Analysis of miRNA and mRNA Environment in Transfected Cell Lines

Simon Blanchard

For many medical conditions caused by genetic mutations, researchers and physicians are well aware of the specific mutation behind the affliction. Additionally, technologies to repair these mutations are established, such as CRISPR-Cas9 gene editing. However, these techniques are far from 100% efficient, especially in cells that come directly from patients. This is because of a variety of cellular processes that protect DNA. One such process is regulation of expression via microRNA. This study aimed to identify microRNAs that are present in commonly used laboratory cell lines and determine their effect on the efficiency of the delivery of genetic engineering products. A bioinformatic analysis was conducted on sequencing data to compare microRNAs to known transfection data. Several microRNAs were identified as potential inhibitors of the effectiveness of genetic engineering techniques. These results could be applied to treatment plans, but future studies must first be conducted to assess how the addition of inhibitors of these microRNAs would impact the efficiency of genetic engineering techniques.

Background

The central dogma of biology is that DNA codes for messenger RNA (mRNA), which is then used to produce proteins (1). There are many types of RNA, but the main type involved in the production of proteins is mRNA, which is a transcript of the cell's genetic code. A second type is ribosomal RNA (rRNA), which comprises the ribosome that synthesizes proteins from mRNA sequences, and a third is transfer RNA (tRNA), which brings the amino acids that actually comprise the protein to the ribosome (1). This is one of the reasons that gene therapy, a therapeutic technique that involves altering the genetic code of the target, is such a promising field, as mutated genes and their products can be manipulated. Mutations in a patient's genetic code can lead to dysfunctional proteins and serious disorders, such as cystic fibrosis or sickle cell anemia, but gene therapy can be used to correct defective proteins that cause disease by replacing the mutated DNA, or by granting a cell new means to produce a desired protein (1). The new gene can be added into the cell through a variety of methods including lipid bundles or viral delivery, through a process known as transfection (1). In reality, the process is not this straightforward, as a variety of factors regulate its steps. Not all of the RNA transcribed from DNA is translated into protein due to post-transcriptional regulatory processes (2). One such method of regulation is miRNA, or microRNA. miRNAs are short strands of RNA, around 20 base pairs in length. When the seed region of a miRNA, which is nucleotides 2-8 from the 5' end, is complementary to the 3'UTR, which is the untranslated region on the 3' end of the strand of an mRNA, the miRNA binds to it and decreases its translation by accelerating its degradation (2). Furthermore, if the entire miRNA is complementary to the mRNA, it can completely silence that mRNA by signaling for cleavage of the mRNA strand (2). These two binding patterns are shown in Figure 1.

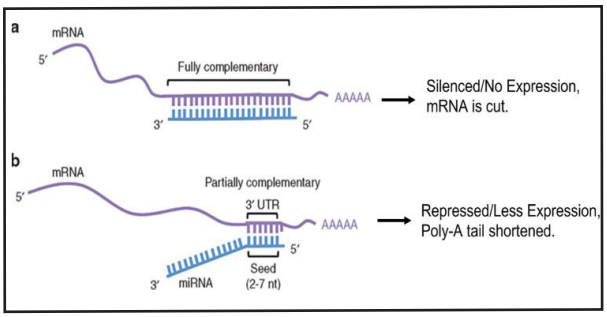


Figure 1. Binding of miRNA to an mRNA transcript demonstrating a) full complementarity and b) partial complementarity (3).

In animal cells, miRNA typically exhibit complementarity only in the seed region and thus reduce the expression of proteins (2). The goal of the project was to determine the interactions between miRNAs and the therapeutic genes (transgenes) used in gene therapy treatments. Initially, the small RNA content of several human cell lines was analyzed to identify any endogenous miRNAs that might interfere with transgene expression by binding to its 3' UTR. An additional goal was to identify any miRNAs that

were differentially expressed after transgenes were introduced into the cell, which would suggest that those miRNAs play a role in the innate immune response to foreign DNA. This was motivated by previous literature in which potentially inhibitory miRNAs were identified, but their function was never validated in cell lines (4).

Methodology

Sequencing data were obtained for the miRNA and mRNA present in several cell lines: human embryonic kidney cells, HEK293-T, prostate cancer cells, PC-3, and a breast cancer cell line, MCF7. These cells were chosen as they are three of the most common cell lines used in gene therapy research. Bioconda, a Python script designed for bioinformatic research, was loaded into a Linux environment to be used for analysis. First, the "chmod" command was used to make the RNA sequencing files executable. Next, the Burrows Wheeler Aligner (bwa) was executed using the "bwa" command to create an index of the reference sequence of the Green Fluorescent Protein (pEF-GFP or GFP) plasmid, which is added to provide a method to confirm successful transfection by the presence of green fluorescence, as well as a method to measure the transfection efficiency based off of the brightness of the fluorescence. Next, the miRNA sequencing data was aligned to the index using the same tool (5). The resulting Sequence Alignment Map (SAM) was used to sort the alignments based on their position in the GFP sequence using samtool's "sort" command, and the "cut" function of samtools was then used to isolate the columns of interest, namely the position of the alignment, the number of alignments in each position, and the sequences of the alignments (5). This data was then ported to Excel for further sorting and analysis. The same procedure was repeated for the mRNA sequencing data, and for the sense and antisense strand of the GFP sequence.

Results and Discussion

The number of miRNA sequences that aligned to the pEF-GFP plasmid sequence in each trial are summarized in Figure 2.

In general, the number of miRNAs found in the cell that aligned with the GFP sequence increased significantly when transfected cell lines were compared to the untransfected controls, according to a t-test with a p value less than 0.05. Figure 3 shows the number of miRNAs aligning to each plasmid feature.

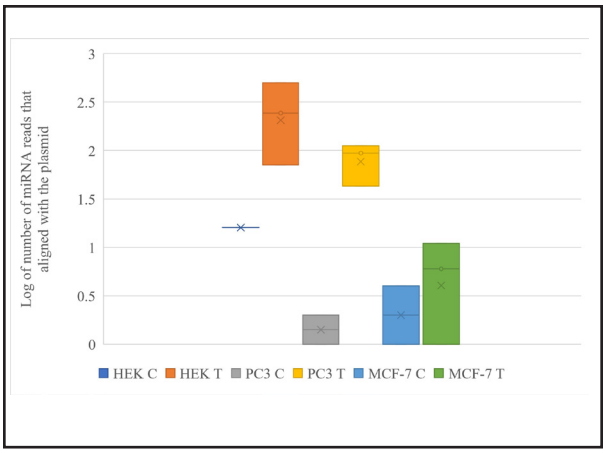


Figure 2. Number of miRNA reads that aligned with the GFP plasmid by cell line. Results were compared using a t-test assuming unequal variance. Each cell line had a sample size of three.

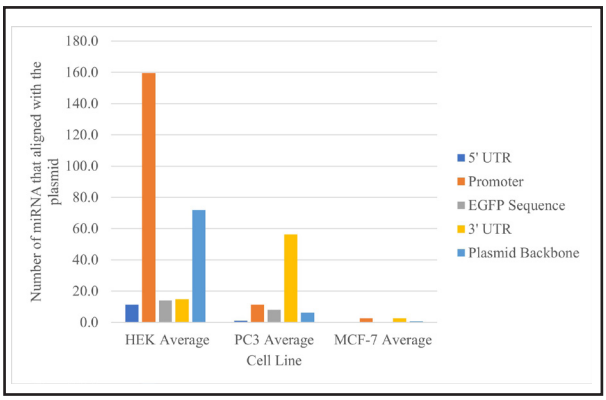


Figure 3. Number of aligned miRNA reads by plasmid feature. Results were compared using a t-test assuming unequal variance. Each cell line had a sample size of three.

Additionally, the number of miRNA reads that mapped to the plasmid was directly correlated with internal data for transfection efficiency. HEK cells, which generally have near 100% efficiency, had the most miRNA alignments. PC3, which tends to have moderate transfection efficiencies of around 50-80%, had the second highest number of reads. Finally, MCF-7 had the lowest number of aligned miRNAs, as well as the lowest transfection efficiencies. The majority of the miRNA in HEK bound to the promoter and the backbone of the plasmid. Since miRNA exhibit their regulatory control at the 3' UTR, these are not expected to have much effect on the transfection efficiency of the plasmid since the backbone of the plasmid is not transcribed or expressed. Most of the miRNA were found to bind to parts of the plasmid in which they would not have a regulatory effect. Additionally, the majority of the miRNA reads that

aligned with the plasmid were identical to the coding strand sequence. This means that they could not bind to the sequence, as reverse complementarity is required for strands to bind together. This, together with many not being known human miRNAs, suggests that most of the reads detected during RNA sequencing may be artifacts of the transcription and translation process rather than regulation of the plasmid. This theory is supported by the number of alignments increasing with increased transfection efficiency, as well as the reads not aligning with the miRNA that would be expected within human cell lines.

While many of the miRNA reads were likely artifacts of transcription and translation, not regulatory miRNAs, there were a handful of reads that exhibited perfect complementarity only in the seed region and were known human miRNAs. The human miRNAs that were upregulated after transfection were human miRNAs 603, 6724, and 6777. miRNA 603 was the most common of the miRNAs but aligned to the EF-1α promoter. The second most common was miRNA 6777, which aligned to the backbone of the plasmid, which likely means it would not have a significant effect on transfection efficiency. miRNA 6724 was found in all three cell lines and was a reverse complement of the 3' UTR. Additionally, miRNA 6724 has been identified in some studies as having an inhibitory effect on mRNA transcripts to which it binds (6). While many of the miRNA reads were likely not true miRNAs, several known regulatory human miRNAs were identified, fulfilling the main goals of the study of identifying regulatory miRNAs and their interactions with the transgene.

While the levels of any given mRNA fluctuate in a cell over time due to regulation, it is generally expected to find mRNAs in levels relative to how much of the protein is present in a cell. A large amount of mRNA is therefore expected for proteins that cells or plasmids are designed to express, in this case the green fluorescent protein for which GFP codes. Transcripts downstream of these genes may also be present if the termination process is not efficient. The sequences following other promoters are also expected to be transcribed, such as the ampicillin resistance gene, since promoters recruit the necessary cellular machinery to transcribe genes. It is also possible that cryptic promoters exist in the backbone that lead to other sequences being transcribed. While normally this may not have a noticeable impact on a cell's functioning, it is possible that an unintended promoter could lead to transcription of the reverse complement of the gene of interest, which could lead to the formation of double-stranded RNA complexes that would then be degraded. The mRNA sequencing alignment showed an increased number of alignments in transfected cells when compared to untransfected cells, and a positive correlation between the number of aligned reads and the average transfection efficiency of the cell line. Unlike the miRNA reads, which were spread over all regions of the sequence, the majority of mRNA reads aligned to either the sequence that coded for GFP, or the sequence that coded for ampicillin resistance, as shown in Figure 4 and Figure 5. This was expected, as the polyA selection in the sequencing process isolates only mRNAs that have a polyA tail. A polyA tail is added when the cellular machinery encounters a polyA site, AATAAA, and while this exact sequence does not occur downstream of the ampicillin resistance gene, there are several analogues that might be sufficient to trigger polyadenylation.

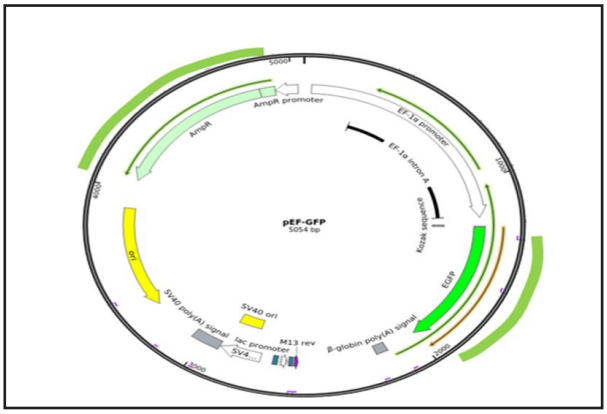


Figure 4. The regions of the pEF-GFP plasmid with high densities, as indicated by levels that were significantly higher than the background level at a p value of 0.05, of mRNA alignments across cell lines.

The quantity of reads aligned to each of these regions is displayed in Figure 5.

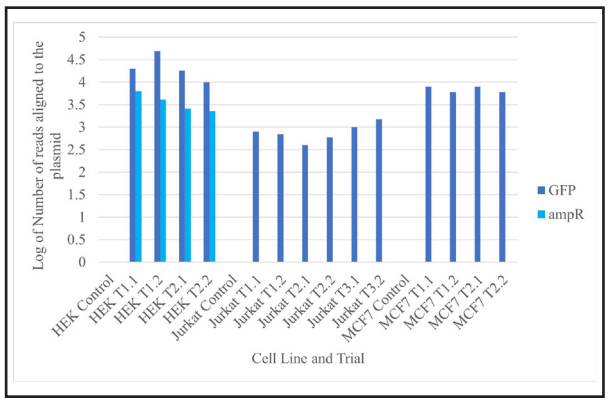


Figure 5. Number of mRNA reads that aligned with the GFP plasmid by cell line. "C" designates an untransfected control and "T" designates trials in which the cell line has been transfected with the GFP Plasmid. Results were compared using a t-test assuming unequal variance. Each cell line had a sample size of three.

Low levels of mRNA that covered the entire plasmid were present, which are likely products of transcription or from the genome, as the backbone regions of the plasmid should not be translated. mRNA that coded for GFP was present in the highest levels, which was expected since that is the protein that the plasmids were designed to express. The presence of high levels of the ampicillin resistance gene relative to the baseline suggests that strong bacterial promoters may be able to function in eukaryotic cells, which raises the possibility of off target effects and reactions from transfected cells if other parts of the plasmid are transcribed to mRNA and translated to proteins.

Further Studies

Further studies should be conducted to verify that the miRNA reads that were present were artifacts of other cellular processes rather than miRNAs regulating translation by using target-directed miRNA degradation to degrade the miRNAs that are present, particularly miRNA 6724, to assess the effect that the lack of this miRNA would have on the transfection efficiency of the cell lines. Additionally, the efficacy of the bacterial promoter in eukaryotic cells could be assessed by splicing it as the promoter for other genes. The findings of these studies could be used in conjunction with gene therapies to help reduce the immune system's action in preventing the expression of transgenes to aid in the treatment of various diseases caused by genetic mutations.

FUNDING

Support was provided by a Villanova Undergraduate Research Fellowship.

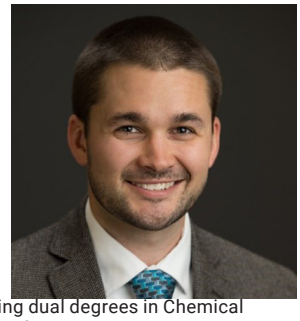
REFERENCES

1. Kaeberlein, M. & Kennedy, B. K. (2007). Protein translation. *Aging Cell*, 6(6), 731-734.
2. Eulalio, A., Huntzinger, E., & Izaurralde, E. (2008). Getting to the root of miRNA-mediated gene silencing. *Cell*, 132(1), 9-14
3. Lam, J. K. W., Chow, M. Y. T., Zhang, Y., & Leung S. W. S. (2015). siRNA Versus miRNA as Therapeutics for Gene Silencing. *Molecular Therapy - Nucleic Acids*, 4, e252, ISSN 2162-2531, https://doi.org/10.1038/mtna.2015.23.
4. Ebert, M. S., Neilson, J. R., & Sharp, P. A. (2007). MicroRNA sponges: competitive inhibitors of small RNAs in mammalian cells. *Nature Methods*, 4(9), 721–726. https://doi.org/10.1038/nmeth1079.
5. Albert, I. (2017). *The Biostar Handbook*. State College, PA: Biostar Genomics.
6. Usuba, W., Urabe, F., Yamamoto, Y., et al. (2019). Circulating miRNA panels for specific and early detection in bladder cancer. *Cancer Sci*, 110: 408– 419. https://doi.org/10.1111/cas.13856.

Author

Simon Blanchard

Simon Blanchard is a member of the Class of 2022. He is a Chemical Engineering major with a global health minor and has been researching in labs within the department since the start of his freshman year. During summers, he had the opportunity to further his research experience through REUs at the University of Texas at Austin and at the Memorial Sloan Kettering Cancer Center. After graduation, Simon plans to obtain an M.D. and engage in clinical research studying the immune system and autoimmune disorders. Simon is a Villanova Presidential Scholar.



Mentor

Dr. Jacob J. Elmer

Dr. Elmer began his career by earning dual degrees in Chemical Engineering and Biological Sciences from the Missouri University of Science & Technology (formerly University of Missouri – Rolla) in 2007. He then earned his Ph.D. from The Ohio State University under the guidance of Dr. Andre F. Palmer in 2011 by completing a dissertation that focused on using recombinant and invertebrate hemoglobins as novel blood substitutes. As a postdoc, he investigated multiple strategies to increase the efficiency of gene therapy treatments in the lab of Dr. Kaushal Rege at Arizona State University. Dr. Elmer in currently continuing his gene therapy research as an Associate Professor at Villanova University. His research mainly focuses on elucidating and manipulating host cell (e.g. T cell) defenses that are activated by the foreign DNA used in gene therapy.