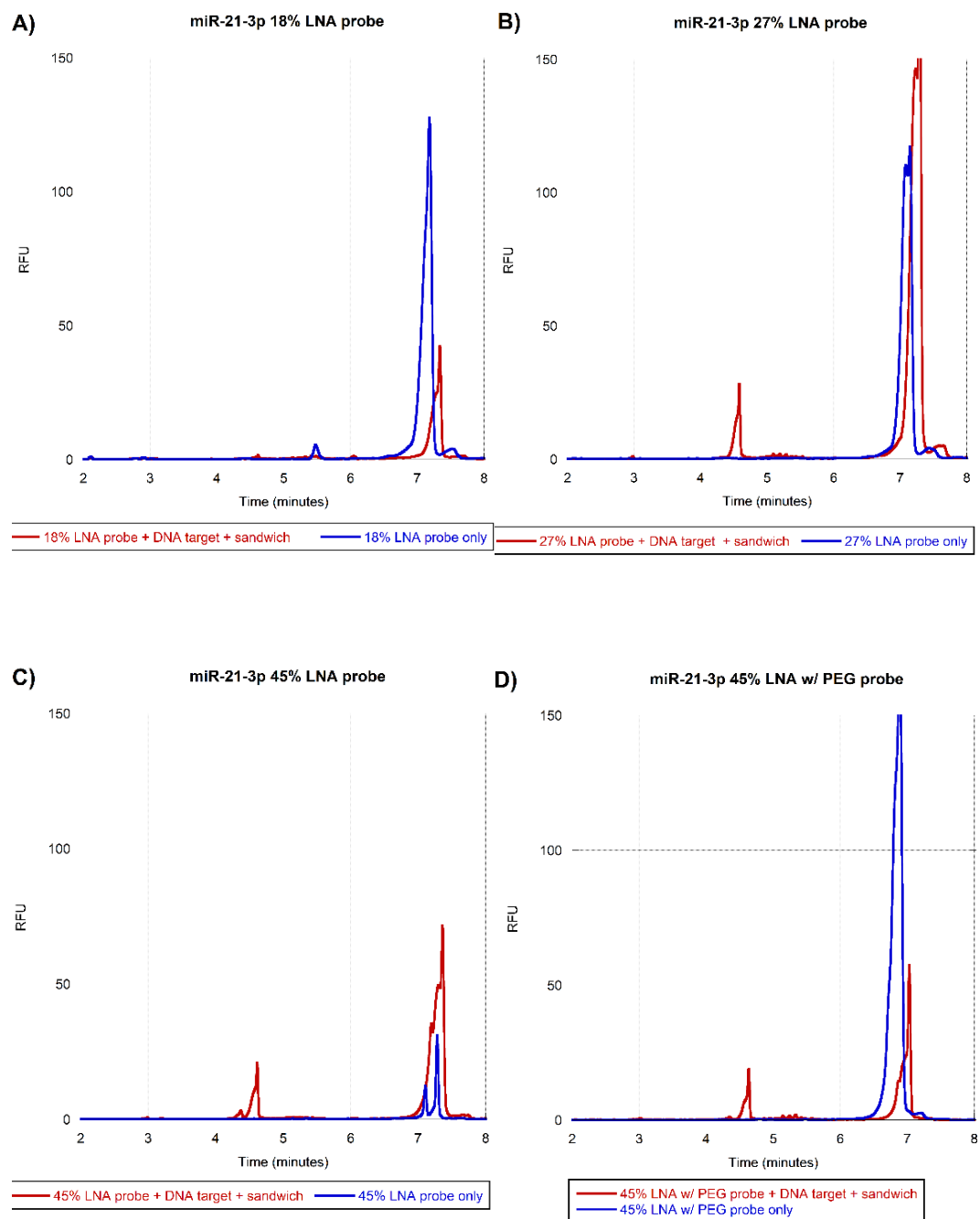
Figure 7.6 Electropherograms of cgr-miR-21-3p sandwich probes hybridized to a DNA target. Micellated probe included an LNA substitution of A) 18% LNA, B) 27% LNA, C) 45% LNA, or D) 45% LNA with a PEG linker for solubility.

7.5.3 Annealing Results

In order to increase hybridization of DNA targets, an annealing step was added as described in the methods. The electropherograms in Figure 7.7 show the results of two different annealing conditions, including a fast cool where the hybridization solution was heated and cooled on ice, and a slow cool where the hybridization solution was heated and allowed to reach room temperature over the span of an hour. First, it is noticeable that there are two separate target peaks with well-defined mobility shifts. The left-most peak around 3 minutes is the DNA target bound to the entire sandwich complex. The long, double-stranded sandwich complex is more negative than the micellated probe only or the micellated probe hybridized to the target and would elute first. The second target peak in the center with elution around 4.5 minutes is the DNA target hybridized to the micellated probe only. It should be noted that the elution times for the hybridized target in Figure 7.6 correlate with this second peak. Although the unannealed LNA probes were capable of invading a structured target, the entire sandwich complex did not form. Since the longer sandwich probes are the drag tags used to tailor mobility shift, this would create a problem for the rapid detection of multiple targets at once.

Similar to the results in Figure 7.6, an annealing step does not improve the aggregation problem with the unmodified 45% LNA probe. The 27% LNA probe still shows optimal invasion which, apart from binding the entire sandwich complex, does not show significant improvement over target invasion without an annealing step. Annealing has the greatest impact on the 18% LNA probe, which shows slight hybridization of the target and formation of the sandwich complex

when annealed and slow-cooled. Ultimately, these results indicate that a 27% or 45% LNA probe would be capable of detecting a structured short nucleic acid target. Separation of the cgr-miR-21-3p was achieved with a CE run time of less than 8 minutes, demonstrating that micellated LNA probes are a viable option for the quick detection of single structured targets.
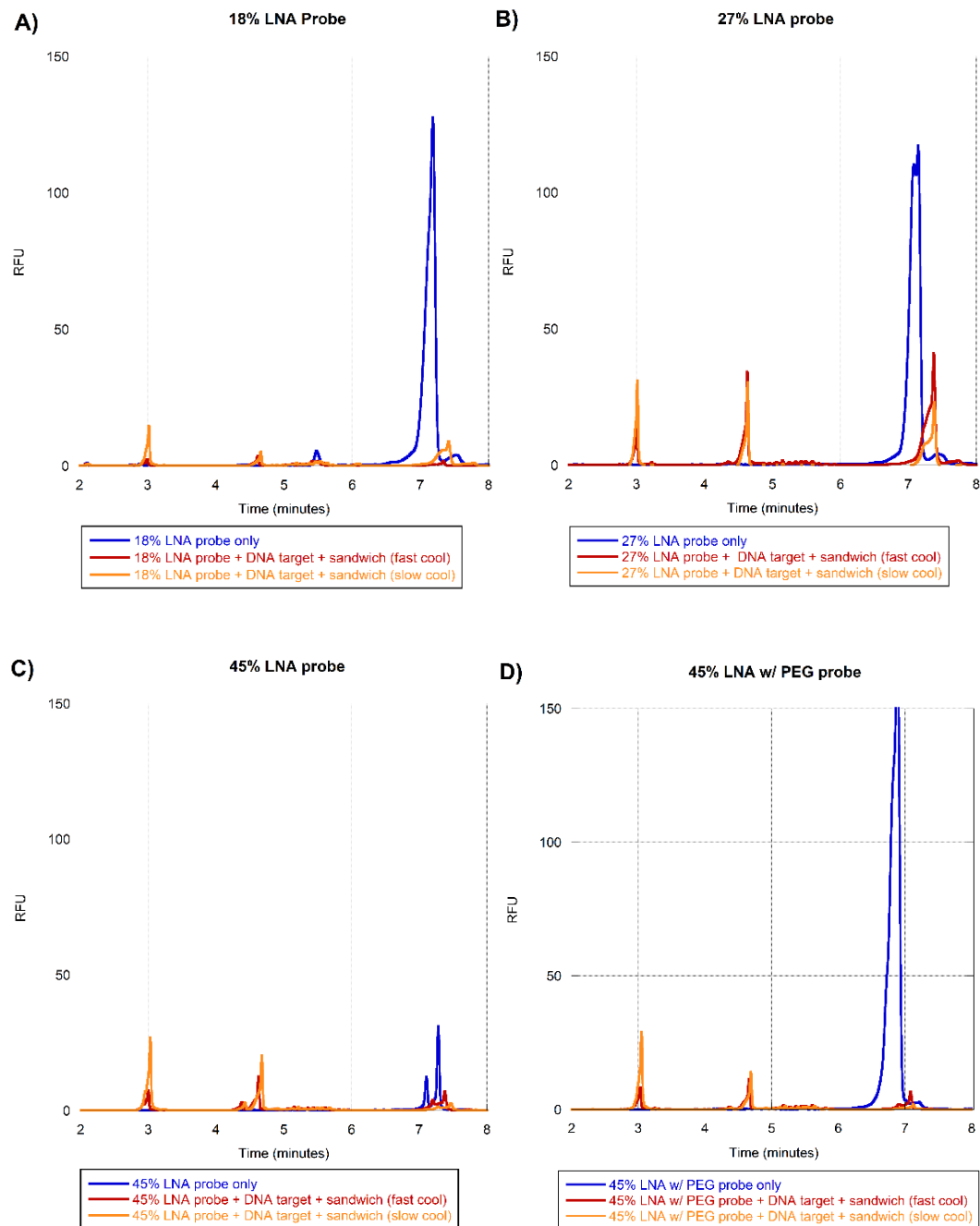
Figure 7.7 Electropherograms of cgr-miR-21-3p sandwich probes hybridized to a cgr-miR-21-3p DNA analogue target. Micellated probe included an LNA substitution of A) 18% LNA, B) 27% LNA, C) 45% LNA, or D) 45% LNA with a PEG linker for solubility.

7.5.4 RNA – LNA Probe Results

Given the ability of LNA probes to invade a structured DNA target, the same conditions were applied with a structured RNA target. Figure 7.8 shows the

electropherogram results of the micellated cgr-miR-21-3p 27% LNA probe hybridized with a cgr-miR-21-3p RNA target. The LNA probe did demonstrate hybridization of the target without an annealing step, however the peak is not well-defined and displays much noise. This may be the result of RNA instability and its predisposition towards degradation. A short annealing did not significantly increase probe-target binding, however, the addition of a long annealing step where the hybridization solution was allowed to slowly cool to the RNA target hybridization did greatly increase binding. Separation between the probe-target peak and unhybridized probe was approximately 3 minutes, which is similar to the separation shown between the same LNA probe and its hybridized complex to a DNA target.
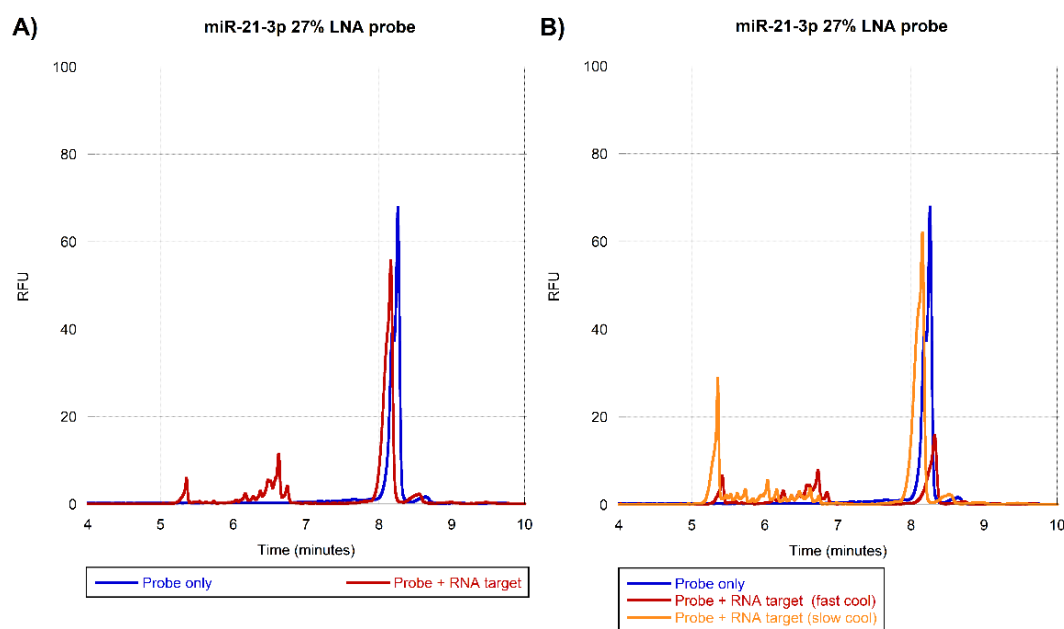
Figure 7.8 Electropherograms of cgr-miR-21-3p micellated probes hybridized to an RNA target with A) no annealing or B) annealing. During the annealing step, hybridization solution was either cooled quickly on ice or allowed to reach room temperature slowly over an hour.

146

7.5.5 Unstructured Target – LNA Probe Results

An LNA probe was also designed for detection of an unstructured cgr-miR-1260 target. Sequences for the sandwich complex probes and DNA target sequence are tabulated in Table 7.8. MTE experiments were run under the same conditions as with the structured cgr-miR-21-3p target.

Probe	Sequence (5'-3')
DNA target	CAACAGCAGTCGATGGGCTGTC
Micellated LNA probe *	TG**GT**G**G**CAG
PCR Primer probe	CGG**T**G**G**GATATGTTCTGAGGGGAGTGAAAATTCCCCTAATTC
Complement	GAATTAGGGGAATTTTCACTCCCCTCAGAACAT

Table 7.8 Sandwich complex probe sequences for unstructured cgr-miR-1260 target. Micellated LNA probe utilized a 33% LNA substitution percentage. LNA substitutions are underlined in red.

As seen in Figure 7.9, a cgr-miR-1260 DNA analogue target showed binding to the micellated probe and the entire sandwich complex without an added annealing step. Despite this, it can be observed that the hybridization of the target to the entire sandwich complex is still significantly less than its hybridization to the micellated probe only. This phenomenon is similar to what was seen with LNA probes binding to a structured target, where the target is more likely to bind to a single micellated LNA probe than the entire sandwich complex.
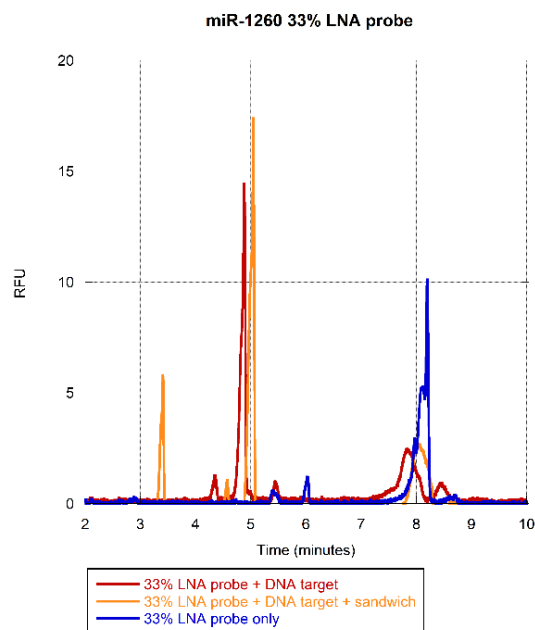
Figure 7.9 Electropheragram of cgr-miR-1260 33% LNA sandwich probes hybridized to a DNA target.

7.6 Conclusions

In conclusion, this work has explored the applications and limits of DNA and LNA probes in conjunction with MTE. We have demonstrated that a micelle tagging electrophoresis approach using only a single DNA probe is capable of detecting solitary, unstructured nucleotide targets. Furthermore, we have shown that these unstructured targets do not require any extra annealing step to hybridize to their probes. Here, it should be noted that out of 2,654 mature human miRNAs listed in the miRBase database, at least 625 contain no secondary structure. Therefore, a minimum of 24% of all human miRNAs would be detectable through this method.[41] However, for structured nucleic acid targets, a more complex sandwich structure probe design with stronger chemistry capable of strand invasion is necessary.

We have also successfully shown that LNA probes can be utilized with MTE for the detection of short nucleic acids. This work has demonstrated that LNA probes are a viable alternative to γPNAA probes and has laid the groundwork for subsequent studies into their limits and applications for rapid detection. Micellated LNA probes at 27% and 45% LNA substitution were shown to bind to structured DNA and RNA targets without annealing. Aggregation difficulties with the 45% LNA probe were overcome by inserting a PEG linker on the 5' end of the micellated probe sequence for solubility. An added annealing step combined with the LNA probe approach did not increase hybridization of the micellated probe to the nucleic acid target, however it was necessary for binding of all three probes in the sandwich structure which would be necessary to create mobility shifts for the multiplex detection of targets. Future work may include strategies for multiplex detection of short nucleotides using LNA probes in addition to the detection of miRNA from biological samples and quantification of the lower limits of detection.

7.7 References

1. Leva, G. Di, Calin, G. A. & Croce, C. M. MicroRNAs: Fundamental facts and involvement in human diseases. *Birth Defects Research Part C - Embryo Today: Reviews* **78**, 180–189 (2006).

2. Leung, A. K. L. & Sharp, P. A. MicroRNA Functions in Stress Responses. *Molecular Cell* **40**, 205–215 (2010).

3. Lu, L. F. & Liston, A. MicroRNA in the immune system, microRNA as an immune system. *Immunology* **127**, 291–298 (2009).

4. Hariharan, M., Scaria, V., Brahmachari, S. K., Pillai, B. & Maiti, S. Host-virus genome interactions: macro roles for microRNAs. *Cellular Microbiology* **9**, 2784–2794 (2007).

5. Skalsky, R. L. & Cullen, B. R. Viruses, microRNAs, and Host Interactions. *Annual Review of Microbiology* **64**, 123–141 (2010).

6.   Marquez, R. T. *et al.* Correlation between microRNA expression levels and clinical parameters associated with chronic hepatitis C viral infection in humans. *Laboratory Investigation* **90**, 1727–1736 (2010).

7.   Guire, D. v. *How close are miRNAs from clinical practice? A perspective on the diagnostic and therapeutic market*.

8.   Cissell, K. A. & Deo, S. K. Trends in microRNA detection. *Analytical and Bioanalytical Chemistry* **394**, 1109–1116 (2009).

9.   Wegman, D. W. *et al.* Highly-sensitive amplification-free analysis of multiple miRNAs by capillary electrophoresis. *Analytical Chemistry* **87**, 1404–1410 (2015).

10.   Wegman, D. W., Cherney, L. T., Yousef, G. M. & Krylov, S. N. Universal drag tag for direct quantitative analysis of multiple microRNAs. *Analytical Chemistry* **85**, 6518–6523 (2013).

11.   Wegman, D. W. & Krylov, S. N. Direct quantitative analysis of multiple miRNAs (DQAMmiR). *Angewandte Chemie - International Edition* **50**, 10335–10339 (2011).

12.   Goldman, J. M., Kim, S., Narburgh, S., Armitage, B. A. & Schneider, J. W. Rapid, multiplexed detection of the let-7 miRNA family using γPNA amphiphiles in micelle-tagging electrophoresis. *Biopolymers* **113**, (2022).

13.   Meagher, R. J. *et al.* End-labeled free-solution electrophoresis of DNA. *Electrophoresis* vol. 26 331–350 Preprint at https://doi.org/10.1002/elps.200410219 (2005).

14.   Goldman, J. M. *et al.* High affinity γpNA sandwich hybridization assay for rapid detection of short nucleic acid targets with single mismatch discrimination. *Biomacromolecules* **14**, 2253–2261 (2013).

15.   Tailhades, J. *et al.* Solid-phase synthesis of difficult purine-rich PNAs through selective hmb incorporation: Application to the total synthesis of cell penetrating peptide-PNAs. *Frontiers in Chemistry* **5**, (2017).

16.   Tolstrup, N. *et al.* OligoDesign: Optimal design of LNA (locked nucleic acid) oligonucleotide capture probes for gene expression profiling. *Nucleic Acids Research* **31**, 3758–3762 (2003).

17.   Válóczi, A. *et al.* Sensitive and specific detection of microRNAs by northern blot analysis using LNA-modified oligonucleotide probes. *Nucleic Acids Res* **32**, (2004).

18.   *User's Guide PA 800 plus Pharmaceutical Analysis System Methods Development*. www.beckmancoulter.com (2014).

19.   Savard, J. M. Micellar and Chromatographic Media for Surfactant-Based DNA Analysis and Purification Processes. (Carnegie Mellon University, 2008).

20. Meagher, R. J. *et al.* End-labeled free-solution electrophoresis of DNA. *Electrophoresis* vol. 26 331–350 Preprint at https://doi.org/10.1002/elps.200410219 (2005).

21. Desruisseaux, C., Long, D., Drouin, G. & Slater, G. W. Electrophoresis of composite molecular objects. 1. Relation between friction, charge, and ionic strength in free solution. *Macromolecules* **34**, 44–52 (2001).

22. Lu, Y., Weers, B. & Stellwagen, N. C. DNA persistence length revisited. *Biopolymers* **61**, 261–275 (2001).

23. Tinland, B., Pluen, A., Sturm, J. & Weill, G. *Persistence Length of Single-Stranded DNA*. https://pubs.acs.org/sharingguidelines (1997).

24. Mathews, D. H. Using an RNA secondary structure partition function to determine confidence in base pairs predicted by free energy minimization. *RNA* **10**, 1178–1190 (2004).

25. Mathews, D. H. *et al. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. PNAS* vol. 101 www.pnas.orgcgidoi10.1073pnas.0401799101 (2004).

26. Mathews, D. H., Sabina, J., Zuker, M. & Turner, D. H. *Expanded Sequence Dependence of Thermodynamic Parameters Improves Prediction of RNA Secondary Structure*. http://www.idealibrary.com.

27. Markham, N. R. & Zuker, M. UNAFold: software for nucleic acid folding and hybridization. *Methods Mol Biol* **453**, 3—31 (2008).

28. Swart, M., Guerra, C. F. & Bickelhaupt, F. M. Hydrogen bonds of RNA are stronger than those of DNA, but NMR monitors only presence of methyl substituent in uracil/thymine. *J Am Chem Soc* **126**, 16718–16719 (2004).

29. Karkare, S. & Bhatnagar, D. Promising nucleic acid analogs and mimics: Characteristic features and applications of PNA, LNA, and morpholino. *Applied Microbiology and Biotechnology* vol. 71 575–586 Preprint at https://doi.org/10.1007/s00253-006-0434-2 (2006).

30. Petersen, M. & Wengel, J. LNA: A versatile tool for therapeutics and genomics. *Trends in Biotechnology* vol. 21 74–81 Preprint at https://doi.org/10.1016/S0167-7799(02)00038-0 (2003).

31. Goldman, J. M. *et al.* High affinity γpNA sandwich hybridization assay for rapid detection of short nucleic acid targets with single mismatch discrimination. *Biomacromolecules* **14**, 2253–2261 (2013).

32. Santalucia, J. A unified view of polymer , dumbbell , and oligonucleotide DNA nearest-neighbor thermodynamics. **95**, 1460–1465 (1998).

33. Allawi, H. T. & Santalucia, J. Thermodynamics of internal C · T mismatches in DNA. **26**, 2694–2701 (1998).

34. Allawi, H. T. & Santalucia, J. Nearest-Neighbor Thermodynamics of Internal A , C Mismatches in DNA : Sequence Dependence and pH Effects. **2960**, 9435–9444 (1998).

35. Allawi, H. T. & Santalucia, J. Nearest Neighbor Thermodynamic Parameters for Internal G , A Mismatches in DNA. **2960**, 2170–2179 (1998).

36. Allawi, H. T. & Santalucia, J. Thermodynamics and NMR of Internal G , T Mismatches in DNA. **2960**, 10581–10594 (1997).

37. Peyret, N., Seneviratne, P. A., Allawi, H. T. & Santalucia, J. Articles Nearest-Neighbor Thermodynamics and NMR of DNA Sequences with Internal. 3468–3477 (1999).

38. Dna, L. N. A., Formation, D., Mctigue, P. M., Peterson, R. J. & Kahn, J. D. Sequence-Dependent Thermodynamic Parameters for Locked Nucleic Acid. 5388–5405 (2021).

39. Banerjee, D. *et al.* Improved nearest-neighbor parameters for the stability of RNA / DNA hybrids under a physiological condition. **48**, 12042–12054 (2020).

40. Chen, J. L. *et al.* Testing the Nearest Neighbor Model for Canonical RNA Base Pairs: Revision of GU Parameters. (2012).

41. Griffiths-Jones, S., Grocock, R. J., van Dongen, S., Bateman, A. & Enright, A. J. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res* **34**, (2006).

Chapter 8 – Conclusions, Challenges, and Future Work

8.1 Conclusions

Overall, this thesis has laid the groundwork for the use of miRNA biomarkers in biomanufacturing. First, we identified the CHO-K1 miRNA transcriptome separate from *Cricetulus griseus*. Second, we quantified the virulence of Reo3, MMV, EMC, and VSV infections in CHO-K1 cells. Using that foundation, we identified both general and specific biomarkers for viral infection in CHO-K1 cultures. MicroRNA biomarkers were identified based on a double filtering process of fold change and p-value to determine differential expression. Subsequently, targets were assessed for their ability to act as classifiers for viral infection through a receiver operating curve analysis and narrowed down to exclude those which were non-discriminant or likely to give false positives or false negatives. General biomarkers for viral infection in CHO-K1 were seen to have excellent discrimination capabilities, with area-under-the-curve values upwards of 0.90.

Additionally, we have identified miRNA biomarkers for an early response to HSV-1 infection in human corneal epithelial (HCLE) cells. The top miRNA biomarker for an early host response to HSV-1 infection in HCLE cells was found to be miR-378f. Additionally, both miR-615-3p and miR-495-3p showed a linear increase in their fold change over time and were good classifiers without a proclivity towards false positives or false negatives. The identification of these biomarkers is preliminary work for the development of a more efficient diagnostic method for ocular surface infection, detected through ocular fluid sampling.

Finally, we explored the use of DNA and LNA probes as a more inexpensive option compared to the γPNAA probes previously used with the MTE method. It was observed that secondary structure of short nucleic acids played a large role in their detection strategy using MTE. Unstructured short nucleic acids could be detected using DNA probes; however, structured targets needed the stronger probe chemistry of LNA to overcome their self-complementarity. Aggregation of LNA probes was overcome by the addition of a soluble PEG linker added to the 5' end, however, the higher substitutions of LNA did not increase probe-target hybridization. The ideal substitution percentage for LNA probes was found to be in the 30% range to optimize binding and prevent aggregation.

8.2 Challenges of miRNA Biomarkers

8.2.1 Lack of Standard Protocols

Due to the relative newness of this field, it can be seen in the literature that the identification of miRNA biomarkers lacks standard protocols. There is a vast array of options for isolation of RNA with different experimental methods, and even more options for how data is analyzed once the RNA is extracted. However, one standard which seems to be consistent is qPCR as the "gold standard" for validation. Despite this, many different types of statistical tests are still used with different cutoffs for significant p-value and fold change to determine differential expression. More recently, some miRNA studies have also been done using machine learning to predict biomarkers. Due to the vast number of algorithm options, a standard protocol for this analysis would be beneficial.

8.2.2 Translation to Commercial Use or Clinical Diagnostics

Although an abundant number of studies have been done on determining miRNA biomarkers, they are not yet widely used for diagnostics either commercially or clinically. As of 2019, there were only 8 companies with products utilizing miRNA diagnostic biomarkers either available commercially or in development.[1] A couple of reasons may account for this. First, is the difficulty of translating research into clinical practice. Much upfront capital must be invested in a product or diagnostic to bring it through the development phase to be commercially available. Additionally, the reproducibility of biomarkers has been shown to be an issue, showing the need for homogeneous protocols for this relatively new field of research.[2] Finally, the variability of clinical subjects impacts the usefulness of miRNA biomarkers in practice. Life may be unlike the controlled nature of a research lab, and a variety of other environmental factors could play a role in cellular expression. This was observed in the experimental work of this thesis, where different detection techniques were shown to change outcomes of our biomarker screening on the same samples.

8.2.3 MicroRNA Research for Non-model Organisms

Many tools have been created to aid in the analysis of miRNA research. These include software for miRNA identification, novel miRNA analysis, target prediction, and functional analysis to name a few categories. A comprehensive list of miRNA software can be found at tools4mirs.org. However, despite the range and availability of numerous software packages, many of these tools are organism specific, particularly those which perform functional analyses. The most common

organisms for miRNA study are human, mouse, and rat, likely due to their overwhelming use in clinical research studies. Although CHO-K1 is a model organism for biomanufacturing, because miRNAs have not been widely studied in this industry, there is a lack of tools which are able to be utilized specifically with CHO-K1. However, as this research continues to grow, new tools may become available or more data for non-model organisms may be incorporated into existing databases.

8.3   Future Work

8.3.1   Fail-Fast Applications in Bioprocessing

In order for the identified biomarkers to be implemented in a biomanufacturing context, guidelines must be created and optimized for consistent monitoring of a host cell culture. This technology is not designed to detect viral infection on a single cell level, but rather for the consistent monitoring of host cell health through repeated, small sampling for fail-fast analysis. Limits of detection and optimal sample sizes must be established. It would also be necessary to gain an understanding of how quickly a virus spreads through the cells in a bioprocessing culture and what fraction of cells need to be infected before a sample will show differential expression of the miRNA biomarkers. The ROC curves previously developed may be utilized to determine an ideal operating point for false positive and false negative rates when compared to industry testing standards and incorporating a time factor. However, if using a panel of more than one miRNA, it would be necessary to develop an ROC curve which incorporates multiple miRNA biomarkers at once. This would be done by using an "and"/"or" condition for

different thresholds of those miRNAs and would ultimately give more accurate results than a single miRNA biomarker. Ideally, the creation of a computational model incorporating false positive/negative rates, a time scale, economic factors, and frequency of testing would allow for the optimization of all these parameters and greatly contribute to the development of this new method for adventitious agent testing.

8.3.2 Validation of General miRNA Biomarkers for Latent and Silent Infections

An important next step for validation of the identified general biomarkers for viral infection in CHO-K1, would be to test these biomarkers for a latent or silent viral infection. For a latent infection, HSV may be utilized as the infecting virus. For a silent infection, PI2 may be further analyzed to determine its infectivity in CHO-K1, or PI3 may be used as it is known to infect CHO-K1 cells and produce more virus without showing any cytopathic effects.[3] This work would help to build a strong case for utilization of miRNA biomarkers in upstream bioprocessing, as it would lead to quick detection of an infecting virus that doesn't show other signs of infection and would overcome the time-consuming challenges of other viral detection techniques.

8.3.3 LNA Probes

Further research is needed to fully understand the boundaries and applications of LNA probes. This may include further optimization of LNA substitution and carefully designed probes which utilize options such as mismatches and the G·U wobble to prevent probe dimers and aggregates from forming while still maintaining enough complementarity to a target to overcome secondary structure.

157

It is still not understood why a micellated LNA probe binds to a structured target without easily forming the whole sandwich complex. Additionally, it is not yet understood why hybridization of LNA probes does not appear to be complete and does not increase with a higher LNA substitution percentage. Detection of multiple miRNAs has been shown for γPNA probes, but not yet LNA probes. Because we have shown LNA sandwich probes to bind to and detect small nucleic acid targets, future work may include the utilization these probes for multiplex miRNA detection.

8.3.4   Implementation of MTE with Biological Samples

Another next step for MTE, is the detection of a target from biological samples. So far, short DNAs and RNAs have been ordered through Integrated DNA Technologies which mimic biological targets to prove the technology works for a model system. However, working with a biological sample introduces a number of complications. First, is the variable concentration of a particular miRNA in a total miRNA isolation or cell lysate solution. This is highly dependent on the particular miRNA target as copies of miRNAs vary widely on the cellular level and change based on cellular stress. Second, an ultimate goal is to be able to utilize cell lysates as an input for MTE, reducing the time, labor, and materials required to isolate and concentrate RNA. Challenges which must be overcome with this include off-target binding and unforeseen impacts of other cellular components which may cause noise or otherwise influence the MTE data.

8.3.5 Development of Standard Protocols for miRNA Biomarker Identification

Ultimately, miRNAs may be ideal biomarker molecules for detecting a variety of conditions. Particularly, from their ability for early detection, specificity, and noninvasive accessibility, miRNAs offer advantages over other types of biomarkers such as mRNA or pathogens. The lack of standard protocols to identify significant miRNAs, has severely impacted the implementation of miRNA biomarkers for clinical and commercial use. Once standard procedures are developed and the challenges of clinical application are increasingly overcome, miRNA biomarkers could become more than just a potential.

8.4 References

1. Bonneau, E., Neveu, B., Kostantin, E., Tsongalis, G. & de Guire, V. How close are miRNAs from clinical practice? A perspective on the diagnostic and therapeutic market. *The Journal of the International Federation of Clinical Chemistry and Laboratory Medicine* 114–127.

2. Condrat, C. E. *et al.* miRNAs as Biomarkers in Disease: Latest Findings Regarding Their Role in Diagnosis and Prognosis. *Cells* vol. 9 Preprint at https://doi.org/10.3390/cells9020276 (2020).

3. Berting, A., Farcet, M. R. & Kreil, T. R. Virus susceptibility of Chinese hamster ovary (CHO) cells and detection of viral contaminations by adventitious agent testing. *Biotechnol Bioeng* **106**, 598–607 (2010).

Appendix A – CHO-K1 mature miRNA Transcriptome

CHO-K1 miRNA	Sequence (5'-3')
cgr-let-7a	TGAGGTAGTAGGTTGTATAGTT
cgr-let-7a-2	CTGTACAGCCTCCTAGCTTTCC
cgr-let-7b	TGAGGTAGTAGGTTGTGTGGTT
cgr-let-7c	CTGTACAACCTTCTAGCTTTCC
cgr-let-7c-2	CTATACAATCTACTGTCTTTC
cgr-let-7d-3p	CTATACGACCTGCTGCCTTTCT
cgr-let-7d-5p	AGAGGTAGTAGGTTGCATAGTT
cgr-let-7f	TGAGGTAGTAGATTGTATAGTT
cgr-let-7g-3p	CTGTACAGGCCACTGCCTTGC
cgr-let-7g-5p	TGAGGTAGTAGTTTGTACAGTT
cgr-miR-1	TGGAATGTAAAGAAGTGTGTAT
cgr-miR-100-3p	CAAGCTTGTGTCTATAGGTATG
cgr-miR-100-5p	AACCCGTAGATCCGAACTTGT
cgr-miR-101a	TCAGTTATCACAGTGCTGATGC
cgr-miR-101b-3p	TACAGTACTGTGATAACTGAAG
cgr-miR-101b-5p	CGGTTATCATGGTACCGATGCTG
cgr-miR-103-3p	AGCAGCATTGTACAGGGCTATGA
cgr-miR-103-5p	AGCTTCTTTACAGTGCTGCCTTGT
cgr-miR-106b-3p	CCGCACTGTGGGTACTTGCTGC
cgr-miR-106b-5p	TAAAGTGCTGACAGTGCAGATA
cgr-miR-107	AGCAGCATTGTACAGGGCTATC
cgr-miR-10a-3p	CAAATTCGTATCTAGGGGAAT
cgr-miR-10a-5p	TACCCTGTAGATCCGAATTTGT
cgr-miR-10b-3p	ACAGATTCGATTCTAGGGGAAT
cgr-miR-10b-5p	TACCCTGTAGAACCGAATTTGT
cgr-miR-122	TGGAGTGTGACAATGGTGTTT
cgr-miR-1224	GTGAGGACTGGGGAGGTGGAG
cgr-miR-124	TAAGGCACGCGGTGAATGCCAA
cgr-miR-1249	ACGCCCTTCCCCCCCTTCTTCA
cgr-miR-125b	ACGGGTTAGGCTCTTGGGAGC
cgr-miR-125b-3p	TCACAAGTCAGGCTCTTGGGAC
cgr-miR-125b-5p	TCCCTGAGACCCTAACTTGTGA
cgr-miR-126a	TCGTACCGTGAGTAATAATGCG
cgr-miR-126b	CGCGTACCAAAAGTAATAATGTG
cgr-miR-127	TCGGATCCGTCTGAGCTTGGCT
cgr-miR-1271	GTGCCTGCTGTGTGCCAAGG
cgr-miR-128	GGGGGCCGATGCACTGTAAGA
cgr-miR-128-3p	TCACAGTGAACCGGTCTCTTT
cgr-miR-128-5p	CGGGGCCGTAGCACTGTCTGAGA
cgr-miR-1285	TGCACTCCAGCCTGGGCAACATAG
cgr-miR-129	CTTTTTGCGGTCTGGGCTTGC
cgr-miR-1306-3p	ACGTTGGCTCTGGTGGTGATG
cgr-miR-1306-5p	CCACCTCCCCTGCAAACGTCC
cgr-miR-130a-3p	CAGTGCAATGTTAAAAGGGCAT
cgr-miR-130a-5p	GCTCTTTTCACATTGTGCTACT
cgr-miR-130b-3p	CAGTGCAATGATGAAAGGGCAT
cgr-miR-130b-5p	ACTCTTTCCCTGTTGCACTACT
cgr-miR-132-3p	TAACAGTCTACAGCCATGGTCG
cgr-miR-132-5p	ACCGTGGCTTTCGATTGTTACT

cgr-miR-134	TGTGACTGGTTGACCAGAGGGG
cgr-miR-1343	CTCCTGGGGCCCGCACTCTCGCT
cgr-miR-137-3p	TTATTGCTTAAGAATACGCGTAG
cgr-miR-137-5p	ACGGGTATTCTTGGGTGGATAAT
cgr-miR-138	AGCTGGTGTTGTGAATCAGGC
cgr-miR-139-3p	TGGAGACGCGGCCCTGTTGGAGT
cgr-miR-139-5p	TCTACAGTGCACGTGTCTCCAG
cgr-miR-140-3p	ACCACAGGGTAGAACCACGGAC
cgr-miR-140-5p	CAGTGGTTTTACCCTATGGTAG
cgr-miR-141	TAACACTGTCTGGTAAAGATGGC
cgr-miR-142-3p	GTAGTGTTTCCTACTTTATGG
cgr-miR-142-5p	CCCATAAAGTAGAAAGCACTAC
cgr-miR-143	TGAGATGAAGCACTGTAGCTC
cgr-miR-144	GGATATCATCATATACTGTAAG
cgr-miR-146a	TGAGAACTGAATTCCATGGGTT
cgr-miR-146b-3p	GCCCTAGGGACTCAGTTCTGGT
cgr-miR-146b-5p	TGAGAACTGAATTCCATAGGCTG
cgr-miR-148a	TCAGTGCACTACAGAACTTTGT
cgr-miR-148b-3p	TCAGTGCATCACAGAACTTTGT
cgr-miR-148b-5p	GAAGTTCTGTTATACACTCAGGC
cgr-miR-149-3p	GAGGGAGGGACGGGGGCGGTGC
cgr-miR-149-5p	TCTGGCTCCGTGTCTTCACTCC
cgr-miR-150	TCTCCCAACCCTTGTACCAGTG
cgr-miR-151-3p	CTAGACTGAGGCTCCTTGAGG
cgr-miR-151-5p	TCGAGGAGCTCACAGTCTAGT
cgr-miR-152-3p	TCAGTGCATGACAGAACTTGG
cgr-miR-152-5p	AGGTTCTGTGATACACTCCGACT
cgr-miR-154-3p	AATCATACACGGTTGACCTATT
cgr-miR-154-5p	TAGGTTATCCGTGTTGCCTTCG
cgr-miR-155	TTAATGCTAATCGTGATAGGGG
cgr-miR-15a-3p	CAGGCCATATTGTGCTGCCTCA
cgr-miR-15a-5p	TAGCAGCACATAATGGTTTGTGGA
cgr-miR-15b-3p	CGAATCATTATTTGCTGCTCT
cgr-miR-15b-5p	TAGCAGCACATCATGGTTTACA
cgr-miR-16	ACCAATATTATTGTGCTGCTTT
cgr-miR-16-3p	CCAGTATTAACTGTGCTGCTGAA
cgr-miR-16-5p	TAGCAGCACGTAAATATTGGCG
cgr-miR-17-3p	ACTGCAGTGCAGGCACTTGTGG
cgr-miR-17-5p	CAAAGTGCTTACAGTGCAGGTAG
cgr-miR-181a-3p	ACCATCGACCGTTGATTGTACC
cgr-miR-181a-5p	AACATTCAACGCTGTCGGTGAGT
cgr-miR-181b-5p	AACATTCATTGCTGTCGGTGG
cgr-miR-181c-3p	ACCATCGACCGTTGAGTGGACC
cgr-miR-181c-5p	AACATTCAACCTGTCGGTGAGT
cgr-miR-181d-3p	CCCACCGGGGGATGAATGTCA
cgr-miR-181d-5p	AACATTCATTGTTGTCGGTGGGT
cgr-miR-182	TTTGGCAATGGTAGAACTCACACC
cgr-miR-183	ATGGCACTGGTAGAATTCACTG
cgr-miR-1839-3p	AGACCTACTTATCTACCAACAG
cgr-miR-1839-5p	AAGGTAGATAGAACAGGTCTTGT
cgr-miR-184	TGGACGGAGAACTGATAAGGGT
cgr-miR-1843	TCTGATCGTTCACCTCCATACA
cgr-miR-185-3p	AGGGGCTGGCTTTCCTCTGGT

cgr-miR-185-5p	TGGAGAGAAAGGCAGTTCCTGA
cgr-miR-186-3p	GCCCAAAGGTGAATTTTTTGG
cgr-miR-186-5p	CAAAGAATTCTCCTTTTGGGCTT
cgr-miR-187	TCGTGTCTTGTGTTGCAGCCG
cgr-miR-188	CATCCCTTGCATGGTGGAGG
cgr-miR-18a-3p	ACTGCCCTAAGTGCTCCTTCTGG
cgr-miR-18a-5p	TAAGGTGCATCTAGTGCAGATAG
cgr-miR-18b	TAAGGTGCATCTAGTGCAGTTAG
cgr-miR-190a	TGATATGTTTGATATATTAGGTTG
cgr-miR-190b	TGATATGTTTGATATTGGGTTG
cgr-miR-191-3p	CTGCGCTTGGATTTCGTTCCC
cgr-miR-191-5p	CAACGGAATCCCAAAAGCAGCTG
cgr-miR-192	CTGACCTATGAATTGACAGCCA
cgr-miR-193b-3p	AACTGGCCCACAAAGTCCCGCT
cgr-miR-193b-5p	CGGGGTTTTGAGGGCGAGATGA
cgr-miR-194	TGTAACAGCAACTCCATGTGGA
cgr-miR-194-2	CCAGTGGGGCTGCTGTTATCTG
cgr-miR-1949	TATACCAGGATGTCAGCATAGTT
cgr-miR-195	TAGCAGCACAGAAATATTGGCA
cgr-miR-1956	AGTCCAGGGCTGAGTCAGCGGA
cgr-miR-196a-3p	TCGGCAACAAGAAACTGCCTGA
cgr-miR-196a-5p	TAGGTAGTTTCATGTTGTTGGG
cgr-miR-196b	TAGGTAGTTTCCTGTTGTTGGG
cgr-miR-196c	TAGGTAGTTTCGTGTTGTTGGG
cgr-miR-199a	CCCAGTGTTCAGACTACCTGTTC
cgr-miR-199a-3p	ACAGTAGTCTGCACATTGGTT
cgr-miR-199a-5p	CCCAGTGTTCAGACTAGCTGTTC
cgr-miR-199b	CCCAGTGTTTAGACTACCTGT
cgr-miR-19a	TGTGCAAATCTATGCAAAACTGA
cgr-miR-19b-3p	TGTGCAAATCCATGCAAAACTGA
cgr-miR-19b-5p	AGTTTTGCAGGTTTGCATCCAGC
cgr-miR-1b	TGGGTACATAAAGAAGTATGTGC
cgr-miR-200a	TAACACTGTCTGGTAACGATGTT
cgr-miR-200b	TAATACTGCCTGGTAATGATGAC
cgr-miR-200c	TAATACTGCCGGGTAATGATGGA
cgr-miR-202	TTCCTATGCATATACTTCTTT
cgr-miR-204	TTCCCTTTGTCATCCTATGCCT
cgr-miR-206	TGGAATGTAAGGAAGTGTGTGG
cgr-miR-20a	TAAAGTGCTTATAGTGCAGGTAG
cgr-miR-20b	CAAAGTGCTCATAGTGCAGGTAG
cgr-miR-21-3p	CAACAGCAGTCGATGGGCTGTC
cgr-miR-21-5p	TAGCTTATCAGACTGATGTTGA
cgr-miR-210-3p	CTGTGCGTGTGACAGCGGCTGA
cgr-miR-210-5p	AGCCACTGCCCACCGCACACTG
cgr-miR-211	TTCCCTTTGTCATCCTTTGCCT
cgr-miR-212	ACCTTGGCTCTAGACTGCTTACT
cgr-miR-214-5p	TGCCTGTCTACACTTGCTGTGC
cgr-miR-215-3p	TCTGTCATTCTGTAGGCCAAT
cgr-miR-215-5p	ATGACCTATGAATTGACAGACA
cgr-miR-216b	AAATCTCTGCAGGCAAATGTGA
cgr-miR-217	TACTGCATCAGGAACTGATTGGA
cgr-miR-218a	TTGTGCTTGATCTAACCATGTG
cgr-miR-219b	AGAATTGCGTTTGGACAATCAGT

cgr-miR-22-3p	AAGCTGCCAGTTGAAGAACTGT
cgr-miR-22-5p	AGTTCTTCAGTGGCAAGCTTT
cgr-miR-221-3p	AGCTACATTGTCTGCTGGGTTTC
cgr-miR-221-5p	ACCTGGCATACAATGTAGATTTCTGT
cgr-miR-222-3p	AGCTACATCTGGCTACTGGGTCTCT
cgr-miR-222-5p	TCAGTAGCCAGTGTAGATCCTG
cgr-miR-23a-3p	ATCACATTGCCAGGGATTTCC
cgr-miR-23a-5p	GGGGTTCCTGGGGATGGGATTT
cgr-miR-23b-3p	ATCACATTGCCAGGGATTACC
cgr-miR-23b-5p	GGGTTCCTGGCATGCTGATT
cgr-miR-24	GTGCCTACTGAGCTGATATCAGT
cgr-miR-24-3p	TGGCTCAGTTCAGCAGGAACAGG
cgr-miR-24-5p	GTGCCTACTGAGCTGAAACAG
cgr-miR-2424	ACAGATCTTTGGTAATCTGATGGCT
cgr-miR-25-3p	CATTGCACTTGTCTCGGTCTGA
cgr-miR-25-5p	AGGCGGAGACTTGGGCAATTGCT
cgr-miR-26a	CCTATTCTTGGTTACTTGCACG
cgr-miR-26a-2	TTCAAGTAATCCAGGATAGG
cgr-miR-26b-3p	CCTGTTCTCCATTACTTGGCTC
cgr-miR-26b-5p	TTCAAGTAATTCAGGATAGGTT
cgr-miR-27a-3p	TTCACAGTGGCTAAGTTCCGC
cgr-miR-27a-5p	AGGGCTTAGCTGCTTGTGAGCA
cgr-miR-27b-3p	TTCACAGTGGCTAAGTTCTGC
cgr-miR-27b-5p	AGAGCTTAGCTGATTGGTGAAC
cgr-miR-28-3p	CACTAGATTGTGAGCTCCTGGA
cgr-miR-28-5p	AAGGAGCTCACAGTCTATTGA
cgr-miR-296	GAGGGTTGGGTGGAGGCTCT
cgr-miR-298-3p	GAGGAACTAGCCTTCTCTCTGC
cgr-miR-298-5p	GGCAGAGGAGGGCTGTTCTTC
cgr-miR-29a-3p	TAGCACCATCTGAAATCGGTT
cgr-miR-29a-5p	ACTGATTTCTTTTGGTGTTCAGAG
cgr-miR-29b	GCTGGTTTCACATGGTGGCTTAGA
cgr-miR-29b-3p	TAGCACCATTTGAAATCAGTGTT
cgr-miR-29b-5p	GCTGGTTTCATATGGTGGTTTAGA
cgr-miR-29c-3p	TAGCACCATTTGAAATCGGTT
cgr-miR-29c-5p	ACCGATTTCTCCTGGTGTTCAGA
cgr-miR-300	TATGCAAGGGCAAGCTCTCTTC
cgr-miR-301a-3p	CAGTGCAATAGTATTGTCAAAGC
cgr-miR-301a-5p	GCTCTGACTTTATTGCACTACT
cgr-miR-301b	CAGTGCAATGATATTGTCAAAGC
cgr-miR-3068	GGTGAATTGCAGTACTCCAACA
cgr-miR-3072-3p	TGCCCCCTCCAGGAAGCCTTCT
cgr-miR-3072-5p	AGGGACCCCGAGGGAGGGCAGG
cgr-miR-3074	GTTCCTGCTGAACTGAGCCAGT
cgr-miR-30a-3p	CTTTCAGTCGGATGTTTGCAGC
cgr-miR-30a-5p	TGTAAACATCCTCGACTGGAAGC
cgr-miR-30b-3p	CTGGGATGTGGATGTTTACGTC
cgr-miR-30b-5p	TGTAAACATCCTACACTCAGCT
cgr-miR-30c	CTGGGAGAGGGTTGTTTACTCC
cgr-miR-30c-2	TGTAAACATCCTACACTCTCAGC
cgr-miR-30d	CTTTCAGTCAGATGTTTGCTGC
cgr-miR-30e-3p	CTTTCAGTCGGATGTTTACAGC
cgr-miR-30e-5p	TGTAAACATCCTTGACTGGAAGC

cgr-miR-31-3p	TGCTATGCCAACATATTGCCATC
cgr-miR-31-5p	AGGCAAGATGCTGGCATAGCTG
cgr-miR-3102	GAGCACCCCATTGGCTACCCACA
cgr-miR-32-3p	CAATTTAGTGTGTGTGATATT
cgr-miR-32-5p	TATTGCACATTACTAAGTTGC
cgr-miR-320a	AAAAGCTGGGTTGAGAGGGCGA
cgr-miR-322-3p	AAACATGAAGCGCTGCAACAC
cgr-miR-322-5p	CAGCAGCAATTCATGTTTTGG
cgr-miR-324-3p	CACTGCCCCAGGTGCTGCTGG
cgr-miR-324-5p	CGCATCCCCTAGGGCATTGGTG
cgr-miR-325-3p	TTTATTGAGCACCTCCTATCA
cgr-miR-325-5p	CCTAGTAGGTGCTCAGTAAGT
cgr-miR-326	CCTCTGGGCCCTTCCTCCAGT
cgr-miR-328	CTGGCCCTCTCTGCCCTTCCGT
cgr-miR-33	GTGCATTGTAGTTGCATTGC
cgr-miR-331-3p	GCCCCTGGGCCTATCCTAGA
cgr-miR-331-5p	CTAGGTATGGTCCCAGGGATC
cgr-miR-338	TCCAGCATCAGTGATTTTGTTGA
cgr-miR-339	TCCCTGTCCTCCAGGAGCTCACG
cgr-miR-340-3p	TCCGTCTCAGTTACTTTATAGCC
cgr-miR-340-5p	TTATAAAGCAATGAGACTGATT
cgr-miR-342-3p	TCTCACACAGAAATCGCACCCGTC
cgr-miR-342-5p	AGGGGTGCTATCTGTGATTGAGG
cgr-miR-344	TGATCTAGCCAAAGCCTGACTGT
cgr-miR-34a	TGGCAGTGTCTTAGCTGGTTGT
cgr-miR-34b-3p	AATCACTAACTCCACTGCCATC
cgr-miR-34b-5p	AGGCAGTGTAATTAGCTGATTGT
cgr-miR-34c-3p	AATCACTAACCACACGGCCAGG
cgr-miR-34c-5p	AGGCAGTGTAGTTAGCTGATTGC
cgr-miR-350-3p	TTCACAAAGCCCATACACTTTCAC
cgr-miR-350-5p	AAAGTGCATGCGCTTTGGGACA
cgr-miR-3535	TGGATATGATGACTGATTACCTGAGA
cgr-miR-361	TTATCAGAATCTCCAGGGGTAC
cgr-miR-362	AATCCTTGGAACCTAGGTGTGAGTGC
cgr-miR-365	AGGGACTTTCAGGGGCAGCTGTG
cgr-miR-365-3p	TAATGCCCCTAAAAATCCTTAT
cgr-miR-365-5p	AGGGACTTTTGGGGGCAGATGTG
cgr-miR-369-3p	AATAATACATGGTTGATCTTT
cgr-miR-369-5p	AGATCGACCGTGTTATATTCGC
cgr-miR-374-3p	CTTATCAGGTTGTATTATCATT
cgr-miR-374-5p	ATATAATACAACCTGCTAAGTG
cgr-miR-377-3p	TGAATCACACAAAGGCAACTTTT
cgr-miR-377-5p	AGAGGTTGCCCTTGGTGAATTCG
cgr-miR-378-3p	ACTGGACTTGGAGTCAGAAGGC
cgr-miR-378-5p	CTCCTGACTCCAGGTCCTGTGT
cgr-miR-379	TGGTAGACTATGGAACGTAGG
cgr-miR-381	TATACAAGGGCAAGCTCTCTGT
cgr-miR-382	GAAGTTGTTCGTGGTGGATTCG
cgr-miR-384-3p	ATTCCTAGAAATTGTTCACAAT
cgr-miR-384-5p	TGTAAACAATTCCTAGGCAATGT
cgr-miR-409-3p	GAATGTTGCTCGGTGAACCCCT
cgr-miR-409-5p	AGGTTACCCGAGCAACTTTGCATC
cgr-miR-410-3p	AATATAACACAGATGGCCTGT

cgr-miR-410-5p	AGGTTGTCTGTGATGAGTTCG
cgr-miR-412-3p	GTACTTCACCTGGTCCACTAGC
cgr-miR-412-5p	TGGTCGACCAGCTGGAAAGTAAT
cgr-miR-423-3p	AGCTCGGTCTGAGGCCCCTCAGT
cgr-miR-423-5p	TGAGGGGCAGAGAGCGAGACTTT
cgr-miR-425-3p	CATCGGGAATGTCGTGTCCGC
cgr-miR-425-5p	AATGACACGATCACTCCCGTTGA
cgr-miR-429	TAATACTGTCTGGTAATGCCGT
cgr-miR-449a	TGGCAGTGTATTGTTAGCTGGT
cgr-miR-450a	TTTTGCGATGTGTTCCTAATAT
cgr-miR-450b-3p	TATTGGGAACATTTTGCATATG
cgr-miR-450b-5p	TTTTGCAGTATGTTCCTGAATA
cgr-miR-451a	AAACCGTTACCATTACTGAGTT
cgr-miR-455-3p	GCAGTCCACGGGCATATACACT
cgr-miR-455-5p	TATGTGCCTTTGGACTACATCG
cgr-miR-484	TCAGGCTCAGTCCCCTCCCGAT
cgr-miR-486-3p	CGGGGCAGCTCAGTACAAGACG
cgr-miR-486-5p	TCCTGTACTGAGCTGCCCCGAG
cgr-miR-496	AGTATTACATGGCCAATCTCC
cgr-miR-497-3p	CAAACCACACTGTGGTGTTAGA
cgr-miR-497-5p	CAGCAGCACACTGTGGTTTGTA
cgr-miR-497b	CACCACAGTGTGGTTTGGACGTGG
cgr-miR-499-3p	GAACATCACAGCAAGTCTGTGC
cgr-miR-499-5p	TTAAGACTTGCAGTGATGTTTA
cgr-miR-501-5p	AATCCTTTGTCCCTGGGTGGAAATGC
cgr-miR-503	AGTATTGTTTCCGCTGCTCGG
cgr-miR-504	AGACCCTGGTCTGCACCTCTATC
cgr-miR-505-3p	GTCAACACTTGCTGGTTTCCTCT
cgr-miR-505-5p	GGGAGCCAGGAAGTATTGATGTT
cgr-miR-532-3p	CCTCCCACACCCAAGGTTTGC
cgr-miR-532-5p	CATGCCTTGAGTGTAGGACCGT
cgr-miR-542-3p	TGTGACAGATTGATAACTGAAAG
cgr-miR-542-5p	CTCGGGGATCATCATGTCACGA
cgr-miR-574	CACGCTCATGCACACACCCACA
cgr-miR-582	TAACCTGTTGAACAACTGAACCC
cgr-miR-598	TACGTCATCGTCGTCATCGTTATC
cgr-miR-6091	TCCTGCGTCGACTTCTCAGGAGA
cgr-miR-6092	TGGACAAGGCTGTGCTCTGAG
cgr-miR-615-3p	TCCGAGCCTGGGTCTCCCTCTT
cgr-miR-615-5p	GGGGGTCCCCGGTGCTCGGATC
cgr-miR-628	ATGCTGACATATTTACTAGAGG
cgr-miR-652-3p	AATGGCGCCACTAGGGTTGTG
cgr-miR-652-5p	CAACCCTAGGAGAGGGTGCCATTCA
cgr-miR-653	TTCACTGGAGTTTGTTTCAATA
cgr-miR-664-3p	TATTCATTTACTCCCCAGCCTAC
cgr-miR-664-5p	CTGGCTGGGGAAAATGATTGGA
cgr-miR-671-3p	TCCGGTTCTCAGGGCTCCACC
cgr-miR-671-5p	AGGAAGCCCTGGAGGGGCTGGAGG
cgr-miR-672	AGGTTGGTGTACTGTGTGTGA
cgr-miR-674	CACAGCTCCCATCTCAGAAC
cgr-miR-702	GTGAGTGGGGTGGTTGGCATG
cgr-miR-708	AAGGAGCTTACAATCTAGCTGG
cgr-miR-744-3p	CTGTTGCCACTAACCTCAACC
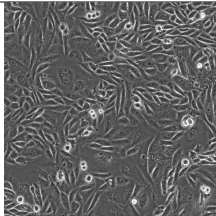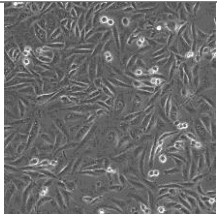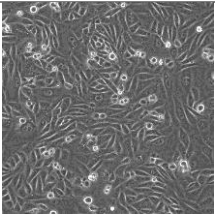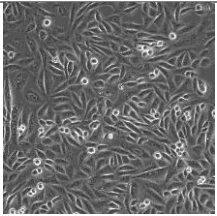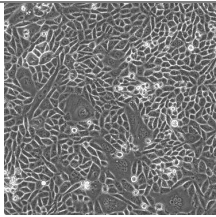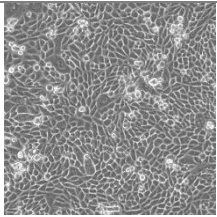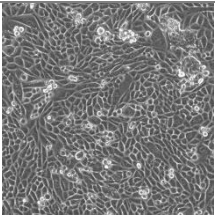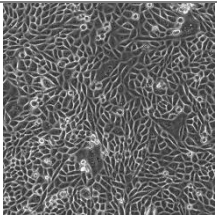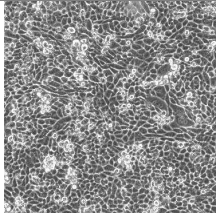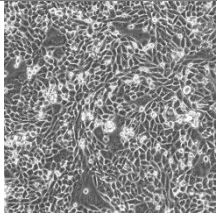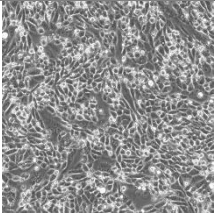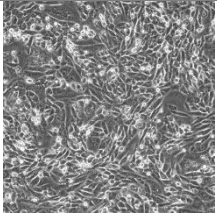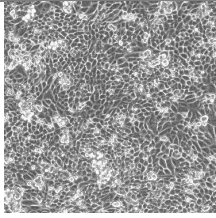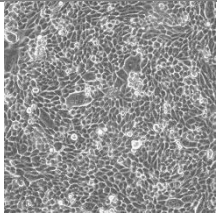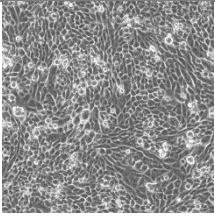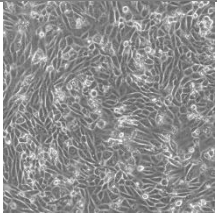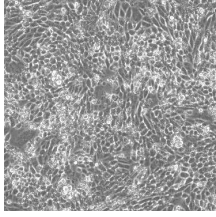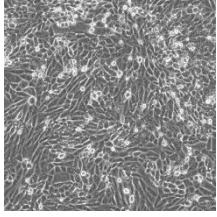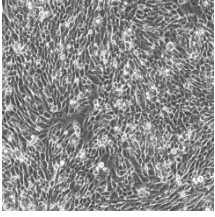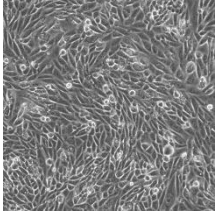
cgr-miR-744-5p	TGCGGGGCTAGGGCTAACAGC
cgr-miR-7a	TGGAAGACTAGTGATTTTGTTGT
cgr-miR-7b	TGGAAGACTTGTGATTTTGTTGTT
cgr-miR-802	TCAGTAACAAAGATTCATCCT
cgr-miR-872-3p	TGAACTATTGCAGTAGCCTCC
cgr-miR-872-5p	AAGGTTACTTGTTAGTTCAGG
cgr-miR-874	CTGCCCTGGCCCGAGGGACCGAC
cgr-miR-9	TCTTTGGTTATCTAGCTGTATG
cgr-miR-92a-3p	TATTGCACTTGTCCCGGCCTGT
cgr-miR-92a-5p	AGGCTGGGATTTGTTGCAATGCT
cgr-miR-92b-3p	TATTGCACTCGTCCCGGCCTCC
cgr-miR-92b-5p	AGGGACGGGACGCGGTGCAGTGTT
cgr-miR-93-3p	ACTGCTGAGCTAGCACTTCCCGA
cgr-miR-93-5p	CAAAGTGCTGTTCGTGCAGGTAG
cgr-miR-98	TGAGGTAGTAAGTTGTATTGTT
cgr-miR-99a-3p	CAAGCTCGCTTCTATGGGTCTG
cgr-miR-99a-5p	AACCCGTAGATCCGATCTTGT

Table A.1 CHO-K1 mature miRNA transcriptome with sequences from blastn-short results

Appendix B – Cell Culture Images for RNA Isolation Samples

B.1 Reo3 Infection

Day	Healthy	Reo3 0.1 TCID$_{50}$	Reo3 1 TCID$_{50}$	Reo3 100 TCID$_{50}$
1				
2				
3				
4				
5				

| 6 | | | | |
| 7 | | | | |
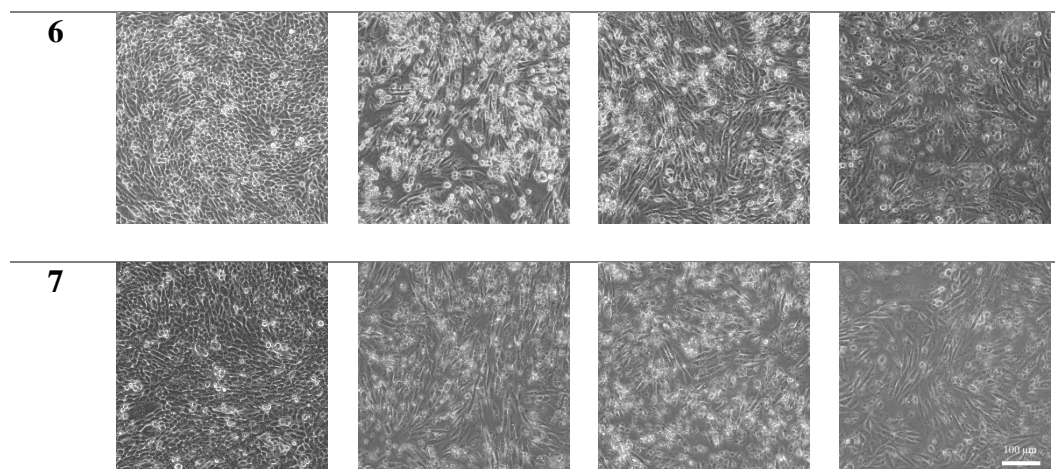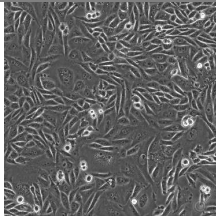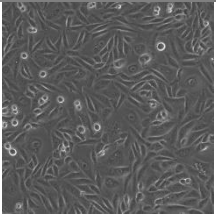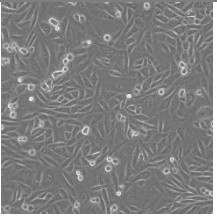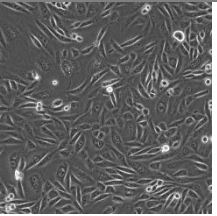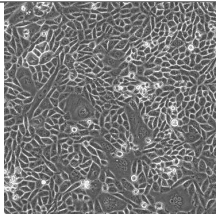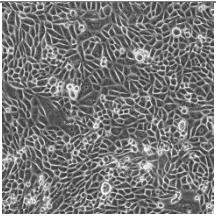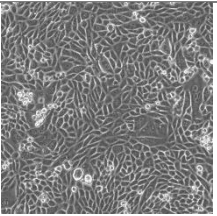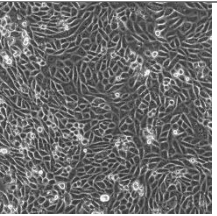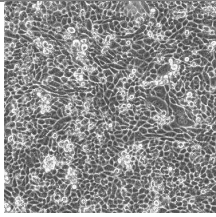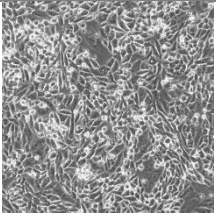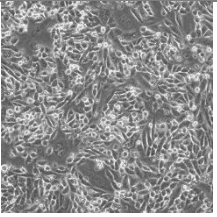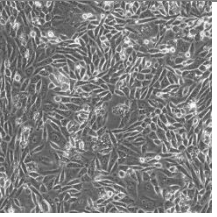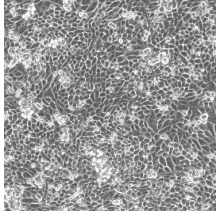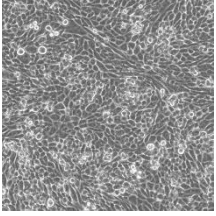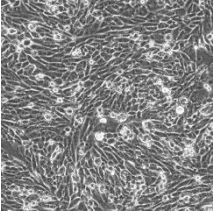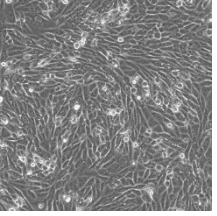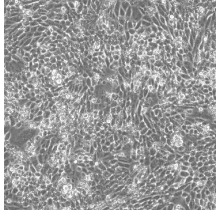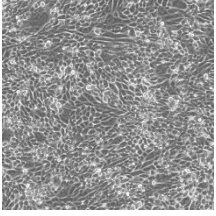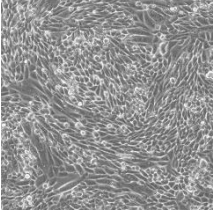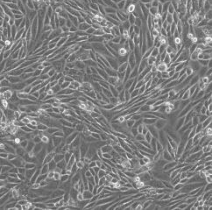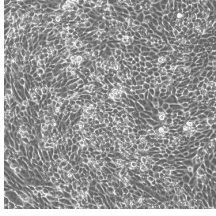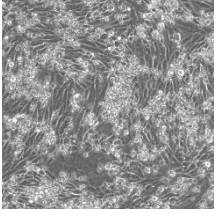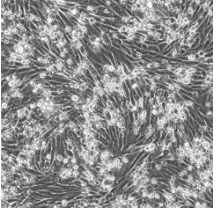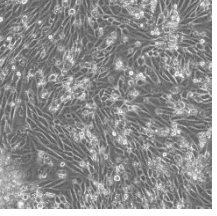
Table B.1 Representative cell culture images of Reo3 infeted samples, days 1 to 7 post-infection, used for RNA isolation and comprehensive CHO-K1 miRNA biomarker screening. Images taken on phase contrast setting at 20x magnification with a Keyance BZ-X series instrument. Scale bar approximate.
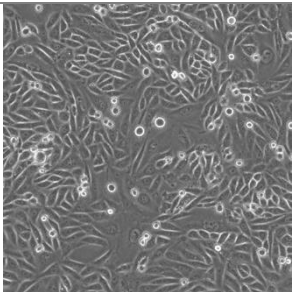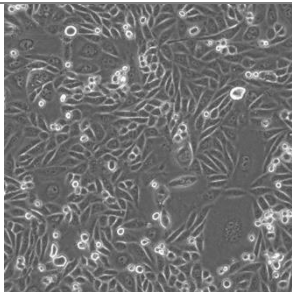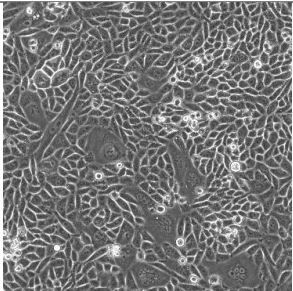
B.2 MMV Infection

Day	Healthy	MMV 0.1 TCID$_{50}$	MMV 1 TCID$_{50}$	MMV 100 TCID$_{50}$
1				
2				
3				
4				
5				
6				

7
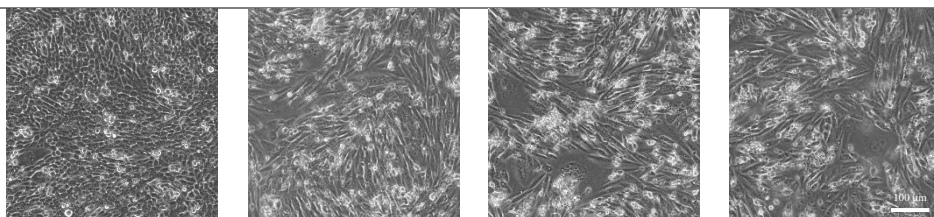
Table B.2 Representative cell culture images of MMV infeted samples, days 1 to 7 post-infection, used for RNA isolation and comprehensive CHO-K1 miRNA biomarker screening. Images taken on phase contrast setting at 20x magnification with a Keyance BZ-X series instrument. Scale bar approximate.

B.3 EMC Infection
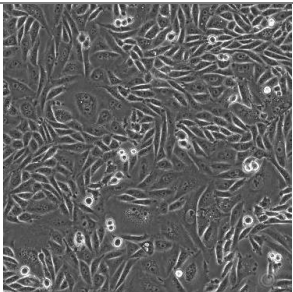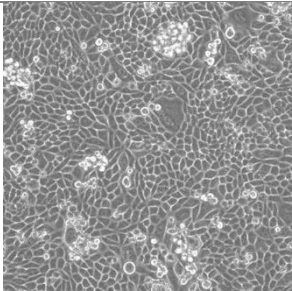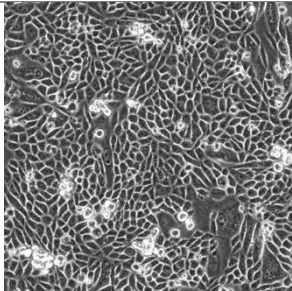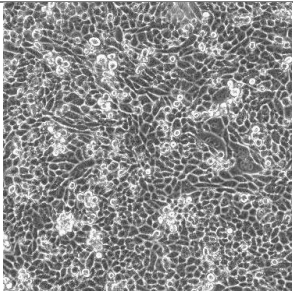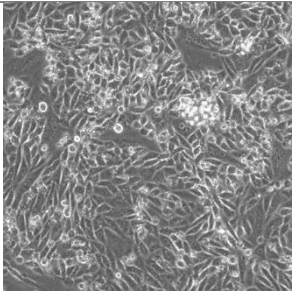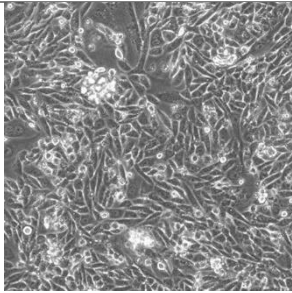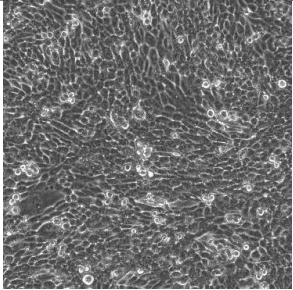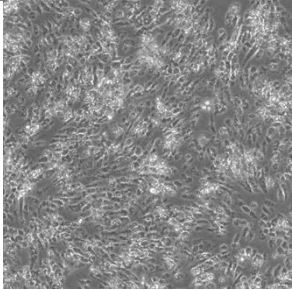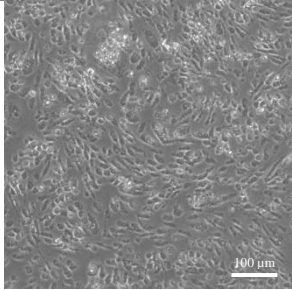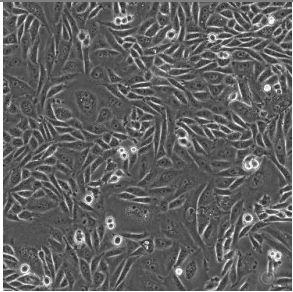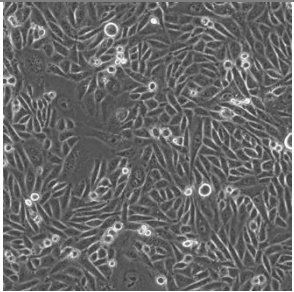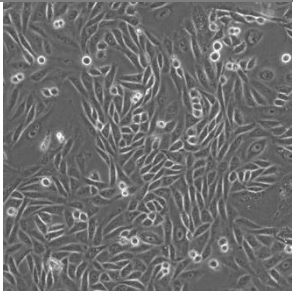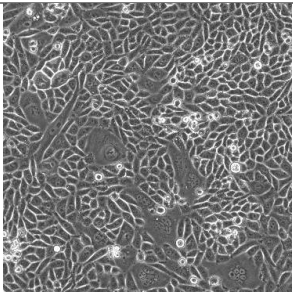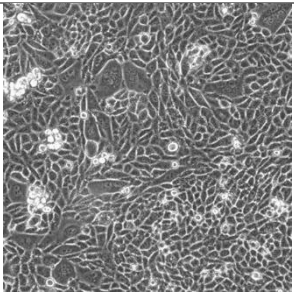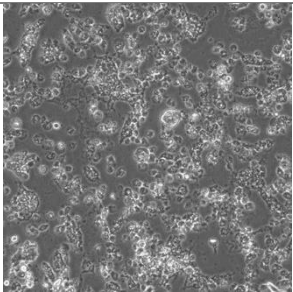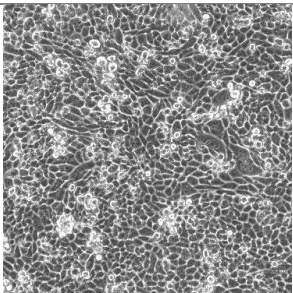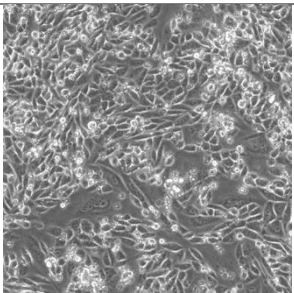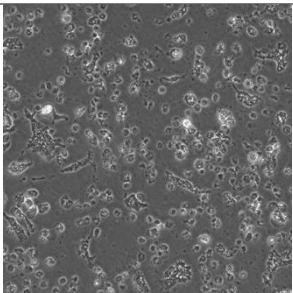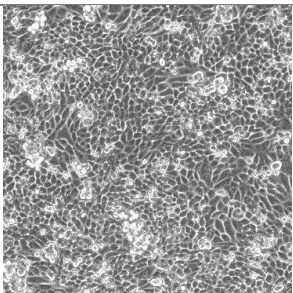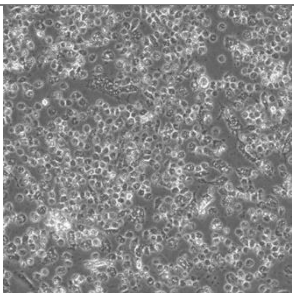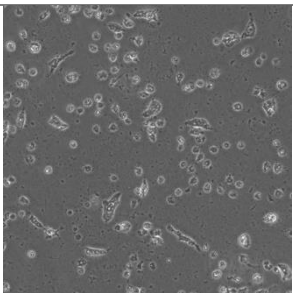
Day	Healthy	EMC 1 TCID$_{50}$	EMC 100 TCID$_{50}$
1			
2			
3			
7			

Table B.3 Representative cell culture images of EMC infeted samples, days 1-3 and 7 post-infection, used for RNA isolation and comprehensive CHO-K1 miRNA biomarker screening. Images taken on phase contrast setting at 20x magnification with a Keyance BZ-X series instrument. Scale bar approximate.

B.4 VSV Infection

Day	Healthy	VSV 1 TCID$_{50}$	VSV 100 TCID$_{50}$
1			
2			
3			
4			
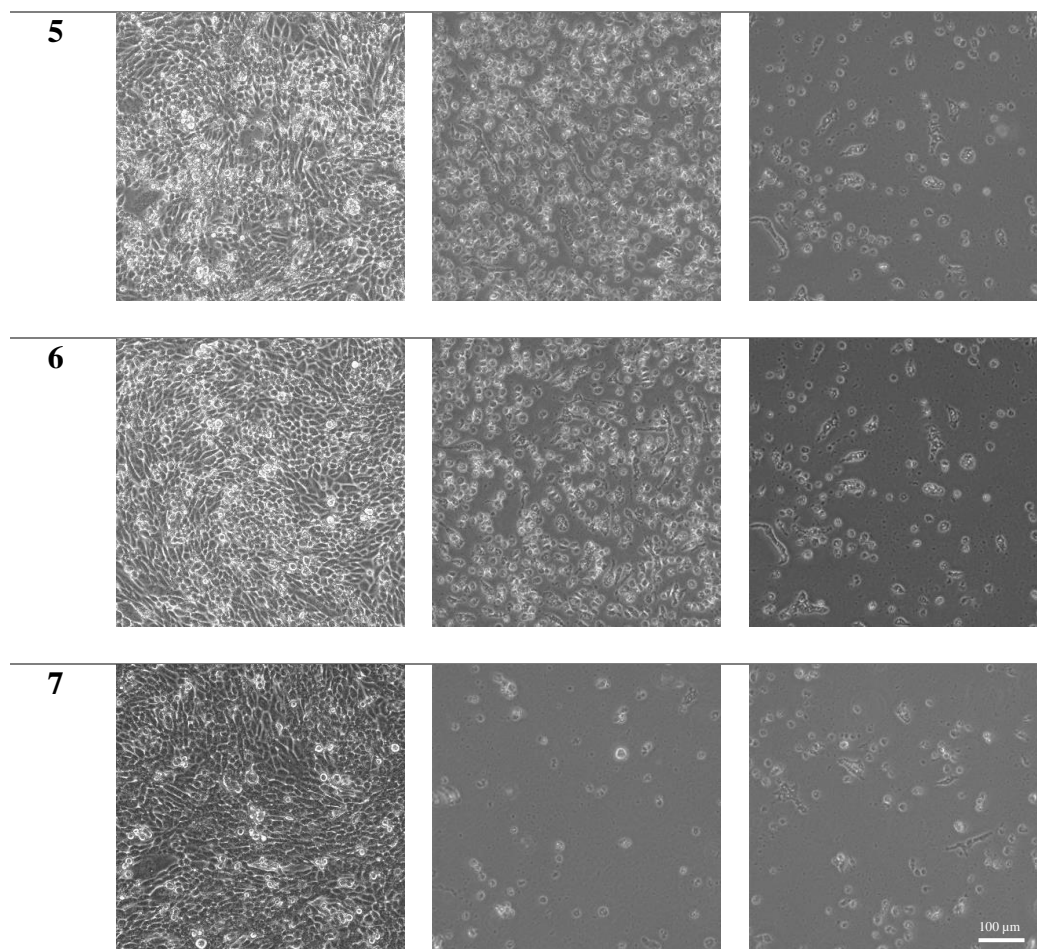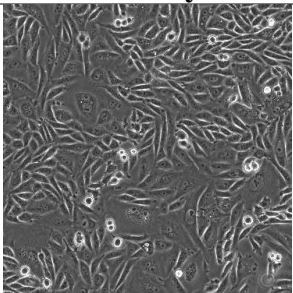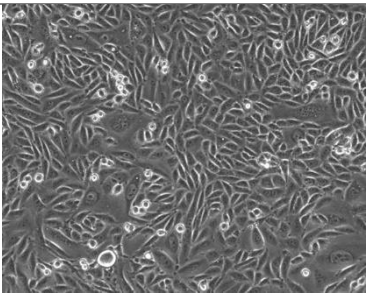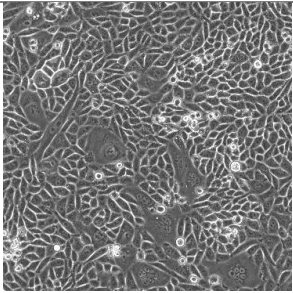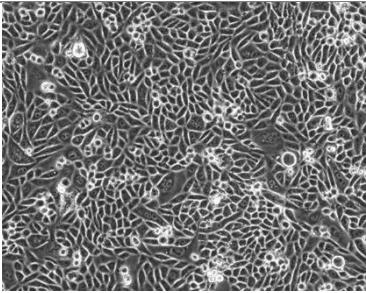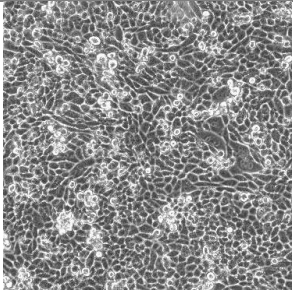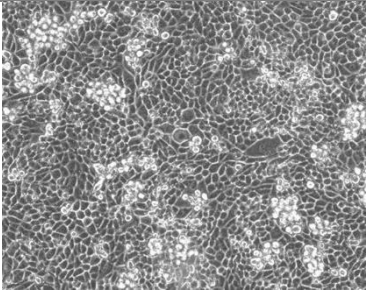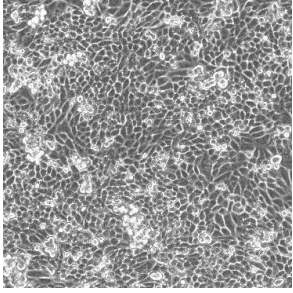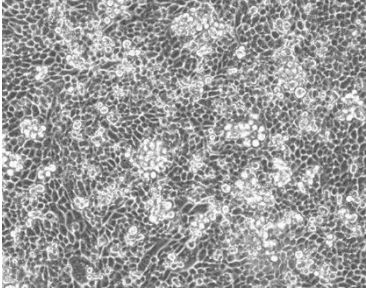
Table B.4 Representative cell culture images of VSV infeted samples, days 1 to 7 post-infection, used for RNA isolation and comprehensive CHO-K1 miRNA biomarker screening. Images taken on phase contrast setting at 20x magnification with a Keyance BZ-X series instrument. Scale bar approximate.

B.5 PI2 Inoculation

Day	Healthy	PI2
1		
2		
3		
4		
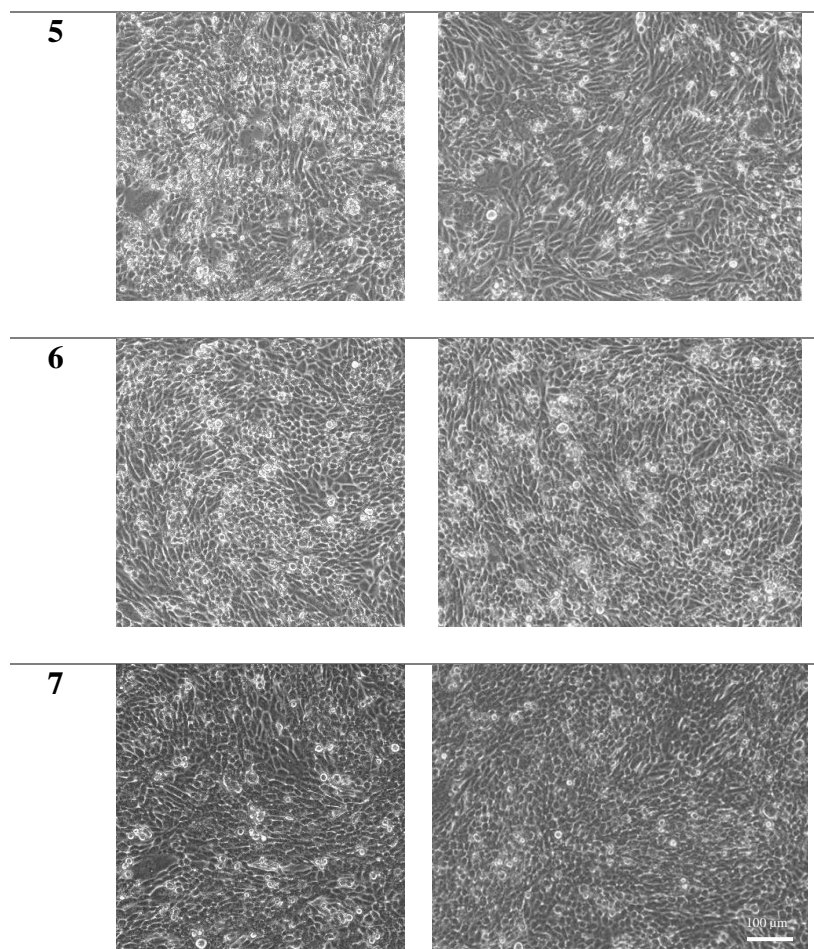
Table B.5 Representative cell culture images of PI2 inlculated samples, days 1 to 7 post-infection, used for RNA isolation and comprehensive CHO-K1 miRNA biomarker screening. Images taken on phase contrast setting at 20x magnification with a Keyance BZ-X series instrument. Scale bar approximate.

Appendix C – Receiver Operating Curve (ROC) Analysis for General Biomarker Screening

C.1 MATLAB Code for ROC Analysis

```matlab
% ROC_analysis.m
%
% Created by Virginia Lane DiNenna
% -------------------------------------------------------------------------

% Clear previous variables
clear
clc

% Load normalized count data
load('GeneralNormCounts.mat');

% Parameters
% Choose infecting virus and miRNA
virus = "all";
    % options: Reo3, MMV, EMC, VSV, all
miRNA = "all";
    % for miRNA options, use GeneralNormCounts.miRNA command
    % to calculate area under the curve (AUC) for all, miRNA = "all"
AUC_matrix = [];
    % Creates a blank matrix to be filled with AUC values
iterations = 101;
    % iterations are the number of calculated values for Sp and Sn, evenly
    % dispersed between the minimum and maximum normalized count values

% Set which columns in the normalized count file are associated with which
% virus
if virus == "Reo3"
    columns = GeneralNormCounts(:,11:37);
elseif virus == "MMV"
    columns = GeneralNormCounts(:,38:64);
elseif virus == "EMC"
    columns = GeneralNormCounts(:,65:76);
elseif virus == "VSV"
    columns = GeneralNormCounts(:,77:85);
elseif virus == "all"
    columns = GeneralNormCounts(:,11:85);
else
    fprintf('error in virus parameter')
end

% Rank miRNAs by AUC value if miRNA = "all"
if miRNA == "all"
    for i = 1:size(GeneralNormCounts.miRNA(:,1))
    % "i" is the number of miRNAs in the normalized count data
    % this for loop goes through each miRNA
    Progress = i % just to monitor how long the code is taking to run

    ROC_matrix = [];
    % Creates a blank matrix to be filled with sensitivity (Sn) and
    % specificity (Sp) values
```

```matlab
% maximum count value
max_count= max(max(GeneralNormCounts{i,2:85}));
j = linspace(0,max_count,iterations);
% creates evenly spaced threshold values

for k = 1:iterations
% "k" is the number of iterations (how many Sp and Sn values are
% calculated per miRNA)
% this for loop calculates Sp and Sn for each miRNA at each normalized
% count threshold "j(k)"

    infected = table2array(columns(i,:)) <= j(k);
    % # of samples that are classifisied infected for "i" miRNA at
    % threshold j(k)
    TP = sum(infected);
    % # of samples that are true positives at threshold j(k)
    FN = size(infected,2)-TP;
    % # of samples that are false negatives at threshold j(k)

    uninfected = table2array(GeneralNormCounts(i,2:10)) > j(k);
    % # of samples that are classifisied infected for "i" miRNA at
    % threshold j(k)
    FP = sum(uninfected);
    % # of samples that are false positives at threshold j(k)
    TN = size(uninfected,2)-FP;
    % # of samples that are true negatives at threshold j(k)

    % Calculations
    TPR = TP/(TP+FN); % Calculates true positive rate, sensitivity Sn
    FPR = FP/(FP+TN); % Calculates false positive rate, specificity Sp
    TNR = 1-FPR; % Calculates true negative rate, 1-Sp

    ROC_matrix = [ROC_matrix; j(k) TPR FPR TNR];
    % This fills the matrix with the following columns:
    % [Threshold, Sensitivity, Specificity, 1-Specificity]

end

[~,idx] = unique(ROC_matrix(:,(2:4)),'stable','rows');
ROC_unique = ROC_matrix(idx,:);
% only keeps unique rows in the ROC_matrix

% Calculate area under curve (AUC) for "i" miRNA
AUC = trapz(ROC_unique(:,4),ROC_unique(:,2));

% If AUC < 0.5, then miRNA is upregulated instead of downregulated.
% The signs for calculation of "infected" and "uninfected" should be
% switched, however subtracting the AUC from 1 will give the same value
if AUC < 0.5
    AUC = 1-AUC;
end

AUC_matrix = [AUC_matrix;...
    GeneralNormCounts.miRNA(i,1) AUC];
% Fills a matrix with all the AUC values, miRNA name in first column

end
AUC_matrix = sortrows(AUC_matrix,2,'descend')
% Sorts the AUC_matrix by AUC value and prints
```

```matlab
% Plots receiver operating curve for a specific miRNA (chosen above)
% if miRNA =/= "all"
else
    ROC_matrix = [];
    i = find(GeneralNormCounts.miRNA==miRNA);
    % This finds which row the parameter miRNA is in and labels it as "i"

    % Previously explained calculations...
    max_count = max(max(GeneralNormCounts{i,2:85}));
    j = linspace(0,max_count,iterations);

    for k = 1:iterations
        infected = table2array(columns(i,:)) <= j(k);
        TP = sum(infected);
        FN = size(infected,2)-TP;
        uninfected = table2array(GeneralNormCounts(i,2:10)) > j(k);
        FP = sum(uninfected);
        TN = size(uninfected,2)-FP;

        TPR = TP/(TP+FN);
        FPR = FP/(FP+TN);
        TNR = 1-FPR;

        ROC_matrix = [ROC_matrix; j(k) TPR FPR TNR];
    end

    [~,idx] = unique(ROC_matrix(:,(2:4)),'stable','rows');
    ROC_unique = ROC_matrix(idx,:);
    AUC = trapz(ROC_unique(:,4),ROC_unique(:,2))

    % Plot receiver operating curve
    figure(2)
    if AUC >= 0.5
        plot(ROC_unique(:,4),ROC_unique(:,2),'linewidth',1)
    else
    % If AUC < 0.5, then miRNA is upregulated instead of downregulated.
    % The signs for calculation of "infected" and "uninfected" should be
    % switched, however subtracting the AUC from 1 will give the same value
    % Additionally, the actual TPR is 1-TPR and TNR is 1-TNR
        AUC2 = 1-AUC
        plot(1-ROC_unique(:,4),1-ROC_unique(:,2),'linewidth',1)
    end

    % plot customization and labels
    ylabel('True Positive Rate (S_n)')
    xlabel('False Positive Rate (1-S_p)')
    xlim([-.05 1.05])
    ylim([-.05 1.05])
    title('Receiver Operating Curve')

    % plot the non-discrimination line
    hold on
    plot([0,1],[0,1],'--','color','r','linewidth',1)
    hold off
    legend(miRNA,'non-discrimination','location','southeast')
end
```

C.2 Ranked Area Under the Curve (AUC) Values for CHO-K1 miRNAs (general biomarkers)

CHO-K1 miRNA	AUC
cgr-miR-1224	1.000
cgr-miR-128	1.000
cgr-miR-1956	1.000
cgr-miR-216b	1.000
cgr-miR-500	1.000
cgr-miR-501-5p	1.000
cgr-miR-501-3p	1.000
cgr-miR-190a	0.953
cgr-miR-615-5p	0.939
cgr-miR-142-3p	0.935
cgr-miR-532-3p	0.935
cgr-miR-26b-5p	0.929
cgr-miR-29b-3p	0.926
cgr-miR-301a-3p	0.925
cgr-miR-9	0.924
cgr-miR-542-3p	0.922
cgr-miR-3154	0.919
cgr-miR-1306-5p	0.910
cgr-miR-192	0.907
cgr-miR-125b-5p_(2)	0.905
cgr-miR-101b-3p	0.902
cgr-miR-101b-3p_(2)	0.902
cgr-miR-134	0.902
cgr-miR-33	0.901
cgr-miR-21-3p	0.900
cgr-miR-455-5p	0.896
cgr-miR-31-3p	0.896
cgr-miR-381	0.895
cgr-miR-1903	0.894
cgr-miR-127	0.893
cgr-miR-196a-5p	0.890
cgr-miR-17-5p	0.889
cgr-miR-24-5p	0.888
cgr-miR-32-5p	0.887
cgr-miR-499-5p	0.884
cgr-miR-384-3p	0.882
cgr-let-7c-2_(2)	0.880
cgr-miR-301b	0.876
cgr-miR-196b	0.873
cgr-miR-378-5p	0.873
cgr-miR-98	0.872
cgr-miR-18a-5p	0.871
cgr-miR-30e-5p	0.867
cgr-miR-450b-5p	0.864
cgr-miR-423-5p	0.855
cgr-miR-106b-5p	0.854
cgr-miR-148b-3p	0.854

cgr-miR-19a	0.852
cgr-miR-674	0.851
cgr-miR-30d	0.849
cgr-miR-3074	0.843
cgr-miR-3074_(2)	0.843
cgr-miR-181d-3p	0.842
cgr-let-7i	0.841
cgr-miR-190b	0.839
cgr-miR-325-5p	0.839
cgr-miR-664-5p	0.838
cgr-miR-450b-3p	0.837
cgr-miR-18b	0.836
cgr-miR-19b-3p	0.834
cgr-miR-760	0.834
cgr-miR-340-5p	0.833
cgr-miR-199b	0.827
cgr-miR-450a	0.825
cgr-miR-139-5p	0.824
cgr-miR-125a-5p	0.824
cgr-miR-365-5p	0.819
cgr-miR-30b-5p	0.818
cgr-let-7f	0.817
cgr-miR-29c-3p	0.816
cgr-miR-412-3p	0.815
cgr-miR-3535	0.813
cgr-miR-130a-3p	0.810
cgr-miR-30a-5p	0.810
cgr-miR-152-3p	0.809
cgr-miR-369-3p	0.806
cgr-miR-139-3p	0.804
cgr-miR-21-5p	0.803
cgr-miR-300	0.803
cgr-miR-671-3p	0.802
cgr-let-7g-5p	0.801
cgr-miR-15a-5p	0.799
cgr-miR-296	0.798
cgr-miR-124	0.796
cgr-miR-125b-5p	0.793
cgr-miR-130b-3p	0.793
cgr-miR-215-5p	0.790
cgr-miR-1249	0.788
cgr-miR-92b-5p	0.784
cgr-miR-328	0.783
cgr-miR-141	0.781
cgr-miR-143	0.781
cgr-miR-34b-5p	0.779
cgr-miR-24-3p_(2)	0.776
cgr-miR-30c-2	0.774
cgr-miR-204	0.772
cgr-miR-107	0.771
cgr-miR-27b-3p	0.770
cgr-miR-429	0.769
cgr-miR-199a-5p	0.767

cgr-miR-320a	0.765
cgr-miR-653	0.764
cgr-let-7d-3p	0.763
cgr-miR-1285	0.763
cgr-miR-872-5p	0.763
cgr-miR-339	0.762
cgr-miR-140-5p	0.760
cgr-miR-151-5p	0.760
cgr-miR-324-3p	0.759
cgr-miR-147	0.756
cgr-miR-210-3p	0.756
cgr-miR-27a-3p	0.756
cgr-miR-615-3p	0.755
cgr-miR-106b-3p	0.754
cgr-miR-652-5p	0.754
cgr-miR-193b-5p	0.752
cgr-miR-423-3p	0.752
cgr-miR-342-3p	0.751
cgr-miR-377-5p	0.750
cgr-miR-154-3p	0.748
cgr-miR-505-5p	0.745
cgr-miR-374-3p	0.742
cgr-miR-6092	0.740
cgr-miR-191-5p	0.739
cgr-miR-30c-2_(2)	0.739
cgr-miR-497-5p	0.739
cgr-miR-182	0.737
cgr-miR-200c	0.737
cgr-miR-1b	0.735
cgr-miR-497-3p	0.735
cgr-miR-92a-5p	0.735
cgr-miR-342-5p	0.734
cgr-miR-298-5p	0.733
cgr-miR-374-5p	0.733
cgr-miR-598	0.732
cgr-miR-186-5p	0.730
cgr-miR-214-5p	0.729
cgr-let-7c-2	0.728
cgr-miR-125a-3p	0.727
cgr-miR-7b	0.727
cgr-miR-34a	0.724
cgr-miR-99b	0.724
cgr-miR-7a	0.722
cgr-miR-27a-5p	0.721
cgr-miR-1343	0.721
cgr-miR-142-5p	0.721
cgr-miR-582	0.721
cgr-miR-361	0.719
cgr-miR-18a-3p	0.718
cgr-miR-31-5p	0.717
cgr-miR-377-3p	0.717
cgr-miR-298-3p	0.716
cgr-miR-30e-3p	0.716

cgr-miR-874	0.716
cgr-miR-664-3p	0.715
cgr-miR-148a	0.714
cgr-miR-148b-5p	0.713
cgr-miR-103-3p	0.713
cgr-miR-218a	0.712
cgr-miR-193a	0.709
cgr-miR-93-5p	0.709
cgr-miR-365-3p_(2)	0.708
cgr-miR-322-5p	0.707
cgr-miR-365-3p	0.707
cgr-miR-484	0.706
cgr-miR-3072-5p	0.704
cgr-miR-137-3p	0.698
cgr-miR-345	0.695
cgr-miR-150	0.692
cgr-miR-30a-3p	0.692
cgr-miR-455-3p	0.692
cgr-miR-504	0.691
cgr-miR-632	0.691
cgr-miR-202	0.690
cgr-miR-326	0.689
cgr-miR-15a-3p	0.687
cgr-miR-194-2	0.687
cgr-miR-221-5p	0.687
cgr-miR-24-3p	0.686
cgr-miR-425-3p	0.686
cgr-miR-140-3p	0.685
cgr-miR-99a-3p	0.684
cgr-miR-130b-5p	0.684
cgr-miR-199a	0.684
cgr-miR-22-3p	0.684
cgr-miR-382	0.683
cgr-miR-154-5p	0.682
cgr-miR-205	0.681
cgr-miR-188	0.680
cgr-miR-30c	0.680
cgr-miR-200a	0.679
cgr-miR-34b-3p	0.679
cgr-miR-384-5p	0.679
cgr-miR-451a	0.676
cgr-miR-802	0.676
cgr-miR-196c	0.672
cgr-miR-3102	0.671
cgr-miR-132-5p	0.669
cgr-miR-128-3p	0.668
cgr-miR-152-5p	0.668
cgr-miR-151-3p	0.667
cgr-miR-92a-3p	0.667
cgr-miR-28-3p	0.667
cgr-miR-1260	0.664
cgr-miR-181c-3p	0.664
cgr-miR-137-5p	0.663

cgr-miR-146b-5p	0.663
cgr-miR-744-3p	0.663
cgr-miR-32-3p	0.662
cgr-miR-409-5p	0.660
cgr-miR-338	0.659
cgr-miR-16-3p	0.658
cgr-miR-872-3p	0.658
cgr-miR-210-5p	0.657
cgr-miR-486-5p	0.653
cgr-miR-744-5p	0.651
cgr-miR-496	0.648
cgr-miR-542-5p	0.647
cgr-miR-17-3p	0.646
cgr-miR-344	0.646
cgr-miR-103-5p	0.642
cgr-miR-350-3p	0.642
cgr-miR-183	0.641
cgr-miR-185-3p	0.640
cgr-miR-195	0.640
cgr-miR-149-5p	0.633
cgr-miR-199a-3p	0.633
cgr-miR-410-5p	0.633
cgr-miR-708	0.633
cgr-let-7b	0.632
cgr-miR-222-3p	0.632
cgr-miR-184	0.631
cgr-miR-1949	0.630
cgr-miR-128-5p	0.628
cgr-miR-222-5p	0.628
cgr-miR-199a-3p_(2)	0.627
cgr-miR-206	0.627
cgr-miR-34c-3p	0.627
cgr-miR-129	0.625
cgr-miR-126b	0.624
cgr-miR-181a-3p	0.624
cgr-miR-219b	0.624
cgr-miR-217	0.623
cgr-miR-196a-3p	0.621
cgr-miR-10a-3p	0.620
cgr-miR-505-3p	0.620
cgr-let-7c	0.615
cgr-let-7a	0.612
cgr-miR-672	0.612
cgr-miR-125b-3p	0.610
cgr-miR-15b-3p	0.610
cgr-miR-212	0.610
cgr-miR-30b-3p	0.609
cgr-miR-425-5p	0.604
cgr-miR-15b-5p	0.602
cgr-miR-146b-3p	0.600
cgr-miR-574	0.600
cgr-miR-350-5p	0.598
cgr-miR-1843	0.597

cgr-miR-34c-5p	0.596
cgr-miR-628	0.596
cgr-miR-186-3p	0.593
cgr-miR-1973	0.593
cgr-miR-23b-5p	0.593
cgr-miR-340-3p	0.593
cgr-miR-93-3p	0.593
cgr-miR-410-3p	0.592
cgr-miR-181a-5p	0.591
cgr-miR-331-3p	0.591
cgr-let-7d-5p	0.590
cgr-miR-181d-5p	0.590
cgr-miR-331-5p	0.589
cgr-miR-20b	0.587
cgr-miR-130a-5p	0.587
cgr-miR-24	0.587
cgr-miR-25-5p	0.587
cgr-miR-1271	0.585
cgr-miR-365	0.582
cgr-let-7a-2	0.580
cgr-miR-101a	0.580
cgr-miR-101b-5p	0.580
cgr-miR-125b	0.580
cgr-miR-16	0.580
cgr-miR-19b-5p	0.580
cgr-miR-215-3p	0.580
cgr-miR-23a-5p	0.580
cgr-miR-26a	0.580
cgr-miR-26b-3p	0.580
cgr-miR-27b-5p	0.580
cgr-miR-330	0.580
cgr-miR-193b-3p	0.579
cgr-miR-499-3p	0.579
cgr-miR-1306-3p	0.576
cgr-miR-652-3p	0.576
cgr-miR-211	0.575
cgr-miR-10b-3p	0.573
cgr-miR-146a	0.571
cgr-miR-185-5p	0.571
cgr-miR-2424	0.570
cgr-miR-181c-5p	0.570
cgr-miR-149-3p	0.567
cgr-miR-379	0.566
cgr-miR-144	0.564
cgr-miR-221-3p	0.563
cgr-miR-194_(2)	0.562
cgr-miR-26a-2_(2)	0.562
cgr-miR-532-5p	0.561
cgr-miR-702	0.559
cgr-miR-6091	0.558
cgr-miR-214-3p	0.553
cgr-miR-29a-5p	0.553
cgr-miR-29a-3p	0.553

cgr-miR-100-3p	0.550
cgr-let-7a_(2)	0.550
cgr-miR-671-5p	0.550
cgr-miR-100-5p	0.548
cgr-miR-3068	0.548
cgr-miR-26a-2	0.547
cgr-miR-29b-5p	0.547
cgr-miR-29c-5p	0.547
cgr-miR-28-5p	0.544
cgr-miR-191-3p	0.544
cgr-miR-132-3p	0.542
cgr-miR-20a	0.542
cgr-miR-138	0.541
cgr-let-7g-3p	0.540
cgr-miR-3072-3p	0.539
cgr-miR-1839-3p	0.537
cgr-miR-412-5p	0.537
cgr-miR-10b-5p	0.535
cgr-miR-29b	0.533
cgr-miR-409-3p	0.533
cgr-miR-486-3p	0.533
cgr-miR-503	0.533
cgr-miR-181b-5p	0.530
cgr-miR-22-5p	0.529
cgr-miR-497b	0.528
cgr-miR-155	0.527
cgr-miR-325-3p	0.527
cgr-miR-322-3p	0.526
cgr-let-7e	0.525
cgr-miR-99a-5p	0.524
cgr-miR-200b	0.522
cgr-miR-10a-5p	0.521
cgr-miR-194	0.521
cgr-miR-369-5p	0.520
cgr-miR-187	0.519
cgr-miR-23a-3p	0.519
cgr-miR-1	0.518
cgr-miR-23b-3p	0.517
cgr-miR-92b-3p	0.516
cgr-miR-126a	0.511
cgr-miR-16-5p_(2)	0.511
cgr-miR-449a	0.510
cgr-miR-1839-5p	0.507
cgr-miR-362	0.505
cgr-miR-122	0.504
cgr-miR-181b-3p	0.504
cgr-miR-378-3p	0.504
cgr-miR-25-3p	0.503
cgr-miR-324-5p	0.501
cgr-miR-16-5p	0.500
cgr-miR-301a-5p	0.500

Table C.1 AUC values of CHO-K1 mature miRNAs from NGS comprehensive miRNA screening for general biomarkers. AUC values encompass Reo3, MMV, EMC, and VSV viruses at all time points and titer levels.